# El Derecho Penal ante la Inteligencia Artificial

## Problemas de imputación y propuesta de solución

**Andrea Bravo Bolado**



Derecho Penal
y Procesal Penal

BOE

# EL DERECHO PENAL ANTE LA INTELIGENCIA ARTIFICIAL. PROBLEMAS DE IMPUTACIÓN Y PROPUESTA DE SOLUCIÓN

## COLECCIÓN DE DERECHO PENAL Y PROCESAL PENAL

Director

**Luis Rodríguez Ramos**
Catedrático de Derecho Penal de la Universidad Nacional de Educación a Distancia

Consejo Asesor

**Nicolás González-Cuéllar Serrano,** catedrático de Derecho Procesal de la Universidad de Castilla-La Mancha.

**Javier Álvarez García,** catedrático de Derecho Penal de la Universidad Carlos III; director de la Sección de Derecho Penal, parte general y parte especial.

**Alicia Gil Gil,** catedrática de Derecho Penal de la Universidad Nacional de Educación a Distancia.

**Silvina Bacigalupo Saggese,** catedrática de Derecho Penal de la Universidad Autónoma de Madrid.

**Adán Nieto Martín,** catedrático de Derecho Penal de la Universidad de Castilla-La Mancha; director de la Sección de Derecho Penal Europeo e Internacional.

**Esteban Mestre Delgado,** catedrático de Derecho Penal de la Universidad de Alcalá de Henares; director de la Sección de Derecho Penitenciario y de Ejecución de Penas y Medidas de Seguridad.

**Jacobo López Barja de Quiroga,** presidente de la Sala Quinta del Tribunal Supremo.

# EL DERECHO PENAL ANTE LA INTELIGENCIA ARTIFICIAL. PROBLEMAS DE IMPUTACIÓN Y PROPUESTA DE SOLUCIÓN

ANDREA BRAVO BOLADO

Prólogo por

FERNANDO MOLINA FERNÁNDEZ

COLECCIÓN DERECHO PENAL Y PROCESAL PENAL 35

AGENCIA ESTATAL BOLETÍN OFICIAL DEL ESTADO

MADRID, 2026

Primera edición: julio de 2026.

En la página web de la Agencia Estatal Boletín Oficial del Estado, www.boe.es, apartado de *La Editorial del BOE/Colecciones,* se incluyen las instrucciones para envío de originales, normas para su presentación y modelo de solicitud de publicación en esta colección que el autor deberá cumplimentar.

La AEBOE no se solidariza con las opiniones sostenidas por los autores de los originales publicados.

© De esta obra, Andrea Bravo Bolado
© Del prólogo, Fernando Molina Fernández
© Agencia Estatal Boletín Oficial del Estado



Esta obra está sujeta a licencia Creative Commons-Atribución-NoComercial-SinDerivadas 4.0 Internacional-CC BY-NC-ND 4.0

https://cpage.mpr.gob.es

NIPO: 144-26-066-7 (edición en papel)
144-26-067-2 (edición en línea, PDF)
144-26-068-8 (edición en línea, ePUB)
ISBN: 978-84-340-3149-4
Depósito legal: M-15544-2026

IMPRENTA NACIONAL DE LA AGENCIA ESTATAL
BOLETÍN OFICIAL DEL ESTADO
Avenida de Manoteras, 54. 28050 Madrid

*A mis abuelos, Ángel, Paquita, Maxi y Pura. Por hacer de mi infancia y juventud un lugar lleno de amor, por mostrarme su sabiduría en su faceta más humana y menos artificial, y por ser pilares de esta preciosa familia que hoy me sostiene y me arropa.*

# ÍNDICE


Páginas

ABREVIATURAS ..... 17

PRÓLOGO ..... 19

NOTA PRELIMINAR ..... 23

INTRODUCCIÓN ..... 25

CAPÍTULO PRIMERO. DE LA DIFÍCIL Y NECESARIA DELIMITACIÓN CONCEPTUAL DE LA INTELIGENCIA ARTIFICIAL ..... 29

I. **El concepto** ..... 30

1. Un breve repaso histórico sobre los orígenes de la inteligencia artificial ..... 31
2. Posibles clasificaciones ..... 37

2.1 La IA en función de la técnica empleada: sistemas expertos *vs.* aprendizaje automático ..... 37
2.2 La IA en función de las competencias: inteligencia *vs.* racionalidad ..... 39
2.3 La IA en función de su alcance: IA débil *vs.* IA fuerte ..... 40
2.4 La IA en función de la mayor o menor intervención humana. Del paradigma del programador al paradigma del demiurgo ..... 42

3. Llegando a puerto: los instrumentos legislativos y la definición que debe importar al Derecho (Penal) ..... 44

Páginas

3.1 La justificación del concepto ..... 44
3.2 Nuestro concepto de IA: desglosando los elementos del concepto propuesto por el Grupo de Expertos de Alto Nivel ..... 46
3.3 Otras definiciones: El Reglamento ..... 53

II. **La autonomía. La nota discordante en las definiciones de IA. ¿De qué hablamos cuando hablamos de autonomía?** ..... 56

1. La autonomía en sentido filosófico o fuerte ..... 59
2. La autonomía en sentido operativo o débil ..... 62

CAPÍTULO SEGUNDO. LAS LAGUNAS O NECESIDADES DEL DERECHO PENAL ANTE LA IA ..... 71

I. **Las lagunas del Derecho Penal** ..... 72

1. ¿Qué entendemos por laguna de responsabilidad? ..... 72
2. ¿Qué entendemos por responsabilidad? ..... 78

2.1 El argumento de las verdaderas lagunas ..... 81
2.2 El cuestionamiento de las lagunas ..... 86

3. Traducción del problema ¿qué necesita el Derecho Penal de hoy? ..... 89

II. **Las propuestas internacionales para paliar estas lagunas** ..... 91

1. El estudio sobre la afectación de las categorías tradicionales del Derecho Penal ..... 92
2. El estudio sobre los viejos y nuevos delitos ..... 101
3. Conclusiones ..... 112

CAPÍTULO TERCERO. DELIMITACIÓN DEL OBJETO DE LA INVESTIGACIÓN ..... 115

I. **Delimitación negativa: consideración sumaria de modelos alternativos** ..... 116

1. La posibilidad de castigo a la máquina ..... 116
2. La responsabilidad dentro de la persona jurídica ..... 127

II. **Delimitación positiva: objeto y foco principal de análisis** ..... 143

1. Las propuestas dogmáticas ..... 143
2. Las propuestas de los países ..... 146
3. Las críticas. Los problemas nucleares como punto de partida ..... 150

Páginas

3.1. La pérdida de importancia del principio de lesividad o la crisis del principio del daño ..... 150
3.2 Tendencia a la objetivación de la culpabilidad ..... 155
3.3 El problema de la pena y el llamado *«retribution gap»* .... 161

CAPÍTULO CUARTO. UNA REVISIÓN DE LOS PARADIGMAS DEL DERECHO PENAL DEL RIESGO ..... 165

I. **Una actualización de la moderna sociedad del riesgo** ..... 166

1. La tradicional discusión sobre la modernidad y la sociedad del riesgo ..... 166

1.1 La década de los 60: un caso que lo cambió todo ..... 167
1.2 La década de los 70: los orígenes de la Escuela de Frankfurt ..... 169
1.3 La década de los 80: la verdadera revolución sociológica del riesgo ..... 171
1.4 La década de los 90: la expansión definitiva del derecho penal ..... 175
1.5 Los años 2000: la discusión española ..... 180

a) Los moderados ..... 181
b) Los favorables ..... 184
c) Los críticos ..... 187

2. Conclusiones sobre la discusión «clásica» ..... 191

II. **La novedosa discusión sobre la ultramoderna sociedad algorítmica** ... 192

III. **Un discurso común** ..... 209

1. Los rasgos propios y genuinos de la ultramoderna sociedad algorítmica ..... 209

1.1 Lógicas estructurales: dataísmo, vigilancia y tecnofeudalismo ..... 209
1.2 Antropología del sujeto algorítmico: individualismo, rendimiento y deshumanización ..... 210
1.3 Desestructuración del espacio jurídico y crisis de deliberación .. 210
1.4 Emergencia de nuevas formas de poder y redefinición de lo político ..... 211

2. Comparativa entre la sociedad moderna y la sociedad algorítmica .... 211

2.1 Elementos comunes ..... 211
2.2 Diferencias esenciales ..... 212

Páginas

3. Reflexión final: ¿Qué le importa esto al derecho penal que surja en la sociedad algorítmica? ..... 213

IV. **El papel del principio de precaución** ..... 214

1. Origen ..... 215
2. Concepto y fundamentos filosóficos ..... 216
3. Los primeros debates sobre el principio de precaución en el derecho penal ..... 219

3.1. El encuadre dogmático del principio de precaución dentro del derecho penal ..... 221

a) Delitos de peligro ..... 221
b) Delitos imprudentes ..... 221
c) Causalidad ..... 222

3.2 Las posturas críticas ..... 222
3.3 Las posturas favorables ..... 225

4. La vuelta al debate de la precaución al hilo de la IA ..... 228
5. Toma de postura: Conclusiones sobre el principio de precaución y su papel ..... 234

CAPÍTULO QUINTO. LOS PROBLEMAS DEL MODELO DE RESPONSABILIDAD PENAL POR EL PRODUCTO ..... 237

I. **Orígenes teóricos y prácticos** ..... 238
II. **Los casos que hicieron nacer la discusión** ..... 240
III. **El problema del encaje del modelo hoy** ..... 244

1. ¿Es la IA un producto? ..... 246
2. ¿Son los errores de la IA un defecto? ..... 250

2.1 ¿Una nueva tipología de defectos? ..... 254

3. Intentos de encajar el modelo tradicional en el actual sistema penal ..... 259
4. Los problemas que subsisten: ..... 262

4.1 Causalidad ..... 262

a) El problema del vínculo causal entonces ..... 263
b) El problema del vínculo causal y la IA ..... 268
c) Toma de postura ..... 275

Páginas

4.2 Imputación subjetiva y objetiva del resultado lesivo del bien jurídico individual ..... 276

a) La preocupación por la previsibilidad, antes ..... 279
b) Los problemas de previsibilidad y la IA: un punto de quiebra estructural ..... 283
c) Propuestas de reconstrucción del concepto de previsibilidad ..... 292
d) Intentos de aplicación práctica del delito imprudente: el ejemplo del contexto médico ..... 296
e) Toma de postura ..... 304

4.3 Dificultades del esquema de la comisión por omisión (las polémicas en torno al deber de retirada) ..... 306

CAPÍTULO SEXTO. VIEJAS POLÉMICAS, NUEVOS RIESGOS: LOS BIENES JURÍDICOS AMENAZADOS Y LA TÉCNICA DEL PELIGRO ABSTRACTO ..... 317

I. **La problemática relectura de los bienes jurídicos colectivos** ..... 318

1. ¿Sigue sirviendo el bien jurídico como aparato conceptual? ..... 318

1.1 La tradicional polémica sobre el bien jurídico colectivo ... 321
1.2 La discusión sobre la salud pública y la seguridad colectiva en el desarrollo del bien jurídico intermedio o de referente individual ..... 327
1.3 Recapitulación ..... 336

2. ¿Necesitamos nuevos bienes jurídicos? La «seguridad colectiva algorítmica o digital» ..... 337

2.1 ¿Nuevos bienes jurídico-penales? ..... 339
2.2 IA y derechos fundamentales ..... 346
2.3 Los nuevos retos constitucionales de la IA ..... 351
2.4 Recapitulación ..... 355

II. **La técnica de los delitos de peligro abstracto: nuevas formas de ataque, nuevas formas de protección** ..... 360

1. El delito de peligro como paradigma de la modernidad ..... 360
2. La triada de los delitos de peligro ..... 361
3. Los problemas del delito de peligro abstracto ..... 365

3.1 Injusto material insuficiente: ¿solo desvalor de acción? ... 365
3.2 Injusto demasiado formal: ¿no hay forma de diferenciarlo de la infracción administrativa? ..... 368
3.3 Vulneración del principio de culpabilidad ..... 369

Páginas

4. Los argumentos para reducir el impacto negativo de los delitos de peligro abstracto. Toma de postura ..... 371

4.1 Los delitos de peligro abstracto como delitos de peligro presunto (una posición superada) ..... 371
4.2 La existencia de un resultado jurídico como daño *sui generis* ... 374
4.3 Los delitos de peligro como tentativas imprudentes ..... 377
4.4 La traslación de la teoría de la imputación objetiva ..... 380
4.5 La resignificación del injusto del delito de peligro abstracto sin referencia a la lesión del bien jurídico ..... 383

5. La utilidad del modelo de delitos contra la seguridad colectiva .... 387

5.1 Delitos relativos a la energía nuclear y a las radiaciones ionizantes ..... 389
5.2 El delito de estragos ..... 393
5.3 Otros delitos de riesgo ..... 395
5.4 El delito de incendio con peligro para la vida o integridad de las personas ..... 398

6. Conclusiones: La defensa de un delito *a la Marie Curie* en el ámbito del riesgo catastrófico tecnológico generado por la IA ... 401

CAPÍTULO SÉPTIMO. EL NUEVO ESCENARIO NORMATIVO EUROPEO ..... 405

I. **El reglamento de IA de la UE como piedra angular** ..... 406

1. Estructura y objetivos generales: el modelo de la gobernanza con legitimidad democrática ..... 406
2. La importancia del Reglamento como marco para el riesgo permitido del sector ..... 411
3. Principales obligaciones impuestas a los agentes económicos y públicos ..... 416

3.1 Introducción general: título competencial y estructura básica .. 416
3.2 Marco de aplicación ..... 418

a) Ámbito objetivo ..... 419
b) Ámbito subjetivo ..... 420
c) Ámbito territorial ..... 421

3.3 La IA prohibida: el núcleo garantista del Reglamento ..... 422

a) Los sistemas de identificación biométrica en tiempo real en espacios de acceso público con fines de garantía de cumplimiento del Derecho [(art. 5.1.h)] ..... 423

Páginas

b) Las técnicas de manipulación subliminal y explotación de las vulnerabilidades (art. 5.1.a y b) ..... 427
c) Los sistemas de clasificación social (5.1.c) ..... 430
d) Sistemas para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de comisión de un delito basándose únicamente en la elaboración de perfiles (5.1.d) ..... 432
e) Sistemas de IA basados en reconocimiento facial con imágenes extraídas de Internet o televisión (5.1.e), reconocimiento de emociones (5.1.f) y categorización biométrica para clasificación individual cuando deduzcan determinados datos sensibles (5.1.g) ..... 434
f) Algunas conclusiones sobre los usos prohibidos ..... 436

3.4 Los sistemas de alto riesgo: el verdadero núcleo del riesgo permitido ..... 436

a) Relación de ámbitos del Anexo III ..... 438
b) Régimen de obligaciones aplicable a los sistemas de IA de alto riesgo ..... 443
c) El núcleo de las obligaciones de los proveedores e implantadores en sistemas de IA de alto riesgo (Sección 2 del Capítulo III) ..... 446
d) El resto de obligaciones y disposiciones del Reglamento ..... 468
e) Recapitulación de los deberes y conclusiones críticas .... 478

II. **Otras normas relevantes** ..... 483

1. La IA como sistema basado en la explotación de datos (el Reglamento Europeo de Protección de Datos) ..... 483
2. Normativa de carácter civil: la Directiva sobre responsabilidad por productos defectuosos y la Propuesta de directiva sobre responsabilidad civil extracontractual en materia de IA ..... 485
3. *Soft law* y principios éticos. Especial mención al Convenio marco del Consejo de Europa ..... 491
4. Normativa de ejecución: El Anteproyecto de ley para el buen uso y la gobernanza de la inteligencia artificial ..... 493

CAPÍTULO OCTAVO. UNA PROPUESTA DE *LEGE FERENDA:* DELITOS CONTRA LA SEGURIDAD COLECTIVA TECNOLÓGICA 499

I. **Introducción** ..... 500

Páginas

II. **La propuesta de *lege ferenda*** ..... 501

1. Los deberes sustantivos ..... 502
2. Los deberes formales ..... 503
3. La exclusión de la condición objetiva de punibilidad ..... 504
4. La exclusión del delito de peligro concreto ..... 506
5. El modelo del delito de idoneidad ..... 508
6. Ubicación sistemática ..... 509
7. Las dificultades del modelo de ley penal en blanco con remisión absoluta ..... 511
8. La proporcionalidad de la pena ..... 514

III. **Epílogo: los sujetos obligados. El papel del usuario supervisor no profesional** ..... 520

1. Exclusión del usuario no profesional por motivos jurídicos ..... 522
2. Exclusión del usuario no profesional por motivos político-criminales ..... 524

2.1 Los humanos tienden a validar las decisiones de las máquinas (sesgo de automatización) ..... 527
2.2 Los humanos acaban por interiorizar inconscientemente el sesgo de los sistemas ..... 528
2.3 Los humanos acaban por perder capacidades cuando interactúan con sistemas que están diseñados para desempeñar tareas que les son propias ..... 529
2.4 Los humanos tienen una capacidad de reacción limitada en contextos que requieren respuestas rápidas ..... 530

3. Reflexión final ..... 532

CONCLUSIONES ..... 535

ANEXO I. PROPUESTA DE REGULACIÓN ..... 541

ANEXO II. BIBLIOGRAFÍA ..... 545

ANEXO III. REFERENCIAS NORMATIVAS ..... 591

ANEXO IV. SENTENCIAS Y RESOLUCIONES ..... 597

# ABREVIATURAS

| | |
|---|---|
| *AAP* | *Auto de la Audiencia Provincial* |
| *AA. VV.* | *Autores varios* |
| *AEPD* | *Agencia Española de Protección de Datos* |
| *AIDP* | *Asociación Internacional de Derecho Penal* |
| *AP* | *Audiencia Provincial* |
| *Art.* | *artículo* |
| *BGH* | *Bundesgerichtshof (Tribunal Federal de Justicia)* |
| *BGHSt* | *Entscheidungen des Bundesgerichtshofs in Strafsachen (Colección de sentencias penales del Tribunal Federal de Justicia alemán)* |
| *BOE* | *Boletín Oficial del Estado* |
| *BVerfG* | *Bundesverfassungsgericht (Tribunal Constitucional Federal alemán)* |
| *CEDH* | *Convenio Europeo de Derechos Humanos* |
| *CEN* | *European Committee for Standardization (Comité Europeo de Normalización)* |
| *CENELEC* | *European Committee for Electrotechnical Standardization (Comité Europeo de Normalización Electrotécnica)* |
| *CMIA* | *Convenio marco sobre la inteligencia artificial, los derechos humanos, la democracia y el Estado de derecho* |
| *Cons.* | *Considerando* |
| *Coord./s.* | *Coordinador/es* |
| *CP* | *Código Penal* |
| *Dir./s.* | *Director/es* |
| *DOUE* | *Diario Oficial de la Unión Europea* |
| *Ed./eds.* | *Editores* |
| *EEMM* | *Estados Miembros* |
| *Et. Al* | *y otros* |
| *ETSI* | *European Telecommunications Standards Institute (Instituto Europeo de Normas de Telecomunicaciones)* |
| *FJ* | *Fundamento Jurídico* |

| | |
|---|---|
| *FRIA* | *Fundamental Rights Impact Assessment (Evaluación de Impacto sobre Derechos Fundamentales)* |
| *HLEG* | *High Level Expert Group (Grupo de Expertos de Alto Nivel)* |
| *IA* | *Inteligencia Artificial* |
| *Ibid.* | *en el mismo lugar* |
| *Idem.* | *el mismo* |
| *ISO* | *Organización Internacional de Normalización* |
| *LO* | *Ley Orgánica* |
| *LOREG* | *Ley Orgánica de Régimen Electoral General* |
| *n.m.* | *número marginal* |
| *Núm.* | *Número* |
| *OCDE* | *Organización para la Cooperación y el Desarrollo Económicos* |
| *ONU* | *Organización de las Naciones Unidas* |
| *Op. cit.* | *obra citada* |
| *Orgs.* | *Organizadores* |
| *Parr.* | *párrafo* |
| *passim* | *en lugares diversos* |
| *PE* | *Parlamento Europeo* |
| *RAE* | *Real Academia Española* |
| *RD* | *Real Decreto* |
| *RDL* | *Real Decreto Ley* |
| *RGPD* | *Reglamento General de Protección de Datos* |
| *RIA* | *Reglamento de Inteligencia Artificial* |
| *SAN* | *Sentencia de la Audiencia Nacional* |
| *STC* | *Sentencia del Tribunal Constitucional* |
| *STEDH* | *Sentencia del Tribunal Europeo de Derechos Humanos* |
| *StGB* | *Strafgesetzbuch (Código penal alemán)* |
| *STJCE* | *Sentencia del Tribunal de Justicia de las Comunidades Europeas* |
| *STS* | *Sentencia del Tribunal Supremo* |
| *TC* | *Tribunal Constitucional* |
| *TEDH* | *Tribunal Europeo de Derechos Humanos* |
| *TFUE* | *Tratado de Funcionamiento de la Unión Europea* |
| *TJUE* | *Tribunal de Justicia de la Unión Europea* |
| *Trad.* | *traductor* |
| *TS* | *Tribunal Supremo* |
| *UE* | *Unión Europea* |
| *Vid.* | *véase* |
| *Vol.* | *volumen* |
| *Vs.* | *versus* |
| *XAI* | *Explainable Artificial Intelligence (Inteligencia Artificial Explicable)* |

# PRÓLOGO

Enfrentado al reto de identificar los «Momentos estelares de la humanidad», cualquier aprendiz de Zweig encabezaría muy probablemente la lista con el desarrollo del lenguaje y el dominio del fuego, y la cerraría con el desarrollo de la Inteligencia Artificial (IA). El éxito evolutivo del ser humano, sin igual para un organismo que consume tantos recursos, se asienta sobre dos características que se potencian mutuamente: la comunicación y la inteligencia. En ambos alcanzamos cotas sin parangón en el resto de la naturaleza conocida. Ningún otro ser vivo en nuestro planeta posee un lenguaje tan complejo, que nos permite construir juntos, y ninguno se aproxima a nuestra inteligencia general, no especializada, capaz de adaptarse a casi cualquier reto. … Ninguno, hasta ahora. La IA amenaza con sacudir, ya lo ha hecho en gran medida, este monótono panorama.

Hasta hace muy poco tiempo, los grandes avances tecnológicos aumentaban nuestra fuerza, nuestra velocidad, nuestra capacidad de comunicación o de cálculo, o extendían o hacían más cómoda nuestra vida, pero ninguna de estas habilidades nos definía antes ni después de su descubrimiento. No somos lo que somos por ser los más rápidos, los más longevos, ni los más fuertes, sino por ser los más inteligentes trabajando juntos. La combinación de inteligencia y cooperación tiene efectos exponenciales: las transformaciones son cada vez más rápidas, lo que es un difícil reto incluso para un ser tan singularmente adaptativo como el ser humano, pero cuya base genética se forjó en un mundo mucho más simple. Pero en esa carrera de progreso aparentemente infinito, de mejora cultural, irrumpe ahora un factor nuevo, un instrumento que apunta a dejar de serlo y convertirse en algo más. Si todo va como vaticinan muchos especialistas, es sólo cuestión de tiempo que la IA se convierta en inteligencia artificial general; la nuestra, pero en grado superior. El efecto que ello tendrá sobre nuestra cultura va a ser seguramente mayor que ningún otro avance del pasado. Pero mientras llega ese momento disruptivo, si es que llega, ya tene-

mos sobre la mesa problemas acuciantes planteados por la IA que el Derecho no puede obviar. De algunos de ellos se ocupa esta ejemplar monografía.

Cuando hace unos años Andrea Bravo comenzó a desarrollar su tesis doctoral sobre IA y Derecho penal, las cejas levantadas eran la tónica. En un mundo tan reacio a lo nuevo como las ciencias jurídicas, aquello a muchos les parecía por lo menos exótico. Hoy, transcurridos unos pocos años, nadie sensato pondría un solo reparo al objeto de investigación. Hemos pasado de una noticia a la semana a muchas todos los días. La transformación de nuestro día a día en un lapso tan corto apenas tiene elementos de comparación. Y acabamos de empezar.

Frente a esta avalancha de novedades se necesitan investigaciones inteligentes y sensatas, y el libro que tengo el placer de prologar abunda en ambas.

Son distintos los enfoques posibles de un tema como el impacto de la IA en el Derecho penal. Los más cinematográficos apuntan a la posible futura aparición de un nuevo sujeto de derechos y deberes, una nueva persona consciente de sí misma que, poseyendo en grado sumo los dos ingredientes del éxito, capacidad de comunicación e inteligencia, pueda alterar el orden biológico que hasta ahora dominamos los seres humanos. Por ahora son sólo cábalas, de las que se ha ocupado con éxito la ciencia ficción (HAL 9000 sería el epítome) y que inquietan a muchos, y no todos son legos sin información suficiente. Pero hay otras cuestiones mucho más urgentes -*Hannibal ad portas*-: entre otras, la cuestión central de la imputación de responsabilidad jurídica, de la que se ocupa esta investigación.

La inteligencia es la herramienta más versátil, y puede usarse en cualquier dirección, para mejorar la vida o para cometer más eficazmente delitos. Pero la IA no es exactamente, o no siempre, una herramienta como las demás. Detrás de un delito hay una persona responsable, que identificamos verificando un nexo de imputación, primero objetiva y luego subjetiva entre el hecho y el autor. Este nexo entraña control del suceder causal, control que, dependiendo de la autonomía del instrumento utilizado, puede ser más o menos amplio, pero que en todo caso permite trazar la conexión del responsable con el hecho lesivo. Pero cuando el instrumento es una IA generativa, básicamente 'autónoma' en sus decisiones, la trazabilidad se vuelve prácticamente imposible. Ni el programador, ni en realidad nadie en la cadena causal puede controlar, con el nivel exigible, el *output*. Ello obliga al Derecho penal a intervenir antes, mediante delitos de peligro, estrategia que es bien conocida en la dogmática. Consciente de este hecho, Andrea Bravo desarrolla a lo largo de esta monografía toda una teoría de la imputación en los delitos de peligro aplicada a la IA, conectando su examen con figuras jurídicas tan conocidas como la responsabilidad por el producto o los delitos de peligro abstracto.

Sólo con eso, la investigación ya estaría más que justificada, pero el libro contiene mucho más. Consciente de que es imposible hablar con autoridad de

IA sin un conocimiento técnico profundo de su realidad, la primera parte de la obra se dedica precisamente a esclarecer su delimitación conceptual. Cualquier lector agradecerá la claridad explicativa y el rigor de la autora, pero si, como sucederá muy probablemente, es además jurista, sabrá ponderar especialmente esta incursión en el mundo técnico, tan poco habitual pero tan necesaria, que indica seguramente mejor que cualquier otro dato la ambición intelectual de la investigadora.

Y por si ello fuera poco, la obra contiene además un análisis inteligente y exhaustivo de la normativa europea sobre IA, entre la que destaca su novedoso Reglamento, y, lo que es especialmente destacable, una propuesta final de regulación que condensa el trabajo previo de la autora y da a la investigación un contenido práctico que merece todos los elogios. No es fácil legislar en general, y menos sobre cuestiones novedosas, así que bajar a la arena y proponer un catálogo de 'Delitos contra la seguridad colectiva tecnológica' es un gesto valiente y honesto, que culmina una obra tan rigurosa y necesaria. Si a todas estas virtudes se suma una redacción excelente, con hallazgos tan originales y atractivos como las extensas citas que encabezan cada capítulo, el resultado es una obra excepcional.

Tuve la suerte de dar clase a Andrea Bravo hace años en el Máster de Derecho penal que cursaba en la UAM. En un grupo excelente, ella destacaba como la mejor alumna. Cuando, al terminar, me confesó que le atraía la carrera académica hice todo lo posible por animarla. Tuve suerte en el empeño, aunque no hizo falta ningún esfuerzo especial para convencerla. Desde entonces ha desarrollado su labor en la UAM y en otros centros, donde se ha ganado el aprecio y la admiración de todos, demostrando lo que apuntaba: que es una excepcional investigadora y docente, que además tiene la virtud de hacer la vida más fácil a los que la rodean. Una genuina inteligencia natural.

Fernando Molina Fernández.
*Catedrático de Derecho penal.*
*Universidad Autónoma de Madrid*

# NOTA PRELIMINAR

El contenido de esta obra coincide, en lo esencial, con el fruto de la investigación desarrollada a efectos de obtener el Título de Doctora en Derecho, Gobierno y Políticas Públicas en la Universidad Autónoma de Madrid el 16 de diciembre de 2025. El tribunal estuvo compuesto por los Profs. Dres. Mirentxu Corcoy Bidasolo, Fernando Miró Llinares, Susana Aires de Sousa, Eduardo Demetrio Crespo y Bernardo Feijoo Sánchez, quienes otorgaron a la misma por unanimidad la calificación de *cum laude* y la mención internacional. A ellos les debo un primer agradecimiento sincero, pues con su calidez y su humanidad hicieron que el día de la defensa quedase, para siempre, como un precioso recuerdo imborrable.

El segundo agradecimiento es, sin duda, a mi director, el profesor Molina Fernández. Porque antes de ser director fue ya maestro, y porque sus clases encendieron en mí la chispa de la curiosidad. Gracias por la libertad, por la confianza ciega, y por su optimismo casi patológico, que resultó ser el antídoto perfecto para una doctoranda sumida en la inseguridad constante. Unas palabras de agradecimiento también a mis supervisores fuera de España, que me hicieron sentir acogida y acompañada lejos de casa. A los profs. Leverick, Chalmers, Fronza y Hilgendorf, y muy especialmente a Lucila y a Leandro.

Un agradecimiento muy especial también a mis compañeros del área de Derecho penal de la Autónoma. Cumplía dieciocho años el día que comencé la carrera, entre nerviosa e ilusionada. Mientras crecía entre los pasillos de la facultad, fascinada por el derecho penal, no hubiera sido capaz de imaginar que los profesores que me hacían salir de clase admirada y estimulada pasarían a ser compañeros con los que compartir el día a día. Compañeros que me harían sentir como una hermana pequeña académica, haciendo que las penas

compartidas fueran menos duras. Confieso que he continuado en ese estado de admiración y fascinación en cada uno de nuestros seminarios y debates improvisados en cada café, comida o encuentro de pasillo.

Gracias, también, a mis amigas y amigos. La palabra amistad se queda corta cuando me tengo que referir a ellos, y es que he tenido la inmensa suerte de encontrarme rodeada de personas que más que amigos son familia. Gracias por estar cada vez que he dudado de mis capacidades y por confiar en que esto saldría adelante mucho más de lo que yo lo he hecho.

A Dani. Gracias por estar aún en la distancia, por darme el tiempo y el espacio, por sacarme una sonrisa ante cada amago de lágrima, por ser familia, y por ofrecerme la tuya propia. Gracias por entender que preferí ser india antes que una importante abogada, gracias por el día a día, y por convencerme de que, al final, todo siempre sale bien.

Gracias infinitas a mi maravillosa familia, y muy especialmente a mis padres y mi hermano. Por ser el hogar al que siempre volver. Cada vez que he dudado sobre el rumbo que debía tomar mi vida, la respuesta ha sido la misma: haz lo que te haga feliz. Su apoyo, confianza y amor incansable han sido refugio en este proceso y en todas las decisiones de mi vida.

# INTRODUCCIÓN

La obra que aquí se presenta trae causa de una investigación que se inicia en el año 2021, cuando la inteligencia artificial era aún un fenómeno incipiente, sobre el que la regulación apenas comenzaba a debatirse. La pregunta inicial era sencilla: ¿cómo debía transformarse el sistema penal frente a una tecnología que la doctrina ya equiparaba a la cuarta revolución industrial? Esa primera formulación, demasiado amplia, se acompañó de un proyecto de investigación titulado *La implantación de la inteligencia artificial en el derecho penal. Predicción de la peligrosidad y crimen en la era de los robots*. Con el tiempo, evolucionó hacia un enfoque más preciso, cristalizado en el título definitivo: *El derecho penal ante la inteligencia artificial: problemas de imputación y propuesta de solución*.

En las etapas iniciales se exploró la IA como herramienta de apoyo a la toma de decisiones penales. Esta aproximación me permitió adentrarme en la realidad de la tecnología, desde la minería de datos y los métodos de inferencia estadística más básicos, hasta los modelos más avanzados de simulación del aprendizaje humano mediante la repetición y la experimentación. Pronto resultó evidente que la pregunta de investigación no podía limitarse al análisis de algunas herramientas predictivas, y que las características inherentes a los sistemas más complejos, esencialmente aquellos basados en el autoaprendizaje, podrían suponer un cambio en las estructuras de imputación del delito. La IA no se usaba solo para tomar decisiones sobre reincidencia o predicción de la peligrosidad, sino que era una silenciosa herramienta de asistencia en tareas tan variadas como la medicina, la conducción, o la concesión de ayudas sociales. La pregunta que guiase la investigación debía ser, pues, más amplia, y menos futurista. El problema no residía en los «robots», sino en los datos.

La investigación se organiza en tres bloques temáticos. El primer bloque, centrado en el estado de la cuestión, caracteriza la tecnología (capítulo primero) y aborda, desde una perspectiva filosófica, los problemas que plantea la imputación de daños (capítulo segundo). A continuación, examina los primeros modelos de imputación desarrollados por la doctrina, poniendo de relieve sus limitaciones (capítulo tercero).

El segundo bloque, dedicado a la sociedad del riesgo y los problemas de imputación de su modelo, parte de la necesidad de revisar el modelo de nueva sociedad, los riesgos que en ella emergen y los principios que la rigen (capítulo cuarto). Sobre esa base, analiza el modelo de imputación penal propio de la sociedad del riesgo y sus deficiencias (capítulo quinto), para luego determinar cuáles son los bienes jurídicos comprometidos en dicho contexto (capítulo sexto).

Finalmente, el tercer bloque, orientado a la propuesta de un modelo alternativo, busca ofrecer una respuesta a las lagunas de responsabilidad detectadas en los modelos de imputación previstos por el derecho penal. Tras un examen exhaustivo de la normativa aplicable (capítulo séptimo), se culmina con la propuesta de un modelo de tipificación innovador (capítulo octavo).

El objeto del capítulo primero es, por tanto, ofrecer al lector un entendimiento sobre la conceptualización de la inteligencia artificial desde un punto de vista técnico. Tras un breve repaso histórico, se ponen de manifiesto los rasgos que componen los diferentes tipos de sistemas, tratando de destacar el salto conceptual que determinados elementos, como la autonomía técnica u operativa, suponen en el entendimiento de los sistemas. Se ofrece, finalmente, un concepto común, que bebe de las definiciones legales.

El segundo capítulo desarrolla los problemas que, desde un punto de vista filosófico, suponen las características inherentes a los sistemas más complejos, explorando la pregunta acerca de las posibles lagunas de responsabilidad que pueden implicar sistemas con gran autonomía y altos niveles de opacidad. Para tratar de responder la pregunta se exploran, también, los planteamientos de diversos países respecto de las necesidades de modificación del derecho penal ante la IA.

En el capítulo tercero se estudian, con el objetivo de encontrar un modelo adecuado para afrontar las lagunas detectadas, diferentes modelos de imputación. En la primera parte se descartan los modelos de imputación del delito directamente al sistema, así como de imputación directa a la persona jurídica, para plantear los problemas que supone el modelo anglosajón de las *minor offences*, similar a nuestro modelo de delito de peligro, si bien con algunas diferencias no desdeñables.

El capítulo cuarto tiene por objeto el estudio de los paradigmas de la sociedad del riesgo, entendiendo que para encontrar un modelo de imputación adecuado es necesario explorar el modelo social en que la tecnología se inserta. Se exploran, en orden cronológico, los postulados económicos y sociales de la bautizada como sociedad del riesgo, para finalmente llegar a los autores más influyentes de la sociedad contemporánea. En la segunda parte se explora uno de los postulados del derecho penal de la modernidad, como es el principio de precaución, todo ello con el objeto de fijar los paradigmas de la sociedad en que se debe insertar el derecho penal que afronte los retos de la IA.

En el capítulo quinto se explora la plasmación de la sociedad del riesgo en un modelo de derecho penal, repasando los postulados de la responsabilidad penal por el producto, estudiando los problemas que surgen en la imputación del delito de resultado, tanto en lo relativo a la causalidad, como en la imputación objetiva y subjetiva del resultado. Para ello, se opta por una metodología comparativa, que pone en contraste el discurso de la sociedad del riesgo, presente en el auge de la modernidad, con el discurso de los autores contemporáneos, con el objetivo de comprobar si los problemas de causalidad e imputación que entonces se planteaban con la multiplicación de riesgos tecnológicos se sostiene hoy en día.

El capítulo sexto sigue una metodología similar, esta vez comparando los riesgos para los bienes jurídico-penales que existían, entonces y ahora. Se parte de la discusión sobre el bien jurídico supraindividual propia de la sociedad del riesgo para realizar un repaso por los posibles bienes afectados en la sociedad tecnológica, con el objeto de responder a la pregunta de si el derecho penal necesita crear nuevos bienes jurídicos ante la IA. Se explora, en la segunda parte del capítulo, el modelo de delito de peligro, por ser la respuesta penal propia de la expansión del derecho penal, instrumento idóneo en su momento para hacer frente al desarrollo tecnológico, estudiando sus críticas más comunes.

El capítulo séptimo estudia la normativa extrapenal directamente aplicable, tanto supranacional (europea), como internacional y nacional. Se trata de responder a la pregunta acerca de cuál debe ser la relación que el derecho penal y el Derecho administrativo tengan en la configuración de deberes en el uso y desarrollo de sistemas de IA.

Finalmente, el capítulo octavo concluye la investigación realizando una propuesta de *lege ferenda.* Tras la detección de lagunas derivadas de las características propias de los sistemas, que hacen pensar que la imputación de los delitos ya existentes en el Código penal puede complicarse, se trata de evitar

la impunidad en los casos más graves, con la propuesta de la creación de un nuevo título en el Código penal.

En definitiva, esta investigación asume una premisa difícil de cuestionar: la impermanencia. La respuesta sobre si el derecho penal debe transformarse frente a la IA es, inevitablemente, afirmativa. Según los escritos budistas, «*todas las cosas condicionadas son como un sueño, un fantasma, una burbuja, una sombra, una gota de rocío o un relámpago*». Fugaces, breves, impermanentes. La apertura al cambio es, pues, un rasgo no solo del derecho penal, sino un postulado indiscutible de nuestra propia existencia.

# CAPÍTULO PRIMERO. DE LA DIFÍCIL Y NECESARIA DELIMITACIÓN CONCEPTUAL DE LA INTELIGENCIA ARTIFICIAL

*En esto, descubrieron treinta o cuarenta molinos de viento que hay en aquel campo, y así como don Quijote los vio, dijo a su escudero:*

*–La ventura va guiando nuestras cosas mejor de lo que acertáramos a desear; porque ves allí, amigo Sancho Panza, donde se descubren treinta o pocos más desaforados gigantes, con quien pienso hacer batalla y quitarles a todos las vidas, con cuyos despojos comenzaremos a enriquecer, que esta es buena guerra, y es gran servicio de Dios quitar tan mala simiente de sobre la faz de la tierra.*

*–¿Qué gigantes? –dijo Sancho Panza.*

*–Aquellos que allí ves –respondió su amo–, de los brazos largos, que los suelen tener algunos de casi dos leguas.*

*–Mire vuestra merced –respondió Sancho– que aquellos que allí se parecen no son gigantes, sino molinos de viento, y lo que en ellos parecen brazos son las aspas, que, volteadas del viento, hacen andar la piedra del molino.*

*–Bien parece –respondió don Quijote– que no estás cursado en esto de las aventuras: ellos son gigantes; y si tienes miedo quítate de ahí, y ponte en oración en el espacio que yo voy a entrar con ellos en fiera y desigual batalla*[1].

[1] De Cervantes, M., *Don Quijote de la Mancha*, Edición del Instituto Cervantes, Dirigida por Francisco Rico, accesible en: <https://cvc.cervantes.es/literatura/clasicos/quijote/default.htm> [Accedido por última vez el 20/01/2026].

## I. EL CONCEPTO

Con estas palabras da comienzo el capítulo VIII del célebre Quijote de Cervantes. En el pasaje se enfrentan dos posturas que pudiéramos fácilmente identificar con la pugna que hoy en día presenciamos al tratar de determinar de qué hablamos cuando nos referimos a la ya manida inteligencia artificial. Vocablo que, pese a su notoria ubicuidad actual, habiéndose convertido en palabra del año consecutivamente en 2022 [2] y 2023 [3], goza de una existencia no demasiado larga y una delimitación en absoluto pacífica. Nos remontaremos, brevemente, a sus orígenes, para tratar de aportar una definición que sea útil a nuestros fines, que no son otros que arrojar algo de luz sobre el conflicto que viene enfrentando a los *Quijotes* y los *Sanchos* de nuestros tiempos, los primeros tratando de convencernos de que la revolución de la IA nos trae una fiera y desigual batalla de la que difícilmente podremos salir ilesos; los segundos tratando de banalizar los temores y reconducir este fenómeno a los confines de lo conocido.

Los molinos de viento eran ya conocidos en la España del Siglo de Oro, pero lo cierto es que aquellos a los que se enfrentó el imaginativo caballero eran muy probablemente una novedad, pues se apunta que se introdujeron hacia 1575, procedentes de Países Bajos [4]. Aquello a lo que se enfrentaba Don Quijote era, pues, a un nuevo artilugio técnico, que tenía algo de familiar, pero cuyos novedosos caracteres le empujaron a pensar que su integridad física podía peligrar, y que era menester quitar tan mala simiente de la faz de la tierra.

Veamos, pues, si la técnica que aquí nos ocupa puede comprenderse de forma algo más clara tras una conceptualización cautelosa, para así, más adelante, determinar si las amenazas a las que nos enfrenta justifican una batalla si no caballeresca, sí legislativa, doctrinal o político-criminal, para evitar que estos nuevos molinos (o gigantes) tecnológicos vengan a sembrar el caos.

---

[2] La Fundación del Español Urgente (FundéuRAE), promovida por la Agencia EFE y la Real Academia Española, otorga el título de palabra del año a la expresión compleja inteligencia artificial. *Vid*.: <https://www.fundeu.es/recomendacion/inteligencia-artificial-es-la-expresion-del-2022-para-la-fundeurae/>, [Accedido por última vez el 20/01/2026].

[3] Declarada así por el Diccionario Collins. Accesible en: <https://www.collinsdictionary.com/es/woty>, [Accedido por última vez el 20/01/2026].

[4] Aclaración contenida en nota al pie n. 2 del capítulo VIII, edición publicada por el Centro Virtual Cervantes, accesible en: <https://cvc.cervantes.es/literatura/clasicos/quijote/edicion/parte1/cap08/default.htm> [Accedido por última vez el 20/01/2026].

## 1. UN BREVE REPASO HISTÓRICO SOBRE LOS ORÍGENES DE LA INTELIGENCIA ARTIFICIAL

La tarea de conceptualizar la inteligencia artificial constituye, sin duda, punto de partida necesario para cualquier reflexión seria sobre las consecuencias y límites de su aplicación. Se trata, sin embargo, de una labor ingrata, pues, como ha sido explicitado por algún autor, es una tarea tan complicada como tratar de alcanzar el horizonte; cuanto más te acercas a él, más se aleja [5]. Lo dinámico y complejo de la conceptualización no es, empero, una realidad que nos exima de la necesaria misión, y a ella dedicaremos este primer capítulo, no sin antes apuntar que, probablemente, lo que, a día de hoy, principios del año 2026, pueda decirse, quedará probablemente superado por el irrefrenable crecimiento de una tecnología tan veloz y cambiante. Esto, no obstante, supone un reto apasionante, pues solo el tiempo dirá si las reflexiones aquí contenidas resisten a la imparable evolución de la tecnología.

Una vez realizada esta advertencia, nos remontamos, ya sí, a los orígenes del concepto que nos ocupa, no sin antes advertir que la prehistoria de la IA quedará fuera de estas páginas, pues todas las reflexiones que, antes de acuñarse el concepto de inteligencia artificial, se realizaron a lo largo de la historia de la humanidad sobre los llamados «autómatas» nos llevarían a extendernos demasiado sobre este tópico. Baste señalar que ya Homero en su Ilíada nos narraba cómo Hefesto, dios del fuego, la metalurgia y la artesanía, insufló vida, a partir del oro, a una pléyade de sirvientas, que no eran más que autómatas con inteligencia y capacidad de habla, encontrando, pues, en una de las primeras obras de referencia de la tradición occidental, una alusión al eterno interés del hombre en crear réplicas de su inteligencia para el desempeño de tareas propiamente humanas. La historia de la humanidad se encuentra, así, impregnada de referencias a estos seres, lo que nos permite concluir que, si bien el concepto es moderno, el interés que está en el fondo de nuestra reflexión es, más bien, atemporal [6].

---

[5] Turner, J., *Robot rules: Regulating artificial intelligence*. Springer, 2019, p. 8.

[6] Para un mayor desarrollo de la importancia en la mitología y la religión de estas figuras: Zarkadakis, G., *In Our Image: Will Artificial Intelligence Save or Destroy Us?*, Rider, Londres, 2015. Homero, La Iliada, Trad. Emeterio Fuentes, Verbum, Madrid, 2020, canto XVIII, verso 410: «*Apartó de la llama los fuelles y puso en un arcón de plata las herramientas con que trabajaba; enjugóse con una esponja el sudor del rostro, de las manos, del vigoroso cuello y del velludo pecho, vistió la túnida, tomó el formido cetro, y salió cojeando, apoyado en dos estatuas de oro que eran semejantes a vivientes jóvenes, pues tenían inteligencia, voz y fuerza, y hallábanse ejercitadas en las obras propias de los inmortales dioses*». Otro antecedente que merece la pena citar por su relevancia filosófica tiene que ver con el debate, desarrollado a partir del siglo XVI, acerca de si los animales son meras máquinas, que tuvo expresiones no solo en la filosofía de Descartes (principalmente en *El discurso del método),* donde el francés sostenía que los

No es posible, tampoco, hacer un recorrido completo por las obras de tinte ficcional que, aún en la «prehistoria» de la IA, plantearon reflexiones profundas sobre problemas que hoy siguen vigentes, aludiendo, sin embargo, a términos más apegados a la ciencia ficción como el de «robot». Es referencia obligada, a estos efectos, la famosa obra de Isaac Asimov «El círculo Vicioso», donde se enuncian las conocidísimas «leyes de la robótica» ya en 1942 [7].

Estas referencias prematuras a los autómatas, androides o robots, máquinas pensantes hechas a la imagen y semejanza del hombre, encuentran su acomodo científico definitivo cuando un grupo de investigadores decide organizar el evento académico fundacional de la disciplina que dio origen a la verdadera historia de la inteligencia artificial.

Nos debemos situar en el verano de 1956 en New Hampshire, Estados Unidos. La Conferencia de Dartmouth es el evento donde se bautiza a la incipiente disciplina con el vocablo «inteligencia artificial», atribuido a uno de los organizadores del ciclo de conferencias, Marvin Minsky, por entonces profesor ayudante de matemáticas en el Dartmouth College. La citada conferencia, que se extendió a lo largo de varias semanas y que dio lugar a las primeras reflexiones académicas sobre la recién estrenada disciplina, reunió a pensadores que provenían del mundo universitario y académico, como es el caso de los matemáticos John McCarthy y Marvin L. Minsky, por entonces matemático y neurólogo por la Universidad de Harvard, pero también del mundo empresarial, como el caso del ingeniero de IBM Nathaniel Rochester y el matemático e ingeniero Claude E. Shannon, considerado padre de la teoría de la información, que trabajaba para Bell Telephone Laboratories [8].

---

animales son esencialmente autómatas guiados por causas físicas, a diferencia del ser humano, único ser dotado de razón y consciencia, sino ya antes en el célebre escrito del español Gómez Pereira (*Antoniana Margarita),* argumentando que los animales carecen de alma racional, siendo su comportamiento meramente automático, trazando una diferencia radical entre bestia y hombre. Se ha considerado que Gómez Pereira, con su pensamiento, anticipa, de hecho, el dualismo cartesiano. *Vid.:* Descartes, R., *Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences,* Jean Marie, Leiden, 1637; Gómez Pereira, *Antoniana Margarita,* Servicio de Publicaciones Universidade de Santiago de Compostela, 2000 (primera edición: 1554).

[7] Asimov, I., «El Círculo Vicioso (Runaround)», *Astounding Science Fiction*, 1942. Resulta indispensable resaltar que el Parlamento Europeo ha utilizado, en recientes resoluciones sobre la regulación de la robótica, referencias a las obras de ficción que encumbraron estas primeras reflexiones, pasando por el Frankenstein de Mary Shelly, el Golem de Praga, hasta el propio Asimov. *Vid.*: Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL), Considerandos A) y T), donde se establece que estas obras demuestran que: «*los seres humanos han fantaseado siempre con la posibilidad de construir máquinas inteligentes, sobre todo androides con características humanas*».

[8] McCarthy, J./Minsky M. L., Rochester, N./Shannon, C. E., «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence», 31 de Agosto de 1955. Disponible en: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> [accedido por última vez el 20/01/2026].

Las primeras definiciones nacen, por tanto, al albur de esta Conferencia, y la disciplina llama la atención de una gran cantidad de científicos de la época, todos dispuestos a contribuir activamente en la delimitación de esta rama de conocimiento que nace unida a las matemáticas, a la neurociencia, y, por supuesto, al mundo empresarial.

Así, por recorrer algunas de las definiciones «históricas», se puede señalar la realizada por el padre ideológico del término, que habla de «la ciencia de crear máquinas que hagan cosas que requerirían la inteligencia si se llevaran a cabo por hombres» [9].

No son casuales las referencias a la inteligencia humana en las primeras reflexiones, pues se produce la influencia notable de un escrito previo a la Conferencia fundacional, pero esencial para entender el porqué de esta antropomorfización de la disciplina. Resulta esencial para entender los primeros intentos de definir la IA el escrito del matemático británico Alan Turing, cuyas aportaciones a la informática, la computación y la inteligencia artificial son innegables aún hoy en día. Su célebre escrito «Computing Machinery and Intelligence», publicado en 1950, mantiene vigencia en la actualidad. Las palabras con las que inicia el ensayo son: «Propongo que se considere la siguiente pregunta, ¿Pueden pensar las máquinas?» [10]. Lo que propugna el matemático es la realización de un test (el bautizado como Test de Turing o Juego de Imitación) que permitirá afirmar que una máquina puede pensar sometiéndola al siguiente juego: un entrevistador se comunica con dos interlocutores, de los cuales uno de ellos es un humano y el otro una máquina. Dicho entrevistador debe ser capaz de discernir, realizando preguntas mediante lenguaje natural, quién es quién. En caso de que el entrevistador no supiera distinguir al humano de la máquina, esta última habría ganado el juego, por lo que se podría interpretar que puede pensar [11].

---

[9] Traducción propia a partir de la referencia en: Kaplan, A., / Haenlein, M., «Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence», *Business horizons, 62(1),* 2019, p. 17. Se refieren estos autores a la obra: Minsky, M. L., *Semantic information processing*. MIT Press, Cambridge, 1968, p. v.

[10] La publicación original es: Turing, A., «Computing machinery and intelligence», *Mind*, 59, 236, 1950, pp. 433-460.

[11] Las críticas a este juego de imitación no tardaron en sucederse, pues en el mismo se evalúa no el conocimiento de la máquina, sino simplemente su capacidad de imitar un comportamiento humano, confundiendo Turing, según ha señalado uno de sus críticos más férreos, John Searle, el plano semántico y el sintáctico. A estos efectos creó Searle el conocido experimento de la habitación china, donde diferencia la mera manipulación de símbolos sin dotarlos de significado (sintaxis), de la capacidad de entender el significado de los mismos (semántica). *Vid*.: Searle, J. R., /Villanueva, L. M. V., *Mentes, cerebros y ciencia*, Cátedra, 2001. Sin embargo, cabe destacar, que, pese a todas las críticas, el juego de imitación de Turing encuentra hoy en día aplicación diaria en el conocido test CAPTCHA (que responde precisamente a las siglas *Completely Automated Public Turing test to tell Computers and Humans Apart*), que es la

Podemos encontrar aquí, por tanto, el origen de una de las maneras de concebir la IA, al estilo de una máquina pensante, que imite los comportamientos de un humano. No es esta, sin embargo, la única de las posibles conceptualizaciones.

Otra de las definiciones clásicas cuando se habla de IA se atribuye a los estadounidenses Rusell y Norvig [12], que realizan una aproximación «moderna» a la inteligencia artificial. Podemos observar aquí un cambio en el enfoque, pues estos autores prefieren una definición basada en la racionalidad, dejando a un lado esa imitación o semejanza con el humano, centrándose más bien en las funciones encomendadas a la máquina. Así lo expresa Gómez Colomer, cuando realiza un magnífico repaso por la historia de la IA y sus diferentes conceptos. Se apunta por este último autor que la definición de Rusell y Norvig pone el énfasis en «*la capacidad de elegir la mejor acción a emprender para alcanzar un determinado objetivo, teniendo en cuenta determinados criterios de optimización y los recursos disponibles*» [13].

Como narra Gómez Colomer [14], las primeras aplicaciones de la IA no iban más allá de tareas simples de búsqueda, clasificación, tratamiento y uso de la información, pero, tras unos años difíciles (la época conocida como «invierno de la IA»), la tecnología consigue despegar de nuevo, dando lugar a nuevas aplicaciones. En los años 80-90, con la generalización de internet, se empieza a producir un desarrollo de las técnicas de computación y la IA empieza a encontrar aplicación práctica, en un primer momento unida a la solución de juegos de tipo lógico, donde el objetivo era optimizar el cálculo. Uno de los hitos de la IA se produce cuando el «Deep blue» de IBM logra vencer al entonces campeón mundial de ajedrez, Garri Kasparov [15]. Se constata entonces la aparición de los sistemas expertos [16], en los que la tecnología trata de asistir

prueba que los usuarios deben realizar para demostrar que son humanos al acceder a determinados contenidos en internet.

[12] Es ya considerada una referencia clásica su manual, cuya primera edición es de 1995. Rusell, S. J., /Norvig, P., *Artificial intelligence: a modern approach*, Prentice Hall, Nueva Jersey, 1995.

[13] Gómez Colomer, J. L., *El juez-robot (La independencia judicial en peligro)*, Tirant Lo Blanch, Valencia, 2023 p. 25.

[14] *Ibid*, p. 29.

[15] Habría que esperar unos años más, concretamente hasta 2016, para que una máquina venciera al campeón mundial de Go, el surcoreano Lee Sedol. Así lo narra Greco, L., *Poder de julgar sem responsabilidade de julgador*, Marcial Pons, São Paulo, 2020, pp. 12, 23 y 34.

[16] Según la definición funcional (basada en el componente humano) que se ofrece por Amador Hidalgo, partiendo del concepto de Feigenbaum, sistema experto es «un programa inteligente de ordenador que utiliza conocimiento y procedimientos inferenciales para solventar problemas lo suficientemente difíciles que exigen la competencia de un experto humano en su resolución». Amador Hidalgo, L., *Inteligencia artificial y sistemas expertos*, Universidad de Córdoba Servicio de Publicaciones, Córdoba, 1996, p. 138.

a la toma de decisiones en contextos complejos mediante la aplicación de reglas lógicas.

A partir de este momento, se comienza a aplicar esta tecnología a diversos campos, donde la predicción y la toma de decisiones más eficientes que la del propio ser humano se convierten en una constante. Es aquí donde, desde el punto de vista jurídico, surgen las primeras preocupaciones relativas al papel del humano y la máquina en la toma de decisiones que afectan a los derechos del ciudadano, pues se empezaron a utilizar estos primigenios sistemas de IA para tomar decisiones con consecuencias jurídicas, tales como la elección sobre la libertad de una persona [17]. Se agudizan también los debates de corte filosófico sobre el papel del humano y su posible sustitución por la máquina en escenarios futuribles [18].

Sin embargo, el paradigma de la IA aún estaba por cambiar, pues, tras unos años de desánimo en la investigación de corte neuro-computacional [19] (es decir, basada en el estudio de las estructuras biológicas del cerebro), comienza el auge de una técnica que cambiará el rumbo y el concepto de la IA: las redes neuronales.

Las redes neuronales imitan el cerebro humano y se componen de muchas «neuronas», que no son más que estados numéricos conectados a través de «enlaces» que sirven de canales de comunicación. Estas redes no solo son una forma de representar de forma compleja el procesamiento y la transmisión de información (datos), sino que además son sistemas adaptativos, con capacidad para ajustar su estructura interna a partir de información nueva, cambiando sus pesos (de una forma similar a lo que ocurre en nuestros procesos de sinapsis), lo que permite afirmar su capacidad de autoaprendizaje [20]. Se produce así,

---

[17] Se produce, así, el surgimiento de los *Risk Assessment Tools*. La literatura en lo relativo a este tipo de herramientas resulta inabarcable, por lo que, a título simplemente ejemplificativo, véase, por todos: Martínez Garay, L./ Montes Suay, F., «El uso de valoraciones del riesgo de violencia en Derecho Penal: algunas cautelas necesarias», *InDret*, 2/2018, pp. 1-47.

[18] Ejemplos: Bostrom, N., *Die Zukunft der Menschheit: Aufsätze (suhrkamp taschenbuch wissenschaft*), trad. Jan-Erik Strasser, Suhrkamp Verlag, Frankfurt a. M., 2018; Weizenbaum, J., *Computer power and human reason: From judgment to calculation,* W. H. Freeman & Co, 1976. También se enmarcan en esta corriente de IA simbólica las preocupaciones en torno al paradigma de «juez robot». La monografía de Greco versa, precisamente, sobre este paradigma, sin ir más lejos. Greco, L., *op. cit.* N. 15.

[19] Una de las causas de este parón en la investigación se debe, precisamente, a la crítica feroz que contra esta nueva técnica vertió el mismo Marvin Minsky. De hecho, en 1969, Minsky y Seymour Papert publican «*Perceptrons: An Introduction to Computational Geometry (Perceptrones: una introducción a la geometría computacional*)», trabajo en el que analizan de manera crítica las limitaciones de las redes neuronales de una sola capa. A menudo se culpa a este trabajo de reducir el interés por las redes neuronales. En la edición de 1988 argumentan que el progreso ya se había estancado. *Vid.:* Mucci, T., «La historia de la IA», IBM blog, 21 de octubre de 2024, disponible en: <https://www.ibm.com/es-es/think/topics/history-of-artificial-intelligence> [Accedido por última vez el 20/01/2026].

[20] Banteka, N., «Artificially intelligent persons», *Houston Law Review*, 58(3), 2021, p. 546.

con la superposición de varias «capas» de neuronas, lo que se conoce como aprendizaje profundo *(«deep learning»* [21]*),* dando lugar a un cambio de paradigma en la inteligencia artificial.

Desde el instante en que las redes neuronales empiezan a utilizarse, el paradigma y la noción de la inteligencia artificial se transforman, hasta el punto de que algunos pensadores sostienen que solo a partir de entonces puede hablarse de auténtica inteligencia artificial en un sentido estricto [22].

En definitiva, que los sistemas de IA sean capaces de aprender (el conocido rasgo del «*machine learning*») implica un cambio de paradigma, y las técnicas de desarrollo se multiplican al albur de este nuevo modo de entender la IA, no ya como un sistema determinista capaz de realizar cálculos y predicciones de la manera más óptima posible, sino como sistemas complejos, dotados de alguna forma de autonomía, donde ya no hay una lógica determinista detrás [23], sino una auténtica toma de decisiones en cada una de sus ramas de aplicación.

El aprendizaje autónomo o «*deep learning*», como explica Hao, es el proceso que impulsa muchos de los servicios que utilizamos hoy en día, desde sistemas de recomendación como los de Netflix, YouTube y Spotify, hasta mo-

---

[21] Para una definición más accesible de este y otros términos relacionados, *vid.:* Hao, K., «What is machine learning?» MIT Technology Review, 2022, disponible en: <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/> [Accedido por última vez el 20/01/2026].

[22] Turner, J. *op. cit.* N. 5, p. 19. El autor habla de que sólo en este momento son significativos (*«salient»*) para el sistema legal.

[23] Se ha de aclarar, en este punto, que cuando nos referimos al determinismo de los sistemas de IA lo hacemos en un sentido computacional, por ser esta la palabra que en la literatura técnica se usa para referirse a los sistemas que, dadas unas entradas específicas y un conjunto de reglas fijas, dan siempre un mismo resultado, sin lugar para la variabilidad. Es el caso, por ejemplo, de un algoritmo clásico básico basado en reglas lógicas que siempre produce la misma salida para una entrada dada. Se oponen estos modelos deterministas en sentido técnico a los que no garantizan una salida fija para la misma entrada pues, o bien se modifican con el aprendizaje o bien pueden incorporar elementos como la aleatoriedad. Estos últimos, si bien están regidos por reglas matemáticas, no tienen un comportamiento totalmente predecible. Este concepto técnico de determinismo no prejuzga, empero, el concepto filosófico, que define el determinismo como la tesis ontológica según la cual todo estado del mundo está causalmente determinado por estados anteriores, sin que exista un margen para el libre albedrío, siguiendo todo proceso una cadena causal ineludible. El determinismo filosófico no depende de que el sistema sea predecible para nosotros como observadores, pues lo importante es que todo ocurre necesariamente como consecuencia de causas anteriores. Por tanto, si tenemos en cuenta ambos conceptos, aunque un sistema de IA como una red neuronal profunda puede ser no determinista desde el punto de vista técnico, puede seguir siendo determinista en el sentido filosófico, pues que algo sea impredecible no implica que sea libre o indeterminado ontológicamente. *Vid.,* para las dos acepciones aquí explicadas: Hoefer, C., «Causal Determinism», *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), disponible en: <https://plato.stanford.edu/entries/determinism-causal/> [Accedido por última vez el 20/01/2026]; «*Determinism in AI: Navigating Predictability and Flexibility*», Jaxon, 2024, Disponible en: <https://jaxon.ai/determinism-in-ai-navigating-predictability-and-flexibility> [Accedido por última vez el 20/01/2026].

tores de búsqueda como Google y Baidu. También redes sociales como Facebook y Twitter y asistentes de voz como Siri y Alexa [24]. Pero también aplicaciones más jurídicamente polémicas como son los vehículos autónomos, las armas autónomas, o las cada vez más generalizadas operaciones de tecnología financiera. Por supuesto, todos los avances en procesamiento de imágenes, voz y texto, utilizados con fines recreativos y de vigilancia, son consecuencia de la aplicación del paradigma del aprendizaje autónomo.

La IA de hoy, por tanto, no es la IA de Dartmouth, y no será, desde luego, la IA del futuro. Debemos tenerlo en cuenta para tratar de acercarnos a su conceptualización.

## 2. POSIBLES CLASIFICACIONES

De cara a clarificar algunos de los conceptos arriba indicados, se procederá a realizar una triple clasificación de la IA en función de tres parámetros, todo ello con la finalidad de apreciar los detalles que pueden ser relevantes al encontrar una definición que nos sea adecuada.

Así, podemos conceptualizar a la inteligencia artificial en función de la técnica de la que se vale, en función de las competencias que desarrolla, o en función de su mayor o menor amplitud. Por simplificar, nos importa cómo opera la IA, para qué opera, y con qué alcance. De esta forma, se hace un esfuerzo por aglutinar las clasificaciones que manejan diferentes autores y que, consideramos, pueden aportar algo de claridad a esta compleja disciplina. Finalmente, se incluye una cuarta clasificación, esta vez no de carácter dicotómico, sino gradual, donde se analizará la importancia de la intervención del ser humano.

### 2.1 **La IA en función de la técnica empleada: sistemas expertos *vs.* aprendizaje automático**

Aunque no se profundizará en las especificidades técnicas de la tecnología, resulta necesario esclarecer que el enfrentamiento de paradigmas apuntado en el apartado primero tiene unas consecuencias importantes a la hora de analizar el fenómeno de la IA. Por ello, resulta justificado que un trabajo centrado principalmente en el aspecto jurídico de la IA incluya alguna referencia a las técnicas que esta utiliza.

---

[24] Hao, K., *op cit*. N. 21.

Las denominaciones que utiliza la doctrina especializada, también al referirse a estos aspectos técnicos, es diversa. De Siles [25], utilizando una conceptualización apegada al riesgo, se refiere a la diferencia entre IA dinámica y estática, es decir, que el sistema pueda modificarse en tiempo real. Pone como ejemplo de IA estática el sistema de predicción de la violencia SAVRY [26] que no puede cambiar en tiempo real los factores con los que trabaja. Sin embargo, un modelo dinámico es mutable en dos aspectos: aprende por medio de mecanismos que no están programados explícitamente en el código, y los datos en que se basa el modelo pueden cambiar. El ejemplo de estos tipos de IA son los sistemas de *machine learning*.

Justifica esta autora la distinción en cuanto a las consecuencias que esto puede tener de cara a la regulación, pues, como principio general, cuanto mayor es la mutabilidad de un sistema, mayor es el riesgo de daños que podrían derivarse de su uso. Así, el uso de IA dinámica presenta un mayor potencial de generar daños.

También Turner [27] pone el énfasis en este aspecto, si bien utilizando otros términos, pues nos habla de IA basada en reglas e IA basada en principios. Así, distingue el paradigma de la IA simbólica, a la que vincula con la idea de «*Good Old Fashion AI*», que identificaría con la aplicación de reglas, donde el contenido es cerrado y tiene forma de silogismo lógico; del paradigma de la IA más avanzada, basada en técnicas de aprendizaje o *machine learning*, que equipara a la idea de los principios, pues aquí la conclusión a la que se llega no es determinista, sino que puede y debe variar en función de las circunstancias. Contrapone dos técnicas, cada una de las cuales se identifica con un paradigma; los árboles de decisión siendo los «procesos deterministas deductivos de tipo lógico» [28], frente al modelo de la IA no determinista, en las que el propio

[25] De Siles, E. L. «AI, on the Law of the Elephant: Toward Understanding Artificial Intelligence», *Buffalo Law Review*, 69(5), 2021, p. 1424 y ss.

[26] SAVRY es el acrónimo utilizado para referirse a *Structured Assessment of Violence Risk in Youth* (Evaluación estructurada del riesgo de violencia en los jóvenes). Más información sobre esta herramienta de valoración del riesgo en: Borum, R., Lodewijks, H., Bartel, P. A.,/ Forth, A. E., «Structured Assessment of Violence Risk in Youth (SAVRY)», en R. K. Otto/ K. S. Douglas (Eds.), *Handbook of violence risk assessment*, Routledge/Taylor & Francis Group, 2010, pp. 63-79.

[27] Turner, J., *op cit*. N. 5, p. 18.

[28] Los árboles de decisión son una técnica de minería de datos para sacar la información oculta en ellos, abordando la solución a problemas de predicción, clasificación y segmentación. Son una forma gráfica y analítica de representar todos los eventos que pueden surgir a partir de una decisión asumida, con el objetivo de tomar la decisión más «acertada», desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Los resultados visuales ayudan a buscar subgrupos específicos y relaciones que tal vez no encontraríamos con estadísticos más tradicionales. Definición extraída de: Berlanga Silvente, V., Rubio Hurtado, M. J., Vilà Baños, R., «Cómo aplicar árboles de decisión en SPSS», *REIRE Revista d'Innovació i Recerca en Educació*, 6(2), 1, 2013, pp. 65-66.

sistema va cambiando el valor de sus inputs a través de procesos «internos», con el ejemplo de la red neuronal como exponente [29].

## 2.2 **La IA en función de las competencias: inteligencia *vs.* racionalidad**

Aunque esta segunda clasificación puede encontrar elementos en común con la señalada anteriormente, merece la pena traerla a colación, pues con ella apreciamos la importancia de la función asociada al sistema. Una cosa es saber cómo funciona y evoluciona la IA, y otra es saber cuáles son los fines o funciones encomendadas a ella.

Al realizar el repaso histórico sobre el nacimiento de la disciplina encontrábamos, ya, dos posturas enfrentadas. Turner contrapone, en este caso, el ideal humanizado («*human-centred*»), con el ideal racionalista.

Así, mientras Turing encarnaría el primero de los ideales, buscando una máquina capaz de imitar lo mejor posible las capacidades humanas, para afirmar su inteligencia, el paradigma de la racionalidad lo situaríamos en Rusell y Norvig que, criticando esa concepción reduccionista de la inteligencia, propusieron una definición basada en las funciones desempeñadas por los sistemas.

De esta forma, continuando con la idea de Turner, lo determinante es atender al fin o meta que persigue el sistema. Se torna así del paradigma de la inteligencia como cualidad humana, a un paradigma «racionalista», donde se debe citar a uno de los padres de la inteligencia artificial: Nils J. Nilsson. Este autor concibe la inteligencia como «*aquella cualidad que permite a una entidad funcionar adecuadamente y con previsión en su entorno*» [30].

El paso de la inteligencia como cualidad humana a la inteligencia como desempeño de funciones permite, en cierta medida, que la definición de la IA sea menos antropocéntrica, y nos lleva a una definición tan aséptica como la que propone Turner: *«La inteligencia artificial es la capacidad de una entidad no natural de tomar decisiones mediante un proceso evaluativo»* [31].

---

[29] El propio Turner reconoce que la distinción no puede ser tajante en exceso, pues en muchos sistemas conviven ambas técnicas. Asimismo, reconoce que los términos por él empleados no son los únicos posibles, y se remite a un trabajo de Karnow donde, hablando de la distinción entre sistemas expertos y «fluidos», propone una clasificación similar, planteando que los últimos deberían tener un tratamiento legal distinto debido a su impredecibilidad. *Vid.*: Karnow, C. E., *op. cit*. N. 177.

[30] Nilsson, N. J., *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge, UK, Cambridge University Press, 2010, Preface.

[31] «*Artificial Intelligence Is the Ability of a Non-natural Entity to Make Choices by an Evaluative* Process», Turner, J., *op cit*. N. 5, p. 16.

En función de si se escoge un enfoque basado en el paradigma de la inteligencia o en el paradigma de la racionalidad, los problemas (y posibles soluciones) asociados a estas tecnologías serán diferentes. Así, las definiciones más apegadas al primer grupo tienden a poner el foco en el riesgo de que el humano se vea sustituido por la IA (recuérdese el auge de las preocupaciones del paradigma del «juez-robot»), mientras que aquellas definiciones que parten de conceptos racionales tienen a concebir la IA como un producto (con potencialidad dañina, ciertamente), sobre el que aplicar soluciones más familiares [32].

### 2.3 **La IA en función de su alcance: IA débil *vs.* IA fuerte**

Otra de las obligadas referencias en la clasificación de la IA tiene que ver con su alcance. Así, distinguimos al menos dos (o incluso tres, si nos colocamos en un escenario de tinte más futurista), posibles paradigmas de la inteligencia artificial.

Así, Kaplan y Haeinlein nos hablan de tres escenarios [33]:

La IA estrecha o débil («*narrow AI*»), que se aplicaría para tareas muy específicas, siendo incapaz de resolver problemas de forma autónoma en otras áreas, pero que puede igualar o incluso mejorar las capacidades de un humano en determinadas áreas. Según aclaran los autores, las aplicaciones de IA de esta «primera generación» son las más presentes, permitiendo el reconocimiento facial en fotografías en Facebook, el reconocimiento de voz de asistentes virtuales, o que Tesla tenga modelos de coches autopilotados [34].

La IA general o fuerte («*general AI*»), que es capaz de resolver problemas de forma autónoma en diferentes áreas, igualando o superando al ser humano en varias áreas de experiencia. Los autores hablan de que en el futuro quizá podamos asistir a la presencia de este tipo de sistemas, capaces de razonar, planificar y resolver problemas de forma autónoma para tareas para las que no fueron diseñados inicialmente [35].

---

[32] Esta influencia se examinará con más detalle en el apartado 3 del capítulo primero, donde se estudiarán las definiciones que los instrumentos legales proponen a día de hoy.

[33] Kaplan, A., / Haenlein, M. *op cit.* N. 9, p. 16.

[34] *Idem.*

[35] El concepto de IA general se ha asociado, tradicionalmente, a un escenario mucho más avanzado entre la doctrina, negando la posibilidad de que los sistemas sean capaces de desempeñar tareas variadas o generales. Así, Del Rosal Blasco, B., «¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la inteligencia artificial?», *Revista electrónica de responsabili-*

Finalmente, en un escenario no actual, en el último de estos tres pasos en la evolución de la IA se situaría la llamada IA superinteligente, nivel en el que estaríamos ante máquinas situadas por encima del ser humano en cualquier tarea, con autoconsciencia y capacidad de resolver problemas en otras áreas de forma instantánea. Estos sistemas podrían aplicar la IA a cualquier área y ser portadores de creatividad científica, habilidades sociales y sabiduría general, razón por la que algunos la han bautizado como la verdadera inteligencia artificial. La posibilidad de que esta realidad se dé ha motivado reflexiones enmarcadas en la filosofía «transhumanista», reflexionando sobre cómo convivir con esta nueva realidad que vendrá a cohabitar con nuestra especie [36].

Son numerosos los autores que han acudido a esta clasificación [37], respecto de la cual hubo poco que añadir en las fases iniciales de negociación del Reglamento Europeo, hasta que el legislador de la Unión decidió incorporar, en las últimas versiones del Reglamento de Inteligencia Artificial, el concepto de «inteligencia artificial de propósito general» entre los sistemas y categorías objeto de regulación. No es de extrañar que, habida cuenta del impacto reciente de las aplicaciones de IA generativa más actuales [38], las instituciones europeas hayan optado por introducir dicho concepto.

---

*dad penal de personas jurídicas y compliance*, 2, 2023, p. 9. Este autor pone en contraste la IA estrecha o débil con la general o fuerte, siendo la primera aquella con un rango limitado de habilidades, mientras la segunda es capaz de «resolver cualquier problema», negando Del Rosal que esto sea un escenario real en el presente. Si bien el escenario en que un sistema pueda resolver cualquier problema queda, evidentemente, lejos, lo cierto es que lejos queda también el paradigma en que las IAs sean entrenadas para desarrollar solo un tipo de función, siendo cada vez más habituales aquellos sistemas que desempeñan tareas muy variadas a modo de «asistentes» complejos. No en vano, el Reglamento de Inteligencia artificial contempla en su artículo 3, apartado 63: *««modelo de IA de uso general»: un modelo de IA, también uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala,* ***que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas****, […]»* (La negrita es nuestra). Lo que hace apenas un par de años constituía un campo de aplicación pequeño (los sistemas capaces de realizar una «gran variedad de tareas distintas»), supone hoy en día uno de los más prometedores sectores ya no de investigación, sino de comercialización, siendo la IA generativa capaz de cumplir con varias tareas la tecnología más utilizada, probablemente, en el día a día.

[36] Para una visión sobre la historia del transhumanismo de la mano de uno de los autores más influyentes: Bostrom, N., «A history of transhumanist thought», *Journal of evolution and technology*, 14(1), 2005. Para una defensa del ideal humanista, poniendo en contraste el humanismo, el transhumanismo y el posthumanismo, Pérez Luño, A. E., «El posthumanismo no es humanismo», *Derechos y libertades*, 44, Época II, 2021, pp. 17-40. Desde un punto de vista feminista y transhumanista, son esenciales los escritos de Donna Haraway. Paradigmáticamente, Haraway, D. J., *Manifiesto Ciborg*, Kaótica Libros, 2020.

[37] Sin ánimo de exahustividad, y por todos: Escott, E., «What Are the 3 Types of AI? A Guide to Narrow, General and Super Artificial Intelligence», *Codebots*, 24 de octubre de 2017, disponible en: <https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible> [Accedido por última vez el 20/01/2026].

[38] La regulación comenzó a abordar explícitamente los modelos de IA de uso general en respuesta al auge de los modelos fundacionales como los *large language models*. Estas disposiciones fueron incluidas en el acuerdo provisional alcanzado entre el Consejo y el Parlamento Europeo en diciembre de 2023, donde se añadieron normas específicas para estos sistemas. Esta información es accesible en el comunica-

Por primera vez, por tanto, se da un paso más al reconocer que es posible que, en el horizonte cercano, haya algún sistema de IA «que muestre una generalidad significativa y sea capaz de realizar de forma competente una amplia gama de tareas distintas» [39]. No significa ello, desde luego, que el camino hacia la generalidad esté abierto sin remedio, pero sí nos debe obligar a llamar la atención, una vez más, sobre la fluidez de estas clasificaciones y sobre la rápida evolución de esta realidad.

### 2.4 La IA en función de la mayor o menor intervención humana. Del paradigma del programador al paradigma del demiurgo

Para cerrar el apartado dedicado a la clasificación, resulta extremadamente útil aludir al papel que desempeña en este concepto el que va a ser protagonista absoluto de la presente obra: el humano. Así, resultará esencial para terminar de formarnos la foto de la inteligencia artificial saber qué papel tiene el hacedor, otrora de sirvientas autómatas de oro, hoy de complejos programas informáticos que cada vez se escapan más de su control.

Resulta útil, a estos efectos, atender a la imagen que nos ofrece el filósofo alemán Andreas Matthias [40], pues nos explica cómo se produce este proceso gradual de lo que él llama el paso del «codificador» al «creador».

Así, el primer papel que desempeña el humano es el propio de un mero programador, que usa lenguajes de programación, formales, orientados a la lógica, donde se producen deducciones a partir de reglas de inferencia. Aunque el flujo de estos sistemas puede ser algo complejo de escribir (y lo será más según aumente el número de datos utilizados para alimentar al sistema), lo cierto es que siempre se podrá verificar cuál es la información de la que se nutre, por lo que el control por parte del programador es posible.

---

do de prensa del Consejo de la Unión Europea tras los diferentes periodos de negociación previos a la aprobación del texto definitivo, accesibles en: https://www.consilium.europa.eu/es/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/ [Accedido por úlitma vez el 20/01/2026]. Para muestra del potencial de esta «nueva» forma de concebir la IA, el informe realizado en junio de 2023 por la consultora McKinsey resulta revelador: «The economic potential of generative AI: The next productivity frontir», 14 de junio de 2023, disponible en: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction> [Accedido por última vez el 20/01/2026].

[39] Artículo 3, apartado 63 del Reglamento, *op cit*. N 35.

[40] Matthias, A., «The Responsibility Gap. Ascribing Responsibility for the Actions of Learning Automata», *Ethics and Information Technology,* 4, 2004, p. 182: «*Thus, we can identify a process in which the designer of a machine increasingly loses control over it, and gradually transfers this control to the machine itself. In a steady progression the programmer role changes from coder to creator of software organisms*.»

En un estadio algo más avanzado de la tecnología, donde aparecen ya las redes neuronales, la representación simbólica de la información desaparece y en lugar de símbolos tendremos «pesos sinápticos», que es conocimiento almacenado en una red pero que sólo se puede conocer o inferir a partir de la experimentación, por lo que no es posible conocer todo el proceso que se lleva a cabo para producir los *outputs* o resultados.

Según se van complejizando las técnicas (Matthias habla de *reinforcement learning* y de métodos de programación de tipo genético, estos últimos generados por máquinas en lugar de por humanos), la posibilidad de control por parte del «creador» es cada vez más remota. En el caso de las redes neuronales, el humano diseña los parámetros del sistema (algo así como su arquitectura, el número de capas de información y salida de la información, o su interpretación), pero pierde el control sobre el proceso interno de procesamiento de la información. En el caso del *reinforcement learning*, incluso, la información almacenada en el sistema cambia de forma continua, y a velocidades vertiginosas, que escapan del todo al escrutinio humano.

Puede tener sentido aquí destacar que el protagonismo del humano es inversamente proporcional a la capacidad de aprendizaje del sistema. De Siles[41] llama la atención sobre la utilidad de considerar, asimismo, el papel que desempeña el humano en función de los modelos de aprendizaje elegidos, pues las posibilidades de aprendizaje reflejan el grado de intervención del humano al entrenar estos sistemas.

Así, se diferencian dos grandes modelos: en el aprendizaje supervisado el programador le expone al sistema conjuntos de datos de entrada (o variables) y datos de salida (o resultados), generalmente ya interpretados y etiquetados, es decir, con una clasificación previa. En palabras de Siles: «*Los entrenadores humanos instruyen al aprendiz diciendo: «si la entrada es A, entonces el resultado es B». A través de este proceso intensivo, el sistema, mediante su lógica algorítmica y estadísticamente fundamentada, «aprende» la función de mapeo por la cual, para un conjunto dado de datos de entrada y características dentro de esos datos, se debe devolver su correspondiente salida*.» [42]

Es decir, podríamos decir que el humano tiene un papel de «instructor activo». Sin embargo, en los modelos de aprendizaje no supervisado, son los propios sistemas los que derivan computacionalmente, esto es, aprenden o encuentran los modelos estadísticos óptimos para lograr el objetivo propuesto, simplemente mediante su exposición a grandes cantidades de datos no catego-

[41] De Siles, E. L., *op cit.* N. 25, p. 1435. Para el autor, de hecho, los modelos de aprendizaje automático supervisados y no supervisados se pueden concebir como ubicados en extremos opuestos de un espectro que refleja los grados de participación humana, si la hay, en el entrenamiento de estos sistemas.

[42] *Idem*., p. 1432 (traducción propia).

rizados. El humano supervisa este procedimiento, sin poder intervenir demasiado en el propio proceso de generación de conocimiento.

En definitiva, esta transformación se produce de una forma paulatina pero imparable, de tal forma que el papel y el protagonismo del humano cada vez queda más diluido y es menos central, produciéndose lo que algunos autores consideran un «alejamiento» o distancia [43] entre la actuación inicial del humano y el output que finalmente se produce.

## 3. LLEGANDO A PUERTO: LOS INSTRUMENTOS LEGISLATIVOS Y LA DEFINICIÓN QUE DEBE IMPORTAR AL DERECHO (PENAL)

### 3.1 **La justificación del concepto**

Una vez realizadas todas las conceptualizaciones previas, llegamos al culmen del intento definitorio, que debe centrar todos los argumentos expuestos para generar algo parecido a una definición operativa.

Y esta operatividad no puede medirse de una forma genérica. Debe ser una operatividad centrada en lo que aquí nos interesa, que no es otra cuestión que los problemas de responsabilidad que se pueden derivar del funcionamiento problemático de estas tecnologías. Así, en definitiva, buscamos una definición que se ajuste a las necesidades jurídicas.

Como acertadamente destaca De Siles al hablarnos de la parábola budista del elefante [44], ser capaces de aportar una definición concisa de un concepto tan

---

[43] Danaher, J., «Robots, Law and Retribution Gap», *Ethics and Information Technology*, 4, 2016, p. 300, traducción propia: «la estipulación de que los robots tienen un alto grado de autonomía es relevante […] porque amenaza con *romper el vínculo* entre los creadores y diseñadores humanos de tales sistemas y sus efectos». (traducción propia, la cursiva es nuestra). También Floridi *et al,* citando a McAllister: *«Los robots autónomos tienen una capacidad única de separarse del acto criminal, donde el humano manifiesta el mens rea y el robot (o agente autónomo) comete el actus reus»* (traducción propia*). Vid.:* Floridi, L./King, T. C./Aggarwal, N./Taddeo, M., «Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions», *Science and Engineering Ethics*, 26, 2020, p. 95. La cita es a: McAllister, A., «Stranger than science fiction: The rise of AI interrogation in the dawn of autonomous robots and the need for an additional protocol to the UN convention against torture», *Minnesota Law Review*, 101, 2017, pp. 2527-2573. Asimismo, hablando de que la actuación de la IA «se encuentra cada vez más alejada del poder de decisión de la persona física»; Benítez Ortuzar, I. F., «Responsabilidad por el hecho cometido por el robot dotado de inteligencia artificial. ¿Existe un paralelismo con la responsabilidad de la persona jurídica? Breves reflexiones», en AA. VV., *Estudios político criminales, jurídicos penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés,* Valencia, Tirant Lo Blanch, 2023, 2.ª ed., p. 512.

[44] De Siles, E. L., *op cit*. N. 25, p. 1395. La parábola de los ciegos y el elefante se encuentra en textos budistas clásicos como el «*Udana»,* parte del Canon Pali, específicamente en el capítulo 6, verso 4. Este texto es una colección de discursos atribuidos al Buda y forma parte de las escrituras *Theravād*a. En esta parábola, varios ciegos tocan diferentes partes de un elefante (colmillo, trompa, cola, etc.) y cada uno describe al elefante de manera distinta según la parte que perciben. La historia ilustra cómo las personas

amplio y cambiante corre el riesgo de que actuemos como lo hicieron los sabios ciegos de la leyenda al tratar de definir un elefante simplemente tocando una de las partes de su cuerpo. Desde luego, no será la misma la definición que nos dé quien tocó la robusta pata, que quien palpó el afilado diente de marfil, como tampoco será la misma la definición que otorguen un ingeniero, un economista o un jurista cuándo les preguntemos sobre este elefante inabarcable que es la IA.

No se tratará empero, en este trabajo, de ser tan sabios como los ciegos de la parábola, pues entendemos que nuestra definición debe ajustarse a nuestras intenciones. No por ello la definición habrá de ser inexacta, pues todo conocedor de la disciplina deberá ver reflejada al menos una parte significativa de la realidad. Sin embargo, no deberá ser tampoco ni demasiado exhaustiva ni excesivamente técnica. Deberá ser, por hablar en términos de penalista, lo que nos exige nuestro principio de legalidad [45].

No me propongo, tampoco, encontrar una definición propia. Son muchos los autores que han intentado definir, con mucho acierto, la inteligencia artificial [46], y muchos de ellos coinciden en adoptar un concepto común como punto de partida. Este concepto no es otro que el contenido en la definición realizada por el Grupo de Expertos de Alto Nivel de la Unión Europea.

El Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (en adelante, HLEG, por sus siglas en inglés) fue nombrado por la Comisión Europea en 2018 con el objetivo de apoyar la aplicación de la Estrategia Europea de Inteligencia Artificial, elaborando recomendaciones sobre el desarrollo de futuras políticas y sobre cuestiones éticas, jurídicas y sociales relacionadas con la IA. Está compuesto por cincuenta y dos representantes del mundo académi-

---

pueden interpretar la realidad de manera parcial y limitada debido a sus perspectivas individuales. Referencias a esta parábola pueden encontrarse en, *The Udāna & The Itivuttaka: Inspired Utterances of the Buddha & The Buddha's sayings*, traducida por John D. Ireland, Pariyatti Publishing, 2018, p. 78.

[45] A estos efectos, y para hacer un repaso sobre los problemas que plantean las manifestaciones del principio de legalidad en la exhaustividad de las definiciones jurídicas; *vid.*, Villaverde Menéndez, I., «Principio de taxatividad. Una reflexión jurisprudencial», en Pérez Manzano, M., Lascuraín Sánchez, J. A., /Mínguez Rosique, M., La tutela multinivel del principio de legalidad penal, Marcial Pons, 2016, pp. 81-116. Así, en p. 105, aclarando que: «*El principio de taxatividad por su parte exige la predeterminación clara y precisa de los tipos sancionadores, solo así el individuo puede predecir razonablemente las consecuencias de su conducta»*. Asimismo, Huerta Tocildo, S., «El derecho fundamental a la legalidad penal», *Revista Española de derecho constitucional,* año 13, n.º 39, 1993. p. 108 y ss. Un análisis de la jurisprudencia constitucional en lo relativo a este mandato en Lascuraín Sánchez, J. A., *Sólo penas legales, precisas y previas. El derecho a la legalidad penal en la jurisprudencia constitucional*, Thomson Reuters Aranzadi, 2009.

[46] Así, en el panorama español, Javier Valls. Valls Prieto, J., *Inteligencia artificial, Derechos humanos y bienes jurídicos*, Thomson Reuters Aranzadi, Cizur Menor, Navarra, 2021, capítulo 1. También se muestra cauto Gómez Colomer, reconociendo la falta de consenso sobre un concepto único, pero realizando un esfuerzo por explicar los conceptos básicos que componen las distintas definiciones. Gómez Colomer, J. L, *op. cit*. N 13, pp. 35 y ss. Asimismo, Del Rosal, reconociendo la labor previa de los autores aquí citados, optando finalmente por acoger la propuesta por el HLEG. Del Rosal Blasco, B., *op cit*. N. 35, p. 8, nota 17.

co, la sociedad civil y la industria. Los dos primeros resultados del HLEG fueron las Directrices éticas para una inteligencia artificial digna de confianza y la definición de IA que aquí se presenta como concepto, elaborada como documento de apoyo para los resultados del HLEG [47].

Es este el concepto escogido, también, como punto de partida, por voces autorizadas de la academia [48], pero también por instituciones como la propia Comisión Europea. Así, en su valioso documento de 2020 «*AI Watch Defining Artificial Intelligence*» [49], un grupo de expertos creado precisamente para la búsqueda de un concepto de IA, eligió el concepto propuesto por el HLEG, tras la revisión de 65 documentos internacionales pertinentes que abarcaban informes políticos e institucionales sobre IA (incluidos esfuerzos de normalización, estrategias nacionales e informes de organizaciones internacionales), publicaciones de investigación e informes de mercado.

Se afirma que este concepto, aunque pueda considerarse muy técnico para distintos públicos y objetivos, es una definición muy completa que incorpora los aspectos de percepción, comprensión, interpretación, interacción, toma de decisiones, adaptación al comportamiento y consecución de objetivos, mientras que otras definiciones no los abordan en su totalidad [50].

## 3.2 **Nuestro concepto de IA: desglosando los elementos del concepto propuesto por el Grupo de Expertos de Alto Nivel**

El HLEG definió en 2018 los sistemas de IA como:

«*Sistemas de software (y posiblemente también de hardware) diseñados por humanos que, dado un objetivo complejo, actúan en una dimensión física*

---

[47] «Ethics guidelines for trustworthy AI», presentadas el 8 de abril de 2019, disponible en: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Accedido por última vez el 20/01/2026]. «A definition of AI: main capabilities and scientific disciplines. Definition developed for the purpose of the deliverables of the High-Level Expert Group on AI», 18 de diciembre de 2018, disponible en: <https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf> [Accedido por última vez el 20/01/2026].

[48] Javier Valls, por ejemplo, se muestra partidario de este concepto por el rendimiento que presenta en cuanto a su posibilidad de proteger derechos fundamentales en peligro con dichas tecnologías. *Vid.*; Valls Prieto, J., *Inteligencia artificial…, op cit*. N.46, y mismo autor: «Sobre la responsabilidad penal por la utilización de sistemas inteligentes» *Revista electrónica de ciencia penal y criminología*, 24-27, 2022, pp. 1-35.

[49] Samoili, S., Lopez Cobo, M., Gomez Gutierrez, E., De Prato, G., Martinez-Plumed, F./ Delipetrev, B., «AI WATCH. Defining Artificial Intelligence», Publications Office of the European Union, 2020, disponible en: <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163> [Accedido por última vez el 20/01/2026].

[50] *Ibid.,* p. 5.

*o digital percibiendo su entorno a través de la obtención de datos, interpretando los datos estructurados o no estructurados recabados, razonando sobre el conocimiento o procesando la información derivada de dichos datos y decidiendo las mejores acciones a tomar para conseguir el objetivo*» [51].

Es pertinente, para terminar esta sección, detallar cuáles son los elementos destacados, y finalizar haciendo referencia a algunos términos conexos que aparecen en el segundo apartado de la definición, quizá menos explorado, que nos servirá para aclarar algunas confusiones terminológicas.

*Sistema.*

Según el diccionario de la Real Academia Española, un sistema es un «conjunto de reglas o principios sobre una materia racionalmente enlazados entre sí» [52]. Cabe resaltar dos aspectos de este término. En primer lugar, al hablar de sistema, se está utilizando un término esencialmente neutro, lo que aleja el concepto de tendencias humanizadoras. En segundo lugar, este término, que nos habla en su acepción más usada de «reglas entrelazadas entre sí», nos puede evocar una relación estrecha con otro de los términos que mayor protagonismo tienen en la literatura sobre IA, como es el de «algoritmo». Una vez más, acudimos al Diccionario de la Real Academia de la Lengua, que nos devuelve la siguiente definición. Un algoritmo es un «conjunto ordenado y finito de operaciones que permite hallar la solución de un problema». Si bien es cierto que esta es una definición simplificada, nos puede servir para entender lo que está en la base del concepto.

*Objetivos definidos por/diseñados por humanos.*

Nos encontramos aquí con uno de los elementos fundamentales para el entendimiento de la IA, que, como se ha resaltado anteriormente, supone poner el énfasis en el enfoque antropomorfizado que estaba en el origen de esta

[51] «*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions*».

[52] Primera acepción del Diccionario de la Real Academia de la Lengua Española para el vocablo «sistema».

tecnología. El sistema de IA debe tener unos objetivos definidos por un humano, y debe ser diseñado por él. Ya se ha visto cómo el papel del humano ha ido variando considerablemente desde los orígenes de la disciplina, pero que aún tiene importancia la intervención del ser humano se hace patente con este elemento en la definición.

De hecho, algunas de las primeras preocupaciones de la comunidad jurídica que se enfrentó a la creciente automatización de los sistemas, fue crear algún tipo de clasificación legal en función del posible control del humano sobre la máquina. De ahí la conocidísima diferenciación entre los sistemas «*human in the loop», «human on the loop», y «human out of the loop*» [53], en función de si el control del ser humano es más cercano, por estar, de algún modo, manejando la máquina desde «dentro», debiendo aprobar cada decisión; si más bien está supervisando, monitoreando y corrigiendo errores; o si, finalmente, el ser humano está fuera de esa máquina, más alejado de un control efectivo sobre las decisiones que esta toma.

Cabría preguntarse, sin embargo (y esto es algo que se trata en las nociones que proponen autores como Turner), si quedarían extramuros del concepto aquellos sistemas que no fueran diseñados por humanos, sino por otras inteligencias artificiales [54].

*Datos.*

La importancia fundamental del dato como nutriente de los sistemas es innegable. Son los datos, y nada más que ellos, los que sirven como punto de partida a la creación de cualquier sistema. Los datos pueden ser, desde luego, de muy diversa índole (pueden ser datos numéricos, estructurados o no estructurados, pueden ser imágenes, sonidos, datos abstractos,

---

[53] Es el modelo que introducen Harbers, M., Peeters, M. M., /Neerincx, M. A.: «Perceived autonomy of robots: Effects of appearance and context», en Aldinhas Ferreira, M. I., Silva Sequeira, J., Tokhi, M. O., Kadar, E., / Virk, G. S. (Eds.), *A World with Robots*, Springer, Cham: Switzerland, 2017. Si bien es una referencia casi obligada en todos los textos sobre la disciplina, quizá la distinción se encuentra, hoy en día, algo superada. *Vid*., en este sentido: Miró Linares, F., «Inteligencia artificial y Justicia Penal. Más allá de los resultados lesivos causados por robots», *Revista de Derecho Penal y Criminología,* 20, 2018, p. 93: «*Una máquina autónoma pertenecería a un cuarto tipo, digamos «No man on the loop», donde o bien el aprendizaje por el que se toma la decisión no hubiera devenido de una acción humana, sino de la propia máquina que ha aprendido por sí misma o de otra máquina «educadora» anterior, o bien el comportamiento de la máquina no dependiera de ese aprendizaje anterior y de las decisiones atribuidas previamente en la IA sino de algo ajeno a ello y propio del mismo sistema».*

[54] Es por este motivo por el que Turner opta por hablar de «entidades no naturales», pues así se da entrada a la opción de que no sea el humano el que cree o diseñe el sistema. Turner, J., *op. cit.* N 5, p. 16.

o incluso datos expresados en lenguaje natural). Una IA no es más que un sistema que recibe determinados datos y produce otros. Precisamente la expansión de disciplinas relacionadas con la ciencia de datos y la mayor capacidad de computación permitieron el auge exponencial de las técnicas más sofisticadas de procesamiento de datos. Así, términos como «*big data*» y «*data mining*» [55] no podrán ser ajenos a una disciplina como la que aquí se estudia.

El dato es, pues, nutriente esencial de la tecnología, y como tal se ha convertido en fuente absoluta de poder y control. De hecho, y como algunos pensadores actuales señalan, el dato se ha convertido ya en paradigma de valor indiscutible en la sociedad de la información, llegándose incluso a cuestionar su papel como elemento disruptor de la sociedad presente y futura [56]. Es por ello esencial la toma de conciencia de las instituciones sobre esta realidad,

[55] Bonis Sanz opta por acudir a la Resolución del Parlamento Europeo, de 14 de marzo de 2017 para encontrar una delimitación jurídica del concepto *«big data»*. Dicha resolución, sobre las implicaciones de los macrodatos en los derechos fundamentales, utiliza el término «macrodatos» como traducción de «*Big Data*» y lo define así: «*Considerando que el concepto de macrodato se refiere a la recopilación, análisis y acumulación constante de grandes cantidades de datos, incluidos datos personales, procedentes de diferentes fuentes y objeto de un tratamiento automatizado mediante algoritmos informáticos y avanzadas técnicas de tratamiento de datos, utilizando tanto datos almacenados como datos transmitidos en flujo continuo, con el fin de generar correlaciones, tendencias y patrones (analítica de macrodatos)»*. Para Bonis, los elementos definitorios de este concepto son, por una parte, un elemento activo de recopilación y análisis de grandes cantidades de datos, mediante un tratamiento automatizado a través de algoritmos informáticos, y otro de carácter tendencial, que sería el fin de ese tratamiento (generar correlaciones, tendencias y patrones) para la predicción de ciertos acontecimientos y comportamientos. Para este autor, es este elemento tendencial el que permite conectar el concepto de «*Big Data*» con el de «minería de datos» que podríamos considerar equivalente a lo que la misma Resolución denomina «analítica de macrodatos». *Vid.:* Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)). Bonis Sanz, A., «Implicaciones jurídicas del uso de técnicas de Big Data y Sistemas de Inteligencia Artificial por las Administraciones Tributarias», *Revista de la Asociación de Abogados del Estado*, 57, mayo 2022, pp. 2-3.

[56] Procaccino habla de la «*infosfera*» como la concepción digitalizada de la realidad, constituida por un conjunto de informaciones, en la que los datos personales asumen un papel, no meramente descriptivo, sino también constitutivo del individuo y fundamental para el funcionamiento de un número cada vez mayor de actividades de la vida cotidiana. Esta autora lo relaciona estrechamente con el fenómeno de la «datificación» («*datificazione*») y la polivalencia de los «datos». Procaccino, A., «Securitizzazione dell'Unione europea e poteri concorrenti. Dall'investigazione, alla prevenzione, all'osservazione», *Diritto Penale Contemporaneo*, 1/2023, p. 147, nota 10.

En su obra *Homo Deus: Breve historia del mañana* (2016), Harari describe el dataísmo como una ideología emergente que considera los datos como el valor supremo (Harari, Y. N., *Homo Deus: Breve historia del mañana*, (trad. R. Joan Domènec), Debate, Barcelona, 2017). También filósofos contemporáneos como Byun Chul Han han puesto de manifiesto la importancia vital de los mismos en nuestras sociedades actuales a lo largo de su obra. Especialmente, en *Infocracia. La digitalización y la crisis de la democracia,* Taurus*,* 2022, donde expone su visión en forma de crítica radical, describiendo cómo la digitalización transforma el espacio público y degrada la democracia representativa, reemplazando la deliberación política por un régimen basado en datos y algoritmos.

pues la gobernanza sobre los datos y su tratamiento se ha convertido en una de las indiscutibles piedras de toque de la nueva realidad tecnológica [57].

*Razonamiento/procesamiento/decisión.*

La definición destaca que el sistema podrá razonar sobre el conocimiento, procesar información y decidir qué acción tomar. Estas tres capacidades han sido tradicionalmente una función exclusiva del razonamiento humano. Que ellas sean, en sí mismas, manifestaciones de la inteligencia es algo que queda extramuros de la definición y que, en opinión de la que escribe estas líneas, debe ser así.

No interesa demasiado, pues probablemente conduciría a disquisiciones intelectuales muy alejadas de la disciplina que nos ocupa, tratar de definir la inteligencia o aquellas características que, al estar presentes en el humano, acaban por afirmar su condición de ser inteligente. Aunque en algunas ocasiones se ha tratado de desempeñar esta tarea para definir la IA [58], no es este el enfoque que adopta la resolución del HLEG, pues, como se ha explicado en los

[57] No es esta una tendencia que esté unida en exclusiva con la disciplina de la IA, pues la importancia de los datos y la privacidad se viene poniendo de manifiesto, desde luego, como corolario necesario del artículo 18.4 de la Constitución, y su reconocimiento ha venido de la mano de cuantas leyes orgánicas de protección de datos se han ido sucediendo (teniendo como antecedente la Ley Orgánica 5/1992 (LORTAD), y como base primigenia en el panorama nacional el ya reconocido derecho a la «autodeterminación informativa» desde 1993 (STC 254/1993, de 20 de julio). Desde luego los fundamentos de la protección de datos que hoy en día afectan al paradigma del *big data* no se encuentran alejados de estos primigenios desarrollos. Para muestra, la reciente sentencia del Tribunal Federal Constitucional alemán declarando inconstitucional una ley por permitir aplicaciones de policía predictiva con IA sin una restricción proporcionada de este derecho. *Vid.:* Bravo Bolado, A., «El guardián de Gotham. De software predictivos y derechos fundamentales: ¿el fin de la inteligencia artificial al servicio del crimen? (Comentario a la Sentencia del Tribunal Constitucional Alemán de 16 de febrero de 2023)». *La ley penal: revista de derecho penal, procesal y penitenciario*, (165), 1, 2023, pp. 1-11. Villaverde Menéndez, I., «Protección de datos personales, derecho a ser informado y autodeterminación informativa del individuo. A propósito de la STC 254/1993», *Revista Española de Derecho Constitucion*al, año 14, 41, 1994. Encontrando el mismo fundamento para las cuestiones de discriminación algorítmica: Rivas Vallejo, P., «Discriminación algorítmica: detección, prevención y tutela», XXXI Jornades Catalanes de Dret Social («Treball, discriminación i COVID»), Barcelona, abril 2021, p. 64. También la doctrina alemana ha llamado ya la atención sobre la relación entre la injerencia en el derecho a la autodeterminación informativa y el empleo de técnicas de IA en contextos policiales. *Vid.*: Golla, S. J., «Lernfähige Systeme, lernfähiges Polizeirecht: Regulierung von künstlicher Intelligenz am Beispiel von Videoüberwachung und Datenabgleich», *Kriminologisches Journal*, 2020, Vol. 52 Issue 2, pp. 149-161.

[58] Así, por ejemplo, De la Cuesta Aguado, hablando de las diferentes acepciones de inteligencia. También, como se ha visto arriba, todos aquellos partidarios de definir la IA en términos de inteligencia humana. *Vid.*: de la Cuesta Aguado, M. P., «Inteligencia artificial y responsabilidad penal», *Revista Penal México*, 9(16-17), 2020, p. 52. La misma autora, en un trabajo reciente, ha cuestionado que el concepto de IA utilizado en el Reglamento sea el correcto, destacando su alto componente de producto comercial (pp. 85 y ss.), y mostrándose partidaria de una adaptación del Código penal para afrontar los riesgos prohibidos regulados en el Reglamento. Esta última cuestión se tratará en profundidad *infra,* en capítulo 7. *Vid.*: De La Cuesta Aguado, M. P., «El derecho penal frente al Reglamento de Inteligencia Artificial», *Ius Criminale*, 1, octubre de 2024, pp. 83-107.

apartados anteriores, no importa tanto la «humanización» de la máquina, tratando de encontrar el paradigma de la «máquina que sea como el humano», sino centrarse en la funcionalidad de la misma. Sea el método que sea el que utilice el sistema, lo hace para acercarse a un objetivo previamente determinado y, en este sentido, la definición tiene en cuenta esta perspectiva finalista o, si se quiere, orientada a un fin.

*Autonomía.*

En esta definición no se encuentra formulado el requisito de autonomía de una forma expresa, como sí ocurre en las definiciones posteriores que las instituciones europeas han fijado normativamente y de forma más reciente[59]. Y, sin embargo, se considera elemento absolutamente necesario para entender la actual inteligencia artificial, pudiéndose derivar de forma implícita del concepto escogido por el Grupo de Expertos. Cuando el HLEG reconoce que el sistema «decidirá las mejores acciones a tomar», podemos entender que, si bien de forma implícita, hay un reconocimiento de la autonomía en esta capacidad de decisión. Baste aquí dejar apuntado por el momento, pues sin duda será necesario ahondar en este elemento, que el poder de tomar una decisión ha de implicar una doble posibilidad: llevar a cabo una determinada actuación o no realizarla. Enseguida examinaremos cómo esta capacidad es, de hecho, no sólo uno de los elementos constitutivos de la definición, sino quizá el más representativo de la tecnología que hoy en día denominamos inteligencia artificial.

El segundo párrafo de la definición que ofrece el HLEG es menos conocido, pero no por ello debe dejarse de lado, pues de él podemos extraer también algunas valiosas conclusiones.

*«Como disciplina científica, la IA incluye varios enfoques y técnicas, como el aprendizaje automático (del que el aprendizaje profundo y el aprendizaje por refuerzo son ejemplos específicos), el razonamiento automático (que incluye la planificación, la programación, la representación y el razonamiento del conocimiento, la búsqueda y la optimización) y la robótica (que incluye el control, la percepción, los sensores y los actuadores, así como la integración de todas las demás técnicas en sistemas ciberfísicos*[60]*)».*

---

[59] *Vid.*, inmediatamente *infra,* en la definición del Reglamento de IA.

[60] «*As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimiza-*

Se reconoce, así, la variabilidad de la disciplina, y se equiparan algunos términos que a menudo aparecen entremezclados en la literatura sobre inteligencia artificial. Merece la pena realizar alguna aclaración sobre la aquí mencionada «robótica». Así, el HLEG habla de robótica como una subdisciplina de la inteligencia artificial, destacando el componente físico o ciberfísico necesario para hablar de esta materia [61]. Cabe señalar que, pese al diferente recorrido de ambas disciplinas, la definición institucional por la que se toma partido, parte de que habrá IA en ambos casos: tanto cuando se trate de una entidad meramente virtual (el referido como «software» en la definición originaria, párrafo primero), como cuando haya un soporte físico (ya sea en forma de «hardware», también nombrado en ese primer párrafo), que puede ser más o menos complejo. Que el sistema actúe en la dimensión física o en la digital no es un elemento diferenciador a tener en cuenta [62].

Cerramos así este apartado aclarando que uno de los efectos de la falta de definición e inseguridad conceptual es, precisamente, la dificultad que plantea la existencia de términos similares, con connotaciones ligeramente diferentes, que vienen a entremezclarse en los escritos académicos y jurídicos sobre la materia. Esto supone una dificultad añadida a la hora de perfilar definiciones, y sin duda cuestiones geográficas y temporales tienen una incidencia innegable en la elección de conceptos. Así, es común que en el entorno germano las referencias se hagan, en general, con relación al término «*Roboter»* en los escritos clásicos, mientras el moderno término «*Künstliche Intelligenz*» se encuentre sólo presente en los textos más recientes. Algo parecido puede decirse del entorno anglosajón, donde, más allá de las referencias iniciales a los ro-

---

*tion), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)».*

[61] También se señala este componente físico o corpóreo en la resolución del Parlamento Europeo sobre robótica, donde están presentes, asimismo, algunos de los elementos definitorios de la IA. *Vid.:* Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica [2015/2103(INL)].

[62] Seguramente esto es algo discutible, y ciertamente hay autores que ponen en duda que la potencialidad lesiva de un sistema con corporeidad y capacidad de interacción en el mundo físico sea igual que la del sistema que solo está presente en el entorno virtual, pero parece ser que de las definiciones institucionales europeas poca diferencia de tratamiento se puede justificar. Así, Calo, señalando la corporeidad (*embodiment*) de los robots como una de las características que pueden dar lugar a un tratamiento jurídico diferenciado; *«Data is not embodied; robots generally are. Sensing, navigating, and acting upon the world generally requires a physical presence, and that physical presence opens up a universe of new possibilities*.» Calo, R., «Robotics and the Lessons of Cyberlaw», *California Law Review,* 103 (3), 2015, p. 532. Nos situamos, sin embargo, más cercanos a la línea expuesta por Agustina cuando parece afirmar que, ante el desdoblamiento que produce el espacio ciber, ahora el yo-real y yo-digital se confunden, de forma que es muy difícil separar con nitidez el mundo offline del online. El punto al que llega este autor es afirmar que el daño, en este contexto tecnológico, depende de mucho más que la corporeidad. Estamos completamente de acuerdo. *Vid.*: Agustina, J. R., «Nuevos retos dogmáticos ante la cibercriminalidad», *Estudios penales y criminológicos*, 41, 2021, pp. 719 y 720.

bots, los debates sobre aplicaciones específicas de la tecnología (véase, destacadamente, el debate en torno a las armas autónomas letales o los vehículos autónomos), han hecho que los vocablos relacionados con la autonomía sean, quizá, más comunes [63]. Así, no es extraño encontrar referencias a términos más complejos, como el de «agentes autónomos artificiales», «agentes inteligentes», especialmente en aquellos textos más apegados a la reflexión jurídica y filosófica [64]. En definitiva, y sin pretender realizar un estudio exhaustivo sobre la mayor o menor prevalencia de unos u otros términos, debe dejarse apuntada esta realidad, pues el lector se sorprenderá, a lo largo de la obra, leyendo referencias a diferentes términos, sin que en muchas ocasiones las diferencias entre ambos sean significativas.

Así, y aunque se pretende utilizar en la mayor medida de lo posible el concepto «sistema de IA», por ser el utilizado en la definición escogida, si se encontraran ulteriores referencias a otras terminologías, y en la medida en que no se especifique que se pretende destacar alguna diferencia significativa, se deberán reconducir al concepto de IA aquí adoptado.

### 3.3 **Otras definiciones: el Reglamento**

Hoy en día, y a pesar de los innumerables documentos de *soft law* que se han ido aprobando al hilo del desarrollo de la IA, sólo uno de los documentos de nuestro entorno cultural puede afirmarse que constituye la primera norma imperativa de contenido general y aplicación obligatoria en la materia. Se trata

---

[63] Las referencias a los términos «*artificial moral agents*» son más comunes en el campo de la ética y la filosofía, donde una de las cuestiones fundamentales ha tenido que ver con la posibilidad de afirmar agencia moral o una ética propia de estos sistemas. Dada la inabarcable discusión al respecto, *vid.,* por todos: Floridi, L., /Sanders, J. W., «On the morality of artificial agents», *Minds and machines*, *14*, 2004, pp. 349-379. En este artículo, los autores introducen el concepto de *Artificial Moral Agents* y exploran las condiciones bajo las cuales los sistemas artificiales pueden ser considerados agentes morales, reconociendo que, aunque no necesariamente tienen intencionalidad como los humanos, estas máquinas pueden desempeñar un rol moral en sistemas complejos.

[64] Sin ir más lejos, Simmler, M., /Markwalder, N., «Roboter in der Verantwortung? –Zur Neuauflage der Debatte um den funktionalen Schuldbegriff», *Zeitschrift für die gesamte Strafrechtswissenschaft*, 129(1), 2017, p. 24, cuando aclaran que utlizarán el término robot («*Roboter*») para referirse tanto a agentes inteligentes («*intelligente Agenten»)* como a autómatas autónomos («*autonome Automaten»),* a sabiendas de que ello representa una simplificación importante de los límites entre estos términos y diversas subcategorías, pero aclarando que el elemento diferenciador es que se trata de verdaderos sistemas inteligentes que actúan con cierto grado de autonomía (*welche große Datenmengen verarbeiten und im Rahmen einer gewissen Autonomie agieren, indem sie in Entscheidungsprozessen auf diese Daten zurückgreifen können*). Se remiten en este aspecto, para la diferenciación, a: Gless, S., /Weigend, T., «Intelligente Agenten und das Strafrecht», *ZSTW* «, 126(3), 2014, pp. 561-591.

del vigente Reglamento de Inteligencia Artificial [65], que en su artículo 3 nos ofrece una definición escueta de lo que debemos entender por «sistemas de IA».

Sin embargo, antes de la aprobación del texto final de la normativa, el documento relativo a la Propuesta de Reglamento, publicado por el Parlamento Europeo en abril de 2021, ofrecía ya las pistas sobre aquellos elementos que debían estar contenidos en la definición.

Se establecía en el considerando 6 que era necesaria una definición capaz de ofrecer seguridad jurídica, al mismo tiempo que flexible, para adaptarse a los futuros avances tecnológicos. La definición debía basarse en las principales características funcionales del *software*, y en particular en su capacidad para generar, en relación con un conjunto concreto de objetivos definidos por seres humanos, contenidos, predicciones, recomendaciones, decisiones u otra información de salida que influyeran en el entorno con el que interactúa el sistema, ya sea en una dimensión física o digital. Se hacía referencia a que los sistemas de IA pueden diseñarse para operar con distintos niveles de autonomía, así como utilizarse de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente de él (integrado) o tiene una funcionalidad en el producto sin formar parte de él (no integrado). La definición de «sistema de IA», decía la Propuesta, debía complementarse con una lista de las técnicas y estrategias concretas que se usen en su desarrollo.

Se puede apreciar cómo están aquí presentes algunos de los elementos arriba identificados. La definición que por entonces se proponía era la siguiente:

*«Sistema de inteligencia artificial (sistema de IA): el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa»* [66].

---

[65] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

[66] Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión, COM/2021/206 final.

La definición que finalmente se plasma en el texto legal no dista demasiado de la de la Propuesta inicial:

*«"sistema de IA": un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales»* [67].

Podemos extraer, de forma similar a como se ha hecho con el concepto propuesto por el HLEG, las características fundamentales:

– Se otorga una importancia primordial, esta vez explícita, a la autonomía, reconociendo que puede ser variable, pero que siempre tiene que estar presente para hablar de IA.

– Se destaca la capacidad de adaptación (es este un elemento que no está presente en la definición propuesta por el Grupo de Expertos, que quizá pueda indicar un papel destacado del aprendizaje, pero que seguramente pueda reconducirse a la capacidad de autonomía, como se explicará inmediatamente después).

– El punto de partida es, una vez más, la existencia previa de objetivos fijados por los humanos.

– Se equipara, definitivamente, la influencia en el mundo físico y en el virtual.

Todos estos elementos requerirán, sin duda, una interpretación adecuada que permita identificar qué resulta esencial para comprender este fenómeno complejo y dinámico. Solo la práctica y la aplicación de esta novedosa herramienta en un futuro próximo (está previsto que entre en vigor de forma escalonada de la forma en que se detalla en el Titulo XII relativo a las disposiciones

[67] Artículo 3. (1) del Reglamento en su version final. *op cit.* N. 65. De ahora en adelante, el *Reglamento de IA*. Más adelante la Comisión publicó unas directrices para la aplicación uniforme de este concepto: European Commission. 2025. *Commission Communication: Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)*, C (2025) 924 final, 6 February 2025.

finales [68]), nos permitirá conocer algo más lo que verdaderamente significa la IA para el derecho europeo [69].

## II. LA AUTONOMÍA. LA NOTA DISCORDANTE EN LAS DEFINICIONES DE IA. ¿DE QUÉ HABLAMOS CUANDO HABLAMOS DE AUTONOMÍA?

Al igual que ocurre con la obra del afamado poeta Raymond Carver («De qué hablamos cuando hablamos de amor» [70]), el complejo término «autonomía» aparece de forma confusa en todos los textos sobre inteligencia artificial, habiéndose convertido en un lugar común de discusión entre los teóricos de las

[68] El *Reglamento de IA* se aplicará completamente a los tres años de su publicación en el DOUE. A los 12 meses de su publicación entrará en vigor lo referente a autoridades notificantes, autoridades nacionales de supervisión, y Comité Europeo de IA (Título III Cap.4 y Título VI). Cada dos años, la Comisión revisará la necesidad de enmienda del Anexo III. Cada cuatro años, informará al Parlamento y el Consejo sobre la aplicación del Reglamento, valorando recursos técnicos y humanos disponibles y sanciones aplicadas. Cada cuatro años, la Comisión evaluará la efectividad de los códigos de conducta. De esta forma, las disposiciones más relevantes quedan configuradas temporalmente de la siguiente forma: el Reglamento de IA entró en vigor el 1 de agosto de 2024. Las prohibiciones en él establecidas surten efecto seis meses más tarde (1 de febrero de 2025). Las normas sobre gobernanza y obligaciones para los modelos de IA de uso general entraron en vigor el 1 de agosto de 2025. El Reglamento será aplicable en su totalidad el 1 de agosto de 2026 (24 meses después de su entrada en vigor).

Nota de actualización: con posterioridad al cierre de redacción de este trabajo, el calendario descrito ha sido objeto de recalibración a través del denominado Digital Omnibus on AI (propuesta COM(2025) 836 final, presentada por la Comisión el 19 de noviembre de 2025 dentro del séptimo paquete ómnibus de simplificación digital), cuyo acuerdo político provisional fue alcanzado en trílogo el 7 de mayo de 2026 –pendiente, al cierre de estas pruebas, de revisión jurídico-lingüística, adopción formal por el Parlamento y aprobación definitiva del Consejo antes de su publicación en el DOUE–. La reforma no altera la entrada en vigor del Reglamento (1 de agosto de 2024), ni la efectividad de las prohibiciones del art. 5 (2 de febrero de 2025), ni el régimen de los modelos de IA de uso general (2 de agosto de 2025), pero difiere la aplicación de las obligaciones relativas a los sistemas de alto riesgo del Anexo III del 2 de agosto de 2026 al 2 de diciembre de 2027, y traslada al 2 de agosto de 2028 la correspondiente a los sistemas de alto riesgo integrados en productos sometidos a la legislación sectorial de seguridad y vigilancia del mercado (Anexo I); fija, además, el 2 de diciembre de 2026 como fecha para las obligaciones de marcado *(watermarking)* de los contenidos generados por IA ya comercializados, e introduce una prohibición expresa de los sistemas generadores de contenido sexual íntimo no consentido y de material de abuso sexual infantil *(nudifier apps)*. En tanto no se publique el texto definitivo en el DOUE, las fechas originales aquí expuestas mantienen plena vigencia jurídica.

[69] Fuera de este entorno europeo, y con un tinte y vocación más internacional, merece destacar, por considerarse adecuada y común en cuanto a los términos fundamentales, la definición propuesta por la OCDE. *«AI system: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment»*. OCDE, « *Recommendation of the Council on Artificial Intelligence*», OECD Legal Instrument, 2021, disponible en: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> [Accedido por última vez el 20/01/2026].

[70] Carver, R., *De qué hablamos cuando hablamos de amor*, Anagrama, Barcelona, 2006.

disciplinas enfrentadas. Tal y como ocurre con la obra de Carver, y quizá para decepción de los lectores, ni el término amor en la obra del primero, ni el término autonomía en el trabajo que aquí se desarrolla responden a las profundidades o romanticismos clásicos que se le presuponen.

Se tratará de desentrañar en esta parte de la obra qué implica hablar de entidades autónomas, tratando de clarificar un concepto que ha sido oscurecido y objeto de demasiada discusión, probablemente, una vez más, por las ansias humanizadoras de esta nueva disciplina científica.

No se pretende profundizar, sin embargo, en todas las aristas filosóficas de un concepto tan intrincado, porque dicha discusión superaría los límites de lo que aquí se pretende, que no es otra cosa que aclarar el concepto de IA que constituirá el marco y punto de partida de esta investigación.

Como se ha podido observar, en las dos definiciones que se toman como válidas –la del Grupo de Expertos de Alto Nivel, por considerarse la más completa, y la del nuevo Reglamento sobre inteligencia artificial, por considerarse la legalmente vinculante–, la autonomía está presente como elemento de diferente forma.

En la definición del Reglamento aparece de forma explícita, donde se define el sistema de IA como*: «un sistema basado en una máquina que está* ***diseñado para funcionar con distintos niveles de autonomía*** *y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.»* [71]

En el caso del HLEG aparece de forma implícita, como se señaló *supra*. («*Sistemas de software (y posiblemente también de hardware) diseñados por humanos que, dado un objetivo complejo, actúan en una dimensión física o digital percibiendo su entorno a través de la obtención de datos, interpretando los datos estructurados o no estructurados recabados, razonando sobre el conocimiento o procesando la información derivada de dichos datos y* ***decidiendo las mejores acciones a tomar*** *para conseguir el objetivo*» [72].) La capacidad para decidir qué acción ejecutar se identifica, de forma pacífica, como una de las manifestaciones

[71] Artículo 3 (1) del *Reglamento de IA*. La negrita es nuestra.

[72] «*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions*». De nuevo, la negrita es nuestra.

de, al menos, un nivel de autonomía en sentido débil, por lo que no parece posible descartar que este elemento tenga protagonismo en esta definición.

Las referencias a la autonomía como elemento caracterizador de la inteligencia artificial son ubicuas [73] y, sin embargo, no dejan de suponer un obstáculo cuando las discusiones se producen en el plano jurídico, por todas las connotaciones que el término «autonomía» posee, más notablemente aún en el Derecho penal, siendo difícil desligar dicho término de conceptos indisolublemente unidos a la dignidad e individualidad humanas. Quizá en el futuro acabe por imponerse otra terminología, y lo cierto es que algún autor ha realizado el intento de huir del concepto por completo, hablando, por ejemplo, de «emergencia» [74]. Sin embargo, ningún concepto alternativo ha conseguido la omnipresencia del anterior, pues no han logrado calar ni en la doctrina ni en los instrumentos legislativos. En estos términos precisamente razona el Grupo Europeo sobre ética en su informe de 2018 sobre el concepto de autonomía [75], reconociendo que la autonomía en el sentido éticamente relevante de la palabra solo puede ser atribuida a seres humanos, siendo inapropiado usarlo para referirse a meros artefactos, pero reconociendo que el término ha ganado gran aceptación en la literatura científica y en los debates públicos, concediendo que se usa para hablar del grado más alto de «automatización» y de independencia en términos operativos.

---

[73] No sería posible listar de forma exhaustiva los trabajos que emplean este término de una u otra manera, aunque sí se puede centrar el origen de esta discusión en uno de los campos de aplicación que supusieron, en su momento, algunos de los mayores retos para la ética y el derecho – el campo de las llamadas «armas autónomas letales». Es al hilo de estas tecnologías donde comienzan las discusiones sobre la autonomía y sus significados, si bien desde una perspectiva más técnica que filosófica. Para estudiar el desarrollo de estas tecnologías: Boulanin, V./ Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems*, Stockholm International Peace Research Institute, *SIPRI*, Stockholm International Peace Research Institute, 2017. Para un repaso sobre el inicio de los problemas y las críticas filosóficas en este campo: Sauer, F., «Autonomous weapons systems,» en Elliott, A. (Ed.), *The Routledge Social Science Handbook of AI*, Routledge, Londres y Nueva York, 2022, pp. 237-250, que ya reconoce que «filosóficamente hablando», el término autonomía es inapropiado porque supone la antropomorfización de las máquinas. El término ha sido utilizado, no obstante, también como concepto crítico, por quienes consideran que, desde el punto de vista del proyecto posthumanista que viene a poner en duda las bases humanistas de nuestras sociedades, «*la autonomía de las personas queda suplantada […] por la autonomía de los sistemas»*, criticando lo anti-humano que tiene que sean «*los algoritmos los que decidan sobre los fines y objetivos futuros del desarrollo de la vida al margen de las elecciones de las personas concernidas por estos designios»*. Así, Pérez Luño, A. E., *op. cit*. N 36, p. 29.

[74] Así, Calo habla de emergencia, aunque ya anuncia antes que es más bien una noción para designar la autonomía o la capacidad de aprendizaje. Y elige el término «emergencia» y no «autonomía en el diseño» porque cree que no se gana nada al usar ese término que puede dar lugar a la imagen confusa de que un robot tenga algo parecido a una intención. Calo, R.*, op. cit.* N 62, p. 538.

[75] European Commission: Directorate-General for Research and Innovation and European Group on Ethics in Science and New Technologies, *Statement on artificial intelligence, robotics and 'autonomous' systems,* Brussels, 9 March 2018, Publications Office, 2018, disponible en: <https://data.europa.eu/doi/10.2777/531856>[Accedido por última vez el 20/01/2026].

Aunque el vocablo no sea quizá el más adecuado dada su larga tradición filosófica, es posible aceptarlo como término ya asentado en la disciplina de la IA y, así, tratar de encontrar una acepción de autonomía más «aséptica» [76], que se refiera exclusivamente a aquellas propiedades diferenciadoras de los sistemas de IA.

Partamos, pues, de dos acepciones del término «autonomía»: la primera, que llamaremos autonomía en sentido filosófico o en sentido fuerte, y la segunda, que llamaremos autonomía en sentido operativo o en sentido débil. Como ya se ha indicado, en este trabajo no se desarrollará ni se utilizará el primero de los conceptos cuando se hable de sistemas de IA dotados de autonomía, pero sí se van a ofrecer unas pinceladas de lo que implica este término de cara a descartar los componentes que no le sirvan a nuestra renovada y desapasionada acepción de autonomía.

## 1. LA AUTONOMÍA EN SENTIDO FILOSÓFICO O FUERTE

Sirviéndonos de una de las definiciones más pacíficas, la contenida en la Enciclopedia de Filosofía de Stanford, la autonomía individual es «una idea que generalmente se entiende que se refiere a la capacidad de ser uno mismo, de vivir la propia vida de acuerdo con razones y motivos propios y no el producto de fuerzas externas manipuladoras o distorsionadoras» [77]. Podemos señalar, por tanto, tres aspectos: la capacidad de ser uno mismo, la posibilidad de vivir la vida de acuerdo con razones o motivos propios, y la no opresión o sujeción a fuerzas externas. De estas tres características, la capacidad de vivir de acuerdo con motivos propios podría considerarse un mínimo imprescindible

[76] Hay quienes se refieren a la autonomía en este campo como «autonomía técnica», o incluso quienes proponen, para evitar confusiones, hablar de «funcionamiento autónomo en un sistema» («*autonomous functioning in a System*»). Se considera aquí, sin embargo, que ya es tarde para evitar el uso generalizado del término «autonomía» en solitario, y que será mejor dedicar nuestros esfuerzos a explicar qué características contiene y cuáles deben quedar fuera, pues el cambio terminológico presenta menos potencial y requiere muchos más esfuerzos (también en lo relativo a la economía del lenguaje). Para el desarrollo y explicación de los términos alternativos propuestos: Simmler, M./ Frischknecht, R., «A taxonomy of human–machine collaboration: capturing automation and technical autonomy», *AI & SOCIETY*, 36, 2020, pp. 239-250, *passim*; y Williams, A. P., «Defining autonomy in systems: challenges and solutions», en: Williams, A. P.,/Scharre, P. D., *Autonomous Systems Issues for Defence Policymakers,* HQ SACT, 2015, pp. 27-62*,* aunque en p. 55 reconoce directamente que «*lo ideal sería eliminar el término "sistema autónomo" de nuestro léxico, pero es una sugerencia poco práctica dado el uso que se hace de él*». [traducción propia].

[77] Christman, J., «Autonomy in Moral and Political Philosophy», en Zalta, E. N., (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), disponible en: <https://plato.stanford.edu/archives/fall2020/entries/autonomy-moral/> [Accedido por última vez el 20/01/2026]. La traducción es propia.

en el entendimiento común del término, que coincide de forma sustancial con la tradición kantiana [78]. Tal y como señalan Hildebrandt [79] o Bradshaw [80], no en vano el término proviene del griego *«auto»* (propia) *«nomos»* (norma). Así, el hecho de «autonormarse» o imponerse a uno mismo las razones vitales, parece central en un significado básico del término.

Qué implicaciones tiene esa asunción o elección propia de valores es algo que debe quedar extramuros de esta reflexión, pero no debemos perder, al menos, la orientación sobre esta idea mínima que nos sirve de referencia.

Silvia Salardi lo formula en unos términos especialmente didácticos en un trabajo sobre los retos para el derecho que presenta la IA [81]. Así, nos habla de la centralidad de autoimposición de una moral propia entendida como posibilidad de hacer planes vitales *(«making life plans»),* que caracteriza como el nivel de autonomía más pleno que se puede afirmar exclusivamente en el caso del ser humano; una libertad de elección entendida en un sentido amplio. Afirma la autora que el sentido profundo que se le reconoce a la autonomía personal en el contexto jurídico europeo no admitiría la interferencia de otra moral externa, heterónoma, no elegida por el propio sujeto [82]. La autonomía en este sentido fuerte queda así configurada como estrechamente vinculada a las opciones de elegir sobre la base de valores que permiten configurar el plan de vida, y no la mera posibilidad de actuar [83].

Resulta, por tanto, que los autores que trabajan sobre los sistemas autónomos de IA son conscientes de las limitaciones del concepto, y manejan una limitación que bebe, a su vez, de los textos más clásicos sobre filosofía moral, y que tiene que ver con la dicotomía entre libertad de acción y libertad de elección. De esta dicotomía se valen los autores que reflexionan, también desde el derecho penal, sobre los problemas relacionados con la autonomía, la libertad y la responsabilidad.

Partiendo de la terminología que Molina toma prestada de Von Wright, podemos considerar, al menos, dos tipos de libertad: la libertad de acción y la

---

[78] Kant, I., *Grounding for the Metaphysics of Morals*, en Kant, I. *Ethical Philosophy*, Ellington, J. W., (trad.), Hackett Publishing Co, Indianapolis, 1983.

[79] Koops, B. J., Hildebrandt, M., / Jaquet-Chiffelle, D. O., «Bridging the accountability gap: Rights for new entities in the information society?», *Minn. JL Sci. & Tech*., 11, 497, 2010, p. 18.

[80] Bradshaw, J, /Hoffman, R./ Johnson, M., /Woods, D., «The seven deadly myths of ‘‘autonomous systems»«, *IEEE Intelligent Systems*, 2013, p. 55.

[81] Salardi, S.: «Robótica e inteligencia artificial: retos para el derecho», *Derechos y libertades*, Núm. 42, Época II, enero 2020, pp. 203-232.

[82] *Idem*, p. 222.

[83] *Idem.*

libertad de elección [84]. Muchos otros autores, si bien no utilizando estos mismos términos, han abrazado esta clasificación, hablando en ocasiones de deseos de primer grado o de segundo grado (en este último caso, referido a desear el hecho de desear algo) [85]; o distinguiendo, como hace De la Cuesta, entre la autonomía de la decisión, que se refiere a la posibilidad de elegir entre una conducta u otra, y la autonomía en la decisión de los fines, que se refiere a la posibilidad de elegir sobre los objetivos a seguir, en un sentido mucho más genérico [86]. Reconoce la misma autora que, por el momento, siempre que nos referimos a decisiones autónomas aludimos a la primera de las acepciones de autonomía, es decir, autonomía para decidir realizar una conducta u otra, pero no a la autonomía que implica decidir sobre los fines propios [87].

La autonomía de la decisión, así entendida, hace referencia a la opción por una conducta en concreto entre una o varias alternativas de conductas posibles, pero, en principio, no hace referencia a la autonomía en la decisión de los fines o los objetivos. Por el momento, y dado el estado de la ciencia, cuando nos refiramos a decisiones autónomas estaremos haciendo referencia a autonomía en cuanto a la conducta o comportamiento más adecuado a un fin u objetivo, excluyendo la autonomía de la decisión sobre el fin o el objetivo (hay quienes hablan de la capacidad de generar sus propios propósitos, metas o intenciones [88]).

Aunque no estamos situados, aún, en la problemática relativa a la responsabilidad, pues solo estamos tratando de afinar lo máximo posible el concepto

---

[84] Molina Fernández, F., «Presupuestos de la responsabilidad jurídica (Análisis de la relación entre libertad y responsabilidad)», *ADFUAM*, 4, 2000, p. 117, y Von Wright. Von Wright, G. H., *Of Human Freedom*, en Von Wright, G. H*, In The Shadow of Descartes*. *Essays in the Philosophy of Mind*, Kluwer Academic Publishers, Dordrecht/Boston/Londres, 1998, p. 9.

[85] Coeckelbergh, M., *Green Leviathan or the poetics of political liberty: navigating freedom in the age of climate change and artificial intelligence*, Routledge, Nueva York, 2021, p. 50 y ss. Por su parte, Blanco Cordero distingue entre «libertad de voluntad» y «libertad de obrar». Blanco Cordero, I: «Un derecho penal de humanos para entidades artificiales inteligentes. ¿Machina delinquere et puniri potest?», en Doval Pais, A., y Gutiérrez Pérez, E. (Dirs.), *Manifestaciones de desigualdad en el sistema de justicia penal. Volumen I,* Aranzadi, Cizur Menor, 2022, p.383. Abbot y Sarch hablan, por su parte, de que una IA sofisticada puede tener autonomía para determinar los medios con los cuales completar los objetivos prefijados. Es decir, no hay libertad de fines, pero sí de medios. *Vid.:* Abbott, R., / Sarch, A., «Punishing artificial intelligence: legal fiction or science fiction», *UC Davis Law Review*, 53(1), 2019, p. 357.

[86] De la Cuesta Aguado, M., P. «Inteligencia artificial y …» *op. cit*. N. 58, p. 53.

[87] *Idem*., p. 53.

[88] Así, Peter Asaro, considerando que solo en el momento en que un agente artificial alcance esta potencialidad estaremos ante un verdadero escenario diferenciador. Por el momento, lo único que reconoce este filósofo es la posibilidad de «agencia causal sin agencia legal». *Vid.*: Asaro, P. M., «The liability problem for autonomous artificial agents», *Association for the Advancement of Artificial Intelligence*, Papers from the 2016 AAAI Spring Symposium, 2016, pp. 191 y 129, disponible en: <https://aaai.org/papers/12699-12699-the-liability-problem-for-autonomous-artificial-agents/>, [Accedido por última vez el 20/01/2026].

que utilizaremos como marco, merece la pena resaltar, ya en este punto, que el hecho de que se afirme de forma bastante pacífica que el sistema de IA no tiene autonomía en un sentido fuerte excluye su consideración como responsable por las actuaciones derivadas de sus decisiones. La IA decide en el sentido débil que a continuación se expondrá, pero no decide de forma autónoma en un sentido fuerte, que es el único que importa para la consideración de agente responsable para el derecho y, desde luego, para el derecho penal [89].

De esta exposición se deriva la consecuencia de que solo la persona que es capaz de fijar esos fines u objetivos en sentido amplio, ya les llamemos deseos de segundo orden, planes vitales, o sistema de valores, será sujeto autónomo en un sentido fuerte, filosófico y jurídico, y, por tanto, capaz de responder [90] por sus actos.

## 2. LA AUTONOMÍA EN SENTIDO OPERATIVO O DÉBIL

Siendo coherentes con la distinción arriba propuesta, si la autonomía en sentido fuerte comprende los dos niveles de decisión mencionados, la libertad de acción y la libertad de elección, la autonomía en sentido débil u operativo implica solamente la primera de estas capacidades.

Pero hablar de libertad de acción es, en este punto, insuficiente, pues, más allá de la genérica concepción de poder tomar decisiones de tipo operativo sobre si actuar en un sentido o en otro, hablar de autonomía en los sistemas de IA requiere profundizar algo más. Tomemos esta libertad de acción como punto de partida y veamos cómo surge este concepto.

Como más arriba se ha señalado, el concepto de autonomía en este campo surge con los primeros desarrollos de las llamadas «armas autónomas letales» (o, por sus siglas en inglés, LAWS, *lethal autonomous weapon systems).* Sin ánimo de profundizar en las distintas definiciones de esta tecnología, baste apuntar que son, a grandes rasgos, sistemas armados capaces de aprender y

[89] Baste dejar afirmada esta consecuencia aquí, aunque en el capítulo 2 se examinará cómo una corriente doctrinal sí ha intentado desmontar este argumento. A nuestro entender, y con Molina Fernández, la autonomía individual es condición imprescindible para la adscripción de responsabilidad: *«**algunas modalidades de responsabilidad,** de las que son paradigma **la responsabilidad moral y la jurídico-penal,** están indisolublemente **vinculadas a una imputación subjetiva en el sentido más estricto del término,** esto es, una imputación que sólo se da en mentes conscientes dotadas de autonomía» (*la negrita está en el original*).* Molina Fernández, F., «La responsabilidad penal de las personas jurídicas y el valor de Pi», Almacén del Derecho, 20 de febrero de 2018, disponible en: <https://almacendederecho.org/la-responsabilidad-penal-las-personas-juridicas-valor-%CF%80> [Accedido por última vez: 20/01/2026].

[90] La responsabilidad a la que se alude aquí es una entendida en sentido genérico.

adaptar su funcionamiento en respuesta a las circunstancias cambiantes del entorno en que se despliegan [91]. Uno de los mayores problemas a la hora de concretar el significado y la potencialidad lesiva de este tipo de herramienta tecnológica ha sido, precisamente, el de concretar el alcance de su autonomía, y por ello el debate desplegado en este concreto campo puede servir a los fines más genéricos de este capítulo.

La autonomía circunscrita a este debate se ha concebido también de forma diversa, aunque lo cierto es que los esfuerzos de algunos autores por explicitar su contenido pueden resultar ilustradores para señalar algunas de las características más importantes. Así, Williams, en su empeño por definir la autonomía en las llamadas armas autónomas letales, afirma que un sistema autónomo es capaz de comprender las direcciones o intenciones suministradas y, a través de esta comprensión y de la percepción de su entorno, tomar las medidas adecuadas para alcanzar el estado deseado. Por tanto, es capaz de decidir un curso de acción a partir de una serie de alternativas sin depender de la supervisión o control humanos [92]. La autonomía implica, por lo tanto, autogobierno, autogestión y toma de decisiones, que se producen en función de condiciones predefinidas o restricciones del sistema. Los resultados obtenidos no son deterministas, sino que dependen de los cambios que el sistema detecte en el entorno [93]. Podemos observar cómo esta referencia a la «toma de decisiones en función de condiciones predefinidas» puede circunscribirse perfectamente a la distinción que arriba se señalaba: es una autonomía que implica capacidad de acción, pero no capacidad de decisión superior o sobre los objetivos últimos.

Bertolini, que parte de una triple caracterización, habla de autonomía como capacidad de realizar tareas sin supervisión humana, y reconoce que esta autonomía lo es en un sentido débil, pues el robot está completando una tarea para lograr una meta establecida por el ser humano, valiéndose del oxímoron de «autonomía heterónoma» o capacidad de elegir entre fines, tomando consciencia del entorno e interactuando con él, pero siendo el humano el que elige

---

[91] Roff, H. M., «Lethal autonomous weapons and jus ad bellum proportionality», *Case W. Res. J. Int'l L.*, 47, 2015, p. 37 (traducción propia). El Departamento de Defensa de EEUU, uno de los más implicados desde el inicio en el desarrollo de esta tecnología, lo define en su directiva sobre la autonomía en los sistemas de armas, simplemente como «*un sistema de armas que, una vez activado, puede seleccionar y atacar objetivos sin la intervención de un operador*» (DoD Directive 3000.09., 2023, traducción propia). Para un mayor desarrollo teórico sobre los matices de esta definición: Boulanin, V./ Verbruggen, M., *op. cit*. N.73.

[92] Williams, A. P., *op. cit*. N 76, p. 33.

[93] *Idem*, p. 39. Una vez más, cuando hablamos de determinismo de los sistemas, lo hacemos en el sentido de determinismo técnico explicado *supra,* en nota 23.

los objetivos finales, negando el autor que el artefacto «actúe» en un sentido filosófico y jurídico [94].

Otros autores han tratado de ser más precisos, intentando diferenciar parámetros. Así, Paul Scharre [95] habla de tres categorías de autonomía en los sistemas autónomos letales, relacionadas con tres variables: el control del humano, la complejidad técnica del sistema y, finalmente, el tipo de decisión que deba tomar:

– En primer lugar, en cuanto al control, se entiende que habrá mayor autonomía cuanto menor sea la posibilidad del humano de controlar la actuación de la máquina [96]. Se puede entender, de hecho, que la supervisión humana (o, más bien, la falta de ella) es uno de los rasgos mínimos imprescindibles para hablar de autonomía en la tecnología, y que al menos esta falta de control en sentido fuerte está presente siempre que se habla de esta característica [97]. Autores como Blanco Cordero afirman que, aunque la autonomía equivale a que no exista control humano o de terceros sobre el ente inteligente, lo cierto es que hay situaciones que no son del todo claras, pues es posible que un robot actúe sin control humano constante, sino que este se dé solo en caso de necesidad [98].

– En segundo lugar, los tipos de decisiones o el tipo de funciones que el sistema realiza con esta autonomía, siendo aquí la idea fuerza la relativa a que

---

[94] Bertolini, A.: «Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules», *Law, Innovation and Technology* 2, 2013, pp. 220-221.

[95] Scharre, P. D., «The Opportunity and Challenge of Autonomous Systems,» en Williams, A. P., /Scharre, P. D., *Autonomous Systems Issues for Defence Policymakers,* HQ SACT, 2015, p. 10 y ss.

[96] De este parámetro derivan todas las reflexiones sobre los niveles de control del humano sobre la herramienta, y las conocidas clasificaciones que hablan de tres posibilidades: el *human in the loop*, el *human on the loop* y human *out of the loop*. Nos remitimos, en este aspecto, a lo dicho *supra* en nota 53.

[97] De hecho, muchos autores, conscientes quizá de la dificultad de hablar de autonomía, son partidarios de afirmar que lo importante no es definir este término, sino profundizar en la relación humano-máquina, es decir, en cómo se desarrolla esta interacción y en determinar quién realiza qué tarea. Así, Sauer, F. *op cit*. N. 73, p. 257. Surge, también en este campo y tras tomar en consideración la importancia de esta interacción, la categoría del «control humano significativo» que pone el énfasis en crear una tecnología fiable y predecible, de la mano de Roff y Moyes. También se empieza a hablar, al hilo de esta falta de control, de los posibles vacíos de responsabilidad que pueden generarse (esta idea se desarrollará *infra* en capítulo 2). En lo relativo al origen del control humano significativo, *vid*.: Roff, H. M./Moyes, R., «Meaningful Human Control, Artificial Intelligence and Autonomous Weapons», Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016, disponible en: <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf> [Accedido por última vez: 20/01/2026]. En cuanto a los primeros problemas ligados al «*accountability gap*» en materia de armas autónomas: Sparrow, R., «Killer Robots», *Journal of Applied Philosophy*, 24 (1), 2007, pp. 62-77. Toda la problemática relativa a estas lagunas se tratará en profundidad *infra*, en capítulo II.

[98] Blanco Cordero, I., *op cit*. N. 85, p. 381.

lo verdaderamente importante no es considerar al sistema autónomo en su conjunto, sino analizar el concreto nivel de autonomía en las funciones desplegadas por el sistema. Este argumento es compartido por Williams, que afirma que solo algunas partes del sistema exhiben algún tipo de autonomía, por lo que sería mejor referirse a «funcionamiento autónomo dentro de un sistema», pues es engañoso o falaz considerar que todo el sistema goza de esta autonomía. Algo parecido señalan Bradshaw *et al* cuando afirman que uno de los siete mitos morales de los sistemas autónomos es que se considere que son, en su totalidad, autónomos, pues la autosuficiencia siempre debe entenderse unida a un específico contexto o tarea [99].

– Finalmente, y desde un punto de vista más técnico, la autonomía puede depender de la complejidad técnica que despliegue el sistema, de forma que se mida cómo de complejo o sofisticado es el proceso técnico de toma de decisiones, teniendo en cuenta la interacción del sistema con su entorno y cómo reacciona frente a las incertidumbres en el mismo [100]. Dentro de esta categoría, referida a la complejidad técnica, es donde la doctrina ha introducido la diferencia entre sistemas automáticos y autónomos (los términos son, una vez más, variables, pues hay quienes introducen otros con ligeras variaciones en cuanto a su significado [101]). Parece haber consenso en que el término «automático» no es el apropiado cuando hablamos de los sistemas más sofisticados técnicamente, pues se refiere simplemente a aquellos sistemas que responden de forma mecánica a la información sensorial, siguiendo procedimientos predefinidos por el humano, sin poder adaptarse a las incertidumbres del entorno [102]. La autonomía, por el contrario, tendrá que ver, en mayor o menor medida, con la capacidad de adaptación a las variaciones del entorno [103], a diferencia de la automatización, la cual implica un comportamiento que, si bien no re-

---

[99] Bradshaw, J., *et al*., *op cit*. N. 80, pp. 56 y 57.

[100] Scharre, P. D., *op. cit*. N. 95, p. 6.

[101] Así, Koops *et al* nos hablan de sistemas «*automatic, autonomic, and autonomous»,* si bien reconoce que las distinciones son más bien analíticas y no suelen ajustarse fácilmente al flujo de una realidad que cambia con rapidez, llegando a afirmar que el problema de categorizar este fenómeno de forma tajante puede recordar a la famosa paradoja Sorites (¿Cuántos granos de arena son necesarios para formar un montón?). *Vid.*: Koops, B. J., Hildebrandt, M., / Jaquet-Chiffelle, D. O., *op. cit.* N. 79, p. 40. Para una revisión en profundidad de los diferentes niveles de autonomía que proponen estos autores y las consecuencias para la imputación penal que pueden tener en la actualidad, *vid.*: Salvadori, I., «Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal», *Cuadernos de política Criminal*, 133, 2011, pp. 147 ss.

[102] Scharre, P. D., *op. cit*. N. 95, p. 6, poniendo el ejemplo de los brazos robóticos usados típicamente en la industria manufacturera.

[103] *Idem.*

quiere de la intervención humana, es determinista y predecible en gran medida, pues no implica adaptación al entorno [104].

Este último rasgo diferenciador de la autonomía de tipo técnico tiene que ver con una capacidad para observar y entender el entorno y actuar en consecuencia, sin necesidad de órdenes explícitas en un sentido o en otro. De alguna forma, puede evocarnos la capacidad de aprendizaje, y, de hecho, aunque no se defina de forma expresa en estos términos, es esta capacidad de adaptación o aprendizaje la que se desprende de la definición a nuestro juicio más completa sobre autonomía en los sistemas.

Así, Williams nos habla, al caracterizar este funcionamiento autónomo en los sistemas, de una capacidad para completar una tarea sin intervención humana, utilizando comportamientos resultantes de la interacción de la programación informática con el entorno externo. Las tareas o funciones pueden realizarse, de hecho, utilizando una variedad de comportamientos, que pueden incluir el razonamiento, la resolución de problemas, la adaptación a situaciones inesperadas y la autodirección, así como el aprendizaje [105].

Aunque la literatura especializada, primordialmente en el campo de las armas autónomas letales, ha desarrollado muchas y muy variadas clasificaciones sobre los posibles niveles de autonomía en función de qué tareas concretas corresponden al humano y cuáles al sistema [106], no parece que aporten mucho al presente trabajo, pues, además de haber sido duramente criticados por la doctrina por su escasa aplicabilidad práctica [107], no tiene demasiado sentido aludir a ellas de cara a terminar de fijar una definición genérica de lo que debe entenderse por sistema de inteligencia artificial, tarea primordial que guía esta

---

[104] Williams, A. P., *op. cit*. N. 76, p. 57. Cabe destacar en este punto que utilizamos la terminología «autonomía operativa» y no «autonomía técnica», vocablo muy común entre la doctrina que niega a la autonomía de la IA un sentido filosófico fuerte (así, Ibold, V., *Künstliche Intelligenz und Strafrecht. Zur strafrechtlichen Produktverantwortung in der Innovationsgesellschaft,* Nomos, Munich, 2024, p. 167 (y con más referencias en nota al pie 820), simplemente con el objeto de destacar que, más allá de la mera técnica, habrá otros rasgos (como el grado de relación con el humano) que determinarán este concepto.

[105] *Idem*., p. 56-57.

[106] National Institute of Standards and Technology (NIST), Hui-Min Huang (Ed.), «Autonomy Levels for Unmanned Systems (ALFUS) Framework», Volume I: Terminology, Version 2.0, *NIST Special Publication 1011-I-2.0*, 2012; Parasuraman, R., Sheridan, T. B./Wickens, C. D., «A Model for Types and Levels of Human Interaction with Automation», *IEEE Transactions on Systems, Man, and Cybernetics*–Part A: Systems and Humans, 30, 3, 2000, pp. 286-297. Para ver una explicación sobre las tipologías de escalas (lineales, multidimensionales y contextuales), propuestas por la doctrina: Williams, A. P., *op. cit*. N. 76, pp. 39-52.

[107] La escasa aplicabilidad práctica de la diferenciación por niveles detallados no ha sido la única crítica. Así, Bradshaw et al, explican en su trabajo las diferentes razones por las cuales hay un consenso entre los técnicos y los académicos al entender que esta no es una tarea útil. Bradshaw, J., *et al. op cit*. N. 80, p. 56.

primera parte de la investigación. Por ello, amén de dejar constancia de esta realidad, no nos detendremos en las diversas tipologías y diferenciaciones en cuanto a los grados de autonomía. Debe reconocerse, pues así lo hace también el Reglamento de IA al definirla, que la autonomía se desarrolla en diferentes niveles, pero más allá de reconocer que sus características pueden darse en mayor o menor grado en función de los parámetros arriba estudiados (control del humano, funciones a desarrollar y complejidad técnica), no parece que sea necesario marcar rangos estáticos en este primer estadio.

Una vez realizadas todas las anteriores consideraciones merece la pena detenerse en la pregunta inicial: ¿qué debemos entender por autonomía cuando nos referimos a un sistema de IA definido en los términos que aquí se proponen?

Recapitulando el contenido de la segunda parte de este capítulo, merece la pena hacer un repaso por los razonamientos ofrecidos y distinguir aquello que debe quedar extramuros de nuestro concepto débil de «autonomía» en sentido operativo. Así, lo que de ninguna forma se cuestiona es que un sistema dotado de autonomía no tiene, (o al menos no actualmente), la capacidad de generar sus propios objetivos o planes vitales, es decir, no puede constituir lo que se entiende de forma consensuada un agente autónomo a efectos filosóficos y con capacidad de responsabilidad moral (y, por añadidura, penal). No se puede afirmar que los sistemas más complejos técnicamente sean capaces de «vivir la vida conforme a sus propias razones y motivos» [108].

¿Y cuáles son, entonces, los rasgos que sí quedan incluidos? Todo lo que tenga que ver con la autonomía entendida como capacidad para actuar, en el sentido propuesto por Scharre y por Willliams, que implica que la autonomía no es un fenómeno que se pueda predicar de forma global para un sistema, sino solo para alguna de sus funciones (Williams), implicando una triple perspectiva:

– Una independencia relativa en cuanto al control que puede ejercitar el humano sobre su actuación.

– La posibilidad de actuar en este sentido solo para un determinado tipo de funcionalidades o tareas (marcadas por el humano), de forma que cuanto más diversos sean los contextos en que se pueden desplegar sus acciones sin ese control humano más autónomo será el sistema [109].

---

[108] Christman, J., *op. cit*. N. 77.
[109] Danaher, J., *op. cit.* N. 43, p. 300.

– La complejidad técnica que permite al sistema completar sus tareas analizando el entorno en el que actúan y tomando decisiones no determinadas y no previstas por el ser humano (autoaprendizaje y adaptabilidad).

Una vez asentado este término, lo único que resta para finalizar este capítulo es comprender por qué se ha hecho tanto hincapié en este rasgo de los sistemas, siendo que es uno de aparición relativamente reciente y que, como se ha visto, supone confrontaciones y discusiones interminables en cuanto a su determinación y su delimitación.

Lo cierto es que se considera, y así se desarrollará en el capítulo siguiente, que precisamente el hecho de que los sistemas tengan esta característica es lo que permite afirmar un salto cualitativo de las tecnologías más avanzadas respecto de las precarias IAs basadas en paradigmas anteriores.

Lo que aquí se defenderá es que precisamente este tipo de autonomía en los sistemas actuales es la que permite afirmar la existencia de un elemento diferenciador con consecuencias jurídicas, pues, tal y como afirma Nyholm, solo habrá problemas relacionados con la responsabilidad cuando de la autonomía del sistema se derive una imprevisibilidad y falta de control por parte del humano [110].

Lo cierto es que el debate sobre la autonomía ha resultado continuamente enturbiado por la querencia de algunos autores de afirmar la responsabilidad propia de estos sistemas, dando por hecho que tener autonomía significa ser responsable por las acciones. Sin embargo, es importante aclarar de forma rotunda desde el inicio de este trabajo que este tipo de autonomía completa se dará solo si hay autonomía entendida en sentido filosófico/jurídico [111] o si lle-

[110] Nyholm, S., «Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility», *Science and Engineering Ethics*, 24, 2018, p. 1206. Para este autor, la autonomía debe ser entendida en un sentido «no trivial», desarrollando un argumentario sobre los diferentes tipos de agencia, concluyendo que el elemento diferencial para saber si existe agencia individual (solo del humano) o colectiva (que implica, inevitablemente, que una parte de esa agencia sea afirmada de la máquina), es precisamente saber quién fija lo que él denomina el «*primary goal*», que se puede identificar con lo que aquí se consideran los objetivos de primer orden o planes vitales.

[111] Se debe compartir en este punto la afirmación de Teubner y Beckers cuando aseveran que la definición de la autonomía dependerá, en todo caso, del campo de estudio que quiera definirla. Así sucede desde los filósofos más apegados a la información, como es el caso de Floridi, que pone el énfasis en las capacidades de interacción (baste aquí este esbozo simplificado), hasta algunos autores más orientados a aproximaciones humanas (que confían en elementos como la inteligencia, la habilidad para aprender, la autoconciencia, la autorregulación moral, la capacidad de sufrir etc.). Lo que destacan los autores es muy interesante: aparte de la necesidad de tomar partido por alguna de las concepciones, es curioso explorar cómo cada disciplina pone el énfasis en algún tipo de razonamiento orientado a la lógica de su materia. Así, explican que los científicos causales interesados en la explicación y la predicción hablan de autonomía solo si el modelo es una *black box,* porque en ese caso ya no pueden analizar las relaciones causales sino solo observar su comportamiento externo. Por el contrario, los que interaccionan con las ciencias

garan a tener capacidad de elección sobre sus objetivos (libertad de elección). Por el momento, no conlleva más que la constatación (bastante pacífica), de que el control humano disminuye según aumenta la autonomía en sentido débil, y esto dificulta la imputación de daños al único sujeto responsable, que es el humano (esto será objeto de discusión en el capítulo segundo).

Por tanto, más que hablar de autonomía en genérico, y con el objetivo de diferenciarla de la tradición filosófica, se aclara que, de aquí en adelante, cuando se hable de «autonomía» en los sistemas, se hará teniendo en cuenta la definición aquí asentada (que se ha venido a denominar autonomía operativa o en sentido débil), asumiendo en todo caso como cierta la crítica de que, filosóficamente hablando, el término «autonomía» no es apropiado y que, por tanto, aunque se use el mismo por la convención ya asentada, no puede estar dotado del mismo significado. Se asume también que, a pesar de que los elementos señalados sirven para caracterizar la autonomía, en ningún momento se pretende afirmar que la misma sea una cualidad de naturaleza dicotómica, que se tenga o no se tenga y que implique un todo o nada en cuanto a su concreta forma de desarrollo, sino que nos situamos en la idea de que debe ser entendida en todo caso como una dimensión de tipo continuo [112].

No pretendemos imputar responsabilidad al propio sistema, en cuyo caso tendría pleno sentido profundizar en cuestiones relacionadas con la adscripción de agencia moral y responsabilidad plena al agente autónomo [113], sino solo describir el tipo de relación humano-máquina que consideramos es sufi-

---

sociales y las humanidades confían en la autonomía constitutiva de los actores; pero una vez más con diferencias. Los economistas destacan las decisiones orientadas a la utilidad, definiendo la autonomía como elección racional, mientras los éticos de la moral tienden a ver la autonomía en forma de consciencia digital. Se comparte la afirmación que sigue, que implica que la ley debe definir la autonomía conforme a sus propias premisas normativas. Beckers, A., /Teubner, G., *op cit.* N. 111, p. 33 y ss. Para estos autores existe ya a día de hoy lo que llaman un «riesgo de autonomía», que surge «de la independencia de las «decisiones» de los agentes de software». Para una aproximación al concepto de Floridi: *vid.*: Floridi, L., /Sanders, J. W., *op. cit*. N 63.

[112] Así lo explica Miró, afirmando que esto es algo pacífico entre los filósofos de la mente. *Vid.*: Miró Linares, F., *op. cit*. N. 53, p. 93 (con más referencias en nota al pie 16).

[113] Como así hacen, de hecho, filósofos de la información como Floridi y Sanders. Floridi, L., / Sanders, J. W., *op. cit.* N 63. Incluso Hildebrandt cuando reconoce la importancia de modificar en cierta parte el concepto de agente, para introducir las reflexiones de la corriente de pensamiento de Bruno Latour, que prefiere hablar de «actuantes» («*actants*») y así incluir a todo aquel sistema, humano o no humano, del que se pueda predicar inteligencia ambiental. *Vid.*: Hildebrandt, M., «Ambient intelligence, criminal liability and democracy», *Criminal Law and Philosophy*, 2, 2008, *passim.* González Granado explica en su monografía cómo estas concepciones beben de una forma de concebir la cognición como inseparable de la interacción con el mundo, bebiendo de la denominada concepción de la «acción distribuida moralmente», que no para de ganar adeptos en el campo de la IA. Para más detalle de esta corriente de pensamiento, *vid.*: González Granado, J., *De la persona a la personalidad algorítmica*, Publicacions i Edicions de la Universitat de Barcelona, 2020.

## CAPÍTULO SEGUNDO. LAS LAGUNAS O NECESIDADES DEL DERECHO PENAL ANTE LA IA

*FRONDOSO:*

*Justo es que honores le den.*
*Pero, decidme, mi amor,*
*¿quién mató al Comendador?*

*LAURENCIA: Fuente Ovejuna, mi bien.*

*FRONDOSO: ¿Quién le mató?*

*LAURENCIA:*

*Dasme espanto.*
*Pues Fuente Ovejuna fue* [114].

Lope de Vega nos narra en su famosa obra teatral una de las más emblemáticas lagunas de responsabilidad que a la historia española concernieran. La muerte del comendador de la aldea, don Fernán Gómez, sucedió en Fuente Ovejuna en 1476, y el dramaturgo adapta aquí una de las versiones que la crónica de la época ofreció, en la que toma como base el dicho que Sebastián de Covarrubias recogía ya en su «Tesoro de la lengua Castellana». Así, sobre este célebre dicho, «Fuente Ovejuna lo hizo», nos explica: *«Fueron enviados jueces pesquisidores, que atormentaron a muchos, así hombres como mujeres; no*

[114] Fragmento de la obra teatral de Lope de Vega *Fuenteovejuna,* acto I. Edición: Fuente Ovejuna, edición de Rinaldo Froldi, apéndice de Abraham Madroñal, 19.ª ed., Madrid, Espasa-Calpe, Colección Austral, 1998.

*les pudieron sacar otra palabra más desta: Fuente Ovejuna lo hizo»*. Toda una población, explica Covarrubias, «*con atrevida y vengativa mano / sin Dios, sin Rey, sin ley, toda se aúna / de hecho a un hecho bárbaro, inhumano»*, confesándose culpable, obligando al juez nombrado para resolver la cuestión a absolver a todos ellos, quedando el terrible crimen sin castigo ni responsable. Se cumple así el conocido emblema de Lucano: *Quidquid multis peccatur, inultum est,* es decir, «el pecado que cometen muchos, queda impune» [115].

Aunque en la historia más reciente lo difícil sería, probablemente, encontrar tan heroicos actos de solidaridad ciudadana y autoinculpación colectiva, el punto de conflicto que a lo largo de este capítulo se tratará tiene mucho que ver con el del célebre pasaje de Lope. Muchos son los autores que hoy en día, tras el avance de la tecnología y la complejidad de los delitos que se vienen cometiendo con estas herramientas, temen que, tras la muerte del comendador, los posibles culpables callen, quizá no en forma heroica y solidaria, sino más bien excusándose en la interposición de fuerzas incontrolables y para ellos desconocidas, llevando a estos nuevos escenarios donde los pecados de muchos, y las muchas manos [116] involucradas en los nuevos delitos, lleve a una impunidad tan insoportable como la generada por la muerte del comendador.

## I. LAS LAGUNAS DEL DERECHO PENAL

### 1. ¿QUÉ ENTENDEMOS POR LAGUNA DE RESPONSABILIDAD?

A lo largo del capítulo mostraremos de qué manera los autores se han ocupado de una de las preocupaciones más comunes cuando se habla de daños genera-

[115] Esta explicación histórica y contextual acerca de la obra la podemos encontrar en la Introducción a la misma a cargo de Rinaldo Froldi. Vega, L.d., Fuenteovejuna, Rinaldo Froldi (Ed.), Alicante, Biblioteca Virtual Miguel de Cervantes, 2002. Otras referencias: Covarrubias, S., /Noydens, R./ Riquer, M. de, *Tesoro de la lengua castellana o española: según la impresión de 1611, con las adiciones de Benito Remigio Noydens publicadas en la de 1674*, Horta, Barcelona, 1943.

[116] El término hace referencia a la expresión, ya muy extendida, de que los supuestos de IA implican lo que se conoce como un problema de las muchas manos *(«many hands»).* A modo de ejemplo y vinculando el fenómeno al ámbito de la responsabilidad empresarial: Diamantis, M. E., «Employed algorithms: labor model of corporate liability for ai», Duke *Law Journal*, 72(4), 2023, p. 806 y ss. También, Coeckelbergh, M., «Artificial intelligence, responsibility attribution, and a relational justification of explainability», *Science and engineering ethics*, 26(4), 2020, p. 2056: «uno de los problemas de la acción tecnológica es que suele haber muchas personas implicadas causalmente en la acción, lo que hace difícil, si no imposible, (a) encontrar al individuo responsable, si sólo hay un individuo responsable, o (b) responsabilizar a cualquier individuo: puede haber más personas responsables, ya que hay muchos agentes implicados. Este problema de atribución y distribución de responsabilidades se conoce a veces como «el problema de las muchas manos».

dos por sistemas de inteligencia artificial, el fenómeno aquí bautizado como «lagunas de responsabilidad» [117], que ha ocupado el debate desde su formulación en 2004, sin que las ideas hayan sido demasiado clarificadas en el ámbito continental.

Dicho concepto peca de amplio y poco delimitado, por lo que el empeño del presente capítulo será tratar de desentrañar qué tiene de relevante para el derecho penal contemporáneo hablar de lagunas en una disciplina donde, en principio, aquello que no esté explícitamente previsto por la *stricta* ley penal [118], debe quedar extramuros de nuestro sistema.

Los autores han venido pensando, desde un incipiente escrito del alemán Andreas Mathhias [119], acerca de este «*Verantwortungslücke*» en términos diversos. Llama la atención, sin embargo, cómo este fenómeno ha sido especialmente tratado por la literatura anglosajona y casi totalmente olvidado por la literatura hispana, que no ha dedicado, por el momento, un esfuerzo explicativo acerca de este fenómeno y sus consecuencias. Se intentará, a lo largo de estas líneas, explicar qué echa en falta la literatura dominante y qué debe preocupar al derecho penal continental de hoy.

La preocupación acerca de las lagunas de responsabilidad que presentan los sistemas actuales de IA tiene que ver con la idea de que, ante la ocurrencia de un resultado dañoso (en un sentido jurídico penal, es decir, un hecho con apariencia delictiva, ya sea en forma de daño o puesta en peligro a un bien jurídico merecedor de protección), en el que de una forma u otra se ha utilizado un sistema dotado de IA, es difícil encontrar, *prima facie,* una persona a la que imputar dicho daño. Danaher lo expresa de manera sencilla diciendo que esta

---

[117] Los términos utilizados son variados, aunque quizá el que se ha establecido con más fuerza es el relativo al *«responsibility gap»,* que hace referencia al fenómeno más genérico, si bien encontramos referencias a *«liability gap», «accountability gap»* o, en un sentido mucho más acotado que se explicará más abajo, *«retribution gap» (vid. infra).* En la literatura en lengua española, si bien escasa, podemos encontrar los términos laguna, vacío o incluso brecha, habiéndose escogido el vocablo laguna por considerarse el más común en el lenguaje jurídico-penal. En todo caso, se elige el término responsabilidad y no otros (autores como Navarro proponen hablar de punibilidad) porque se entiende que es el que captura de forma más genérica todo el espectro de problemas que implica este fenómeno. Más adelante se explicarán las diferentes acepciones de esta genérica «responsabilidad» para tratar de aclarar qué aspectos realmente importan. Así, Navarro Dolmestch, realizando un repaso por los autores que han tratado este fenómeno y aportando un intento de delimitación del término. *Vid.*: Navarro Dolmestch, R., «Brechas de responsabilidad penal por la actuación de máquinas dotadas de inteligencia artificial». *Polít. Crim*., vol. 20, 40, 2025, pp. 77-107.

[118] Nos referimos, desde luego, a la formula latina con la cual se enuncia uno de los mandatos del principio de legalidad, principio básico de nuestro derecho penal constitucional. Para una profundización en este principio esencial, por todos: Pérez Manzano, M., Lascuraín Sánchez, J. A., (Dirs.)., *La tutela multinivel del principio de legalidad penal,* Marcial Pons, 2016.

[119] Matthias, A., «The Responsibility Gap., *op. cit.* N. 40. Más adelante, desarrollando de forma más completa su tesis: Matthias, A., *Automaten als Träger von Rechten, Plädoyer für eine Gesetzänderung*, Humboldt Universität, Berlin, Logos Verlag, 2007.

laguna aparece cuando la gente busca un infractor al que culpar por un hecho perjudicial, pero no se puede encontrar [120].

Merece la pena realizar aquí otra aclaración terminológica previa, para sortear una confusión que puede resultar común. Danaher y otros autores no hablan propiamente de laguna en la responsabilidad, sino de laguna en la retribución *(retribution gap).* No debemos confundir los argumentos como el expuesto arriba, que hablan de retribución como forma de expresar que lo que se pretende es imponer un castigo (moviéndonos, por tanto, dentro de los paradigmas de la responsabilidad jurídica penal, eminentemente subjetiva y dependiente de las finalidades atribuidas a la pena), con los argumentos que tienen que ver con esta laguna en un plano meramente reparatorio como paradigmáticamente el civil (aquí sí, se utiliza el genérico vocablo de laguna de responsabilidad; *«liability gap»).* Tampoco se debe confundir con los autores que hablan de laguna de retribución en un sentido completamente ajeno al aquí explorado, relacionado con la posibilidad de que sea el propio sistema de IA quien reciba el castigo [121]. A lo largo de este capítulo solo se hablará de laguna en cuanto a la posible dificultad de encontrar un humano al que someter a responsabilidad penal. Así, para el presente trabajo, la laguna de responsabilidad será laguna jurídico-penal y siempre referida al castigo a imponer al ser humano.

La idea fuerza que se deriva en todos los autores que han tratado el tema tiene que ver con que el nivel de autonomía alcanzado por los sistemas más complejos, que dota de imprevisibilidad a las acciones derivadas de sus actuaciones, hace que se rompa el vínculo que requiere la atribución de responsabilidad [122]. Formulado de manera simple: es ese vínculo entre el humano y el

---

[120] Danaher, J., *op. cit*. N. 43.

[121] Así, paradigmáticamente, ocupándose de este problema de imponer castigo a la propia máquina, y poniendo de manifiesto sus limitaciones: Asaro, P., «A Body To Kick, but Still No Soul to Damn: Legal Perspectives on Robotics», en Lin E. P./ Abney, P. K./Bekey G. A., (Eds.), *Robot Ethics. The Ethical and Social Implications of Robotics*, Cambridge, 2012, p.182. También, reflexionando sobre las cuestiones relativas a las necesidades que requiere el castigo, y proponiendo que un botón de apagado en las máquinas fuera suficiente para cumplir con la función expresiva del castigo: Turner, J., *op. cit.* N. 5, p. 360. También Salvadori, I., *op. cit*. N. 101, p. 154.

[122] Así, Koops *et al.*, que afirman que «las nuevas entidades de la sociedad de la información, […], crean un «vacío de responsabilidad», porque operan a una distancia cada vez mayor de sus mandantes» (traducción propia). En: Koops, B. J., Hildebrandt, M., / Jaquet-Chiffelle, D. O., *op. cit.* N. 79, p. 7. También hablando de la ruptura de ese vínculo entre la decisión humana y el acontecimiento por la presencia de la IA que provoca un acontecimiento imprevisto, Consulich. P. 1037: «*La ragione dell'insufficienza della regolazione contemporanea è purtroppo semplice. La presenza di un'AI è un grado di recidere il nesso tra decisione umana ed evento, tute le volte in cui il soggetto artificiale abbia agito discostandosi di propria iniziativa dalle istruzione ricevute, tanto da realizzare un fatto non voluto o addirittua diverso da quello avuto di mira da parte dell'agente umano. Ciò è reso possibile, anzi frequente, dalla complessità del software e dalla capacità di apprendimento dall'ambiente sempre più spiccata (è il denso capitulo del machine learning)*». Consulich, F., «Flash offenders. Le prospettive di accountability penale nel contrasto alle Intelligenze Artificiali devianti», *Rivista italiana di diritto e procedura penale*, 2022, p. 1037.

fruto de su actuación el que fundamenta la responsabilidad. Si ese vínculo se debilita, la afirmación de la responsabilidad será más costosa. Así, Salvadori afirma que, en el caso de los agentes artificiales más autónomos, la dificultad (o incluso imposibilidad) de que los hombres ejerzan control sobre su trabajo puede llevar a un vacío de responsabilidad incluso en materia penal [123].

Otros autores destacan otros elementos de la tecnología como conformadores de esta laguna, pero no dejan de ser elementos derivados de esta autonomía (en los términos explicados en el capítulo anterior). Así, la adaptabilidad [124], la complejidad de sus tareas y la proliferación de contextos de uso [125], la falta de control [126], y, finalmente, la impredecibilidad.

Para el filósofo Andreas Mathhias, que es quien abre la discusión acerca de esta problemática, hay un vacío de responsabilidad en una clase creciente de accidentes causados por máquinas, en los que las formas tradicionales para atribuir responsabilidad no son compatibles con el sentimiento de justicia, ya que nadie tiene suficiente control sobre las acciones de la máquina para poder asumir la responsabilidad. De esta forma, los agentes de software presentan una brecha de responsabilidad porque son impredecibles *(unberechenbar)* y porque actúan fuera del horizonte de visibilidad de su creador [127].

No solo la doctrina penalista y civilista ha puesto de manifiesto esta preocupación acerca de la afectación a la responsabilidad tal y como la conocemos con la irrupción de estas entidades, sino que las propias instituciones europeas han manifestado inquietud al constatar que el desarrollo de las herramientas tecnológicas cada vez más complejas podría suponer problemas para las estructuras de imputación clásicas.

---

[123] Salvadori, I., *op. cit*. N. 101, p. 149.

[124] Así, Busato habla de la posibilidad que tiene la IA de evolucionar por sí sola, según su propia experiencia y dentro del ambiente real, como el punto en el que empiezan a surgir los supuestos de «*responsibility gap*». Busato, P. C., «De máquinas y seres vivos: ¿Quién actúa en los resultados delictivos derivados de decisiones cibernéticas?», en Demetrio Crespo, E. (Dir.), *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant Lo Blanch, Valencia, 2022. p. 351.

[125] González Granado, J., *op. cit.* N. 113 p. 73.

[126] En este sentido, muchos son los autores. A título de ejemplo, en el ámbito español, Quintero Olivares, que se refiere a este problema, aunque de forma no explícita, cuando habla de la existencia de un vacío de respuesta jurídica a las desviaciones incontroladas de los sistemas, estudiando como escenario el de la máquina que lleva a cabo acciones no deseadas pero cuya evitación no es posible por el humano, partiendo de la autonomía como capacidad de desviarse de la actuación y realizar acciones no previsibles. *Vid.:* Quintero Olivares, G., «La robótica ante el derecho penal el vacío de respuesta jurídica a las desviaciones incontroladas», *Revista Electrónica de Estudios penales y de la seguridad, REEPS,* (1), 1, 2017*, passim.* También, Beckers, A., /Teubner, G., *op cit.* N. 111, p. 3.

[127] Dicho fragmento se encuentra traducido al inglés en: Koops, B. J., Hildebrandt, M., / Jaquet-Chiffelle, D. O., *op. cit.* N. 79, a partir de p. 36. El original, en alemán, puede encontrarse en Matthias, A. (2007), *Automaten als Träger von Rechten*, *op cit*. N. 119, p. 22. Traducción al español propia.

Uno de los más llamativos intentos del Parlamento Europeo por poner a la Comisión sobre aviso de este fenómeno se produjo con la conocida Resolución de 16 de febrero de 2017 con Recomendaciones a la Comisión sobre normas de Derecho civil en materia de robótica [128]. En dicho documento, la Eurocámara expresó toda una serie de preocupaciones acerca de cómo se podrían modificar los mecanismos de responsabilidad habida cuenta de la expansión de la robótica.

A lo largo de los considerandos del texto se pone de manifiesto que esta tecnología ha ido desarrollando una serie de rasgos que antes eran exclusivos del ser humano, entre los cuales destaca la capacidad de aprender de la experiencia y tomar decisiones cuasi independientes (considerando Z), haciendo que se asimilen cada vez más a agentes que interactúan con su entorno y pueden modificarlo de forma significativa. Se señala como crucial la cuestión de la responsabilidad jurídica por los daños que pueda ocasionar la actuación de los robots. El problema que se suscita es que cuanta más autonomía, menor será la posibilidad de considerarlos meros instrumentos, y ello da lugar a la pregunta sobre si la normativa de responsabilidad es suficiente o si se requieren normas específicas para aquellos «actos y omisiones de los robots cuya causa no pueda atribuirse a un agente humano concreto» [129].

Hasta aquí la preocupación manifestada por el Parlamento parece estar en línea con lo puesto de manifiesto por la doctrina, coincidiendo en lo esencial con los postulados teóricos de este *«responsibility gap»*. Sin embargo, la Eurocámara va mucho más allá cuando solicita que, en consideración a los rasgos anteriormente mencionados, y teniendo en cuenta la posible insuficiencia de las normas actuales, con el temor de que no se pueda encontrar la forma de reparar los daños ocasionados [130], se tome en consideración la opción de dotar a los robots de personalidad electrónica [131].

---

[128] Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL).

[129] *Idem.,* Considerando AB.

[130] Lo que se considera insuficiente en este texto es la realidad de la normativa sobre responsabilidad contractual y extracontractual en los casos de daños de carácter civil. Las repercusiones de estas advertencias no han tardado en notarse, porque desde las instituciones se ha impulsado la creación de nuevas leyes de responsabilidad extracontractual para poder incluir los daños causados por sistemas de IA. Para más información *vid.:* Gómez Ligüerre, C., «La Propuesta de Directiva sobre responsabilidad por daños causados por productos defectuosos*», InDret,* 4/2022. Sin embargo, y a pesar de que la UE no se ha pronunciado sobre asuntos penales, se considera que estos documentos son expresivos del sentir generalizado de los gobernantes europeos. *Vid.:* Parlamento Europeo, Civil Liability Regime for Artificial Intelligence, Resolución de 20 de octubre de 2020, 2020/2012(INL), Informe a la Comisión de la UE sobre la responsabilidad de la IA: Comisión Europea, «Informe sobre las implicaciones en materia de seguridad y responsabilidad de la inteligencia artificial, Internet de las cosas y la robótica», COM(2020) 64 final, 16.

[131] Considerando 59, apartado f): *«Pide a la Comisión que, cuando realice una evaluación de impacto de su futuro instrumento legislativo, explore, analice y considere las implicaciones de todas las posibles soluciones jurídicas, tales como: f) crear a largo plazo una personalidad jurídica específica para los ro-*

Es a partir de aquí donde podemos apreciar una división en la doctrina, pues, aunque la reacción mayoritaria es una clara negativa a aceptar esta posibilidad, lo cierto es que han ido surgiendo voces (principalmente dentro del campo de la filosofía y de la ética de la información, pero también desde el punto de vista jurídico), que recientemente se han manifestado a favor de esta posibilidad [132].

Sin embargo, esta pretensión es abandonada de forma temprana, pues en una posterior resolución de 2020 el Parlamento parece haber cambiado su percepción en cuanto a la personalidad jurídica. No obstante, sigue manteniendo esa preocupación por todos aquellos atributos novedosos de la IA que pueden requerir alguna modificación con el objetivo de evitar que ningún perjudicado o dañado por estas tecnologías quede sin reparación.

Así, manifiesta la Eurocámara que «*la complejidad, la conectividad, la opacidad, la vulnerabilidad, la capacidad de ser modificados mediante actualizaciones, la capacidad de autoaprendizaje y la autonomía potencial de los sistemas de IA, así como la multitud de agentes involucrados representan un reto importante para la eficacia de las disposiciones del marco de responsabilidad civil de la Unión y nacional*» [133]. Sin embargo, esta vez reconoce que todas las actividades de los sistemas son casi siempre el resultado de que una persona los haya construido, y por lo tanto, no es necesario atribuir personalidad a los mismos, sino que se puede responsabilizar a las diferentes personas de toda la cadena de valor: los que crean, los que mantienen y los que controlan el riesgo asociado al sistema [134].

---

*bots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente.»* Por supuesto, las reacciones en contra de dicha petición no tardaron en sucederse, y, de hecho, ha sido la última vez que el Parlamento plantea dicha cuestión, no así en la doctrina.

[132] Solo a efectos civiles, Teuber. *Vid.:* Teubner, G., «¿Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten», *Archiv Für Die Civilistische Praxis*, 218, 2018, pp. 7-8. También, aunque con un enfoque cauteloso: Koops, B. J., Hildebrandt, M., / Jaquet-Chiffelle, D. O., *op. cit.* N. 79, *passim*. Incluso el propio Matthias, A., «The Responsibility Gap., *op. cit.* N. 40. También, Chopra, S./White, L. F., *op. cit. N. 132*, p. 156. En general, los argumentos a favor de afirmar la posibilidad de esta personalidad algorítmica se circunscriben, sin embargo, al ámbito civil, donde la cuestión a determinar es quién es responsable de los daños. En el ámbito penal, en cambio, son pocas las voces que han planteado esta posibilidad. Por parte de todos, y muy especialmente de Gabriel Hallevy a lo largo de toda su obra (se profundiza en ella *infra,* en capítulo 2). También: Hu, Ying, «Robot criminals», *University of Michigan Journal of Law Reform,* 52(2)*,* 2019, pp. 487-532, y Mulligan, C., «Revenge against robots», *South Carolina Law Review*, 69(3), 2018, pp. 579-596.

[133] Parlamento Europeo, Civil Liability Regime for Artificial Intelligence, Resolución de 20 de octubre de 2020, 2020/2012(INL., Recomendación n.º 6.

[134] *Idem*, recomendación n.º 7.

El problema queda, entonces, reducido a la cuestión que al inicio se apuntaba: ante un hecho que, *prima facie,* es perjudicial para algún interés jurídico penal, necesitamos encontrar la forma de imputar responsabilidad al humano, que aparece alejado o desvinculado de alguna forma del hecho original, debido principalmente a las características de autonomía y descontrol de la herramienta.

Para ahondar más en el fondo del problema, resulta útil acotar sus términos. ¿Qué tipo de responsabilidad es aquella que se ve dificultada con la interposición de un sistema dotado de IA? Resulta esencial, pues, clarificar, aunque sea someramente, los tipos de responsabilidad a que la literatura se refiere cuando se refiere a estas lagunas.

## 2. ¿QUÉ ENTENDEMOS POR RESPONSABILIDAD?

Partimos, aquí, de la innegable idea de que la responsabilidad es un «genérico y forzosamente vago criterio anudador» [135], y no podemos más que compartir, con Ricoeur [136], la «perplejidad en que [nos] ha sumido el examen de los empleos contextuales contemporáneos del término responsabilidad». Expresa Figueroa, estudiando la obra de Ricoeur, que, *«aunque en el uso jurídico clásico del término, que se remonta al siglo xix, parece bien establecido –con la «responsabilidad» caracterizada en términos de obligaciones tanto civiles como penales (ya sea la obligación de compensar los daños o la obligación de someterse al castigo)– el concepto filosófico es confuso y algo vago»* [137].

Aunque no es objeto de este trabajo realizar un estudio en profundidad sobre el término, por exceder con mucho las pretensiones de este capítulo, sí se considerarán algunas de las acepciones más comunes, exploradas en diversas obras de la literatura sobre inteligencia artificial y derecho, pues se considera que son capaces de explicitar las razones que hay detrás de las preocupaciones relacionadas con las lagunas que genera la IA.

---

[135] Molina Fernández, F., «Presupuestos de la responsabilidad …», *op. cit*. N. 84, p. 58.

[136] Ricoeur, P., *Reflections on the just*, trad. David Pellauer Chicago, University of Chicago Press, 2007. p. 11.

[137] Figueroa Rubio, S., *Adscripción y reacción: Responsabilidad jurídica y moral desde una perspectiva interpersonal*, Marcial Pons, Ediciones Jurídicas y Sociales, Madrid, 2019, p. 28. Asimismo, Gunkel, haciendo un repaso por los orígenes del término, se sorprende de que «un término con un sentido tan firme en el plano jurídico sea de origen tan reciente y no esté realmente bien asentado dentro de la tradición filosófica» (traducción propia). Gunkel, D. J., *«Mind the gap: responsible robotics and the problem of responsibility», Ethics and Information Technology,* 22, 2020, p. 307.

El autor más comúnmente citado cuando se trata de cartografiar la responsabilidad es H. L. A. Hart.[138] Este filósofo, en palabras de Gardner, supo entender, junto con Feinberg, que toda la bibliografía existente hasta el momento estaba sumida en la confusión, tanto la relativa al plano jurídico como al plano moral[139]. Distingue Hart hasta cuatro sentidos de responsabilidad:

– Responsabilidad por el rol *(Role-Responsibility):* es la responsabilidad entendida como competencia sobre un determinado ámbito, rol o posición social, determinada por las relaciones con otras personas y la posición que da lugar a determinados deberes. Es una responsabilidad de tipo prospectivo, es decir, obliga a realizar tareas en un futuro[140]. No tiene demasiada relevancia en la discusión sobre las lagunas de responsabilidad causadas por los sistemas de IA, pues el problema aquí presente tiene que ver con cómo asignar responsabilidad por un hecho que ya ha ocurrido, por lo que nos movemos en el campo de una responsabilidad de tipo retrospectivo[141]. No ha sido obviada, sin embargo, este tipo de responsabilidad en la literatura sobre IA[142].

– Responsabilidad causal *(causal responsibility):* se refiere a la responsabilidad en el sentido de relación entre dos eventos, de tal forma que uno no hubiera acontecido sin que el otro se hubiera producido. Veremos cómo uno de los problemas de estas pretendidas lagunas puede tener que ver con la opción de que, dadas las características de autonomía y complejidad de las herramien-

---

[138] Hart, H. L. A., *Punishment and Responsibility. Essays in the Philosophy of Law (with an Introduction by John Gardner,* Oxford, Oxford University Press, Oxford, 2.ª ed. 2008.

[139] Esta opinión es compartida por Figueroa Rubio que, pese a proponer algún otro criterio o tipología, no deja de reconocer las cuatro categorías propuestas por Hart como punto de partida esencial para conceptualizar la responsabilidad. *Vid.*: Figueroa Rubio, S. *op cit.* N. 137, p. 41, nota 17.

[140] Figueroa Rubio, S., *ibid*, p. 43.

[141] Esta es, en esencia, la distinción (responsabilidad prospectiva versus retrospectiva) que Molina cree que aporta mayor claridad, y la que utiliza en su trabajo sobre responsabilidad y agencia. *Vid.:* Molina Fernández, F., «Presupuestos de la responsabilidad …», *op. cit*. N. 84, p. 58.

[142] La denominada vertiente «relacional» de la responsabilidad es defendida por quienes entienden que las relaciones personales deben estar en el centro de este esquema. Así: Liu, H. Y. /Zawieska, C., «From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence», *Ethics and Information Technology* 22 (4), 2020, pp. 321-333. Filósofos como Coecklebergh han sido asimismo partidarios de esta concepción, siguiendo los postulados esenciales de Duff. *Vid,:* Duff, A. R., «Who is responsible, for what and to whom?», *Ohio State Journal Of Criminal Law*, 2, 2005, pp. 441-461. El mismo: «Legal and moral responsibility», *Philosophy Compass* 4 (6), 2009, pp. 978-986. Coeckelbergh, M., *op. cit*. N. 116. Al ser esta, sin embargo, una postura centrada en los deberes futuros no debe tratarse, todavía, en esta sede. En todo caso, baste dejar apuntado que la responsabilidad relacional entendida como *«answerability»,* es decir, como necesidad de dar explicaciones sobre el acaecimiento de un hecho, la sitúa cerca de la responsabilidad causal. Por ello, es necesario entender que las categorías aquí presentadas no conforman compartimentos estancos con perfiles nítidamente definidos, sino que deben servir para clarificar un concepto altamente abstracto y, *a priori,* inaprensible.

tas de IA, se dificulte la trazabilidad entre el hecho y su antecedente causal humano [143].

– Responsabilidad como sometimiento o sujeción (*Liability-responsibility):* esta sujeción o sometimiento puede ser legal o moral, aunque lo que en esencia implica es la posibilidad de que una persona quede sujeta al actuar de otras, concretamente a la reacción o consecuencias derivadas de una previa incorrección [144]. En definitiva, ser responsable aquí significa ser capaz de soportar alguna consecuencia negativa, y ello presupone estar en condiciones de soportar esa reacción. Esta responsabilidad en el sentido de «asumir las consecuencias» o hacerse cargo de ellas es, asimismo, importante para el debate sobre las lagunas de responsabilidad en los delitos cometidos con IA, pues la primera de las condiciones (ser sujeto apto para la responsabilidad) es negada en el caso de los sistemas, pero debe ser afirmada para los humanos, y precisamente algunas características de la IA pueden complicar la afirmación de esta aptitud.

En este sentido, puede decirse que esta responsabilidad implica la «atribución» o «adscripción» de un hecho y, como derivado, de sus consecuencias. Para poder asumir las consecuencias es necesario que ese hecho se considere como propio (es decir, pueda considerarse, utilizando la metáfora de la que se sirve Feinberg, que algo se puede «cargar a la cuenta de alguien o incluirse en su inventario» [145]). Veremos de qué manera la IA puede hacer que esta atribución se tambalee.

– Responsabilidad como capacidad *(Capacity-Responsibility):* ser responsable en este sentido es contar con todas las características, atributos o capacidades para ser considerado un agente. Sobre los problemas relacionados con la agencia se han escrito ríos de tinta, tanto en la filosofía general como en la literatura sobre inteligencia artificial, y desde luego en la dogmática penal. No es necesario extendernos aquí sobre esta realidad, pues la postura esencial

---

[143] Algunos autores han denominado a este fenómeno fallos de causalidad (*«failures of causation»). Vid.:* Giannini, A., *Criminal behavior and accountability of artificial intelligence systems*, [Tesis doctoral, Universidad de Maastricht, Universidad de Florencia], Eleven Publishers, 2023, p. 11 y ss. «*Pagallo aplica la expresión «fallo de causalidad» para describir el efecto perturbador de la autonomía de los agentes de IA sobre el vínculo entre la agencia y el resultado negativo ocurrido (en un análisis de causa y efecto: si A, entonces B)»* (traducción propia). Se refiere al trabajo de Pagallo, U., *The Laws of Robots. Crimes, Contracts, and Torts,* Springer, Nueva York, 2013, p. 73.

[144] Figueroa Rubio, S., *op. cit*. N 137, p. 44.

[145] Figueroa Rubio, S., *op. cit*. N. 137, p. 35, nota 8. Originalmente en: Feinberg, J., «Action and Responsibility», en Feinberg, J., *Doing and deserving, Essays in the theory of responsibility,* Princeton University Press, Princeton, Nueva Jersey, 1970, pp. 124-125.

que se adopta, y que se ha defendido en el capítulo primero, es que no puede hablarse de agencia en el sentido fuerte cuando se trata de sistemas de IA. No es, por tanto, esta, una cuestión que se vaya a discutir en esta sede [146].

Observamos, por tanto, que, de las acepciones más genéricamente aceptadas dentro de la literatura filosófica, solo dos de ellas importan a los fines de este capítulo.

En primer lugar, los problemas que tienen que ver con la causalidad, y, en segundo lugar, los que tienen que ver con la adscripción de responsabilidad en un sentido de sujeción a las consecuencias.

La doctrina ha tratado estas dos manifestaciones de forma variada, pudiendo distinguir entre aquellos que consideran que, efectivamente, de las características propias de la IA se genera en algún caso un verdadero vacío o laguna de responsabilidad, y quienes, por el contrario, consideran que esta laguna es aparente y que solo se requieren algunos ajustes dogmáticos para seguir afirmando que nuestro esquema de responsabilidad funciona a pesar de las innovadoras características de esta tecnología.

## 2.1 **El argumento de las verdaderas lagunas**

Empezaremos por localizar los argumentos de quienes sí consideran que existe este vacío. En este sentido, Danaher asume que las lagunas se producen en dos campos, partiendo de la obra de Hart: «se necesita una conexión de un acto con un agente, en el sentido de causalidad» y «una conexión con el hecho en el sentido de capacidades para que se pueda adscribir para dar lugar a la imposición de unas consecuencias» [147]. Alude así a la responsabilidad-causalidad y la responsabilidad-sujeción [148].

---

[146] A estos efectos, se comparten los argumentos esgrimidos por Navarro Dolmestch cuando asevera que no hay brecha de responsabilidad en los supuestos en que el legislador ha decidido no pedir explicaciones a la IA como sujeto del derecho penal. El autor ofrece diversos argumentos. Si no hay sujeto al que pedir explicaciones o responsabilizar la idea es que esto se debe a una decisión previa del legislador. Es decir, aunque el hecho no sea atípico hay una valoración (fruto de un consenso social previo) según la cual los entes artificiales no son sujetos a los que atribuir personalidad. Navarro Dolmestch, R., «Brechas de responsabilidad…», *op cit*. N., 117, p. 11 y ss. La postura propia y los argumentos por los cuales se rechaza en esta investigación la posibilidad de la atribución de responsabilidad al sistema se desarrollan en mayor profundidad *infra,* en el capítulo tercero.

[147] Danaher, J., *op. cit.* N. 43, p. 4 y ss.

[148] El punto que defiende Danaher es que la laguna no tendrá las mismas connotaciones cuando de lo que se trata es de imponer responsabilidad penal o civil. Así, afirma, si es civil será compensatoria, si es penal el autor parte de la base de que hay muchos estudios que ponen de manifiesto que el castigo siempre tiene una lógica retribucionista. Danaher se muestra preocupado por el hecho de que se ha prestado demasiada atención a la laguna generada en el campo civil, pero poco a la generada en el campo penal, y que

Los filósofos De Sio y Mecacci se preguntan, para ir más allá de la mera afirmación de una laguna en la responsabilidad, acerca de qué tipo de responsabilidad se amenaza con qué tipo de característica de la IA [149]. Mencionan, de nuevo, las cuatro categorías propuestas por Hart [150], aunque se valen de otras clasificaciones. Estos autores hablan, así, de responsabilidad como culpabilidad y de responsabilidad moral *(«culpability» y «moral responsibility»,* respectivamente [151]).

La responsabilidad en el sentido de culpabilidad se define como «culpa por un delito basada en la intención, el conocimiento o el control», y presenta problemas en relación con la IA en el sentido de que «la IA hace más difícil la predicción y el control, dando lugar a nuevas razones o excusas legítimas para actuar mal, por ejemplo, un accidente de tráfico evitable en el que esté implicado un sistema de conducción automatizada y que nadie pueda predecir o prevenir individualmente» [152]. En el caso de la responsabilidad como responsabilidad moral, se refiere al deber de las personas de explicar sus razones y acciones a los demás. Esta responsabilidad, cuando la IA se ve relacionada, tiene que ver con un problema de explicabilidad, es decir, debido a la complejidad del sistema, la persona que lo usa no es capaz de saber las razones que han llevado a la actuación.

En cierta forma, ambas nociones pueden reconducirse a las más genéricas arriba identificadas como responsabilidad como sujeción y responsabilidad causal, pues se trata de adscribir o imputar un hecho a una persona (en el caso de la culpabilidad, los requisitos para afirmar dicha adscripción vienen determinados por la legislación penal), y ser capaces de explicar cuál es el proceso por el cual un resultado es causalmente reconducible a una determinada actuación humana.

Estos dos aspectos de la responsabilidad están presentes también en un reciente trabajo de Del Rosal, que, implícitamente, centra en ellos los pro-

---

esto puede tener consecuencias importantes para las teorías de la pena. El autor se muestra partidario de un neo-retribucionismo de base empírica, y teme que la lógica utilitarista de la IA nos lleve hacia penas esencialmente prevencionistas, donde encajarían mejor «[S]us llamamientos en favor de un enfoque más consecuencialista y reduccionista del daño en nuestras prácticas de castigo y culpabilización podrían ser mejor escuchadas». (traducción propia). *Idem*., p. 21

[149] Santoni de Sio, F / Mecacci, G.*,* «Four responsibility gaps with artificial intelligence: Why they matter and how to address them», *Philosophy & technology*, 34(4), 2021, pp. 1058-1059.

[150] Estas referencias pueden encontrarse en: Hart, H. L. A., *Punishment and Responsibility, Essays in the Philosophy of Law*, Clarendon Press, Oxford, 1970, pp. 210 y ss.

[151] Santoni de Sio, F / Mecacci, G., *op. cit*. N 149. En realidad se distinguen hasta cuatro tipos de responsabilidad, que se reproducen y definen en el cuadro en p. 61, pero se considera que las que se señalan son aquellas más comunes en la discusión, pues la «*public accountability*» y la «*active responsibility*» pueden reconducirse a fenómenos mucho más estrechos (la falta de explicación de las instituciones públicas y el establecimiento de deberes activos en operadores de la IA, respectivamente), cuestiones que, si bien tienen relación con el tema de la investigación, son reconducibles a cuestiones que se tratarán más adelante.

[152] *Idem,* p. 5.

blemas de las lagunas de responsabilidad, afirmando que son un problema de imputación a la conducta de unos resultados, pues puede que no se pueda afirmar ni siquiera la causalidad entre dichos resultados y la conducta humana [153]. Vemos, aquí, como de lo que se trata es de imputar unos hechos, y que esa adscripción tendrá dificultades tanto a nivel causal como a nivel normativo [154].

Resulta clarificador, una vez centrado el esfuerzo conceptual en torno a qué responsabilidad es aquella que preocupa a la doctrina, saber si realmente tales problemas existen. El trabajo de Köhler, Roughley y Sauer es, en este sentido, clarificador. El punto de partida de esos autores es que la complejidad que han alcanzado las tecnologías es tal que dificulta la atribución de responsabilidad moral [155]. Para estos, el problema está en la responsabilidad en el sentido de responder por los hechos pasados (en términos de rendición de cuentas, *«accountability»),* en un sentido prospectivo, es decir, un hecho, ya ocurrido, debe ser (para que alguien sea responsable por él) atribuible a esa persona (es decir, pertenecer a dicha persona en un sentido moral), debe poder dar explicaciones por él, y además debe poder asumir las consecuencias que se le impongan. La hipótesis en la que se mueve el trabajo es que puede haber situaciones en las que parezca que alguien debe ser considerado responsable, pero o no hay un candidato adecuado o bien no hay una distribución no arbitraria de la culpa entre los candidatos disponibles [156]. Se trata, por tanto, de evitar escoger como candidato a alguien que sirva meramente como un chivo expiatorio moral [157].

Para verificar si esta hipótesis es verdadera exploran dos tipos de razones, reconociendo que nuestra concepción de la responsabilidad implica tanto elementos causales (generar actitudes que controlen las acciones) como elementos epistéi-

---

[153] Del Rosal Blasco, B., *op cit*. N 35, p. 15.

[154] Aunque los vocablos adscribir e imputar se están usando aquí como sinónimos, no es cierto que en todos los contextos deban serlo. Para una aproximación a las diferencias, Figueroa Rubio, S., *op. cit*. N. 137, p. 35.

[155] Köhler, S., Roughley, N., /Sauer, H., «Technologically blurred accountability? Technology, responsibility gaps and the robustness of our everyday conceptual scheme», en: Ulbert, C., Finkenbusch, P., Sondermann, E., T. Debiel, T. (Eds.), *Moral Agency and the Politics of Responsibility*, Routledge, 1.ª ed., 2017, p. 51.

[156] *Ibid.,* pp. 53-54. Citan en este punto a Sparrow, que fue el filósofo pionero a la hora de destacar la existencia de un vacío de responsabilidad de tipo moral en el uso de armas autónomas letales. Sparrow, R., *op. cit.* N. 97.

[157] Paradigmáticmente, Danaher, hablando del riesgo del «moral scapegoating»: «*If there is a deep human desire to find appropriate targets of retributive blame, but none really exist, then there is a danger that people will try to fulfil that desire in inappropriate ways*». Danaher, J., *op. cit.* N.43, p. 20. Más tarde, adoptando el mismo término, Giannini, A., *op. cit*. N. 143, p. 162: «*It might even be the case that instead of an actual oversight on the action of the machine, the human is only capable of «rubber-stamping» the results produced. The risk is to create a (human) scapegoat: the contrast with common principles of criminal law is self-evident.»* También, Santoni de Sio, F / Mecacci, G., *op. cit*. N. 149, p. 1078.

micos o relativos a las consecuencias probables de nuestras acciones [158]. Es decir, traduciendo esta afirmación a un lenguaje que resultará más familiar al derecho penal, para ser responsables en un sentido pleno se considera que debemos tener un control sobre nuestras acciones, así como sobre las consecuencias de las mismas [159]. A continuación, los autores enumeran los argumentos por los cuales las modernas tecnologías de IA pueden suponer un problema en estos dos planos:

En el plano de la causalidad o control sobre las acciones, los problemas que merece la pena destacar son dos:

– En primer lugar, cuando los daños son causados por entidades tan independientes de los agentes, su comportamiento deja de estar determinado y se puede afirmar que no hay un control sobre sus acciones, sino que actúan sobre la base de sus propias decisiones, por lo que la cadena causal necesaria para la responsabilidad por el resultado se rompe, debido a esta autonomía. En este caso los autores niegan el problema, pues no consideran que sea cierto que de esta autonomía y descontrol derive la irresponsabilidad del agente, pues se mantiene aún un control relativo al haber situado al sistema autónomo en ese escenario o contexto, siendo este un tipo de control, si bien algo debilitado, suficiente para afirmar la responsabilidad. Por tanto, no hay laguna de responsabilidad aquí, sino una necesidad de aplicar categorías normativas de imputación previstas para supuestos donde los deberes se retrotraen a un momento anterior. Los autores hablan del modelo de delitos imprudentes, pero se puede entender que caben esquemas donde se adelante, quizá, la barrera de punición [160].

– En segundo lugar, se pone de manifiesto el conocido como *«many hands problem»,* es decir, la contribución causal de cada uno de los individuos causalmente implicados en el hecho final es tan pequeña que resulta difícil culpar a alguien. Se señala una vez más por los autores que aquí no se produce un verdadero vacío de responsabilidad, pues, de hecho, no hay impedimento

---

[158] Köhler, S., Roughley, N., /Sauer, H., *op cit*. N. 155, p. 53.

[159] De los problemas que genera la suerte moral en la configuración del resultado son conscientes estos autores, pero reconocen que la forma más común de operar es aceptando mayor responsabilidad cuando el resultado dañoso efectivamente se produce.

[160] Köhler, S., Roughley, N., /Sauer, H., *op cit*. N. 155, p. 57. Los autores, al situarse en el contexto del *common law*, hablan de formas de «*negligence»* o *«recklessness»,* pero otros autores del ámbito continental sitúan la solución de la imprudencia como la más adecuada para resolver estos casos. Así, Zhao, S., *Principle of Criminal Imputation for Negligence Crime Involving Artificial Intelligence,* Springer, Singapore, 2024, *passim*. También, Consulich, F., *op cit.* N. 122, pp. 1015-1055. No obstante, los partidarios de elevar el umbral de responsabilidad también sostienen que la acción humana se produce necesariamente en una fase anterior. *Vid.:* Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., (Eds.) *«Criminalisation of AI-related offences»,* International Colloqium, Bucharest, Romania, 14-16 June 2023, *RIDP*, 95,1, 2024, p. 48.

para considerar a cada individuo responsable por el hecho de que su contribución sea de pequeño calado, sino que el problema reside en cuanto a la graduación de esa responsabilidad, pues la sanción a imponer a cada uno puede no resultar satisfactoria globalmente [161]. Se explica, asimismo, que desde el plano epistémico se genera el problema de determinar cómo distribuir esa culpa de forma adecuada entre los distintos participantes de la acción conjunta, y aquí la mayor dificultad reside en encontrar argumentos sólidos o de carácter objetivo, o, en palabras de los autores, «razones normativas claras». En este sentido se acepta que se debe renunciar a la idea de que el nivel de responsabilidad individualizada se basará en este tipo de criterios claros o «hechos metafísicos sólidos», y que inevitablemente, en la determinación social del grado apropiado de responsabilidad habrá algo de convención y algo de arbitrariedad.

En definitiva, en este primer grupo de problemas no hay verdaderos vacíos o lagunas, sino más bien problemas de tipo práctico relacionados con el modelo que mejor puede servir para imputar daños por colocar un objeto sobre el que se tiene un control relativo, así como encontrar una sanción individual adecuada a la gravedad de los hechos. Esta cuestión es puesta de manifiesto, en términos similares, por Danaher, cuando habla de lo que él denomina problema de la atenuación [162]. El problema reside en que, aunque exista algún vínculo residual entre los creadores y el robot, es probable que esté atenuado hasta el punto de que no sería moralmente apropiado atribuir a esos creadores un nivel de culpa que cubra toda la gravedad del daño moral causado [163]. Nos encontramos en este punto ante una genuina preocupación por la justicia y proporcionalidad del castigo.

En cuanto a los problemas de tipo epistémico, relativos al conocimiento y control sobre las consecuencias, el problema principal que se plantea es el de la previsibilidad. Es decir, la idea de que solo es legítimo responsabilizar a un

---

[161] Köhler, S., Roughley, N., /Sauer, H., *op cit*. N. 155, p. 59. En definitiva, torna esta cuestión en un problema relativo a la necesidad de encontrar una sanción proporcionada en todas sus vertientes. En todo caso, el problema de valorar un hecho de muy pequeña entidad y la necesidad de castigo no es exclusivo en este campo, pues es objeto de discusión central en los llamados delitos acumulativos, y también en el campo de la filosofía moral ha dado que hablar. Así, Parfit defiende desde una perspectiva filosófica la legitimidad de castigar pequeñas acciones nocivas en: Parfit, D., *Razones y personas.* Trad. De Mariano L Rodríguez González, A. Machado Libros, Boadilla del Monte, 2004.

[162] Danaher, J., *op. cit.* N.43, p. 13 y 14.

[163] Sobre esta situación y algunas variaciones han tratado Abbott y Sarch al referirse a los escenarios en que, de un delito primigenio cometido con intención deriva otro no deseado y absolutamente imprevisible, que agrava el daño y que dificulta, por tanto, la imposición de un castigo adecuado. Estos autores tratan de trasladar la doctrina de las «*constructive offences*» del *common law* a los escenarios tecnológicos aquí descritos, para resolver los casos de lagunas que ellos califican de conductas irreducibles. *Vid.*: Abbott, R., / Sarch, A., «Punishing artificial intelligence: legal fiction or science fiction», *op. cit. N. 85*, pp. 323-384.

agente cuando el estado de cosas resultante de su actuación es algo que pudo prever [164]. Sobre el argumento de la previsibilidad los autores responden con el denominado como modelo de los perros [165], argumentando que no hay laguna de responsabilidad porque se puede entender que «si dejas suelto un dispositivo autónomo portador de armas mortales en un entorno determinado en el que existe el riesgo de que haga algo dañino, aunque no sea tu intención, sigues siendo responsable» [166]. La cuestión será determinar qué tipo de responsabilidad se deriva, pero lo que está claro es que se podrá derivar legítimamente.

En resumen, ciertas características de los sistemas de IA nos obligan a repensar las formas de imputación: el control se aleja según la complejidad crece; la previsibilidad disminuye a medida que la autonomía aumenta; y la multiplicidad de agentes implicados en las acciones colectivas nos sitúan en un escenario con contornos diferentes que nos obliga a repensar los modelos de imputación. Köhler, Roughley y Sauer hablan de una responsabilidad difuminada *(«blurred»* [167]*)* por la interacción de todos estos factores, y por la constatación de que las normas que fijen los estándares de comportamiento y cuidado están, en este campo, aún por determinar.

## 2.2 **El cuestionamiento de las lagunas**

En la línea de lo apuntado anteriormente, si bien se ha convertido en lenguaje común el hablar de lagunas, a negar su existencia han ido dirigidos los esfuerzos

---

[164] Köhler, S., Roughley, N., /Sauer, H., *op cit*. N. 155, p. 60. Desde luego, la previsibilidad como base sobre la que se asienta la responsabilidad penal es un postulado que queda fuera de toda duda para el derecho penal. Así lo expone, con claridad brillante, Zhao cuando resalta que este es uno de los mayores problemas a que se enfrenta la teoría penal contemporánea, por su fuerte dependencia en este elemento de «foreseeability». Zhao, S., *op cit.* N. 160, *passim*. Si bien aquí se exploran los problemas de causalidad y previsibilidad desde un punto de vista filosófico, se discutirán en su vertiente penal en el capítulo quinto.

[165] *Ibid.,* p. 63. Dsouza también propone la aplicación del modelo establecido en la Ley de Perros Peligrosos de 1991, pero lo hace para negar que existan cambios en las estructuras tradicionales del derecho penal como resultado de la llegada de la IA. Dsouza, M., *Don't panic: Artificial intelligence and Criminal Law,* en Baker, D. J./ Robinson, P. H*., Artificial Intelligence and the Law. Cibercrime and Criminal Liability,* Routledge, 2020, pp. 247-264.

[166] Köhler, S., Roughley, N., /Sauer, H., *op cit*. N. 155, p. 63 (traducción propia).

[167] No es esta la única ocasión en que se alude a este término para caracterizar los borrosos contornos en que nos movemos en esta nueva realidad. Para una explicación de lo que significa esta «difuminación» en genérico, *vid*.: Pinto Neto, M., «Blur», en Sabariego, J., Amaral, A. J., / Salles, E. B. (Orgs.) *Algoritarismos*, Tirant Lo Blanch Sao Paulo, 2020, p. 141-157. Para profundizar en el modo en que la IA está desdibujando las distintas categorías del Derecho penal y las sanciones: Bravo Bolado, A., «De las cosas a las no-cosas. Una transformación de los sujetos, los paradigmas de responsabilidad y las sanciones en la era del crimen tecnológico. Desdibujando las fronteras de lo conocido en tres movimientos», en Rodríguez Horcajo, D., Basso, G. J., (Dirs.), *Repensando las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2025, pp. 161-184.

recientes de una parte de la literatura, que considera que más que vacíos de responsabilidad lo que enfrentamos son problemas epistémicos, pues, en última instancia, siempre podrá responder un ser humano [168]. Lo único que debemos averiguar es cuál es la forma más justa de exigir esa responsabilidad, habida cuenta de las características innovadoras de estas nuevas herramientas perfeccionadas.

Para Königs, que parte de que el supuesto vacío de responsabilidad se debe a la autonomía del sistema, entendida como capacidad para funcionar de forma independiente a la intervención humana en entornos complejos y tomar decisiones no basadas en instrucciones concretas, la existencia de una laguna real resulta complicada teóricamente.

Si los modelos de responsabilidad encuentran la forma de culpar a los humanos cuando se producen conductas imprudentes o dolosas, solo será posible teóricamente la laguna en caso de que el humano haya cumplido debidamente con sus obligaciones y haya observado un deber de cuidado, y aun así se produzca el daño, siendo necesario castigar. El problema que trata de poner de manifiesto esta doctrina de las lagunas de responsabilidad es que no se debe desincentivar el cuidado en el uso (en sentido amplio) de estas peligrosas tecnologías. Pero, entonces, solo hay dos escenarios posibles:

– O el humano cumple con todos sus deberes de cuidado en el uso de estas tecnologías, en cuyo caso no hay comportamiento reprochable ni necesidad de exigir responsabilidad. Es decir, no hay laguna;

– O bien el humano incumple esos deberes y se puede afirmar su responsabilidad, ya sea por su conducta dolosa o culposa. En cuyo caso, tampoco hay laguna.

---

[168] Así lo afirma, claramente, Busato, P. C., *op. cit*. N. 124, p. 372. Una variante de esta forma de cuestionamiento de las lagunas la encontramos en los filósofos Shannon Vallor y Tillmann Vierkant que sostienen que el verdadero problema de responsabilidad en los sistemas autónomos no reside en la falta de conocimiento o control sobre lo que hace la máquina, pues esos déficits ya existían en la acción humana ordinaria, sino en un nuevo «gap de vulnerabilidad», según el cual los sistemas sociotécnicos de IA crean una fuerte asimetría entre usuarios muy expuestos a los daños, y agentes (desarrolladores, empresas, reguladores) estructuralmente protegidos frente a las consecuencias morales y prácticas de esos daños. A partir de teorías instrumentalistas de la responsabilidad, que conciben las prácticas de reproche, rendición de cuentas y sanción como herramientas para cultivar agencia responsable dentro de una «ecología moral» compartida, argumentan que la IA rompe estas dinámicas de vulnerabilidad recíproca que hacen efectivas dichas prácticas: ni las máquinas son moralmente vulnerables, ni los humanos situados en la cadena de decisión suelen estar en la relación adecuada con las personas afectadas para ser realmente interpelados. *Vid.*: Vallor, S./Vierkant, T., "Find the Gap: AI, Responsible Agency and Vulnerability", *Minds and Machines,* 34, 2024, pp. 14 y ss.

Sin embargo, las cosas pueden no ser tan fáciles como las plantea Königs [169], porque el principal obstáculo con el que nos encontramos en este campo es saber, en efecto, si se cumple la primera premisa, es decir, si el humano que actúa lo hace en cumplimiento de todos los deberes que le son exigibles, habida cuenta del entorno tecnológico. Y en este sentido el mismo autor reconoce que el problema no es, pues, metafísico, en el sentido de que exista o no un agente responsable, sino de carácter epistémico, es decir, acerca de cómo podemos conocer o determinar de forma correcta el grado de responsabilidad de las personas [170]. Aunque al autor se muestra optimista ante el pretendido problema de las lagunas, no por ello considera que sea de fácil solución, pues apunta que no altera el hecho de que lograr la justicia de la responsabilidad es difícil en contextos de autonomía [171].

Por volver a nuestro ejemplo inicial, no es que no exista la persona que en efecto asesinó al comendador, sino que, dadas las limitaciones contextuales (en el ejemplo histórico, el contexto de solidaridad y encubrimiento mutuo del pueblo, y en el contexto tecnológico las características inherentes y diferenciadoras de estas tecnologías), quien debe imponer el castigo por el hecho dañoso no sabe a qué criterios atenerse, ni qué exigir a cada una de las personas envueltas en el embrollo. Es decir, no es que no exista una persona responsable, es que no sabemos cómo culparla de una manera justa.

Los filósofos De Sio y Mecacci nos ponen sobre la pista, asimismo, de esta realidad, poniendo el énfasis en que las diferentes «lagunas» se producen por causas que tienen relación con las propias características inherentes a la tecnología. Algunas de ellas son antiguas, tales como la complejidad de la interacción en los sistemas sociales y técnicos, pero otras son más nuevas, como es el caso del aprendizaje basado en datos. Algunas tienen carácter técnico, como la opacidad intrínseca al proceso de toma de decisiones algorítmico, mientras otras son más de tipo político y económico (se refieren, en este punto, a la muy acertada observación acerca de la privatización implícita de la administración pública en la gestión de la información tecnológica), y, finalmente sociales [172]. De estas últimas merece la pena destacar una idea: los actores implicados en el ciclo de vida de esta nueva tecnología (desde los ingenieros hasta los productores, empresarios, fabricantes, o usuarios), pueden aun no ser conscientes de las nuevas obligaciones morales, legales y sociales que este contexto requiere [173].

---

[169] Könings, P., «Artificial intelligence and responsibility gaps: what is the problem?», *Ethics and Information Technology,* 2022, *passim*.

[170] *Idem.,* p. 9.

[171] *Idem*., p. 10.

[172] Santoni de Sio, F / Mecacci, G., *op. cit*. N. 149, p. 1068.

[173] *Idem.* En un sentido similar, criticando que el énfasis se ponga en la discusión sobre las pretendidas lagunas y apostando porque se dediquen los esfuerzos a individualizar la responsabilidad dentro de todos los seres

El problema que demandan estas nuevas realidades es, precisamente, reconocer en qué espacios se pueden tolerar las lagunas y preguntarse hasta qué punto es recomendable establecer normas de diligencia razonable de la misma forma en que se ha hecho con otras técnicas a lo largo de la historia [174].

Conocer las características que diferencian a esta técnica de las anteriores nos permitirá, pues, abordar los problemas con conocimiento de causa, es decir, localizar lo que Miró ha denominado elementos que cambian las reglas del juego *(«game-changers»* [175]).

## 3. TRADUCCIÓN DEL PROBLEMA ¿QUÉ NECESITA EL DERECHO PENAL DE HOY?

Utilicemos la fórmula que utilicemos, la verdadera necesidad de nuestro sistema de adscripción de responsabilidad reside en saber si, ante un hecho dañoso, aparentemente penal y aparentemente necesitado de castigo, seremos capaces de identificar a una (o varias) personas a las que responsabilizar de forma justa, cuando en la actuación haya contribuido causalmente, de una u otra forma, un sistema de inteligencia artificial.

Si existe una verdadera laguna, en el sentido de que es imposible que esta pretendida persona sea responsable, o si esta laguna es, en realidad, aparente, pues lo que ocurre es que no sabemos cuáles son los criterios con los que adscribir esa responsabilidad, es algo que no tiene, en principio, demasiada importancia, pues de lo que se trata es de encontrar criterios para una distribución justa de una responsabilidad que, *prima facie,* parece ser algo difusa.

Se ha puesto de manifiesto a lo largo de este capítulo que resulta esencial, para ello, conocer la verdadera naturaleza de la tecnología implicada, pues solo teniendo en cuenta sus características esenciales podremos saber de qué manera afecta a la imputación de responsabilidad.

---

humanos que pueden estar imbuidos en el contexto de los sistemas autónomos: Nyholm, S., *op. cit.* N. 110.

[174] Santoni de Sio, F / Mecacci, G., *op. cit*. N. 149, p. 1070. Esta es una idea que exponen como crítica a las perspectivas que ellos denominan «solucionistas», que son aquellas que vienen a afirmar que la solución es tan simple como establecer deberes de cuidado razonables como se ha hecho en el pasado. La crítica apunta a que esta solución naif desconoce las particularidades que hacen única a esta tecnología, y que conocer esta dinámica es esencial para afrontar los problemas en la responsabilidad que se derivan. Para explorar la solución que estos autores proponen, basada en el control significativo y en un seguimiento y rastreo durante todo el proceso productivo, *vid.* p. 1077. También: Santoni de Sio, S./Van den Hoven, J., «Meaningful Human Control Over Autonomous Systems: A Philosophical Account», *Frontiers in Robotics and AI,* 2018, *passim.*

[175] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M.*, op cit.* N. 160*,* p. 47.

Aunque son muchos los rasgos que se han ido destacando a lo largo de estas páginas, en aras de simplificar la discusión se escogen aquí dos grandes grupos de problemas, al hilo de las dos grandes características que modelan la realidad aplicativa de los sistemas de IA.

En primer lugar, la autonomía, en los términos descritos en el capítulo primero, de la cual se derivan rasgos como la capacidad de aprendizaje, la adaptabilidad y la impredecibilidad. En segundo lugar, la multiplicidad de interacciones, en escenarios complejos donde no solo humanos, sino humanos y sistemas, operan conjuntamente en el seno de procesos de producción [176] dotados de complejidad innegable [177].

Ello conduce, en definitiva, a un rasgo omnipresente y común a todos estos factores: la falta de control. Y esto conlleva toda una serie de implicaciones para el derecho penal. No es ya que las acciones de los sistemas de IA no se puedan explicar o prever, sino que las acciones humanas ya no son exclusivamente humanas, sino en muchísimos ámbitos mediadas por la tecnología, lo que hace que la acción humana en sí misma sea, en cierto modo, algo menos explicable, algo menos controlable y algo menos predecible, siempre que esté mediada por este tipo de tecnologías [178].

Si el derecho penal quiere seguir castigando estas acciones poco predecibles, poco controlables y poco explicables (y esta es solo una de las opciones posibles), debe profundizar en cómo esta falta de control en sentido amplio juega en sus estructuras de imputación tradicionales, con el objeto de comprobar si siguen siendo válidas.

---

[176] El estudio de la estructura productiva empresarial no agota, desde luego, el ciclo de vida de la IA, donde la interacción con los usuarios finales y otros actores en red puede llegar a alcanzar complejidades enormes. *Vid.*: Beckers, A., /Teubner, G., *op cit.* N. 111. Sin embargo, merece la pena resaltar que el proceso circunscrito al modelo productivo y económico de la IA merece una atención especial, pues dista mucho del modelo de empresa tradicional, en el que, si bien la complejidad y las interacciones son punto común, se presupone la posibilidad de una organización. Habrá que esperar a ver qué estructuras de negocio y desarrollo se acaban implantando en el sector de la IA, pero baste dejar apuntado que, por el momento, el proceso dista de ser un entorno de complejidad organizada, pareciéndose más bien a un paradigma de Salvaje Oeste donde reinan aún la libertad, el intercambio de modelos de forma abierta y la auto regulación (en ocasiones equivalente a no regulación). Queda muy lejos, por tanto, el paradigma de la empresa convencional.

[177] Entre la literatura especializada este rasgo se ha llegado a caracterizar como «inteligencia ambiental», o incluso «inteligencia en red», y son numerosos los autores que han puesto de manifiesto la importancia de esta «acción colectiva o distribuida». Paradigmáticamente: Floridi, L., «Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions», *Royal Society's Philosophical Transactions A: Mathematical, Physical and Engineering Sciences,* 374*,* 2016, pp. 1-22. Hildebrandt, M., *op. cit*. N. 113, pp. 163-180; Koops, B. J., Hildebrandt, M., / Jaquet-Chiffelle, D. O., *op. cit.* N. 79. La referencia a la inteligencia distribuida aparece ya en un famoso texto de 1996. *Vid.:* Karnow, C. E., «Liability for distributed artificial intelligences», *Berkely Tech. LJ*, 11, 1996, p. 147.

[178] Hay quienes afirman, de hecho, que los humanos perdemos, en este camino, nuestra percepción sobre la agencia, que percibimos debilitada por lo que se conoce como «*moral disengagement».* El concepto proviene de Bandura, A., «Moral Disengagement in the Perpetration of Inhumanities», *Personality and Social Psychology Review*, Vol. 3, No. 3, 1999, pp.193-209. Para el autor se trata de un conjunto de técnicas mediante las cuales los seres humanos se desentienden selectivamente de las autosanciones morales cuando hacen algo que perciben como malo.

No es casualidad que uno de los focos más importantes en la literatura jurídica, algorítmica y de las propias instituciones en lo relativo a los principios que deben gobernar la inteligencia artificial haya sido, precisamente, el del denominado *control humano significativo* [179]. Sin embargo, este principio general, ampliamente desarrollado a nivel ético y con grandes defensores en la doctrina, debe analizarse, también, desde el prisma del derecho penal, donde uno de los presupuestos básicos de la acción humana, como primer elemento de un delito, es, precisamente, un control humano significativo.

El objetivo de los siguientes capítulos será, por tanto, tratar de desentrañar cómo algunas estructuras de imputación tradicionales pueden verse moldeadas por las realidades aquí descritas.

## II. LAS PROPUESTAS INTERNACIONALES PARA PALIAR ESTAS LAGUNAS

El problema relativo a las lagunas que la IA puede estar produciendo en nuestros sistemas contemporáneos de derecho penal ha levantado preocupaciones no solo a nivel teórico, sino, desde luego, desde una perspectiva práctica. Por ello, en un capítulo dedicado a los problemas y las lagunas que hasta el momento se han detectado, no podría faltar una alusión a la realidad actual de los distintos ordenamientos jurídicos. El segundo capítulo tiene como objeto, por tanto, una doble reflexión: la primera centrada en las discusiones de carácter filosófico sobre lo que implica que la IA genere lagunas de responsabilidad; y la segunda centrada en cómo los distintos ordenamientos de los países de nuestro entorno están tratando de solucionar estas lagunas, en una etapa todavía muy incipiente de este proceso regulatorio.

---

[179] Este principio, que surgió, una vez más, al hilo del debate sobre las armas autónomas letales, ha experimentado un gran apoyo en diferentes campos de aplicación (tales como los coches autopilotados o las aplicaciones de uso médico). Así, algunas obras de referencia: Amoroso, D., /Tamburrini, G., «What makes human control over weapon systems «meaningful»?», *ICRAC Working Paper Series,* 4, 2019.; Ekelhof, M., «Moving beyond semantics on autonomous weapons: Meaningful human control in operation», *Global Policy*, 10(3), 2019, pp. 343-348. Horowitz, M. C., /Scharre, P., «Meaningful human control in weapons systems: A primer», *Center for a New American Security*, 2015. Mecacci, G./Santoni de Sio, F., «Meaningful human control as reason-responsiveness: The case of dual-mode vehicles», *Ethics and Information Technology*, 22, 2020, pp. 103-115. En el panorama español, dos autores especialmente defensores de este principio han sido Valls y Romeo Casabona. *Vid.*: Valls Prieto, J., *Inteligencia artificial…, op cit.* N. 46,; Romeo Casabona, C. M., «La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robots y tecnologías conexas«, en Sánchez, J. M./García Pérez, O./ Cerezo Domínguez, A. I./García España, E. (Dirs.), *Estudios político-criminales, jurídico-penales y criminológicos. Libro homenaje al profesor José Luis Díez Ripollés,* 2023, p. 706.

Como muestra de las preocupaciones que suscita la IA dentro de la doctrina penal, la Asociación Internacional de Derecho penal (en adelante, AIDP) decidió dedicarle sus jornadas de 2024, como forma de abordar, desde diferentes perspectivas (parte general y especial, proceso y derecho internacional, respectivamente) el problema de la responsabilidad penal y la inteligencia artificial [180]. Resulta valioso, pues, más que analizar toda la legislación que sobre la materia viene proponiéndose en los últimos tiempos (por lo inabarcable de la misma, amén de configurarse como ejemplos concretos de la regulación sectorial de determinados campos de aplicación), estudiar los informes que, fruto de los estudios nacionales, recogen los relatores generales nombrados al efecto, haciendo hincapié en aquellos datos que dejan entrever tendencias generales en la forma de abordar el fenómeno.

Nos centraremos en dos de las cuatro líneas de trabajo desarrolladas por la AIDP [181], cuyos informes generales se analizarán algo más pormenorizadamente, sin perjuicio de señalar las particularidades de algunos informes nacionales. Estas dos líneas de trabajo son: la afectación a las categorías tradicionales del derecho penal con la IA *(Traditional Criminal Law Categories and AI: Crisis or Palingenesis?)* y la pregunta sobre la evolución de los tipos penales y los instrumentos de protección a las víctimas *(Old and New Criminal Offences: AI Systems as Instruments and Victims)* [182].

## 1. EL ESTUDIO SOBRE LA AFECTACIÓN DE LAS CATEGORÍAS TRADICIONALES DEL DERECHO PENAL [183]

El informe del profesor Picotti toma como base 22 informes nacionales de diferentes países [184]. Tras tratar las cuestiones relativas a la definición (des-

[180] Así, en el XXI Congreso Internacional de Derecho Penal (Paris, 26-28 junio 2024), la temática fue: «Artificial Intelligence and Criminal Law».

[181] Se excluye, por tanto, el análisis de los trabajos desarrollados dentro de la Sección 3 *(AI and Administration of Justice: Predictive Policing and Predictive Justice),* y la Sección 4 (*International Perspectives on AI: Challenges for Judicial Cooperation and International Humanitarian/Criminal Law*), que abordan los problemas que la IA presenta desde un punto de vista más centrado en el proceso y la cooperación judicial y humanitaria, respectivamente, y que se encuentran, por tanto, más alejados de nuestro objeto de estudio.

[182] Secciones 1 y 2 del trabajo desarrollado por las comisiones de la AIDP generadas en los años 2023 y 2024, y que se plasman, finalmente, en estudios publicados en la *Revue Internationale de Droit Pénal* (en adelante, por sus siglas, RIDP), en los números 94 y 95, respectivamente.

[183] Este informe se encuentra publicado en Picotti, L./ Panattoni, B. (Eds.)., «Traditional criminal law categories and AI: Crisis or palingenesis?», *RIDP*, 94, 1, 2023. En el mismo se recogen los trabajos realizados en el desarrollo de la sección 1 *(Traditional Criminal Law Categories and AI: Crisis or Palingenesis?).*

[184] Dichos informes, explica Picotti, fueron elaborados antes de junio de 2022 por los relatores de 19 países de Europa continental (Alemania, Italia, Francia, Países Bajos, Bélgica, Grecia, Portugal, República Checa), Europa del este (Rumanía, Hungría, Croacia y Ucrania), uno del norte de Europa (Finlandia),

tacando que no existe en la mayoría de los países una definición oficial), lo cual puede causar problemas de legalidad, y destacando la ausencia general de marcos legales que regulen cuestiones penales, salvo contadas excepciones, los países suelen considerar aplicables figuras delictivas afines a la ciberdelincuencia en casos de ataques sobre sistemas informáticos.

Destaca el relator que, sin embargo, aunque la mayoría de delitos existentes en las legislaciones nacionales se consideran aplicables a actos ilícitos que se realicen o bien a través o bien contra sistemas de IA, sí se pueden identificar ciertos límites en la protección penal existente, así como lagunas en la legislación [185].

Los delitos existentes, explica Picotti, pueden aplicarse cuando se trate de sistemas de IA automatizados con un nivel *limitado* de autonomía, donde siempre existirá un agente humano que controle y pueda intervenir activamente en el sistema. Sin embargo, continúa, en el caso de los sistemas con *alto grado* de autonomía, la contribución del usuario podría reducirse a la mera activación del sistema, y, en esos casos, para que los delitos existentes puedan aplicarse, debe determinarse caso por caso si las contribuciones generadas por el sistema de IA pueden atribuirse a este usuario, lo cual dependerá del objetivo y alcance de su funcionamiento. Esto es algo que se pone de manifiesto en el informe alemán, en los siguientes términos:

*«Mientras los sistemas utilizados sean solo parcialmente autónomos, la atribución no se ve impedida por la autonomía otorgada a los sistemas, ya que estos siguen operando, al menos, dentro de un marco determinista predefinido. Sin embargo, en el futuro podrían surgir lagunas en la responsabilidad penal, ya que, cuanto mayor sea el grado de autonomía del sistema, más probable será que se rechace su atribución al usuario* [186].»

En general, el informe destaca que no se han creado nuevos delitos en los distintos ordenamientos, y que algunos países (como Portugal) han destacado la necesidad de primero generar una regulación general previa a la intervención penal [187]. Sí que se ha producido mucho debate sobre la regulación extrapenal en

---

tres informes de Sudamérica (Argentina, Brasil y Chile), y otros dos de China y Egipto. Más tarde se preparó un informe de Estados Unidos y finalmente se añadieron los de Rusia, Turquía y Japón. La enorme variedad geográfica y cultural que recoge el trabajo que aquí analizamos es, por tanto, un valioso activo, y justifica su estudio detallado.

[185] Picotti, L./ Panattoni, B., *op cit*. N 183, p. 23.

[186] Beck, S./Gerndt, S., «German report on traditional criminal law categories and AI», *RIDP*, 195, p. 203.

[187] Aires De Sousa, S., «Portuguese report on traditional criminal law categories and AI», *RIDP*, Vol. 94 issue 1, 2023, pp. 319-330. Destaca, no obstante, Picotti, que países como China, Rusia o Turquía

algunos sectores (reconocidamente, en sectores como la conducción autónoma, la tecnología de reconocimiento facial, o el servicio de salud), así como la conveniencia de prohibir ciertos usos (más allá del manido debate sobre las armas autónomas, se destaca en el informe la cuestión de la vigilancia biométrica, así como el uso en áreas como la justicia [188]). Las referencias a la jurisprudencia recaída hasta el momento son escasas, y se centran principalmente en accidentes con coches automáticos y accesos no consentidos en sistemas informáticos, donde no ha habido modificaciones reseñables en cuanto a la aplicación de la ley.

El informe descarta, pues así lo hacen en general los informes nacionales, la posibilidad de que la IA sea penalmente responsable, pero sí se detiene más en explorar los modelos de responsabilidad de los humanos que actúan detrás de la IA.

En lo relativo a la posibilidad de explorar la responsabilidad penal de los productores y proveedores, se destaca la importancia del deber de controlar las fuentes de peligro como fuente de obligaciones más relevante [189]. Se pone el énfasis en la relevante posición de los desarrolladores en el caso de los sistemas más autónomos, señalando el informe japonés que es esencial examinar cuidadosamente el proceso de desarrollo de la función de aprendizaje no supervisado de la IA, pues, ya que no podemos controlar la función propia de la IA, debemos poder detener toda investigación o desarrollo de su aprendizaje cuando vaya más allá de los límites del control humano [190].

Es interesante cómo se establece, de forma implícita, esa relación inversa entre control y autonomía, dando a entender que, si en un determinado punto del proceso de aprendizaje no se puede afirmar un control, la responsabilidad debe centrarse justo en ese momento en que se mantiene, aún, una parte de ese control.

---

sí se ha propuesto regular penalmente algunas cuestiones (el abuso de IA, el tráfico ilegal de robots, y la creación de circunstancias agravantes, respectivamente). Picotti, L./ Panattoni, B., *op cit*. N 183, p. 25.

Nota de actualización: con posterioridad al cierre de redacción de esta monografía, el panorama regulatorio internacional ha experimentado una notable aceleración. Tras la entrada en vigor del Reglamento, diversas jurisdicciones han aprobado reformas penales sustantivas para incorporar el uso de inteligencia artificial como circunstancia agravante o como elemento de nuevos tipos autónomos, destacando especialmente la Ley italiana n.º 132/2025, de 23 de septiembre –primera norma orgánica de un Estado miembro de la Unión Europea en esta materia, que introduce una agravante común en el art. 61 CP, agravantes especiales y el nuevo delito de difusión ilícita de contenidos generados o alterados con IA–.

[188] *Ibid.,* p. 26.

[189] En este punto se remite a un informe específico: Grandi, C., «Positive Obligations (*Garantestellung*) Grounding Criminal Responsibility for not having Avoided an Illegal Result Connected to the AI Functioning», 2023, 1 *RIDP*, 59, pp. 67-76.

[190] *«As we cannot control the AI that fulfills the function, the only remaining solution would be to stop the research of AI's unsupervised learning function when it seems to reach the limit of control of the human agent»* Imai, T., «Japanese report on traditional criminal law categories and AI», *RIDP*, 94, 1, 2023, p. 316.

Concede Picotti que, a la vista de los informes, dado que la delegación en sistemas de IA aumentará, y el control de los usuarios sobre el funcionamiento del sistema será limitado, la responsabilidad de los programadores está llamada a desempeñar un papel crucial en las futuras políticas penales [191]. La advertencia final de Picotti en este punto viene de la mano de la propuesta de EEUU, que propone en su informe un nuevo tipo de «responsabilidad de los productores» como una forma de infracción penal de responsabilidad estricta en los casos en que las violaciones normativas sean cometidas por sistemas de IA. El relator general advierte de que esta no puede ser, en ningún caso, la tendencia en los ordenamientos occidentales, donde una responsabilidad por la posición no es viable [192].

Dejando de lado la responsabilidad por falta de control de fuentes de peligro, Picotti se centra en la posibilidad de derivar responsabilidad basada en obligaciones positivas (es decir, explorando la posibilidad de imputar delitos por infracción de deberes positivos de hacer). Es en este punto donde se resaltan algunas advertencias esenciales a los efectos del presente capítulo.

Se señala que la responsabilidad basada en obligaciones positivas presenta deficiencias relevantes precisamente debido a las características peculiares de los sistemas de IA. El hecho de que su mecanismo de creación involucre a un gran número de sujetos con diferentes tareas, condicionándose mutuamente, y que implique en ocasiones a muchas organizaciones, dificulta la individualización de responsables, así como la determinación de sus deberes de control [193].

Sumado a esto (que no es más que la reproducción de los problemas presentes en las organizaciones complejas), Picotti va más allá, pues afirma que la autonomía e imprevisibilidad de los sistemas de IA agravan estos problemas, dificultando la identificación del factor que inicialmente da lugar a la cadena causal que produce el daño. Estas características hacen que la descripción legislativa abstracta de las posiciones de garantía sea particularmente un desafío, pues en tales escenarios la atribución de responsabilidad penal a un único agente humano resulta problemática, al menos de acuerdo con los principios constitucionales, en especial el principio de responsabilidad personal y el de legalidad, y así se sugiere explorar el modelo de responsabilidad por productos defectuosos [194]. Conecta esta preocu-

---

[191] Se señalan algunos ejemplos de regulaciones que podrían aplicar, aunque no exentas de dificultades. Así, el ordenamiento húngaro prevé la penalización del deber de cuidado profesional cuando pueda causar daños sobre la vida o integridad, *(the offence regulated by paragraph 165 (1) of the CC might apply),* pero lo que requiere esta infracción es un peligro inminente y concreto para una persona específica, lo que en el momento de la producción aún no puede predicarse. PICOTTI, L./ PANATTONI, B., *op cit*. N. 183, p. 33.

[192] *Ibid.,* p. 34.

[193] *Idem.,* p. 34.

[194] *Idem.*

pación directamente con la expuesta en la primera parte de este capítulo, pues la falta de explicabilidad de la cadena causal sugiere una dificultad aplicativa derivada de forma directa de las peculiaridades de esta tecnología.

Pasa a explorar el relator general cómo esta situación descrita puede generar alguna distorsión en las categorías tradicionales. Pone de manifiesto que muchos de los informes nacionales han señalado la existencia de una posible brecha de responsabilidad en estos contextos (Portugal, Alemania, Francia, Italia y China) [195], señalando que esta brecha se refiere principalmente a dos aspectos: la causalidad y el problema del *black box,* por una parte, y el problema de la culpabilidad y los resultados impredecibles, por otra. Aquí se señala, de nuevo, una diferencia que consideramos esencial: en los sistemas donde todo el contexto viene dado por los humanos, que determinan todas las acciones del sistema (pone el ejemplo de los algoritmos predictivos usados para recomendar a partir de un conjunto de datos existente, identificando patrones y probabilidades), no es necesario realizar cambio alguno en el sistema de atribución de responsabilidad. Pero sí podrían ser necesarias adaptaciones en cuanto los sistemas de IA adquieran características más amplias o autónomas [196].

En lo relativo al *actus reus* (esta es la categoría utilizada en el informe), se considera la posibilidad de anticipar la protección penal a través de la criminalización de conductas peligrosas. De hecho, el informe griego lista toda una serie de conductas peligrosas relacionadas directamente con la fase del diseño del sistema [197]. Aunque se reconoce que ese *actus reus* humano puede estar constituido no solo por acciones, sino también por omisiones, se reconoce una laguna en este punto, pues en la actualidad no se ha definido un marco legal claro que contenga disposiciones expresas sobre tales obligaciones de actuar en contextos de aplicaciones de IA, o que haya sido regulado por fuentes formales, siendo solo de aplicación las reglas generales de seguridad y protección para productos técnicos, que no aportan la seguridad jurídica requerida [198].

---

[195] *Idem.,* p. 37.

[196] *Idem:* «*adaptations could be required in the future if AI systems acquire wider or fully autonomous features*».

[197] Kaiafa Gbandi, M., Sachoulidou, A., / Lima, D., «Greek report on traditional criminal law categories and AI», *RIDP*, 94, 1, 2023, p. 231. «Puede ser el caso de diseñar un robot que no garantice la estabilidad de sus movimientos, con el resultado de que se caiga durante su uso y cause daños corporales a su usuario o a las personas que se mueven a su alrededor. Otro ejemplo sería programar un sistema de IA de forma que reaccione de forma exagerada a las entradas, con el mismo efecto perjudicial, ya que el usuario no puede controlar dicha reacción exagerada. Del mismo modo, un error de producción sería no incorporar los sensores adecuados para la función del sistema de IA, con el efecto de que la máquina no reaccione como se espera y, por tanto, cause daños» (traducción propia).

[198] Picotti, L./ Panattoni, B., *op cit*. N. 183, p. 38. Debe tenerse en cuenta que este informe es anterior a la entrada en vigor del Reglamento de IA, aunque el mismo se encontraba ya en sus fases finales de negociación. Ninguno de los informes nacionales destaca, sin embargo, la existencia de estos deberes

En lo relativo a la causalidad, las opiniones de los informes nacionales son algo diversas. El informe belga destaca que no habrá problema en afirmar la causalidad según la teoría de la *conditio sine qua non*, pues incluso las consecuencias inicialmente no previstas pueden atribuirse a una persona que haya realizado un comportamiento indebido, pues si se puede entender que el daño causado por culpa de un agente humano se ve agravado por el mal funcionamiento de un sistema de IA, ello no servirá para liberar al humano de su responsabilidad. Dicha liberación solo se podrá producir cuando el daño hubiera ocurrido independientemente de la acción u omisión de la persona humana [199].

Sin embargo, se destaca por el informe griego [200] que reconocer un acto peligroso objetivamente negligente en una etapa muy temprana de la cadena causal como válido puede llevar a afirmar un patrón de causalidad demasiado amplio. También considera el informe griego que el esquema de delegación de responsabilidades es diferente en los casos de sistemas de IA, y que en ningún caso puede ser igual asignar ciertas tareas a una persona, que asignar tareas a un sistema, siendo irracional considerar responsable a cualquiera que actúe dentro de esta compleja cadena de producción.

El informe italiano, por su parte, señala que, si la contribución del sistema de IA se considera una desviación autónoma, en el sentido de ser ajena al diseño operativo original del programador, esto podría constituir un factor causal sobrevenido que interrumpe el vínculo etiológico [201].

Esta posición no es sin embargo mayoritaria en muchos países, que consideran en sus informes nacionales que solo factores causales extraordinarios con «eficacia causal abrumadora» podrían romper la cadena causal, llegando incluso otros informes a negar que el comportamiento autónomo de un sistema de IA pueda romper la cadena causal de atribución. Se apunta por Picotti que, para dar solución a estos problemas de causalidad, se ha propuesto tipificar por el aumento del peligro, incluso adoptando el modelo del delito de peligro abstracto [202], modelo que al relator general le parece cuestionable.

---

generales de cuidado a nivel nacional, y Picotti pone de manifiesto esta incertidumbre, que se puede conectar de nuevo con el principio de legalidad, y que tiene que ver con esta idea de laguna o *gap* epistémico *supra* explicado (1.ª parte del capítulo segundo).

[199] *Idem.*

[200] Kaiafa Gbandi, M., Sachoulidou, A./ Lima, D, *op. cit*. N. 197, pp. 238-240. Lo cierto es que el problema señalado por el informe griego no es más que una expresión singular del problema general de la teoría de la equivalencia de las condiciones, común en el sistema germánico continental. Sobre los problemas en torno al concepto de causalidad en el derecho penal continental se profundizará infra, en capítulo quinto.

[201] Barresi, O., «Italian report on traditional criminal law categories and AI», *RIDP*, vol. 94, 1, pp. 269-302.

[202] Es el caso de Rusia, que propone en su informe aplicar este paradigma, aludiendo a un delito de su código penal que castiga la liberación de un animal peligroso. Picotti, L./ Panattoni, B., *op cit*. N. 183, p. 37.

El informe alemán destaca que es relevante la reducción de la toma de decisiones de la persona, pues esto plantea el riesgo de que el humano se reduzca a un mero objeto al que se adjunta responsabilidad, pero que carece de posibilidades esenciales de control [203].

En lo relativo al segundo grupo de problemas (principio de culpabilidad o *«state of mind»),* Picotti vuelve a resaltar, con alusión al informe belga, que la pérdida de control de los agentes humanos sobre el funcionamiento de los sistemas de IA podría inducir a la adopción de un enfoque de responsabilidad legal menos orientado hacia la culpa y el reproche [204]. En el contexto de los crímenes relacionados con la IA, donde toda o una parte de la conducta relevante es realizada por el sistema, puede ser difícil determinar si el estado mental del agente humano incluyó el *modus operandi* desarrollado. Es necesario que todos los actores involucrados en el ciclo de vida del sistema (programadores, productores y usuarios) puedan conocer de alguna forma este *modus operandi* de manera que puedan predecir la ocurrencia del resultado dañino tipificado [205].

Llama la atención Picotti, en este punto, sobre la solución que se propone por el informe italiano para abordar este problema sobre el conocimiento: aplicar la estructura de la *actio libera in causa,* en caso de que el humano no se representara o no quisiera representarse las operaciones específicamente llevadas a cabo por el sistema de IA. Picotti destaca que esta teoría no podría tener cabida, pues se aplica solo a la imputabilidad, por lo que sería una aplicación analógica *in malam partem* [206].

---

[203] Beck, S/Gerndt, S., *op cit.* N.186, pp. 209 y 210. Esta es la idea que Beck ha puesto de manifiesto en muchos de sus trabajos, dando a entender que hay una tensión esencial entre la falta de control y la responsabilización sin atender a este cambio. Beck, S., «Google-Cars, Software-Agents, Autonome Waffensysteme - neue Herausforderungen für das Strafrecht?» en Beck, S., Meier, B.-D.,/ Momsen, C. (eds), *Cybercrime und Cyberin-vestigation,* Nomos, 2015, 14, pp. 32 y ss.

[204] Picotti, L./ Panattoni, B., *op cit*. N. 183, p. 41.

[205] *Idem: «The private actors involved in the programming and production phases of the AI applications involved in the crime, as well as their end user, need to have somehow the possibility and the duty to know its modus operandi, at least to an extent that makes them able to predict the causation of the typified harmful result».*

[206] Barresi, O., *op cit*. N. 201, p. 287, también propone la posibilidad teórica de aplicar la estructura de la *actio libera in causa,* pero solo para los casos en que haya una intención del autor de perpetrar el delito valiéndose de la IA simplemente como herramienta, Zhao, S., *op cit.* N. 160, p. 11. Lo que se está proponiendo, en esencia, con esta estructura, cuando se aplica en los casos en que el humano no tiene una intención clara de usar la herramienta para cometer el delito, es estirar los límites del dolo hasta extremos insospechados, rozando la presunción de dolo o una concepción absolutamente normativa del mismo. Cabe añadir, asimismo, que no es cierto, al menos no en nuestro ordenamiento, que la estructura de la *actio libera in causa* esté exclusivamente prevista para los supuestos de inimputabilidad, pues de hecho se aplica sin mayores problemas en los casos de ausencia de acción y en cuanto estructura de imputación (lo que Silva Sánchez ha venido a llamar como «doctrina» de la *actio libera in causa,* en contraposición a los «casos» de *actio libera in causa).* De hecho, se ha planteado su aplicación respecto de estructuras muy diversas; así, en los casos de provocación de las causas de justificación, respecto del modelo de autoría

Se comienza a tratar la pregunta relativa a la culpabilidad en los casos en que el resultado dañino generado por la IA no es querido ni predecible para el agente humano, siendo exclusivamente atribuible a la evaluación realizada por el sistema. Queda claro que en los casos en que las personas involucradas conocían la capacidad de la IA de aprender y tomar decisiones, pero no pudieron prever la realización de un daño ni lo desearon, la intención queda excluida (habrá que ver si hay espacio para la conducta imprudente). En todo caso, la mayoría de los informes nacionales señala que cuando el funcionamiento del sistema que generó el daño no era previsible para el agente humano normal, razonable y cuidadoso, no puede afirmarse la culpabilidad, y por tanto debe excluirse la responsabilidad criminal. La clave está aquí, por tanto, en la distinción entre precaución y obligación [207].

Pero muchos de los informes nacionales destacan que podría surgir el caso en que la característica de imprevisibilidad de los sistemas de IA más sofisticados excluya directamente cualquier forma de *mens rea* (no solo en los usuarios, sino en las personas responsables de su fabricación). Así se señala por Bélgica, Grecia, Italia, Alemania, Francia, Portugal, Finlandia, Hungría y Ucrania [208]. Habla entonces Picotti, explícitamente, de *«responsibility gap»,* y enumera las posibles soluciones que la literatura ha planteado en los diferentes países.

– La prohibición directa de ciertas prácticas de IA con base en el principio de precaución [209].

– El castigo a programadores, diseñadores y usuarios que no cumplan con una serie de medidas de seguridad pensadas para limitar la peligrosidad, con base en su deber de neutralizar el riesgo [210].

---

mediata o incluso para fundamentar la responsabilidad basada en delitos de peligro (*actio praecedens*). En todo caso, la decisión sobre la pena a imponer para cada uno de los casos (exenta, plena o rebajada), ha sido objeto de debate, tanto político-criminal como dogmático en la larga tradición que ha representado esta figura, no exenta de polémicas en cuanto a su posible contrariedad con el principio de culpabilidad o legalidad, identificándose tradicionalmente dos modelos (el de la excepción y el de la tipicidad). Sin ánimo de exahustividad, explicando adecuadamente los términos del debate: Silva Sánchez, J. M., «A vueltas con la actio libera in causa», *InDret*, 4/2016.

[207] El informe francés, de hecho, pone de manifiesto sus reticencias respecto de la aplicación de la lógica del principio de precaución (que se desarrollará *infra),* pues las reglas de prudencia no se inspiran en la precaución, sino en la experiencia. Lacaze, M./ Walther, J., «French report on traditional criminal law categories and AI», RIDP, 94, 1, 2023 p. 175.

[208] Picotti, L./ Panattoni, B., *op cit*. N. 183, p. 43.

[209] *Ibid.* Aunque en esta parte del informe solo da algunos ejemplos no circunscritos a países concretos, en la p. 26 del informe concreta algunas propuestas de Estados ya firmes en este sentido (así, Francia y su prohibición de armas autónomas, Bélgica y el debate acerca de la prohibición del reconocimiento facial).

[210] En este sentido alude al informe francés. Aunque no se especifica aquí si estas sanciones deben ser penales, hemos de entender que sí, pues Picotti separa en un apartado diferente la posibilidad de que

– El reconocimiento de áreas de riesgo permitido [211].

– Recurrir al modelo de responsabilidad penal por el producto, con base en el principio de precaución [212].

– Desplazar el problema regulatorio al ámbito del derecho administrativo [213].

Las conclusiones a las que llega el relator general implican que hay dos posibles formas de concebir a la IA: o bien considerarla simplemente como una nueva generación de medios técnicos con los que realizar comportamientos criminales (en cuyo caso el conjunto de categorías tradicionales del derecho penal sigue siendo adecuado), o bien considerarla como un nuevo agente social artificial en el que se delegan total o parcialmente algunas decisiones relevantes. Desde esta perspectiva sí podrían necesitarse nuevos modelos de responsabilidad. Reconoce Picotti que las políticas actuales se centran en asegurar el papel protagonista del agente humano durante todas las fases del ciclo de vida del sistema. Sin embargo, las políticas legislativas deben proponer soluciones y regímenes de responsabilidad diferentes según el grado de autonomía de las aplicaciones de IA [214].

Las conclusiones finales destacan, en efecto, cuáles son las tareas pendientes, señalando que, aunque algunos delitos existentes podrían aplicarse a los crímenes relacionados con la IA, la literatura ha comenzado a señalar posibles vacíos en la protección penal, específicamente en sistemas de IA con un alto nivel de autonomía. En esta fase preliminar, el análisis de los informes nacionales y especiales sugiere que, en el futuro cercano, los legisladores deben ajustar y actualizar el modelo de responsabilidad penal, estableciendo una regulación detallada de sus respectivas obligaciones para evaluar y prevenir los riesgos en este campo [215].

---

se trate solo de sanciones administrativas. Asimismo, en las conclusiones del informe francés, puede leerse: *«when the autonomous decision of AI is involved, it seems inappropriate to exclude any possibility of criminal liability for users, and even more so for programmers and manufacturers»*. LACAZE, M./ WALTHER, J., *op cit*. N 207, pp. 188-189.

[211] En este punto se cita el informe alemán (BECK, S./GERNDT, S., *op cit*. N.186), pero se reconoce que podría dar lugar a discrepancias entre los diferentes sistemas nacionales, dificultando el tratamiento de casos transfronterizos.

[212] AIRES DE SOUSA, S., *op cit.* N. 187, pp. 327-328.

[213] LACAZE, M./ WALTHER, J., *op cit*. N 207.

[214] PICOTTI, L./ PANATTONI, B., *op cit*. N. 183, p. 50.

[215] *Ibid.,* pp. 50-51.

## 2. EL ESTUDIO SOBRE LOS VIEJOS Y NUEVOS DELITOS

El segundo de los informes objeto de estudio, el General Report del profesor Miró-Llinares [216], nos permitirá situarnos en una perspectiva de «vista de pájaro», pues se considera más útil, aún en este punto incipiente, conocer cuáles son las propuestas de reacción al fenómeno de la IA (tanto a nivel legislativo, como dogmático y jurisprudencial), que centrarnos en legislaciones concretas ya aprobadas cuya aplicación, es, en todo caso, de muy corto recorrido y relativa a cuestiones demasiado fragmentadas.

El informe general que Fernando Miró, en calidad de relator, escribe, se basa a su vez en informes nacionales de 22 países, la mayoría de Europa continental (España, Francia, Italia, Austria, Países Bajos, Bélgica, Grecia, República Checa, y Alemania), de Europa del Este (Polonia, Hungría, Croacia, Rusia, Turquía, Ucrania y Rumanía), y del norte de Europa (Finlandia), pero también se incluyen informes de Latinoamérica (Argentina, Chile y Perú), África (Egipto) y Asia (China).

Partiendo de una definición amplia de IA, basada, en cierto modo, en la ofrecida por la OCDE [217], el informe, que en principio se centra en la necesidad de reforma de la parte especial de los códigos penales, no puede dejar de lado la afectación que se produce sobre la parte general, ya que la IA trae consigo, inevitablemente, una afectación de las estructuras de imputación y una necesidad de reflexionar sobre la legitimidad de la intervención penal. El informe enumera algunos de los incidentes reportados por los países relacionados con la IA, destacando que el país que más reporta es China [218], debiéndose esto a que es, quizá, el país que ha implantado de manera más general esta tecnología. Es necesario, sin embargo, resaltar que hay dos tipos de daños reportados: los maliciosos, es decir, aquellos en que las tecnologías se utilizan con el propósito de generar un daño, y aquellos que producen daños no intencionados, que a su vez pueden deberse a un mal uso o a defectos en el diseño.

En el primer grupo de casos, los daños *intencionados* causados con una herramienta de IA, se reportan, especialmente, delitos relacionados con el fraude (de forma destacada cometidos en el ciberespacio, pero no solo), así

---

[216] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160.

[217] «*AI system: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment*», OECD Recommendation of the Council on Artificial Intelligence, 22 de mayo de 2019, disponible en: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449> [Accedido por última vez el 21/01/2026].

[218] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 26.

como crímenes de tipo económico o informático. Especialmente destacadas son las referencias a los delitos de *hacking* (ahondando, en este caso, en la posible aplicación de los delitos denominados informáticos a los supuestos en que interviene la IA) y a la expansión del fenómeno del *deep fake* [219], que se relaciona, a su vez, con afectación a intereses diversos: la desinformación, la

[219] En lo relativo a las propuestas de tipificación del contenido delictivo relacionado con el *Deep fake* en España, cabe resaltar la existencia de una Propuesta de Ley Orgánica que prevé, en España, la modificación de ciertos contenidos del CP, entre ellos, el delito de injurias y los delitos electorales. Sin embargo, esta propuesta, de 2023, sigue en *standby,* habiéndose señalado ya por la doctrina numerosas contradicciones respecto de las obligaciones internacionales que atañen a España: por ejemplo, respecto de la Directiva (UE) 2024/1385 del Parlamento Europeo y del Consejo, de 14 de mayo de 2024, sobre la lucha contra la violencia contra las mujeres y la violencia doméstica. Así, la Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial, presentada por el Grupo Parlamentario Plurinacional SUMAR, de 13 de octubre de 2023. Crítico con esta forma de abordar el fenómeno: Santisteban Galarza, M., «La criminalización de las ultrafalsificaciones (con especial atención a las implicaciones de la normativa europea de servicios digitales e inteligencia artificial), *RDPC*, 31, 2024. A esta propuesta cabe sumar también otras dos recientes, ambas aún en proceso de discusión. La primera, también en el ámbito penal, si bien aplicable solo a los menores de edad, propone la tipificación de la difusión de contenido *deepfake* no consentido dentro de los delitos contra la integridad moral, mientras que la segunda, también en forma de Anteproyecto, propone reformar la Ley Orgánica de protección del derecho al honor y propia imagen, reconociendo, si bien en el campo civil, que el fenómeno *deepfake* se enmarca dentro de los daños contra el honor y la propia imagen. Nos referimos, respectivamente, a: Proyecto de Ley Orgánica para la protección integral de las personas menores de edad en entornos digitales (aprobado por Consejo de Ministros el 21 de marzo de 2025), y Anteproyecto de Ley Orgánica de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen (aprobado por Consejo de Ministros el 13 de enero de 2026). Comentando la confusión actual de la regulación española: Rodríguez López, S., «Hacia un nuevo delito de difusión no consentida de material íntimo manipulado (deepfakes)», en AAVV., *Un derecho penal moderno para una democracia del siglo XXI, Libro homenaje al Prof. Dr. Carlos Martínez-Buján Pérez*, Tirant Lo Blanch, Valencia, 2025, pp. 628-636.

Nota complementaria de actualización: con posterioridad al cierre de redacción, debe advertirse que el acuerdo político provisional alcanzado en trílogo el 7 de mayo de 2026 sobre el Digital Omnibus on AI (COM(2025) 836 final) incorpora –junto a la recalibración del calendario de aplicación del Reglamento (UE) 2024/1689– la prohibición expresa, como práctica vedada en sede del art. 5, de los sistemas de IA destinados a generar contenido sexualmente explícito e íntimo sin consentimiento y material de abuso sexual infantil, comprendiendo paradigmáticamente las denominadas nudifier apps. Esta prohibición, llamada a desplegar plena eficacia tras la publicación definitiva del texto en el DOUE, se proyecta precisamente sobre el ámbito sustantivo que las propuestas legislativas españolas aquí citadas intentan abordar desde la perspectiva penal y civil, abriendo un nuevo escenario de articulación entre prohibición administrativa europea ex ante y  respuesta nacional ex post que, previsiblemente,condicionará la tramitación pendiente de las iniciativas en curso.

Nota complementaria de actualización: con posterioridad al cierre de redacción, debe advertirse que el acuerdo político provisional alcanzado en trílogo el 7 de mayo de 2026 sobre el Digital Omnibus on AI (COM(2025) 836 final) incorpora –junto a la recalibración del calendario de aplicación del Reglamento (UE) 2024/1689– la prohibición expresa, como práctica vedada en sede del artículo 5, de los sistemas de IA destinados a generar contenido sexualmente explícito e íntimo sin consentimiento y material de abuso sexual infantil, comprendiendo paradigmáticamente las denominadas nudifier apps. Esta prohibición, llamada a desplegar plena eficacia tras la publicación definitiva del texto en el DOUE, se proyecta precisamente sobre el ámbito sustantivo que las propuestas legislativas españolas aquí citadas intentan abordar desde la perspectiva penal y civil, abriendo un nuevo escenario de articulación entre prohibición administrativa europea ex ante y respuesta nacional ex post que, previsiblemente, condicionará la tramitación pendiente de las iniciativas en curso.

influencia en la política (a título de ejemplo, Francia ha sido uno de los primeros países en castigar los fraudes electorales), así como en la violencia de tipo sexual (este último fenómeno comienza, también, a ser regulado de forma incipiente en las normativas nacionales, así como en la normativa europea).

En el segundo grupo de casos, los de daños *no intencionados*, se enfatizan los problemas generados por los vehículos autónomos, destacando que la regulación sí se va haciendo presente en algunos países (paradigmáticamente Alemania, que es la nación que por el momento más ha regulado a este respecto, y Francia, que ha previsto un régimen de responsabilidad penal y también extrapenal; aunque en EEUU y Finlandia ha llegado a haber condenas por homicidio involuntario y delito de tráfico, respectivamente, a las personas que conducían dichos vehículos [220]).

También ha llamado la atención de las autoridades nacionales encargadas de informar sobre estos hechos la proliferación de casos de discriminación algorítmica por el uso, cada vez más omnipresente, de algoritmos predictivos [221].

Finalmente, los daños a la privacidad también son destacados como un ámbito de preocupación entre los diferentes países, especialmente con el uso de herramientas que usan datos de forma masiva e indiscriminada, con Chat-

---

[220] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M.*, op cit.* N. 160, p. 35. Precisamente el sector de la conducción autónoma ha sido uno de los que han despertado mayor interés no solo a nivel teórico (siendo ya paradigmáticos los trabajos que tratan de resolver los dilemas morales de la conducción autónoma; por todos: Coca Vila, I., «Coches autopilotados en situaciones de necesidad: una aproximación desde la teoría de la justificación penal», *Cuadernos de Política Criminal:* 122, II, 2017, pp. 235-275). En el ámbito germano: Hilgendorf, E., «The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems», en Hilgendorf, E/Feldle, J. (Eds.), *Digitization and the Law* Nomos, 2018, pp. 57-90; Greco, L., «Vehículos de motor autónomos y situaciones de colisión», en Basso, G. /Cancio Meliá, M., /Maraver Gómez, M. /Fakhouri, C., (Coords.), *Libro Homenaje al prof. Dr. Agustín Jorge Barreiro*, Ed. Universidad Autónoma de Madrid, 2019; y Hörnle, T./Wohlers, W., «The Trolley Problem Reloaded, ¿Cómo deben programarse los vehículos autónomos para los dilemas de ¨vida contra vida" ? », (trad. por José Béguelin), en Hörnle, T., *Criminalización, castigo, y dilemas morales en la obra de Tatjana Hörnle,* Buenos Aires, Editores del Sur, 2022. Para un repaso de la normativa más reciente sobre conducción autónoma: Giannini, A./ Kwik, J., «Negligence failures and negligence fixes. A comparative analysis of criminal regulation of AI and autonomous vehicles», *Criminal law fórum*, 34, 1, 2023, pp. 43-85.

[221] La bibliografía sobre los riesgos de la discriminación algorítmica resulta, a estas alturas, inabarcable. Algunos de los trabajos que, a título de ejemplo, se han encargado de esta problemática: Xenidis, R., «Tuning EU equality law to algorithmic discrimination: Three pathways to resilience», *Maastricht Journal of European and Comparative Law*, 27, 2020, pp. 736-758. También: Hacker, P., «Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law», *Common Market Law Review*, 55(4), 2018, pp.1143-1186. Flores Anarte, L., «Sesgos de género en la Inteligencia Artificial: El Estado de derecho frente a la discriminación algorítmica por razón de sexo», *Revista Internacional De Pensamiento Político*, 18(18), 2023, pp. 95-120. Bravo Bolado, A., «Justicia algorítmica. Un enfoque sociotécnico», *Revista de Estudios Penales y Criminológicos,* n. 44, 2023, pp. 1-43. Senent Vidal, M. J./Ventura Franch A., «Inteligencia artificial e igualdad de mujeres y hombres: una regulación jurídica necesaria. Especial referencia al proyecto de reglamento europeo», *IgualdadES*, 10, 2024, pp. 45-71.

GPT como ejemplo paradigmático. Asimismo relevante, en lo relativo a la protección de la privacidad en el actual panorama, es la discusión sobre las polémicas técnicas de vigilancia biométrica y reconocimiento facial, que han sido objeto de encarnizado debate en las negociaciones del Reglamento de IA. Destaca, en este sentido, la sentencia del TEDH del caso Glukhin contra Rusia [222].

Lo cierto es que, pese a todos los casos reportados y las preocupaciones expresadas por los informes, ninguno de los países ha adoptado, aún, una regulación general vinculante respecto a esta tecnología, y como mucho se ha procedido a regular algún sector muy concreto o se han ido aprobando estrategias nacionales de carácter no vinculante (estas, en cambio, sí son mucho más omnipresentes).

Las conclusiones que del informe se derivan en lo relativo a la afectación a la parte general resultan interesantes.

En primer lugar, es posible que haya que plantearse la creación de nuevos intereses dignos de tutela penal según van surgiendo nuevas formas de criminalidad, en cuyo caso no quedará más remedio que tipificar delitos *ex novo*. Se ofrecen algunos ejemplos de estos posibles nuevos intereses [223]: quizá sea necesario proteger a los propios sistemas de IA cuando estos sean objeto de ataque, dado que cada vez más se usarán como prestadores de servicios de toda clase; también se debate sobre la cuestión relativa a la protección de los contenidos creados mediante IA, aunque en este extremo no parece haber consenso ninguno entre países.

La realidad es que, aún hoy, no hay acuerdo a la hora de determinar si con la irrupción de la IA aparecen genuinamente nuevos intereses que proteger, o si no son más que los intereses tradicionales atacados por medio de técnicas novedosas.

Hay quienes piensan en el primer sentido, apostando por la necesidad de proteger ciertas características valiosas derivadas de los sistemas (a título de ejemplo, Turquía apuesta por proteger la transparencia y la responsabilidad de la IA en sí mismas, a través de genuinos tipos penales al efecto). Asimismo, el

---

[222] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 40. Se refieren al caso Glukhin vs. Rusia (2023). Núm. 11519/20, Tribunal Europeo de Derechos Humanos, [TEDH, 4 de julio de 2023]. En este caso el TEDH condenó el uso de tecnologías de reconocimiento facial por parte de Rusia para identificar a un manifestante solitario y localizarlo en el metro de Moscú, considerando que vulneró los artículos 8 (privacidad) y 10 (libertad de expresión) del Convenio. El fallo subrayó que la intimidad no se restringe al ámbito físico privado, sino que incluye aspectos sociales de la identidad, y que, aunque existía una base legal para la identificación biométrica, la injerencia no cumplió el criterio de «necesidad en una sociedad democrática». El TEDH destacó la ausencia de una urgencia social justificada para emplear inteligencia artificial en este contexto, ya que el caso no involucraba delincuencia organizada o terrorismo, sino una infracción administrativa menor. Además, criticó la vigilancia masiva sin control judicial previo, marcando una línea clara entre objetivos legítimos (como combatir crímenes graves) y usos desproporcionados que reprimen protestas pacíficas, reforzando así límites al despliegue indiscriminado de estas tecnologías.

[223] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 42.

debate sobre si será beneficioso proteger la realidad como valor frente a la ficción, ante la creciente aparición de mecanismos que son capaces de alterar nuestra percepción sobre lo que es realidad y lo que es falso, se ha puesto sobre la mesa [224]. Otro de los ejemplos de posibles nuevos bienes jurídicos tiene que ver con los usos prohibidos de los sistemas de IA contemplados en el novedoso Reglamento, pues parece discutible que algunos de ellos puedan encontrar un bien cuya vulneración esté ya cubierta por un delito previo (así, en el caso de las técnicas subliminales de manipulación, o la explotación de vulnerabilidades personales o grupales, así como los sistemas de categorización o *social scoring)* [225]. ¿Qué delito o bien jurídico protege al humano frente a una posible clasificación social o puntaje realizado por un sistema de IA?

En segundo lugar, hay un acuerdo común en considerar que con la utilización de este tipo de sistemas se producirá un *aumento del riesgo o del nivel de daño* causado al cometer delitos tradicionales, es decir, se producirá un «cambio en la dimensión del daño y del riesgo» [226].

Estas dos consecuencias pueden dar lugar a la creación de lo que se conoce como *«AI-related offences»* [227]. Esta nomenclatura es interesante, porque aglutina toda una serie de delitos entendidos de una forma genérica. Es decir, no se trata de una tipología concreta de delitos que puedan cometerse solo con IA, sino que abarca todos aquellos delitos cuya tipificación responde a las nuevas necesidades de intervención del Derecho penal, relacionadas con todos los riesgos y daños que derivan del desarrollo de los sistemas de IA. Puede comprender delitos creados *ex novo,* así como delitos ya existentes que se deben reformar en algún extremo para cubrir las nuevas características de la tecnología.

Lo que nos interesa llegados a este punto es destacar cuáles son las técnicas que, según la doctrina y la opinión de los expertos encargados del estudio

---

[224] En todo caso, y de forma preliminar, pues será en el capítulo sexto donde se profundice sobre la creación de posibles bienes jurídicos de nueva creación, baste dejar aquí apuntado que no resulta en absoluto novedoso que los códigos penales protejan de alguna forma los atentados a la verdad. Así, es el caso de los delitos de falso testimonio, calumnias o injurias, donde podemos afirmar que el objeto de protección no es la verdad en sí misma, sino ciertos bienes jurídico penales finales, por lo que estos delitos se configuran, así, como una especie de delitos de peligro, con una estructura que se explicará *infra*.

[225] Previstos en el artículo 5 del Reglamento de IA. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

[226] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 40.

[227] *Ibid.,* p. 40 y ss. Se aclara que este término no comprende una tipología de delitos que sólo puedan cometerse con el uso de IA y que se usa de forma similar al término *«AI Crimes»,* utilizado para referirse a los actos delictivos en los que la inteligencia artificial puede desempeñar un papel cada vez más esencial en el futuro. Floridi, L./King, T. C./Aggarwal, N./Taddeo, M., *op. cit.* N. 43, pp. 89-120.

pormenorizado de este fenómeno en cada país, son más adecuadas para afrontar esta nueva realidad delictiva. Miró enumera algunas de ellas [228]:

a) el diseño, la producción o el uso de sistemas de IA que se consideren directamente nocivos o peligrosos;
b) nuevos delitos que puedan sancionar actos de infracción de deberes administrativos, o conductas preparatorias que, en otro caso, pueda considerarse que entrañan riesgo suficiente para ser castigadas anticipada y autónomamente;
c) revisión de los delitos que sólo admiten dolo, contemplando los delitos por imprudencia;
d) modificación de los delitos que no responden con penas proporcionales a la gravedad de las conductas lesivas o peligrosas cuando se realizan utilizando sistemas de IA;
e) delitos agravados que castigan conductas que son castigadas con menor pena cuando se realizan sin IA, por el mayor disvalor que supondría el uso de esta tecnología.

La mayor laguna, según se deriva del informe, no estará, por tanto, en encontrar tipos delictivos que encajen con las conductas dolosas tipificadas ya en la mayoría de los Códigos penales, sino más bien se trata de preguntarse si se va a poder imputar responsabilidad en los casos de sistemas de IA no considerados nocivos o maliciosos en sí mismos.

Se reconoce que ya el Reglamento fija un marco de referencia, por lo que los legisladores podrán dejarse guiar por las indicaciones que, por ejemplo, se realizan en cuanto a las tecnologías de uso prohibido, pues en estos casos será más sencillo penalizar su simple uso o creación, siendo más complejo determinar si cabe la penalización en el resto de los sistemas (paradigmáticamente en los sistemas de alto riesgo). Destaca Miró en este punto que no queda muy claro si, en utilización de su margen de apreciación como Estados soberanos y competentes en materia penal, los Estados miembros pueden prohibir o penalizar aquellos sistemas que no están prohibidos por el Reglamento [229]. Que una prohibición penal de tales características pueda considerarse una restricción al desarrollo prohibida por la competencia europea o bien una manifestación del *ius puniendi* soberano dirigido a proteger intereses valiosos es algo que, aún, está pendiente de debatir. Lo cierto es que al menos algún Estado miembro ha restringido, ya, un uso no expresamente pros-

[228] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 41.
[229] Este argumento lo basa precisamente en el Considerando primero del Reglamento de IA, que establece que los Estados no pueden imponer restricciones al desarrollo de los sistemas de IA, salvo autorización expresa del Reglamento. *Ibid.* p. 46.

crito por la regulación europea –Francia ha prohibido las aplicaciones de predicción de las decisiones judiciales, e Italia prohibió, en su momento, el uso de modelos de lenguaje como Chat-GPT–, por lo que habrá que esperar a las reacciones que surjan en adelante para poder analizar este problema.

Entrando a analizar la primera de las propuestas genéricas, que es un lugar común en la doctrina, se analiza en el informe la posibilidad de crear nuevos delitos de peligro para castigar actos preparatorios relacionados con el diseño, producción o uso de los sistemas de IA. Esta es una de las alternativas propuestas en numerosos informes nacionales y contempladas por una buena parte de la doctrina como forma de reaccionar a los riesgos comunes de los denominados como *«AI-related crimes»*. De cara a regular estos nuevos delitos, es necesario, según Miró, atender a las características que son, en sus propias palabras, aquellas que cambian las reglas del juego («*game-changers*» [230]).

En este sentido, pone el foco en el papel de la automatización de la toma de decisiones, destacando que la IA ha supuesto un cambio en el momento en que la agencia humana se vuelve esencial y esto debe tener implicaciones en cuanto al momento en que se inicia la responsabilidad penal [231]. La justificación teórica que ofrece Miró acerca de la utilidad y justificación de los delitos de peligro con la punición adelantada tiene que ver con que el momento en que la responsabilidad humana es esencial no está ya cercano a la exteriorización del daño, sino en los primeros pasos del ciclo de vida del sistema, tales como el diseño o la selección de datos. Por ello, estos momentos previos, (en los que se mantiene aún un control más férreo por parte de la persona física), son los relevantes para el derecho penal, por ser los únicos en que el daño puede ser previsto y prevenido. En definitiva, lo que concluye, con cita a Danaher [232], es que la habilidad esencial de estos sistemas de actuar de forma autónoma y su capacidad de aprendizaje nos obliga a repensar las normas que fijan la responsabilidad para evitar que se disminuya la prevención del daño. Es decir, en sistemas pensados para desincentivar un comportamiento (criminal), la responsabilidad debe ser transferida al momento en que la persona realmente puede motivarse, que es cuando la IA puede ser controlada.

Esta idea que aquí se expresa por el relator general casa perfectamente con lo señalado arriba, al explicar lo que ocurre con las denominadas lagunas de respon-

---

[230] *Ibid.,* p. 47.

[231] *Idem.: «we shall focus now on the role of automation on decision making. If looked carefully, it can be inferred that IA implies a change in the «moment» in which human agency becomes crucial. Accordingly, this could shift the stage in which criminal liability is established».*

[232] Danaher, J., *Automation and Utopia: Human Flourishing in a World without Work,* Harvard University Press, 2019.

sabilidad en el plano dogmático. A medida que la autonomía aumenta, el control humano disminuye, y para responsabilizar debe haber control por parte del agente.

Esto implica, según Miró, una necesidad de tipificar acciones anteriores al daño, en forma de delitos autónomos por un diseño deficiente o que no permita el control sobre los sistemas [233], siempre fundamentados en el riesgo que conlleve el incumplimiento de deberes de diligencia propios del sector.

De hecho, no solo la doctrina ha sido partidaria de esta solución, sino que varios países han expresado su interés en este modelo, especialmente en lo relativo a los sistemas de alto riesgo (así, Chile, Finlandia o República Checa). Repasando algunas propuestas nacionales:

– El informe checo apuesta por penalizar delitos que castiguen el mero hecho de no mantener unos niveles de seguridad óptimos en ciertos momentos de la vida de la IA, sin necesidad de que se produzca un daño, y propone, asimismo la creación de un delito de peligro en sentido muy amplio [234] («poner en peligro a otras personas utilizando un dispositivo de inteligencia artificial sin el debido mantenimiento o actualización, o siguiendo utilizándolo a pesar de la advertencia de que se ha producido un fallo fatal en el modelo concreto y de la llamada al servicio de mantenimiento»).

– El informe argentino también propone tipificar las violaciones de este tipo de deberes, y el italiano apoya un sistema parecido a los modelos de gestión de riesgo. De hecho, Italia es proclive a tipificar conductas consideradas «precursoras» de un delito más lesivo posterior, explorando la posibilidad de tipificar delitos relacionados con el mero diseño o control de algunos sistemas de IA, paradigmáticamente los de alto riesgo.

– El informe croata, por su parte, apunta que se deben introducir nuevos delitos que tipifiquen un mal diseño o falta de control en algunos sectores de la IA (concretamente, se refiere a las armas autónomas, las IA para el uso médico y las herramientas financieras).

– Asimismo, China es partidaria de un enfoque basado en el modelo de imprudencia en el diseño programación y uso [235].

---

[233] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 48.

[234] *Ibid.,* p. 49. Provaznick, J.: «Czech Republic national report on criminalisation of AI-related offences», *RIDP,* 95, 2024, p. 257: *«criminalization of endangering others by operating an AI-device without due maintenance or actualization, continuance of using it despite a warning on an occurrence of a fatal flaw in the particular model and call to maintenance service».*

[235] Así, proponiendo un modelo completo de imprudencia, analizado desde el punto de vista dogmático, Zhao, S., *op cit*. N. 160.

Aunque, como se ha visto, muchos son los países y autores partidarios de este adelantamiento de las barreras de punición, hay quienes consideran que no son necesarias nuevas tipificaciones, bastando con aplicar la ya existente legislación penal sobre productos defectuosos (en este sentido, el informe nacional holandés, pero también muchos autores, entre los que cabe destacar, por su elaboración del modelo, a Aires de Sousa [236]).

También se reportan casos de países no tan proclives a este adelantamiento, aunque suponen la excepción. Así, el informe polaco, aunque reticente en un principio a adelantar tanto el momento de atribución de responsabilidad, acaba por aceptar esta opción si los bienes protegidos son reconocibles de forma objetiva ya en el momento de crear la herramienta [237]. Para Miró, sin embargo, este reconocimiento previo de la posibilidad de crear un daño futuro puede estar cubierto ya por las formas de comisión imprudentes, por lo que, quizá, no sería necesario ese adelantamiento. Lo que queda, pues, es un trabajo de concreción de hasta dónde deben llegar las obligaciones y si estas deben tener repercusión penal, o, como apuntan otros autores, realizar una reflexión sobre cuál debe ser el conocimiento exigido para hablar de imprudencia en estos escenarios. El Reglamento, continúa Miró, es una base excelente en este sentido, facilitando la creación de posibles normas penales en blanco o bien transponiendo los deberes expresamente previstos en la regulación europea a la regulación penal nacional, asegurándose siempre de que exista un interés digno de protección penal.

Lo cierto es que en la doctrina se encuentra también mucho apoyo a la aplicación del modelo de delitos imprudentes, y muchos de los informes nacionales señalan que sería conveniente que algunos delitos que solo prevén la modalidad dolosa se castiguen en su forma imprudente [238]. Así se manifiesta, de hecho, en el informe español, donde se pone de relieve que esto supone un problema para la comisión de delitos por parte de productores, distribuidores o usuarios. Esta realidad ha sido advertida por China, por Francia, e incluso por el Consejo de Europa, hablando en este caso de un verdadero *«gap»* o laguna de responsabilidad.

Desde luego que la discusión, en este caso, se deberá centrar en la concreción de los deberes de cuidado, destacándose que no se debe caer en enfoques basados en responsabilidad objetiva, aunque existe un cierto riesgo laten-

---

[236] Aires de Sousa, S., *op cit*. N. 187. Más allá de su exposición como relatora nacional, la autora se ha mostrado partidaria de este modelo en todas sus publicaciones. Paradigmáticamente, Aires de Sousa, S., «Não fui eu, foi a máquina: teoria do crime, responsabilidade e inteligência artificial», en Miranda Rodrigues, A. (Coord.), *A Inteligência Artificial no Direito Penal*, Almedina, Coimbra, 2020. Dicho modelo será objeto de análisis pormenorizado *infra* (capítulo quinto).

[237] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M., *op cit.* N. 160, p. 50.

[238] *Idem.*

te de que esto ocurra, ante la falta de previsibilidad del daño y un desarrollo científico prematuro y lleno de incertidumbres [239]. Resulta, pues, una cuestión de política criminal determinar si es legítimo perseguir este tipo de infracciones imprudentes, cuestión que es destacada por el informe español [240].

Finalmente, se explora la posibilidad de crear algún tipo de agravación específica por el uso de la IA, pues hay bastante consenso en considerar que el daño es más significativo, afectando a más sujetos, facilitando la comisión y dificultando la persecución (así lo consideran Rumanía, Italia, Alemania, España o Finlandia) [241]. Sin embargo, algunos países ponen de manifiesto el peligro de que la agravación se aplique de forma automática, sin entrar a analizar la mayor potencialidad lesiva en el caso concreto, lo cual llevaría a una forma objetiva de responsabilidad. Se entiende que esta solución se aplicaría, sin mayores problemas, cuando, analizado caso por caso y nunca de forma automática, se escogiera de forma dolosa una herramienta de IA con conocimiento de esta mayor potencialidad lesiva, pero no sería una solución aplicable a los supuestos imprudentes o de daño remoto.

Se realiza una pequeña reflexión final acerca de la posibilidad de que los regímenes administrativo y penal deban convivir, lo cual puede suponer algún tipo de conflicto (así, para Finlandia o Austria es un problema que viene unido con la tendencia expansiva del derecho penal del riesgo), pero para otros países, como es el caso de España o China, no hay problema en esta convivencia, que no es en absoluto novedosa y que además salvaría todos los inconvenientes manifestados por la doctrina si fuera capaz de proteger por vía penal solo los intereses más importantes, respetando el principio de *ultima ratio*, que es mencionado por prácticamente todos los informes nacionales.

Finalmente, el informe general se encarga de recopilar la información más relevante sobre los delitos que los relatores nacionales consideran que se verán más afectados con el uso de la IA. En este sentido, se destaca que algunas de las tecnologías más peligrosas para la salud y la vida son las denominadas armas autónomas letales (que, pese a haber sido objeto de discusión desde

---

[239] Se cita en este punto el informe de Picotti. Picotti, L./ Panattoni, B., *op cit*. N 183.

[240] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M.*, op cit.* N. 160, p. 52.

[241] *Ibid.,* p. 54. En este sentido, cabe mencionar la existencia de una reciente ley italiana, primera intervención legislativa tras la aprobación del Reglamento Europeo, que dispone, entre otros contenidos, la modificación del código penal, con la creación de una agravante genérica por haber cometido el hecho mediante uso de sistemas de IA cuando por su naturaleza o modalidad de uso hayan constituido medio insidioso o haya obstaculizado la defensa pública o privada, o agravado las consecuencias del delito (modificando el artículo 61.11-undecies del *codice penale italiano*). La norma a la que nos referimos es: *Legge 23 settembre 2025, n. 132: Disposizioni e deleghe al Governo in materia di intelligenza artificiale*. Gazzetta Ufficiale, n. 223. Asimismo, se prevén agravaciones específicas para algunos delitos. Para un comentario crítico sobre esta norma: Scollo, L., «*'intelligenza artificiale entra nel Codice penale*», *Diritto Penale e processo,* 6/2025.

muy temprano no han alcanzado ningún tipo de contexto regulativo general [242]). En cuanto a los bienes jurídicos de carácter personal, se destaca la privacidad como el que se ve más afectado por las nuevas técnicas de IA, destacando en todos los informes la importancia de la legislación de protección de datos y una posible transformación de la protección de la privacidad, que puede pasar a considerarse un bien de tipo colectivo. Se plantea también si la discriminación algorítmica (fenómeno de sobra constatado en todos los sistemas), podría y/o debería encontrar reflejo en la legislación penal, sin que se consiga un consenso. Se destaca como problema específico en todos los informes el daño que se puede causar a otros bienes de carácter eminentemente personal con las aplicaciones de la tecnología *Deep fake*. Las sanciones que se pueden aplicar en los diferentes ordenamientos son muy variadas, y no está muy claro si llegan a captar el desvalor adecuado según el daño causado [243].

En lo relativo al cibercrimen y la criminalidad económica, la persecución, ya de por sí difícil, por las características propias de esta tecnología, se torna aún más complicada, aunque en general hay consenso al entender que sí hay regulación penal adecuada, con la sola duda de si es aceptable aplicar el Convenio de Budapest [244] a las IAs, considerándolas meramente un programa informático más.

En lo relativo a los delitos de propiedad intelectual, falta la regulación administrativa que establezca si hay o no algún tipo de derecho reconocido a las creaciones de los sistemas de IA. Entre tanto, esta cuestión sigue siendo polémica y diversa en los ordenamientos.

En cuanto a los delitos contra el mercado y los consumidores, resalta la preocupación relativa a la negociación algorítmica de alta frecuencia, donde destaca el caso Coscia, por ser la primera condena penal por fraude en el mercado utilizando un algoritmo de negociación de alta frecuencia [245]. No habrá

---

[242] La literatura en torno a la problemática de las armas autónomas letales resulta, asimismo, inabarcable. Entiéndanse las siguientes referencias a título ejemplificativo: Jain N., «Autonomous weapons systems: new frameworks for individual responsibility», en Bhuta, N. *et al.* (Eds.), *Autonomous weapons systems*, Law, ethics, policy, Cambridge University Press, 2016. En el panorama español: Gil Gil., A., «Sistemas de armas autónomos letales. Comentario a la posición española, con especial atención a algunos problemas jurídico-penales», 38, 2022, pp. 1-50.

[243] Precisamente en este sentido, Santisteban Galarza, M., *op. cit.* N.219.

[244] Nos referimos al Convenio de Budapest sobre la Ciberdelincuencia, adoptado por el Consejo de Europa en Budapest el 23/11/2011 y en vigor en España por el Instrumento de Ratificación del Convenio sobre la Ciberdelincuencia, hecho en Budapest el 23 de noviembre de 2001, publicado en el BOE núm. 226, de 17 de septiembre de 2010.

[245] Para más detalles sobre el caso Coscia, Aires de Sousa, S., «The importance of AI for financial crimes», *Revue Internationale de Droit Pénal*, 95/1, 2024, p. 164-165. Para un repaso sobre la problemática que suponen las aplicaciones más recientes de inteligencia artificial en el derecho penal del mercado de valores: Bravo Bolado, A., «Negociando en la sombra: Inteligencia artificial, alta frecuencia y desa-

problemas, en principio, para imputar responsabilidad cuando quien use el algoritmo lo haga con una intención fraudulenta, pero, una vez más, si nos alejamos de los escenarios dolosos, los daños para el mercado pueden ser difíciles de castigar [246].

Finalmente, se habla de los casos de delitos relacionados con el fenómeno de la desinformación, campo que, si bien está en auge, no encuentra aún una respuesta penal consolidada. Numerosos países ponen de manifiesto la potencialidad lesiva de la desinformación, pero solo algunos tipifican estos actos (Francia, prototípicamente, con una ley que prevé multa o prisión de un año para la desinformación en periodo electoral, o Grecia, que prevé penas de hasta tres años de prisión en casos de daños señalados [247]). Sin embargo, estos modelos son poco aplicados, pues confrontan directamente con el derecho fundamental a la libertad de expresión, que suele pesar más, al menos en el ámbito occidental. Llama la atención, en este sentido, cómo los países del contexto asiático (así, China, Vietnam, Singapur o Malasia) sí tipifican y aplican sanciones penales especialmente elevadas en estos casos.

## 3. CONCLUSIONES

El repaso por los informes nacionales que analizan el estado de la cuestión (tanto desde el punto de vista dogmático como legislativo) en hasta veintidós ordenamientos jurídicos sirve como muestra del estado de la cuestión. Es posible encontrar una continuidad entre las preocupaciones teóricas sobre el pretendido *responsibility gap* y la discusión dogmática-penal y legislativa de países de entornos geográficos y tradiciones jurídicas variadas. Las conclusiones que podemos extraer son valiosas, pues nos sirven para obtener un panorama general sobre el estado de la discusión científica. Así, las preocupaciones

---

fíos para el derecho penal económico», en Feijoo Sánchez, B/Peñaranda Ramos, E. (Dirs.), *Teoría y práctica de los Delitos Económicos. Hacia una racionalización de la protección penal del orden socioeconómico*, BdeF, Buenos Aires, 2025, pp. 243-288.

[246] En este sentido, Fletcher, G. G. S., «Deterring Algorithmic Manipulation», *Vanderbilt Law Review*, 74/2, p. 310; Azzutti, A., «AI trading and the limits of EU law enforcement in deterring market manipulation», *Computer Law & Security Review,* Volume 45*,* July 2022.

[247] Miró-Llinares, F., Duvac, C., Toader, T. /Santisteban Galarza, M.*, op cit.* N. 160, pp. 81-82. Así, a título de ejemplo, el artículo 191 del Código Penal griego, en la redacción dada por el artículo 36 de la Ley 4855/2021 (publicada en el FEK de 12 de noviembre de 2021), castiga con pena de prisión de al menos tres meses y multa a quien «pública o telemáticamente difunda o divulgue, por cualquier medio, noticias falsas susceptibles de causar preocupación o temor a los ciudadanos o de menoscabar la confianza pública en la economía nacional, en la capacidad de defensa del país o en la salud pública», elevándose la pena mínima a seis meses de prisión cuando los hechos se cometan reiteradamente a través de la prensa o por internet, y hasta cinco años para el editor o propietario del medio responsable.

de los teóricos y los legisladores parecen confluir: algunos rasgos concretos de las nuevas tecnologías de IA están poniendo en cuestión la validez de los delitos hasta ahora contenidos en los Códigos penales, cuya aplicación comienza a cuestionarse en los casos en que los sistemas ganan en autonomía y los humanos pierden el control. Esta dicotomía vuelve a estar presente en la discusión, y se traduce en dos reconocimientos:

Primero, que inevitablemente las categorías tradicionales del Derecho penal van a verse afectadas por determinados rasgos propios de las tecnologías de IA (paradigmáticamente, la autonomía). Y segundo, que los viejos delitos tradicionalmente tipificados en los códigos penales corren el riesgo de quedarse obsoletos, pues no existe campo de aplicación donde las herramientas de IA no puedan generar daños (no intencionados) que no encajen en las figuras delictivas clásicas, obligándonos a repensar cómo debe procederse a la creación de esas nuevas formas de delito.

Solo una vez examinadas las alternativas propuestas y defendidas por los distintos países y autores, tanto desde el punto de vista de propuestas legislativas como dogmáticas, podemos comenzar a examinar las bondades y dificultades de los modelos de tipificación propuestos.

Por tanto, una vez descrito el estado de la cuestión, los próximos capítulos se dedicarán a examinar aquellos modelos de reacción penal que, a nuestro juicio, resultan más adecuados y acordes con nuestros principios tradicionales.

# CAPÍTULO TERCERO. DELIMITACIÓN DEL OBJETO DE LA INVESTIGACIÓN

*– Oiga, capitán Achab– […] -. He oído hablar de Moby Dick. ¿No fue esa ballena la que… se llevó su pierna, señor?*

*– ¿Quién te ha dicho todo eso? ¿Cómo lo sabes, Starbuck? –contestó el capitán animándose a medida que hablaba–. ¡En efecto, muchachos, fue esa demoníaca ballena llamada Moby Dick la que se llevó mi pierna y me dejó inválido para el resto de mi vida! ¡Desde luego, hijos –sollozó, levantando ambos brazos al cielo–! ¡Pero os juro que he de perseguirla hasta más allá del Cabo de Hornos y más allá del de Buena Esperanza y del Maelstrom noruego, y más allá de los fuegos del averno si es preciso!*

*[…] – Perdone, capitán, pero sigue pareciéndome una locura y hasta una blasfemia todo eso de querer vengarse de un animal. ¡De un animal que se limitó a atacarle por instinto, porque es su deber!*

*No creáis que los arponeros y pescadores dedicados a la navegación tras las ballenas son gente impresionable o sensible al miedo. Si os digo que tales hombres desconocen realmente el miedo, me quedo corto. Pero lo que sí es cierto es el hecho de que ellos, que se hallan siempre en contacto con lo más bello y con lo más aterrador del mar, están por lo mismo más predispuestos a la influencia supersticiosa que los demás marineros. Completamente solos en aguas remotas, generalmente desérticas, se abandonan lentamente a su imaginación, enfebrecida contra su propia voluntad, llena toda ella de fantásticas imágenes de un mundo fabuloso del que, en cierto modo, acaban por formar parte… […] Pensando solamente en su venganza audaz, dura, sobrenatural, implacable, Achab ha conseguido transmitir a todos y cada uno de sus hombres la resolución heroica y absurda de aniquilar a la ballena blanca o morir en el empeño. ¿Qué clase de hechizo hizo reaccionar a la tripulación, convirtiendo a Moby Dick en un mortal enemigo común?* [248]

---

[248] Extractos de la obra *Moby Dick,* de Herman Melville (Editorial Bruguera, 5.ª edición, Colección Historias n.º 7, Barcelona, 1973), pp. 68-80. Publicada originalmente en 1851.

La lección que Melville nos enseña en estos bellos pasajes la podemos bautizar con el nombre de «error de la antropomorfización». Esta es la idea que da origen y justificación a este capítulo intermedio, que se dedica tan solo a demostrar por qué tiene sentido que quede fuera de esta investigación lo que no se va a tratar en ella.

En el libro «*When Robots Kill: Artificial Intelligence under Criminal Law*», el profesor Gabriel Hallevy narra en el prefacio una historia real sucedida en 1981 [249], tan solo ciento treinta años después de la publicación de Moby Dick. En esta ocasión la víctima es un trabajador japonés de 37 años que, mientras trabajaba en una fábrica de motocicletas, resulta muerto por un robot que trabajaba con él, al ser identificado como una amenaza a su trabajo y ser machacado con su brazo hidráulico. La pregunta que da origen a la monografía de Hallevy, tras este caso real, es la siguiente: ¿Quién debe ser considerado responsable por este homicidio? El resto de la obra constituye toda una reivindicación de la necesidad de responsabilizar a los robots.

Tras la sucesión de trágicos episodios protagonizados esta vez por robots y sistemas dotados de IA en cada vez más contextos, son numerosas las voces de aquellos «capitanes» que pretenden ver en estos sistemas aquellos tintes de inteligencia y maldad que ya Ahab predicaba de su archienemiga ballena.

El punto de partida es erróneo en ambos casos, pues cuando proyectamos nuestras narrativas y expectativas en lo no humano, arriesgamos a perder de vista su alteridad (ya sea esta una temible criatura marina o una sofisticada herramienta de IA) y a situar el problema lejos de los confines de la realidad.

## I. DELIMITACIÓN NEGATIVA: CONSIDERACIÓN SUMARIA DE MODELOS ALTERNATIVOS

### 1. LA POSIBILIDAD DE CASTIGO A LA MÁQUINA

Tan pronto como empezaron a surgir las cuestiones relativas a la IA y su potencialidad lesiva, algunos autores señalaron la posibilidad de lo que se ha

[249] HALLEVY, G., *When Robots Kill. Artificial Intelligence under Criminal Law,* Northeastern University Press, Boston, 2013, prefacio. El episodio al que se refiere Hallevy es el que tuvo como víctima a Kenji Urada, trabajador de Kawasaki Heavy Industries. No ha sido este el único de los incidentes en que, por deficientes condiciones de seguridad en el trabajo, robots industriales acaban con la vida de empleados. Un repaso por los primeros casos y los problemas judiciales que plantearon (fuera, desde luego, de la hipótesis sostenida por Hallevy), puede verse en *El Confidencial,* 19/09/2024: <https://www.elconfidencial.com/alma-corazon-vida/2024-09-19/la-primera-persona-murio-robot-a-manos-historia-1qrt_3963922/> [Accedido por última vez el 21/01/2026].

venido a bautizar por su más fiel defensor como «modelo de responsabilidad penal directa» [250], sin que haya sido posible esquivar, desde entonces, la alusión a los problemas relativos a la adscripción de responsabilidad penal al propio sistema de IA.

Es de referencia obligada en este punto la obra de Hallevy, que, a lo largo de los años y en sucesivas publicaciones [251], ha ido dando forma a su propuesta. Lo cierto es que Hallevy presenta hasta tres modelos de responsabilidad. El primero de ellos (el de responsabilidad directa), es el de preferente aplicación según el propio autor, concibiendo los otros solo como respuestas subsidiarias.

De forma subsidiaria, el segundo modelo se aplicará en caso de producirse daños y no poder imputárselos penalmente de forma directa al sistema de IA, proponiendo el denominado *«perpetration by another»* [252], según el cual se ha de considerar que los sistemas son instrumentalizados por un autor que se sirve de él, a modo de autor mediato, de forma que lo único que hay que probar es la intención por parte del humano para poder imputar el daño [253].

---

[250] Una explicación completa de este modelo por el autor en Hallevy, G., «The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control», *Akron Intellectual Property Journal Akron Law Journals March 2010*, pp. 186 y ss.

[251] «The Criminal Liability of Artificial Intelligence Entities – From Science Fiction to Legal Social Control, *Akron Intellectual Property Journal*, 4, 2010; «I, Robot – I, Criminal – When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offences», *Syracuse Sci. & Tech. L. Rep.,* 2010; «Virtual Criminal Responsibility», *Original Law Review,* 6 (1), 2011; «Unmanned Vehicles. Subordination to Criminal Law Under the Modern Concept of Criminal Liability», *Journal of Law, Information and Science,* Vol 21, 2012; *When Robots Kill, op cit*. N. 249*; Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; «Dangerous Robots – Artificial Intelligence vs. Human Intelligence», 2018, disponible en: SSRN: <https://ssrn.com/abstract=3121905 >, [Accedido por última vez el 21/01/2026];; «The Basic Models of Criminal Liability of AI Systems and Outer Circles», 2019, disponible en: <https://ssrn.com/abstract=3402527> [Accedido por última vez el 21/01/2026].

[252] «Unmanned Vehicles. Subordination to Criminal Law Under the Modern Concept of Criminal Liability», Journal of Law, Information and Science, Vol 21, 2012, pp. 201 y ss.

[253] No ha sido el único en tratar la posibilidad de una autoría mediata, aunque su propuesta no ha estado, tampoco, exenta de críticas. Entre otras, Domínguez Peco, E. M., «Los robots en el derecho penal», en Moisés Barrio, A. (Dir)., *Derecho de los robots*, Wolters Kluwer, Madrid, 2018 pp. 131-150. El argumentado de Hallevy simplifica y deforma al extremo la institución de la autoría mediata, que requiere, en todo caso, ciertas precondiciones o capacidades previas en el sujeto instrumentalizado. De lo contrario, no se alcanza a entender la diferencia entre el uso de un sujeto instrumentalizado (al que faltan algunos elementos para imputar el delito, pero en todo caso no la capacidad mínima para cometerlo). A título de ejemplo, es pacífico en la doctrina penalista entender que no cabrá autoría mediata sobre una persona que incurre en una causa de ausencia de acción, por ser esta un mero «objeto» en manos del perpetrador, que no es autor mediato, sino autor directo del hecho. Molina Fernández, F., «Acción y tipicidad en los delitos activos» *Memento Práctico penal*, Francis Lefebvre, 2025, nm. 1258, «*El tercero que interviene provocando o favoreciendo dolosa o imprudentemente el hecho, no lo hace a título de partícipe en un hecho ajeno, sino de autor, ya que no hay un hecho principal al que sumarse. Por la misma razón, la autoría del tercero es directa y no mediata. El cuerpo del forzado o del inconsciente es para el tercero un instrumento físico que le sirve para cometer el delito, igual que puede serlo un arma. Al utilizarlo se ejecuta directamente el hecho. La autoría mediata, a diferencia de la directa, requiere utilizar a una persona como instrumento, y no solo el cuerpo de una persona*». Sin embargo, recientemente, una forma de autoría mediata de las máquinas, afirmando su condición de «actuantes» como superación de la dicotomía

Finalmente, apunta un tercer modelo, solo previsto en caso de no poder aplicar los otros dos, que bautiza como modelo de la consecuencia natural probable *(natural-probable-consequence* [254]). Solo en este último escenario se plantea el autor la posibilidad de que el humano que interactúa con el sistema no haya tenido intención o conocimiento de perpetrar un hecho delictivo, pero sí pueda demostrarse que se ha producido un comportamiento que comporte desprecio o negligencia *(negligence o recklessness),* unido a la materialización de un daño que sea consecuencia probable de este actuar descuidado.

Más allá de estas dos últimas propuestas, subsidiarias y escasamente desarrolladas por el autor israelí, el modelo que supone una verdadera provocación jurídica viene de la mano de su pionero trabajo de 2010 [255]. Según el modelo de responsabilidad directa, todo lo que se debe comprobar ante la sucesión de un evento lesivo de un bien jurídico penal es la existencia de los elementos sobre los que pivota el derecho penal anglosajón, esto es, el *actus reus* y la *mens rea* [256]. En la afirmación del *actus reus*, el autor ofrece una simplificada noción de acción, desprovista de cualquier exigencia referida a la voluntariedad. En tanto en cuanto la IA tiene capacidad para producir modificaciones en el mundo exterior (cuestión incontrovertida), se podrá afirmar su capacidad de acción. En lo relati-

---

sujeto/objeto, basado en el concepto de Latour, ha sido defendida por Navarro Dolmestch, R., «Inteligencia artificial como «actuante» en el derecho penal. Una primera aproximación», *Revista de Internet, Derecho y Política*, (43), 2025, pp. 1-16.

[254] El propio Hallevy explica que este modelo, originalmente, se utilizaba para hacer responsables a los cómplices cuando uno de los participantes cometía un delito que no había sido planeado por el resto, ni formaba parte de una conspiración previa. Hallevy, G., «The Criminal Liability of Artificial Intelligence Entities…», *op. cit*. N.250, p. 183: «*Originally, the natural-probable-consequence liability model was used to impose criminal liability upon accomplices, when one committed an offense, which had not been planned by all of them and which was not part of a conspiracy. The established rule prescribed by courts and commentators is that accomplice liability extends to acts of a perpetrator that were a «natural and probable consequence» of a criminal scheme that the accomplice encouraged or aided. Natural probable-consequence liability has been widely accepted in accomplice liability statutes and codifications*». Para más detalle: Bird, K. R, «Natural and Probable Consequences Doctrine: Your Acts Are My Acts», *W. St. UL Rev.*, *43*, 2006, pp. 43-66.

[255] Hallevy, G., «The Criminal Liability of Artificial Intelligence Entities… *op cit*. N. 251, págs. 171 y ss.

[256] Freitas *et al.,* los caracterizan asimilando el *actus reus* con la agencia o acto (*Handlung* para la tradición germana), mientras que la *mens rea* se identifica con la *blameworthiness* o culpabilidad (la germana *Schuld*). *Vid*.: Freitas, P. M./ Andrade, F./ Novais, P., «Criminal liability of autonomous agents: From the unthinkable to the plausible» en Casanovas, P. / Pagallo, U., /Palmirani, M./ Sartor, G. (Eds.), *AI approaches to the complexity of legal systems*. AICOL 2013. Lecture notes in computer science (Vol. 8929). Berlin, Springer, 2014, p. 147. Aclaran en la misma página los autores que la equivalencia no es, en todo caso, total, pues «debe tenerse en cuenta que, dependiendo de la perspectiva que se asuma sobre el concepto de delito, más concretamente, si se trata de un enfoque clásico (de autores como Liszt, Beling y Berner), un enfoque neoclásico (Mezger), una teoría teleológica (Welzel) o un sistema funcional-teleológico y racional (Schünemann y Roxin), la traducción de *mens rea* al alemán puede ser *subjektiver Tatbestand* y/o *Schuld*» (traducción propia).

vo al elemento mental o subjetivo *(mens rea),* Hallevy ofrece una interpretación extravagante sobre los requisitos de conocimiento y voluntad, reduciendo el elemento cognitivo a la percepción de la realidad en forma sensorial, y justificando el elemento volitivo en la capacidad de evaluar las probabilidades y opciones de conducta tras haber procesado la información del entorno. El autor basa la capacidad delictiva en la evaluación precisa de la información.

Pocos son los autores que se han sumado, sin matices, a esta tendencia de la antropomorfización absoluta. Desde luego, pocos son los que lo hacen desde una perspectiva penal [257], y los que tímidamente, lo hacen desde otros campos del derecho (el civil por antonomasia [258], aunque también desde una reflexión más genérica [259]), acaban por reconducir la cuestión a otras estructuras ya existentes [260]. Lejos parece quedar la primigenia propuesta de un Parlamento Europeo

[257] Así, destacadamente, Hu, centrando la cuestión en la capacidad de tomar decisiones relevantes moralmente, o Mulligan, que pone el énfasis en la necesidad de que el castigo satisfaga psicológicamente a las víctimas. Referencias completas obras. *Vid.*: Hu, Ying., *op. cit.* N. 132, pp. 487-532, y Mulligan, C., *op. cit*. N. 132.

[258] Pionero a estos efectos el texto de 1992 de Solum. Solum, L. B.: «Legal Personhood for Artificial Intelligences», 70 North Carolina Law Review, 1992, pp. 1231-1287.

[259] Beck sugiere que, quizá, la mejor solución fuera crear una especie de personalidad electrónica, algo así como una entidad que agrupe a todas las personas que conforman la tecnología, pero con entidad propia, sin necesidad de asimilarlo a los humanos moralmente, señalando que este concepto legal no estaría pensado para afirmar que los robots tienen las características propias de un humano. La idea, en esencia, es que al construir intencionadamente máquinas que toman decisiones por nosotros, y al dotarlas de personalidad jurídica les damos parte de nuestra identidad (social), reconstruyendo nuestra identidad de forma que incluya a las máquinas porque de antemano hemos decidido utilizarlas para una parte específica de nuestro espacio de autonomía. *Vid.*: Beck, S., «Intelligent agents and criminal law-Negligence, diffusion of liability and electronic personhood», *Robotics and Autonomous Systems,* Volume 86, 2016, pp. 141-142.

[260] Teubner es el ejemplo más claro de esto último. Teubner, G., *op. cit.* N. 132. Asimismo, en una monografía posterior, Beckers, A., /Teubner, G., *op cit.* N. 111. Los autores proponen en la monografía un enfoque diferenciado para abordar las brechas de responsabilidad generadas por los sistemas de inteligencia artificial, identificando tres regímenes de responsabilidad, cada uno adaptado a contextos sociodigitales específicos: el régimen de responsabilidad vicaria para agentes autónomos (*actants*), previsto para los escenarios en los que los algoritmos actúan como «asistentes digitales» autónomos –sustituyendo decisiones humanas–, se aplica la responsabilidad vicaria, reconociendo personalidad jurídica parcial a los algoritmos, similar a la de agentes humanos, atribuyendo responsabilidad al principal humano por acciones no autorizadas del agente algorítmico, basándose en principios del derecho de agencia. El segundo modelo es el de responsabilidad empresarial para híbridos humano-IA, en contextos de interacciones donde humanos y algoritmos operan de forma inseparable (hibridación digital), la responsabilidad recae en la asociación colectiva. Se trata de sistemas con propiedades emergentes donde no es posible delimitar acciones individuales. La responsabilidad se distribuye entre los actores clave detrás de la asociación, aplicando principios de responsabilidad empresarial. El último escenario es el de los fondos colectivos para sistemas interconectados (multitudes), pensado para redes complejas de algoritmos interdependientes (interconectividad digital). En este caso se propone crear agrupaciones de riesgo financiados por los participantes (empresas, desarrolladores, etc.) para cubrir daños. Estos fondos permiten compensar víctimas sin identificar responsabilidades individuales, dada la opacidad de las interacciones algorítmicas. Los regímenes se basan en un análisis interdisciplinario que combina estudios tecnológicos, teoría social (principalmente de influencia *luhmaniana*) y derecho comparado. Se realiza una crítica a las soluciones uniformes, abogando por adaptar el estatus jurídico de la IA a su concreto contexto institucional. En conclusión, el libro aboga por una regulación institucionalmente receptiva, evitando tanto la personificación total de

que, allá por 2017, (periodo que en tiempos tecnológicos supone una eternidad), lanzaba la polémica propuesta que fue rechazada de plano por la doctrina de la época [261]. No es de extrañar que una petición de semejantes características desencadenara toda una serie de reflexiones y críticas (la mayoría de las cuales, como se puede acertar a pensar, de corte negativo) [262]. Es difícil conocer si ese largo plazo del que hablaba el Parlamento Europeo en 2017 se puede considerar ya superado, pero lo que parece evidente es que, pese a las discusiones teóricas que han derivado de semejante propuesta, ni siquiera en el derecho civil ha dado lugar a una modificación de los sujetos con personalidad jurídica.

No parece adecuado ni oportuno que una investigación sobre derecho penal parta de una premisa que, a nuestro parecer, está alejada de la realidad tanto teórica como práctica. Ni los hechos lesivos que protagonizan los sistemas más actuales ni los fundamentos teóricos de la ciencia penal hacen pensar que adscribir capacidad jurídica (mucho menos capacidad jurídico-penal) sea adecuado.

Consideramos que basta un brevísimo repaso por los argumentos que se han esgrimido por la doctrina penalista más autorizada que se ha dedicado a estudiar en profundidad estas propuestas. A ellas nos adscribimos para dejar de lado un escenario que, sin negar que en una perspectiva de largo plazo y desde el punto de vista filosófico pueda albergar algún interés, no nos sirve para resolver el actual problema de los *«hard AI crimes»* objeto del presente estudio. Se opta por utilizar el término anglosajón *«hard AI crimes»* por encontrar en él un con-

---

la IA como su tratamiento como mera herramienta, y enfocándose en mitigar riesgos sociales específicos mediante regímenes diferenciados. A pesar de lo profundo del enfoque, y de la correcta identificación de algunos riesgos y su caracterización, no se alcanza a entender demasiado bien cuál es la diferencia entre estos regímenes, que afirman la necesidad de personificación en algún extremo, y un modelo de responsabilidad objetiva de corte más clásico, *(strict liability),* pues no se trata más que de atribuir decisiones a actores colectivos. Pese al esfuerzo clasificador y la puesta de manifiesto de la complejidad que presenta la autonomía e interoperatividad en contextos complejos, no creemos que este régimen presente novedades muy reseñables, sino que aplica, en esencia, conceptos propios de la delegación empresarial (resulta similar, por ello, al modelo propuesto por Diamantis, *infra).*

[261] Nos referimos a la Resolución del Parlamento Europeo, de 16 de febrero de 2017, comentada *supra* (capítulo segundo).

[262] Para un repaso completo por la evolución histórica de los argumentos en contra y a favor de reconocer personalidad jurídica a la IA: Turner, J., *op. cit.* N 5. En el entorno español, sin ánimo exhaustivo: Mozo Seoane, A. «La revolución tecnológica y sus retos: medios de control, fallos de los sistemas y ciberdelincuencia», en Rogel Vide C. (coord..), *Los robots y el Derecho,* Reus, Madrid, 2020, pp. 79-98. Asimismo, Lacruz Mantecón, M., «Potencialidades de los robots y capacidades de las personas» en Rogel Vide, C. (Coord..), *Los robots y el Derecho,* Reus, Madrid, 2020, pp. 25-77. También, en una clara oposición por el peligro contra los postulados del humanismo: Nevejans, N., «Study for Juri Committee, European Civil Law RulesIn Robotics, Policy Department, Citizens' Rights and Constitutional Affairs», European Parliament, Brussels», 2016, pp.14 y ss., accesible en:<https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf>. [Accedido por última vez el 21/01/2026]. Asimismo, Nevejans, N., *Open letter to the European Commission: Artificial Intelligence and Robotics. Robotics Openletter,* 2018, accessible en: <http://www.robotics-openletter.eu/> [Accedido por última vez el 21/01/2026].

cepto útil (y común entre la doctrina especializada), que es capaz de afinar y centrar el tiro en la problemática que nos interesa. Utilizaremos, de aquí en adelante, este concepto, otorgándole el significado que sus propios creadores, Abbott y Sarch [263], le dan. Así, serán *«hard AI crimes»* aquellos delitos cometidos funcionalmente por sistemas de IA en los que no resulte fácil identificar a una persona que haya actuado con plena culpabilidad penal o no se pueda reconducir el comportamiento del sistema a una actuación individualizada.

Debemos inferir de esta definición que quedan excluidos, por tanto, de este grupo conceptual, los supuestos en los que la implicación del sistema de IA en el delito no implique especiales dificultades en la imputación de responsabilidad, o bien por haber utilizado el usuario humano una herramienta sin ninguna capacidad de autonomía operativa (elemento de imprevisibilidad respecto de la conducta), o bien porque el operador humano ha actuado con dolo completo respecto de la ejecución del resultado final (operando, por tanto, el sistema, como mero objeto o herramienta en manos de un autor directo), siendo un actor perfectamente individualizable [264].

Pasemos, resumidamente, a presentar los valiosos argumentos que hasta ahora se han ofrecido, a nuestro juicio acertadamente, y sin ánimo de exhaustividad, pues nuestra única intención en este punto es adherirnos a ellos (con los matices que abajo se expondrán), y justificar la elusión de una discusión que no consideramos prioritaria.

Para Del Rosal, que ventila la cuestión en apenas unas líneas, la atribución del castigo a la máquina es, al menos hoy por hoy, inviable, pues además de contravenir los requerimientos más asentados de la teoría continental del delito (tanto en lo relativo a la capacidad de acción [265], como de culpabilidad y

---

[263] Utilizamos el término «*hard AI crimes*» propuesto en: Abbott, R., /Sarch, A. F., «Punishing artificial intelligence: legal fiction or science fiction», *op. cit. N. 85*, p. 328. «*scenarios where crimes are functionally committed by machines and there is no identifiable person who has acted with criminal culpability. We call these Hard AI Crimes. This can occur when no person has acted with criminal culpability, or when it is not practicably defensible to reduce an Al's behavior to bad actors*». Cuando se haga referencia a este término en lo que sigue, se utilizará únicamente en el contexto de esta definición.

[264] Para Aires de Sousa la característica esencial de estos «*Hard AI Crimes»,* es que «el resultado dañoso aparece como opaco a un juicio *ex ante* de previsibilidad humana. Esta opacidad está asociada necesariamente a técnicas de inteligencia artificial que actúan con una cierta capacidad de autonomía, sin que sea posible explicar el proceso que lleva a la máquina a alcanzar un determinado output». (traducción propia). Sigue la autora explicando que en estos contextos es posible afirmar que la máquina actúa, en cierto sentido, como resultado de la contribución de actores diversos, pero que el resultado al que llega es de alguna forma propio (*seu*). *Vid*.: Aires de Sousa, S., «Não fui eu...», *op. cit*. N. 236, p, 62.

[265] En este sentido, y negando cualquier posibilidad de afirmar la capacidad de acción a las IA con rotundidad en aplicación de un concepto de acción personal, Roxin y Greco: «*Es evidente que las máquinas en sí mismas no pueden cometer delitos. Basándonos en el concepto de acción personal, la situación es clara: la máquina no tiene personalidad que pueda expresarse a través de sus movimientos; por lo tanto, no actúa*» (traducción propia). Roxin, C./ Greco, L., *Strafrecht. Allgemeiner Teil. Band I: Grundlagen. Der Aufbau der*

pena), no puede fijarse paralelismo válido respecto de la personalidad jurídico-penal de las personas jurídicas [266].

Mucho más profunda es la reflexión que realiza Giannini [267] al respecto, pues la cuestión en torno a la que pivota su investigación doctoral tiene que ver, precisamente, con negar la posibilidad de esta agencia criminal de los sistemas de IA. La autora italiana (que realiza un estudio comparativo que combina las propuestas anglosajonas con las de ámbito germánico e italiano), analiza en profundidad lo que considera el prerrequisito básico para la atribución de responsabilidad penal (al que denomina *«criminal capacity»),* derivando de él dos subrequisitos: las capacidades cognitivas o epistémicas y las capacidades de control. Tras plantear sus dudas acerca de si lo relevante sobre el conocimiento de la ilicitud (elemento epistémico) debe ser la vertiente de ilicitud legal o moral, pasa a analizar las capacidades de control, resaltando la idea de que no es tan importante poner el énfasis en la dimensión interna (relativa a cómo se configura el proceso de decisión), como en la externa (es decir, cómo se percibe desde fuera, tanto por el sistema como por la sociedad, esa capacidad de actuación moral) [268]. En todo caso, Giannini plantea la posibili-

*Verbrechenslehre. C. H. Beck, Munich, 5.ª ed.,* 2020, § 8, núm. 66g. También respecto de la capacidad de culpabilidad, excluyendo que el concepto material de culpa se aplicable a máquinas autónomas. *Ídem.*

[266] Del Rosal Blasco, B., *op cit*. N. 35, p. 20, utiliza un argumento a nuestro juicio débil para señalar las diferencias. Se basa en las posibles lagunas de impunidad que pueden darse en la persona jurídica por la dificultad de individualización del comportamiento en el seno de las corporaciones, resaltando que este no es el riesgo en el caso de los sistemas de IA, que califica como autómatas que generan riesgos que «nada tienen que ver» con los generados por las personas jurídicas. No podemos compartir esta afirmación. El tópico comparativo no ha sido, desde luego, infrecuente, y son numerosos los autores que han dedicado su tiempo a desmentir o afirmar las bondades de una comparación tan manida. Así, sin ánimo de exhaustividad: Blanco Cordero, I., *op. cit*. N. 85, pp. 371-391, aludiendo a la ya célebre expresión «*societas delinquere non potest*».

[267] Giannini, A., *op. cit*. N. 143.

[268] Resulta interesante la perspectiva de la autora italiana que consigue, así, rehuyendo del examen de la perspectiva interna, dejar de lado los eternos problemas sobre el libre albedrío, no ajenos a la discusión en este campo, pero acaba por aproximarse, a nuestro juicio, a la idea desarrollada por Simmler y Markwalder, a las que Giannini califica de moderadas (pp. 76 y ss.). La idea del reconocimiento externo (ya sea por parte del sistema legal o de la sociedad), aparece como punto clave en la propuesta de las alemanas Simmler y Markwalder que, desde una perspectiva *jakobsiana* del derecho penal, y aferrándose a la idea del reconocimiento de la personalidad por otros agentes sociales, acaban por dejar la puerta abierta (si bien no en el momento presente), al reconocimiento de personalidad jurídico penal para los agentes. *Vid.*: Simmler, M./ Markwalder, N., «Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artifcial Intelligence», *Criminal Law Forum*, 30, 2019, *passim*. Para una visión general y actual sobre libre albedrío y derecho penal, y sin ánimo de exhaustividad: Feijoo Sánchez, B., «El Derecho penal de la culpabilidad ante el neurodeterminismo», en AA. VV., *Un modelo integral de Derecho penal: libro homenaje a la profesora Mirentxu Corcoy Bidasolo*, BOE, Madrid, 2022, pp. 601-613, 2022. El mismo: «Neurociencias y Derecho Penal de un sistema de libertades», en Demetrio Crespo, E., (Dir.), *Derecho penal y comportamiento humano: avances desde la neurociencia y la inteligencia artificial*, Valencia, Tirant Lo Blanch, 2022, pp. 127-146. Pérez Manzano, M., «El tiempo de la consciencia y la libertad de decisión: Bases para una reflexión sobre neurociencia y responsabilidad penal», *Doxa, Cuadernos de Filosofía del Derecho*, 2012, 35, 2012, pp. 471-498. Molina Fernández, F., «Pre-

dad de que estos argumentos pudieran parecer insuficientes, y que, incluso aunque cupiera afirmar esta *«criminal capacity»* por una interpretación conforme de las capacidades de conocer el entorno, las normas (conocimiento sobre la ilicitud) y sus acciones fueran reconocidas como morales, aún seguiría siendo necesaria la alusión a las finalidades del castigo [269].

Por su parte, Aires de Sousa [270] califica las propuestas teóricas de Hallevy como «asépticas» (*assética*), en el sentido de que se encuentran vacías del contenido valorativo propio de las categorías del derecho penal, viendo en ellas una concepción puramente formalista «en la que el ilícito se confunde con el daño causado», confundiendo «responsabilidad con causa o fuente de daño» [271]. Parte la autora portuguesa de la idea de Pagallo [272] sobre la importancia del entendimiento moral, afirmando que solo puede haber responsabilidad allí donde exista un agente consciente, moral y con capacidad de entender el sentido de la norma como destinatario jurídico. Centra su argumento, de forma muy acertada, en la función de reprochabilidad o censurabilidad asociada a la culpa (entendida esta desde un concepto amplio, aludiendo a los requisitos del principio de culpabilidad en su sentido más esencial). Dicha ausencia de reprochabilidad (de culpabilidad en sentido amplio), minaría la función del derecho penal, tanto en lo relativo a la propia realización del hecho (entendemos que se

supuestos de la responsabilidad …», *op. cit*. N. 84, pp. 169-284. Demetrio Crespo, E., «Libertad de voluntad, investigación sobre el cerebro y responsabilidad penal», *InDret*, 2/2011. Aunque con matices diferenciadores, todos los autores coinciden en que los avances en neurociencia plantean desafíos al concepto tradicional del libre albedrío y de responsabilidad penal, reconociendo la necesidad de adaptar el derecho penal a estos descubrimientos. El punto común a todos ellos reside en subrayar las limitaciones metodológicas de la neurociencia para resolver cuestiones filosóficas fundamentales, por lo que proponen enfoques normativos o pragmáticos para integrar estos avances sin abandonar la noción de imputabilidad.

[269] En este punto la autora italiana se muestra irremediablemente retributivista, reconociendo que, aunque no sea esta la única y primordial finalidad del castigo, será innegable la existencia de un «*gap*» o vacío si ningún tipo de retribución podemos encontrar a la hora de castigar. En este sentido alude la autora a Danaher y su *«retribution gap»,* (Danaher, J., *op. cit.* N.43, pp. 299 y ss.) A pesar de las muchas matizaciones que puedan realizarse desde el punto de vista de la finalidad del castigo, podemos mostrarnos de acuerdo con la autora y con Danaher en un sentido amplio. Así, el autor irlandés habla de los deseos de pena que la sociedad manifiesta basándose en estudios sociológicos, sin atribuir al concepto «retribución» un significado estrictamente penal. De esta forma, con independencia de la importancia que tenga la retribución en cuando fin que justifique el castigo penal, la idea que parece derivarse puede reconducirse a un argumento más simple: la discordancia entre lo que la sociedad demanda y lo que los estándares de imputación de responsabilidad normalmente utilizados requieren. Esta idea básica es una que aquí se considera aceptable sin mayores dificultades, siempre que se enmarque debidamente. Para una caracterización de la retribución como «marcador somático» o mecanismo evolutivo, pero vacío de contenido en su papel como justificación penal; *vid.*: Rodríguez Horcajo, D. *Comportamiento humano y pena estatal: disuasión, cooperación y equidad*, Marcial Pons, 2017, p. 309.

[270] Aires de Sousa, S., *op cit*. N. 264, p. 77.

[271] *Idem*, p. 70.

[272] Pagallo, U., *The laws of robots, op. cit*. N. 143.

refiere a la capacidad de actuación criminal, en el sentido defendido por Giannini), como en lo relativo al fin de la pena.

Ofrece finalmente Aires de Sousa un argumento adicional relacionado con la idea humanista puesta de manifiesto por Nevejans [273] y defendida por autores como Navarro Dolmetsch [274], más orientada a ideales vinculados a la singularidad y dignidad humana, destacando la «naturaleza humanista y esencia ética de un derecho público punitivo como el derecho penal» [275].

Finalmente, y como exponente anglosajón de argumentos contrarios a la expansión, podemos destacar los trabajos de Dsouza [276]. El autor niega la agencia cualificada *(qualifying agency),* que toma como presupuesto esencial y que podría ser el equivalente a la «capacidad criminal» de Giannini, explicando que los sistemas de IA no tienen la capacidad de reflexionar sobre el valor moral de sus acciones ni sobre las normas legales que pueden infringir. Observamos una remisión, una vez más, a argumentos relativos a esta capacidad mínima de entendimiento sobre la ilicitud y la moralidad.

Señalando los elementos básicos de la teoría del delito, destaca el autor que todo acto criminal requiere, para ser delictivo, una conexión entre el acto y el elemento subjetivo *(actus reus y mens rea),* conexión que no puede afirmarse en estos casos. Profundiza Dsouza en la discusión en una publicación más reciente [277], e insiste en la importancia de la agencia moral: no debe confundirse la realización de actos con aparente intención, con la tenencia de la cualidad moral necesaria para ser sujeto del derecho penal. Añade, también, argumentos relativos a la finalidad del castigo penal, que implica necesariamente un castigo en sentido

---

[273] Nevejans, N., «Study for Juri Committee …», *op cit*. N. 262, p. 16: «*creating a new type of person– an electronic person– sends a strong signal which could not only reignite the fear of artificial beings but also call into question Europe's humanist foundations. Assigning person status to a nonliving, non-conscious entity would therefore be an error since, in the end, humankind would likely be demoted to the rank of a machine. Robots should serve humanity and should have no other role, except in the realms of science-fiction*.».

[274] Navarro Dolmestch, R., «¿Qué haremos cuando la inteligencia artificial mate?», blog UCM (Universidad Católica de Maule), 3 de enero de 2024, disponible en: <https://portal.ucm.cl/noticias/opinion-haremos-cuando-la-inteligencia-artificial-mate> [Accedido por última vez el 21/01/2026]. *«El segundo aspecto del problema pasa por recuperar el concepto de dignidad humana. Aun con sus escasos rendimientos, un derecho penal democrático aspira a construirse en torno a la idea de dignidad humana, ya que es una parte del ordenamiento jurídico que debe diseñarse y aplicarse considerando que sus destinatarios son personas».*

[275] Aclarando así la autora que el problema de la responsabilización penal de las máquinas trae causa de una discusión de fondo mucho más profunda, que requeriría un análisis mucho mayor, «*por contrariar a natureza humanista e essência ética de um direito público punitivo como é o direito penal».* Aires de Sousa, S., *op. cit*., N 264, p. 79.

[276] Dsouza, M. *op. cit*. N. 165, p. 248.

[277] En tono algo más pedagógico, Dsouza, M., «Why Autonomous AI cannot be a Criminal Law agent», 29 de noviembre de 2024, disponible en: <https://criminaljusticetheoryblog.wordpress.com/2024/11/29/why-autonomous-ai-cannot-be-a-criminal-law-agent/> [Accedido por última vez el 21/01/2026].

moral y comunicativo [278]. Finaliza su argumento apelando al carácter intrínsecamente moral del derecho penal; sin moralidad (exclusivamente humana), no hay derecho penal, sino, a lo sumo, algo que se le parece [279].

Los autores de la tradición germánica han sido, también, contundentes. Así, Gless, Silverman y Weigend [280], negando con rotundidad la noción de acto y la de culpabilidad.

En definitiva, y a la vista de las posiciones mayoritarias aquí expuestas, parece que los argumentos en contra de la responsabilidad directa de los sistemas de IA pueden pivotar en torno a tres ideas principales, todas ellas complementarias y no excluyentes, y que, a nuestro modo de ver, responden a fundamentos distintos.

La primera de ellas, de corte más ético o filosófico, apela a la idea de dignidad y singularidad humana. La segunda y la tercera, respectivamente, responden a argumentos de tipo dogmático penal, y por ello debieran suponer, a nuestro parecer, el argumento más definitivo. Tanto las ideas relativas al fundamento mismo de la capacidad de culpabilidad, nombrado de manera variada por los autores (así, «*criminal capacity*», «*qualify agency*», capacidades epistémicas y de control, agencia moral o reprochabilidad ética), pueden reconducirse sin mayores problemas a la idea mínima o esencial de la culpabilidad, algo así como el mínimo común denominador de la capacidad de imputabilidad, que se conforma necesariamente (no solo en los ordenamientos continentales, sino también en los

[278] Comprobamos, por tanto, y en la línea de lo afirmado por Dsouza, que la posibilidad de imponer un castigo a los sistemas no contraviene solo las teorías de la pena de corte más retributivo, sino que las objeciones son aplicables también desde concepciones comunicativas (o, si se prefiere, en un sentido más amplio, preventivo-generales) del castigo. Las teorías comunicativas de la pena son, de hecho, familiares al entorno jurídico germano, pues beben fundamentalmente de la teoría de la pena de Jakobs. *Vid.*: Jakobs, G., *La pena estatal: significado y finalidad* (traducción y estudio preliminar de M. Cancio Meliá y B. Feijóo Sánchez), Thomson-Civitas, Madrid, 2006.

[279] Responde también Dsouza al argumento de la analogía con la personalidad de la persona jurídica señalando que: «hay un sentido claro en el que las corporaciones se distinguen de las entidades de IA. Si traspasáramos el velo corporativo, encontraríamos (eventualmente) a un ser humano manejando los hilos. Somos reacios a traspasar el velo corporativo por razones que tienen que ver con el valor económico de mantener la separación de la persona corporativa, pero el hecho de que en algún lugar detrás del velo haya un controlador humano nos da motivos para querer responsabilizar penalmente a la empresa.» (traducción propia). La afirmación de Dsouza resulta, en este punto, algo cuestionable, pues si es obvio que aceptamos esa ficción de personalidad por razones de tipo económico, ¿por qué no hacerlo por razones de otro tipo? Puede parecer algo contradictorio afirmar la posibilidad de agencia de las personas jurídicas por razones de tipo consecuencialista, pero no hacerlo en el caso de los sistemas de IA. Sin que tenga apenas importancia en lo que atañe a esta investigación, debemos dejar explicitado, en pos de la coherencia, que no nos parece asumible mantener al mismo tiempo la imposibilidad de afirmar la responsabilidad penal de las personas jurídicas y de los sistemas de IA, al menos si aceptamos, como Dsouza hace, como válido, un argumento puramente consecuencialista.

[280] Gless, S., Silverman, E., /Weigend, T., «If robots cause harm, who is to blame? Self-driving cars and criminal liability», *New Criminal Law Review*, 19(3), 2016, p. 418 y ss.

de corte anglosajón), como capacidades mínimas de obrar con conocimiento jurídico, ético y capacidad de actuación conforme a ese conocimiento [281].

Somos conscientes de que sobre esta capacidad de culpabilidad y la interpretación de sus elementos han corrido ríos de tinta, pero reiteramos que no es objeto de nuestra investigación discutir estas premisas, sino presentar como válidas estas ideas y tratar de avanzar en lo que consideramos un escenario más realista.

Finalmente, la tercera de las ideas, también de corte dogmático, tiene que ver con la finalidad propia del castigo penal. [282] Acogemos, asimismo, las ideas expuestas, todas ellas coincidentes en que la aflicción que todo castigo penal persigue no puede encontrar sentido cuando de lo que se trata es de imponer un castigo a entidades que no son capaces, al menos por el momento, de sufrir las consecuencias del mismo. La idea, considero, es sencilla: si lo que queremos conseguir imponiendo el castigo al sistema de IA es causar en otros una determinada reacción (sea esta la de satisfacción de la sensación de venganza, o bien de prevención en el resto, o bien de sufrimiento o ejemplo), ¿por qué castigar a ese agente intermedio? Castigar al instrumento para generar algún tipo de reacción en las personas supone dar un rodeo innecesario (además de,

---

[281] Baste tomar como referencia nuestro artículo 20. 1.º CP, que, al recoger las causas de exención de la imputabilidad se refiere, como requisito mínimo, a la capacidad de «*comprender la ilicitud del hecho o actuar conforme a esa comprensión*».

[282] A propósito de la finalidad del castigo encuentra Giannini la razón más poderosa, a nuestro juicio, para poder afirmar la responsabilidad penal de las personas jurídicas, pero negarla en el caso de los sistemas de IA. La idea central es que, en el caso de las corporaciones, las penas a ellas dirigidas (de carácter fundamentalmente económico y prohibitivo de desarrollar sus actividades económicas) van dirigidas directamente a atacar su fin principal y común a todas ellas. Sin embargo, en el caso de los sistemas de IA, dicho fin genérico no parece tan claro, por lo que más allá de tener una entidad corpórea, lo esencial es poseer unos intereses propios y no ajenos en los que poder basar el elemento aflictivo que fundamenta cualquier castigo. *Vid.:* Giannini, A., *op. cit*. N. 143, p. 95: «*corporations are created to make money, hence monetary sanctions would be effective because they would target a corporation's essential purpose. The same cannot be said about robots, as their purposes are not as straightforward.»* Las referencias a la innecesariedad del daño corpóreo se basan en la ya clásica obra de Asaro, P., *op cit.* N. 121, p. 181, que a su vez bebe de la previa reflexión al hilo de las personas jurídicas planteada por Coffee. *Vid.*: Coffee, J., C., Jr., «No Soul to Damn: No Body to Kick»: An Unscandalised Inquiry into the Problem of Corporate Punishment», *79 MICH. L. REV*, 386, 198, pp. 386-4591. En un sentido similar, Gless, advirtiendo que se ha superado la objeción fundamental de que una empresa en sí misma no puede experimentar un castigo al considerar que resulta intuitivamente evidente que una multa económica «duele» (*schmerzt*) a una empresa orientada a la maximización de beneficios. Sin embargo, los procesos penales económicos (como se ha observado en empresas europeas en los EE. UU.) han demostrado que el umbral del dolor debe establecerse muy alto para que realmente tenga un impacto y sea tomado en serio. Gless, S., «Von der Verantwortung einer E-Person. Würdigung eines unveröffentlichten Regelungsentwurfs von Professor Dr. Dr. h.c. Kurt Seelmann anlässlich seines 70. Geburtstags», *Goltdammer's Archiv für Strafrecht* (*GA),* 6, 164, 2017, p. 328.

por todo lo anteriormente afirmado, contraintuitivo según todos los principios dogmáticos) [283].

## 2. LA RESPONSABILIDAD DENTRO DE LA PERSONA JURÍDICA

Más allá de la provocativa idea de acoger la personalidad jurídica de los sistemas de IA a modo de personalidad jurídica electrónica [284], cuestión que aquí, en aplicación de los motivos arriba señalados, se descarta, muchos son los autores que se han propuesto ahondar en la estructura corporativa para tratar de desentrañar los problemas de imputación de los *«hard AI crimes»,* es decir, el problema de transferir la responsabilidad por un acto desarrollado por una IA a una persona (ya sea física o jurídica).

El elemento común a ambas estructuras (la corporativa y el sistema de IA) lo constituye el ya célebre *«many hands problem»* [285]. Este problema tiene que

---

[283] Una idea similar expresa Molina en su rotunda exposición contra la posibilidad de afirmar una verdadera responsabilidad penal de las personas jurídicas, usando estos argumentos. Por la claridad de su exposición, y lo coincidente, en esencia, con las ideas de Giannini, merece la pena su transcripción literal: *«Dolo natural e imprudencia, y su reverso, el error invencible de tipo (art. 14.1y 2 CP); conocimiento de la prohibición/dolo malo, y su reverso, el error de prohibición (art. 14.3 CP); imputabilidad y las causas que la excluyen –minoría de edad (art. 3 LORPM), anomalía o alteración psíquica (art. 20.1 CP), intoxicación plena (art. 20.2 CP), y alteración grave de la conciencia de la realidad debido a alteración en la percepción desde el nacimiento o la infancia (art. 20.3 CP)–; y capacidad de comprender el sentido de la pena, y su reverso, la exclusión de la punibilidad por trastorno mental sobrevenido tras el delito (art. 60 CP), son todas circunstancias que hacen referencia a estados subjetivo-mentales que desarrollan el principio de responsabilidad que comenzó a fijarse en la acción voluntaria, y que expresa una misma, simple y poderosa idea: el delito no es sólo un hecho objetivamente lesivo para intereses ajenos, sino que también es un hecho que sucede en una mente consciente y autónoma que sabe y decide lo que hace; y la pena no es simplemente algo malo que le sucede o se le impone a alguien, sino un mal que se impone a quien puede sufrirlo y comprender su sentido.»,* Molina Fernández, F., «Societas peccare non potest...: nec delinquere», en Bacigalupo Sagesse, S., Feijoo Sánchez, B., Echano Asaldua, J. I., (Coords.), *Estudios de Derecho Penal: homenaje al profesor Miguel Bajo,* Ed. Ramón Areces, Madrid, 2017, p. 377. Sobre la relevancia del artículo 60 CP, en general poco explorado, para apoyar estos argumentos, p. 390.

[284] En ocasiones los autores se refieren a la terminología «*e-personhood*». Así: Manes, V., «L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia», en Ruffolo, U. (ed.), *Intelligenza Artificiale – Il diritto, i diritti, l'etica*, Giuffré, Milano, 2020, p. 547-567.

[285] Este problema, que es ya referencia obligada en la literatura sobre IA no es, desde luego, exclusivo a este campo. Otros autores lo han tratado en contextos tales como el daño medioambiental (así, Van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., Royakkers, L., «The problem of many hands: Climate change as an example», *Science and Engineering Ethics,* 18(1),2012, pp. 49-67), así como desde el ámbito empresarial o de la filosofía moral (Thompson, D. F., «Designing responsibility: The problem of many hands in complex organizations», en Van den Hoven, J./ Miller, S./ Pogge, T. (Eds.), *Designing in Ethics,* Cambridge University Press, 2017, pp. 32-56). Más adelante, y ya en el contexto de la atribución de responsabilidad por actos de sistemas de IA; Coeckelbergh, M., *op. cit.* N. 116, p. 2056: «uno de los problemas de la acción tecnológica es que suele haber muchas personas implicadas causalmente en la acción, lo que hace difícil, si no imposible, (a) encontrar al individuo responsable, si sólo hay un individuo responsable, o (b) responsabilizar a cualquier individuo: puede haber más personas responsables al haber tantos agentes implicados. Este problema de atribución y distribución de responsabilidades

ver con la existencia de una gran cantidad de agentes implicados, de forma simultánea, en la concepción, producción, comercialización y uso de sistemas de IA, así como a la compartimentación de los papeles, roles y responsabilidad de los agentes humanos individuales a lo largo de este proceso. Los sistemas de IA son el producto final de contribuciones individuales de una multiplicidad de individuos, organizaciones, componentes de máquinas, algoritmos de software y usuarios humanos, a menudo en entornos altamente complejos y dinámicos [286].

Hemos de distinguir aquí, sin embargo, al menos tres niveles de problemas. Al sostener la relación entre la IA y la persona jurídica cabe realizar tres afirmaciones:

(i) Que la IA, en tanto estructura compleja, es asimilable a una persona jurídica, por lo que debe reconocérsele personalidad jurídica. Esta cuestión ha sido descartada en el apartado anterior.

(ii) Que al producirse un hecho por una IA este hecho sea atribuible a un ente con personalidad jurídica (ya sea este una persona física o una persona jurídica). Es aquí donde entran los debates que a continuación se tratarán de si el hecho de la IA es directamente atribuible a la persona jurídica (modelo de autorresponsabilidad) o si es atribuible, como aquí se defenderá, a una persona física.

(iii) Que, una vez atribuido un hecho a una persona (con capacidad criminal), podamos individualizar la responsabilidad dentro de la entidad, tratando de discernir quién, de entre todos los actores implicados en el seno de una empresa, ha de responder por cada hecho. Es decir, los problemas propios de la individualización de responsabilidad (personal) dentro de una empresa, problemas, al fin y al cabo, de autoría y participación.

No puede negarse que la individualización de la responsabilidad dentro del seno de la empresa (problema iii) será, debido a la interacción del *many hands problem* y la complejidad y opacidad de los sistemas de IA, compleja. Sin embargo, esta reflexión quedará extramuros de la presente investigación porque, a nuestro parecer, forma parte de un problema ulterior. Solo una vez determinada la existencia de delitos imputables a un individuo con capacidad criminal (necesariamente, según lo aquí defendido, solo una persona, nunca un sistema), y una vez que se hayan asentado las características de esta infracción

---

se conoce a veces como «el problema de las muchas manos» (traducción propia). Asimismo, hablando de esta «responsabilidad distribuida» que surge como consecuencia de esta interacción múltiple: Taddeo, M., /Floridi, L., «How AI can be a force for good», *Science*, 361(6404), 2018, pp. 751- 752.

[286] Giannini, A., *op. cit*. N. 143, p. 170.

criminal y el modelo más adecuado de imputación, podremos determinar cómo individualizar la responsabilidad en caso de que esta se desarrolle (como entendemos será lo más habitual) en el seno de una corporación. Los problemas, según los concebimos aquí, son sucesivos, y solo trataremos el primero de ellos (es decir, el relativo a la individualización de la responsabilidad penal original, con independencia de que, más adelante, entren en juego las posibles delegaciones que, como en todo delito cometido en el seno de una organización empresarial, puedan dar lugar a las particularidades de autoría y participación) [287].

Entremos, pues, a desarrollar brevemente el segundo de los problemas apuntados, es decir, la discusión relativa a la posible imputación de un hecho en principio cometido por una IA, debiendo decidir a quién se transfiere, imputa o atribuye. Las alternativas son, esencialmente, dos (descartado que sea un hecho propio de la IA); o es un hecho de la corporación o es un hecho de una persona que trabaja en el seno de una organización. Para ello resultará útil realizar un breve repaso del surgimiento de sendos modelos.

## Las propuestas tradicionales de imputación a la persona jurídica

En el origen de la discusión acerca de la posibilidad de que se adscribiera responsabilidad a las personas jurídicas, dejando al margen la en absoluto cuestión baladí acerca de su conveniencia filosófica o moral [288], se encuentran

[287] Una breve explicación sobre el modelo de individualización derivado del *many hands problem* es tratado por Ibold V., *op. cit*. N. 104, pp. 342 y ss. La alemana, en aplicación de un enfoque *top down,* parte de que, en la búsqueda de la individualización de responsabilidad, la obligación genérica pertenecerá al empresario, pero más adelante se deberá buscar a la persona individual dentro de la empresa, donde no quedará más remedio que aplicar los ya conocidos modelos de división vertical y horizontal del trabajo, reflexionando sobre el papel del principio de confianza, tal y como ha venido haciéndose tradicionalmente. La verdadera aportación original de Ibold, creemos, tiene que ver con la preocupación por una específica laguna de responsabilidad difícilmente resoluble, pues una de las características específicas del modelo productivo de los sistemas de IA tiene que ver con que los ámbitos profesionales son cada vez más especializados y compartimentalizados, conduciendo a un desconocimiento *de facto* de los aspectos no relativos a la experticia, lo que va a impedir, en la práctica, la imputación. Aunque este es un problema especialmente reseñable, no consideramos que deba reducirse a la individualización en el seno de la empresa, y se tratará de forma particularizada en el capítulo quinto, cuando hablemos de las modificaciones respecto del régimen de la imprudencia y, muy particularmente, cuando discutamos sobre la exigencia (a nuestro juicio equivocada) de mayores estándares de conocimiento a profesionales, pues a nuestro juicio forma parte de un inconveniente mucho más amplio, que se debe tratar extramuros de la individualización de la responsabilidad en el seno de la empresa.

[288] No se desarrollarán en profundidad los argumentos de quienes, aún a día de hoy, y a pesar del reconocimiento positivo de los ordenamientos, siguen siendo absolutamente críticos con la posibilidad de afirmar una genuina responsabilidad penal a este tipo de entidades, pero vale la pena al menos dejar apuntado que en el ámbito continental son muchas y muy autorizadas las voces de quienes no aceptan esta

esencialmente dos posturas enfrentadas acerca de cómo concebir a la entidad o corporación. Por una parte, la visión individualista o atómica (la empresa no es más que la suma de sus miembros), y por otra, la orgánica o realista (la organización tiene una entidad propia y diferente a la suma de sus partes) [289]. Estas dos visiones, llevadas a sus últimas consecuencias, han llevado a dibujar al menos dos grandes modelos a la hora de atribuir responsabilidad penal a las entidades.

El primero de ellos, que apareció antes quizá por la menor abstracción intelectual que exigía, es el denominado modelo de responsabilidad vicarial o de heterorresponsabilidad. En este caso, partiendo de la idea de que la entidad se compone de una serie de personas que actúan en interés de otras, los primeros modelos de responsabilidad corporativa nacen en el ámbito estadounidense precisamente al amparo de figuras provenientes del derecho de daños, paradigmáticamente tomando el modelo del *«respondeat superior»* como ejemplo. Así, las empresas responden por los actos cometidos por sus empleados o agentes [290]. En el ámbito estadounidense, la teoría se llega a denominar *«broad*

construcción como una acorde a los requisitos mínimos de un derecho penal conforme a sus principios esenciales vertebradores. Sí puede apuntarse, no obstante, que las razones aducidas por estos autores tienen que ver con la imposibilidad de afirmar capacidad criminal (en los términos arriba expuestos) en entidades que carecen por completo de estados mentales. Así, con una exposición magistral de las razones para la crítica: Molina Fernández, F., «La responsabilidad penal de las personas jurídicas y el valor de π», blog de la facultad de derecho de la UAM, 14 de febrero de 2018, disponible en: <https://www.blog.fder.uam.es/la-responsabilidad-penal-de-las-personas-juridicas-y-el-valor-de-%CF%80-fernando-molina-fernandez/> [Accedido por última vez el 21/01/2026] o, este mismo autor: Molina Fernández, F., «Societas peccare non potest...: nec delinquere», *op. cit*. N. 283. También, dentro de la academia española, entre otros: Gracia Martín, L., «Crítica de las modernas construcciones de una mal llamada responsabilidad penal de la persona jurídica», *Revista Electrónica de Ciencia Penal y Criminología* (en línea). 2016, n. 18-05, pp. 1-95. Robles Planas, R., «¿Delitos de personas jurídicas?», *InDret*, 2/2006, pp. 1-25, o el mismo autor, «Pena y persona jurídica. Crítica del artículo 31 bis CP», *Diario La Ley,* n. 7705, 2011. En un sentido también escéptico, Silva Sánchez, J. M., «La responsabilidad penal de las personas jurídicas en Derecho español», en Silva Sánchez, J. M., (Dir.), Montaner Fernández, R. (Coord.) *Criminalidad de empresa y Compliance*, Atelier, Barcelona, 2013, pp. 15-39.

[289] Gobert, J., / Punch, M., *Rethinking corporate crime,* Cambridge University Press, 2003, p. 46. Estos últimos sitúan como el origen de esta categorización a Wolf. Wolf, S., «The legal an Moral Responsiblity of Organisations» en Pernock, J./ Chapman, J. (eds.), *Criminal Justice*, New York University Press, Nueva York, 1985.

[290] Se puede considerar que esta idea básica del modelo «*respondeat superior*» (superior que responde), bebe efectivamente de los primigenios modelos agente/principal que, al albur de instituciones como la esclavitud y la servidumbre, vuelven a encontrar una aplicación en los contextos de «esclavos o agentes electrónicos». Así, Bathae trae a colación los orígenes de esta institución, explicando que la revolución industrial trajo consigo problemas para la institución de la agencia, muchos de los cuales surgieron de la mayor independencia de los agentes y la falta de supervisión directa de los principales. Precisamente una de las doctrinas creadas para abordar el problema de un agente empleado con una tarea general de cumplir ciertos objetivos de su principal es la del *respondeat superior*, una forma de responsabilidad vicaria que hace al principal responsable de la conducta de un agente que emplea. Esta problemática ya estaba formulada en los escritos de Blackstone, que en sus comentarios establece que la razón de ser de esta doctrina se encuentra en los casos en que el uso de un agente por parte de un principal equivale a una «orden general»

*agency»,* pues se entiende que el acto de cualquiera de las personas vinculadas con la empresa (empleados o directivos), es capaz de generar esta responsabilidad en el «otro», es decir, en la propia organización [291]. La evolución del derecho inglés sigue un camino parecido, siendo sus orígenes los mismos, pero separándose en lo relativo a las personas que podían vincular a la persona jurídica. Así, y todavía en el seno de un modelo vicarial, la jurisprudencia inglesa [292] crea su propia teoría («teoría de la identificación» o «*alter ego theory*»), con la que se afirma que no todos los actos de cualquier persona que trabaje para la empresa pueden dar lugar a esa atribución de su responsabilidad, sino que solo lo podrán hacer los de las personas que sean el «cerebro» de la misma [293], que no son otros que los altos directivos. Se crea así una ficción capaz de limitar la derivación de responsabilidad, que será la que se aplique por los tribunales sin mayores complicaciones (principalmente por el hecho de que los delitos a los que se aplicaba eran típicamente de *strict liability* [294], donde no era necesario buscar un estado mental que adscribir), hasta la llegada de casos más complejos que empiezan a cuestionar los fundamentos de esta construcción [295].

Sin embargo, el derecho penal corporativo tomó otra perspectiva en su desarrollo en otro de los territorios del *common law:* Australia [296]. Con su Código Penal Federal de 1995, dicho país integró un modelo basado precisamen-

---

para lograr los objetivos del principal. Bajo esta regla, si, por ejemplo, un «camarero en una taberna vende a un hombre vino en mal estado, causando daños a su salud, este puede presentar una demanda contra el dueño», porque «aunque el dueño no ordenó expresamente al sirviente que lo vendiera a esa persona en particular», existía «implícitamente una orden general» de vender el vino. De hecho, los primeros casos de *respondeat superior* y responsabilidad vicaria enfrentaron dificultades precisamente con los agravios intencionales causados por un agente en uso de su autonomía. Bathaee, Y., «The artificial intelligence black box and the failure of intent and causation», *Harvard Journal of Law & Technology*, 31(2), 2018, p. 934. De hecho, en nota 194, aclara que Blackstone se refirió a la relación maestro-sirviente, pero la estructura de principal-agente es funcionalmente la misma para los propósitos de la discusión de este artículo (que versa, precisamente, sobre los problemas para imputar daños en los casos de herramientas de IA complejas). En consecuencia, se usan de manera intercambiable los términos «principal», «dueño» (*master*) y «empleador». Del mismo modo, también hay una referencia indistinta a agentes, sirvientes y empleados.

[291] Wells, C., *Corporations and criminal responsibility*, Oxford University Press, 2001, Cap. 5, donde realiza un repaso completo de la historia de la responsabilidad corporativa en el ámbito anglosajón.

[292] El caso de obligada referencia es: United Kingdom House of Lords: Tesco Supermarkets Ltd. v. Nattrass, 1971.

[293] Gobert, J., / Punch, M, *op. cit*. N. 289, p. 63.

[294] Un resumen de la problemática derivada de las *strict liability offences infra*.

[295] Estos casos son aquellos en que se producen muertes en el desarrollo de la actividad empresarial (conocidos como *«corporate manslaughter»),* en los que los tribunales encontraban problemas para castigar sucesos de gravedad palmaria que no cuadraban con su teoría de la identificación. El *leading case* paradigmático, que marcó un punto de inflexión en el desarrollo de esta teoría, fue R v P&O European Ferres (Dover) Ltd (1991) 93 Crim App R 72. Las consecuencias legislativas que se siguieron de este accidente fatal de ferry que supuso la muerte de 193 pasajeros, son narradas en Gobert, J., / Punch, M, *op. cit*. N. 289, p. 103 y ss.

[296] Wells, C., *op. cit*. N. 291, Cap. 7.

te en la otra visión de las corporaciones, no ya como agregación de personas individuales (con la necesidad de buscar, por ello, la atribución de los elementos típicos del delito, buscando adscribir tanto el *actus reus* como la *mens rea* por medio de complicadas ficciones), sino como una entidad con individualidad propia, al margen de la de sus componentes. Se trata, por tanto, del primer modelo de autorresponsabilidad, donde por primera vez la empresa es el actor en el que buscar un acto y una culpa propias.

En el ámbito continental, como es sabido, el desarrollo de la responsabilidad corporativa fue más tardío, pero lo cierto es que el dilema que se plantea es esencialmente el mismo. Las alternativas son, pues, idénticas: optar por un modelo vicarial o de heterorresponsabilidad, bajo la idea de que la sociedad no puede considerarse en ningún caso una organización con entidad propia ni nada similar a una persona, con la necesidad de buscar los elementos del delito en las personas físicas que la componen; o, por el contrario, optar por un modelo donde, afirmando la entidad propia de la organización, se afirme la «autorresponsabilidad» de este nuevo actor, creando una culpabilidad específica de la persona jurídica.

En el caso de los modelos de heterorresponsabilidad, Alemania representa el mejor ejemplo. Con su reticencia mantenida en el tiempo a afirmar que un ente no humano pueda delinquir, (bajo la máxima latina *societas delinquere non potest*[297]), el único modelo que ha acabado adoptando es uno vicarial, donde, por medio de sanciones de naturaleza administrativa (impuestas, eso sí, por órganos penales), se sigue la lógica de imputar un delito individual a la corporación[298]. Países como Portugal han seguido también una lógica vicarial[299]. Pero lo cierto

[297] Para Wells la reticencia tiene que ver también, en el contexto continental, con la asociación histórica de la responsabilidad grupal con regímenes represivos. *Ibid*, p. 140.

[298] Feijoo ha analizado los intentos más recientes de modificar la normativa germana, sin que pueda afirmarse que suponen un cambio en su fundamento, que sigue basado en lo que ya en 1979 expusieran autores como Schünemann. Vid: Feijoo Sánchez, B., «La función de la responsabilidad penal de las personas jurídicas en el derecho penal español«, REDEPEC: Revista Electrónica de Responsabilidad Penal de Personas Jurídicas y Compliance, vol. 1, n. 1, 2023, p. 17 y ss. Asimismo, la obra del autor alemán a la que se refiere Feijoo: Schünemann, B., *Unternehmenskriminalität und Strafrecht: eine Untersuchung der Verantwortlichkeit der Unternehmen und ihrer Führungskräfte nach geltendem und geplantem Straf-und Ordnungswidrigkeitenrecht*, Heymann, Köln, 1979. Un examen reciente de estas propuestas de introducción de un modelo de responsabilidad penal de la persona jurídica en Alemania, en Lohmann, A., *Strafrecht im Zeitalter von Künstlicher Intelligenzp. Der Einfluss von autonomen Systemen und KI auf die tradierten strafrechtlichen Verantwortungsstrukturen*, Nomos, Baden-Baden, 2021, pp. 235 y ss., evidenciando que todos los problemas que impiden reconocer personalidad jurídica a la IA surgen inevitablemente en el debate alemán cuando se trata de reconocer capacidad de culpabilidad a la empresa (capacidad de obrar, de culpabilidad y capacidad de pena). Especialmente en p. 242.

[299] Así, el Código Penal Portugués, en el artículo 11.º, n. 2, acoge, directamente, una teoría de «identificación» que se puede poner en relación, directamente, con la doctrina inglesa arriba expuesta. *Vid.*: Januário, T. F. X., *op. cit*. N. 309, p. 12.

es que también en el ámbito continental nos encontramos con ordenamientos penales que han acabado reconociendo abiertamente otra posibilidad. Así, la apuesta por el modelo de autorresponsabilidad a la «australiana», ha sido influencia clara de, en primer lugar, la pionera legislación italiana [300] y, más adelante, incipientemente en 2010 y de forma rotunda en 2015, en la española [301]. Así, en la actualidad, ambos ordenamientos [302] beben de forma clara de la inspiración del derecho australiano y de una concepción muy concreta de la persona jurídica. Asistimos, en este campo, a un claro punto de contacto entre ordenamientos con tradiciones dispares que, quizá por motivos diferentes, llegan a converger en puntos comunes. Cabe destacar, de hecho, que el actual modelo estadounidense es muy cercano al español/italiano, pues el original sistema vicarial ha ido cediendo, de forma casi natural [303] por las dificultades aplicativas que suponía y los vacíos de responsabilidad que presentaba.

Si bien no puede afirmarse que exista un acuerdo absoluto acerca de los fundamentos de esta responsabilidad corporativa [304], lo cierto es que poco margen de interpretación queda en cuanto a que todos estos sistemas (ya los continentales, ya los angloamericanos), en la actualidad hacen depender la responsabilidad en un enfoque común, que no es otro que un enfoque basado en el riesgo u orientado al riesgo de incumplimiento. Muy brevemente, la idea imperante es que la corporación tiene, en tanto actor de la vida social, unos deberes relativos

---

[300] Decreto legislativo italiano de 8 de junio de 2001, n. 231, que vino a regular la «*Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica*».

[301] La responsabilidad penal de las personas jurídicas se introduce en España por primera vez mediante la LO 5/2010, y se modifica posteriormente mediante LO 1/2015, ambas de modificación del Código Penal.

[302] Y no sólo ellos porque, de hecho, la expansión se sigue produciendo hacia el ámbito latinoamericano donde al menos en Chile y México se están acogiendo los aspectos sustanciales de este modelo. *Vid.*: Feijoo, explicando la evolución y la impronta del primigenio decreto italiano en todos estos modelos. Feijoo Sánchez, B., «La función de la responsabilidad penal de las personas jurídicas en el derecho penal español…, *op. cit*. N. 298, p. 24 y ss.

[303] Así, autores como Gómez-Jara, al hacer una comparación con el derecho americano, hablan de esta tendencia a dejar de lado ese origen más cercano a modelos del derecho civil, para buscar modelos basados en teorías organizativas. Para más desarrollo, *Vid.*: Gómez-Jara Díez, C., *La responsabilidad penal de las empresas en los EE. UU*, Madrid, Ed. Universitaria Ramón Areces, Madrid, 2006. También partidarios de esta visión y preferencia por las teorías organizativas en el ámbito anglosajón, Wells, C., *op. cit*. N. 291, y Gobert, J., / Punch, M., *op. cit*., N. 289. Y especialmente muy favorables estos autores con el modelo italiano, p. 111.

[304] Aún en el ámbito español hay autores que encuentran el fundamento de la organización en otro tipo de argumento, vinculado a la culpabilidad basada en el «defecto de organización» y no tanto en la «cultura corporativa». Ello no impide, no obstante, afirmar la importancia del paradigma del *compliance* y la gestión corporativa de riesgos como piedra de toque del sistema. *Vid.,* a título de ejemplo: Lascuraín, J, A, «Compliance, debido control y unos refrescos», en Arroyo Zapatero, L., Nieto Martín, A. (Dirs.), *El derecho penal económico en la era compliance*, Tirant Lo Blanch, Valencia, 2013, pp. 111-136. Este modelo bebe de las ideas pioneras del autor alemán Klaus Tiedemann.

al cumplimiento de la legalidad, y que dentro de las funciones corporativas que la empresa debe desempeñar de forma diligente se encuentra, entre otras más aceptadas como la labor fiscal, financiera o contable, la de asegurar la legalidad penal [305]. El hecho penal es, por tanto, no de las personas que trabajan para la empresa o la dirigen, sino de la empresa misma, que debe desplegar toda una serie de actuaciones que garanticen unos niveles de cumplimiento con la legalidad. El cómo se deban concretar estos deberes es algo que depende de cada sistema, pero lo cierto es que no son tan lejanos sistemas como el español o el italiano, que exigen de forma expresa y detallada ciertas características en estos programas de cumplimiento, y el angloamericano que, por medio de la *«due diligence defence»,* permiten exonerar de responsabilidad a la persona jurídica que demuestre haber diseñado e implementado estos métodos de cumplimiento [306].

En definitiva, este breve recorrido puede servir para dar cuenta de cómo la evolución de la responsabilidad penal empresarial se ha ido viendo influenciada por visiones comunes, y su enriquecimiento viene de la mano de propuestas que han ido confluyendo hacia un modelo donde el foco se pone en la gestión de riesgos, lo cual deja todo el peso en la concreción de deberes o programas de cumplimiento que permitan tener un control real de las actividades peligrosas que se desarrollen en el seno de la misma. Es en esta indiscutible labor de gestión de riesgos donde surge, de nuevo, la pregunta clave, aun sobre la mesa: ¿a quién compete, propiamente, esta gestión? Es decir, ¿quién es el agente obligado penalmente a gestionar estos riesgos? Si asumimos un modelo de autorresponsabilidad lo será la propia persona jurídica, mientras que si asumimos un modelo vicarial habrá de ser una persona física.

El debate entre estos dos grandes modelos, lejos de estar superado, aparece con fuerza cuando la IA empieza a desempeñar tareas dentro de las empresas, por lo que merece la pena realizar un recorrido por los argumentos de los autores que recientemente han tratado estos problemas.

---

[305] Feijoo Sánchez, B., «La función de la responsabilidad penal de las personas jurídicas en el derecho penal español…, *op. cit*. N. 298, nota a pie de página 71.

[306] Los matices, desde luego, son importantes, y no será igual que la concurrencia de estos protocolos de cumplimiento sirva para exonerar de toda responsabilidad, como ocurre en el caso español e italiano, que sirvan para suspender la pena de multa, como ocurre en el caso estadounidense. También resulta, desde luego, interesante, ver cómo opera la carga de la prueba a la hora de demostrar esta debida diligencia o control por parte de la empresa. Una reflexión interesante en este sentido: Shiner, R. A., «Corporations and the Presumption of Innocence», *Criminal Law and Philosophy*, *8*, 2014, pp. 485-503. También, profundizando sobre en qué deberes debe concretarse la «*due diligence*» y el juego de la inversión de la carga de la prueba en varios ordenamientos: Gobert, J./Punch, M., *op. cit*. N. 289, pp. 204 y ss.

No es difícil convenir, como muchos autores señalan, que será un modelo de gestión de riesgos el que venga a aplicarse en el caso de las corporaciones que usen sistemas de IA, pues no cabe duda de que el Reglamento de Inteligencia Artificial ha apostado por un enfoque basado en el riesgo. No prejuzga esto, empero, la solución que se haya de dar a la pregunta anterior. Existe una obligación de gestionar nuevos riesgos provenientes de sistemas. Cuando estos riesgos se den en el seno de corporaciones (como será muy habitual), sin embargo, todavía queda por establecer a quién atribuir estas obligaciones y, sobre todo, a quién imputar el incumplimiento de las mismas.

### Las propuestas de imputación IA-corporación-persona física

Quizá uno de los autores pioneros en proponer un esquema de responsabilidad robótica de tipo empresarial fuera Lehman-Wilzig [307], que ya en 1981 propone como solución a los futuros crímenes donde estén implicados robots adoptar un modelo similar al del «*agent/servant*» (agente/sirviente), como una forma de extender, por analogía, figuras legales preexistentes donde el énfasis se ponía en que el agente fuera una persona empleada por otra para hacer un trabajo para él, estando sujeto al control y las direcciones del empleador. Si bien las referencias de este autor son a figuras que pueden resultar extravagantes en una época como la actual (hablando de instituciones como la responsabilidad del dueño por su esclavo, del padre sobre sus hijos o del principal sobre su agente), nos pone sobre la pista de un camino que ya se ha recorrido, pues, cuando en su momento hubo que buscar soluciones innovadoras para hacer responder a nuevos entes (en este caso, a las corporaciones), no hubo más remedio que acudir a figuras de responsabilidad vicarial, esencialmente basadas en instituciones de responsabilidad civil, como la que este autor propone. Es este un camino que, de hecho, está siendo explorado por diversos autores, que no dudan en acudir a instituciones como la esclavitud para explicar estas nuevas tendencias [308].

Fuera del contexto anglosajón, también encontramos propuestas de aplicación del modelo de responsabilidad de las corporaciones como el más viable en un momento como el presente. Así, Januário [309], que, aunque señala que el

[307] Lehman-Wilzig, S. N., «Frankenstein unbound: Towards a legal definition of artificial intelligence», *Futures*, *13*(6), 1981, p. 542.

[308] Algunos autores han propuesto, directamente, la aplicación de figuras como el «*peculio*» romano. *Vid*., con referencias a estos autores en el ámbito anglosajón: Chesterman, S., *We, ¿the robots?*, Cambridge University Press, Cambridge, 2021, p. 129.

[309] Januário, T. F. X., «Inteligencia artificial y responsabilidad penal de personas jurídicas: un análisis de sus aspectos materiales y procesales», *Estudios Penales y Criminológicos*, *44* (Especial «Inteligen-

ordenamiento penal portugués sigue apostando por un modelo de responsabilidad vicarial, afirma que un modelo de autorresponsabilidad de la persona jurídica, con un injusto propio basado en una cultura de cumplimiento normativo, es el que mejor encaja con la necesidad de «identificar y gestionar los riesgos derivados de las nuevas tecnologías» [310]. Los argumentos esgrimidos por el autor tienen que ver, esencialmente, con las dificultades que presentaría encontrar un hecho de conexión propio de una persona individual, por las dificultades que la complejidad de los sistemas de IA supone para la identificación del nexo causal entre la acción y el daño y la problemática individualización de una eventual autoría individual [311].

En el ámbito español, Llonín Blasco [312], siendo aún más concreto, apuesta directamente por la aplicación de mecanismos de cumplimiento normativo como forma de atribuir responsabilidad a las personas jurídicas o corporaciones que tengan el control de las IAs, y que por tanto asuman toda la gestión de riesgos asociadas a las mismas. Para este autor, la solución pasará necesariamente por trasladar las decisiones que, a nivel de gestión de riesgos, formen parte del Reglamento Europeo de IA, de tal forma que el incumplimiento de alguno de los deberes propios de la normativa administrativa podrá dar lugar a la responsabilidad penal [313]. Otros autores españoles han reconocido la posibilidad de esta aplicación directa de los deberes de control contenidos en la normativa comunitaria [314].

No solucionan demasiado, nos tememos, estas propuestas, pues el incumplimiento de deberes extrapenales no puede constituir motivo suficiente para activar la responsabilidad penal de la persona jurídica. Entendemos que se trata, en esencia, de propuestas que objetivan la responsabilidad de la persona jurídica hasta unos límites insospechados, pues el mero incumplimiento de deberes administrativos parece fundamentar aquí la responsabilidad del ente, sin plantearse la cuestión de a quién compete cumplir con cada uno de los de-

---

cia artificial y sistema penal»), 2023, pp. 1-39.

[310] *Ibid.* p. 21.

[311] *Ibid.,* pp. 14 y 15.

[312] Llonín Blasco, L., «Acerca de la relación entre inteligencia artificial y responsabilidad penal empresarial», *Revista Sistema Penal Crítico*, 3, 2022, pp. 27-48.

[313] A nuestro juicio, peca esta propuesta de una simplicidad y automatismo no justificables, pues de ninguna forma sería posible identificar, sin ulteriores referencias, el injusto administrativo con el penal, pues supondría una clara contravención a las exigencias de antijuridicidad material requeridas por el principio de lesividad *(vid. infra).*

[314] Romeo Casabona, C. M., «Criminal responsability of robots and autonomous artificial intelligent systems?», *Comunicaciones en propiedad industrial y derecho de la competencia,* n.º 91, 2020, p. 173. Puede afirmarse respecto de esta propuesta lo mismo que se ha dicho para la de Llonín Blasco *supra.*

beres y sin ahondar, por tanto, en la verdadera problemática que aquí nos ocupa, cual es la individualización de estos deberes.

Una propuesta algo más profunda en este sentido es la de De la Mata Barranco y Pérez González, que han explorado recientemente la posibilidad de aplicar el modelo de cumplimiento normativo en los contextos de IA [315], si bien aceptando las dificultades que pueden suponer la autonomía y las capacidades de aprendizaje de estos sistemas en la imputación a las personas jurídicas. Se muestran partidarios de aplicar el artículo 31 bis del CP (si bien reconocen que no parece ser pensado para entidades no humanas), entendiendo que el agente o sistema es asimilable a una persona autorizada para tomar decisiones en nombre de la persona jurídica (apartado 1.a) del citado artículo), o bien que es similar a un empleado sobre el que se han incumplido los deberes de supervisión (apartado 1.b), destacando la necesidad de que las empresas implementen también en este campo programas de cumplimiento normativo que sean capaces de abordar los riesgos propios de los sistemas autónomos. No obstante, los autores son conscientes de que su propuesta, además de forzar los términos de la ley, podría conducir a «virtualizar» en exceso la responsabilidad de la persona física [316]. No dejan de plantear, por ello, que otra opción para responsabilizar a ésta nos obligaría a entender, cuando de inteligencia artificial autónoma se trate, que aquí siempre hay una «persona física» detrás, pues toda inteligencia artificial tiene quien la crea o quien la debe controlar. [317]

Algo más atrevida, aunque en esta misma línea de interpretación analógica, es la propuesta de Diamantis, cuya idea fuerza consiste en considerar al sistema de IA como un empleado más de la empresa, a todos los efectos. En su trabajo de 2023 [318] aboga por que se considere a los algoritmos autónomos de IA como empleados de las organizaciones empresariales. El autor se centra en la defensa del modelo del *respondeat superior,* si bien tratando de salvar las desventajas que presentan los modelos vicariales, aludiendo para ello a los

---

[315] De La Mata Barranco, N./Pérez González, S., «Inteligencia artificial autónoma y responsabilidad penal de las personas jurídicas», blog Almacén de Derecho, 12 de septiembre de 2013, disponible en: <https://almacendederecho.org/inteligencia-artificial-autonoma-y-responsabilidad-penal-de-las-personas-juridicas> [Accedido por última vez el 21/01/2026].

[316] Conceden que: *«el artículo 31 ter 1 CP precisa que «la responsabilidad penal de las personas jurídicas será exigible siempre que se constate la comisión de un delito que haya tenido que cometerse por quien ostente los cargos o funciones aludidas en el artículo anterior, aun cuando la concreta persona física responsable no haya sido individualizada o no haya sido posible dirigir el procedimiento contra ella». Persona física, se dice. Y de nuevo de personas se habla en los apartados 2 y 4 del artículo 31 bis. ¿Hay que sustituir las alusiones a las personas físicas por «agentes humanos o artificiales»? Ayudaría, por más que pueda entenderse que estamos virtualizando en exceso la responsabilidad de la persona jurídica.»*

[317] *Ibid*.

[318] Diamantis, M. E., *op. cit.* N. 116.

requisitos de que la IA actúe en beneficio de la empresa y en el marco de las tareas propias o bajo su control. A pesar de reconocer sus desventajas, acepta que es el modelo que más fácilmente se puede implementar, presentando una serie de razones que lo hacen prevalecer frente a otras propuestas más cercanas a la responsabilidad objetiva [319]. Para este autor, los robots son parte de la fuerza productiva de las empresas, sin que tenga sentido ni aplicación ya la segunda de las leyes Robóticas de Asimov, según la cual un robot siempre tiene que obedecer las órdenes dadas por un humano en un contexto donde los algoritmos son más eficientes o incluso justos sin que dependan de las órdenes humanas, pues su capacidad de actuación autónoma ha hecho que se supere esta necesidad de recibir órdenes humanas en el desempeño de sus tareas «laborales». De hecho, ya en un trabajo anterior [320] sostiene que afirmar la existencia de una mente corporativa tiene sentido desde un punto de vista psicológico. La idea no es crear un nuevo modelo que rompa con los esquemas aceptados y establecidos, sino explorar los límites de considerar que es la propia corporación la que tiene, a modo de ficción ya aceptada, una mente, y cómo esta mente puede estar configurada o afectada también por las actuaciones de los algoritmos [321]. La propuesta es extender esta teoría a las mentes empresariales, de tal forma que no solo lo que saben los empleados, sino también el conocimiento del que disponen los algoritmos que se usan bajo los criterios del *respondeat superior*, pueden formar parte de la *«mente empresarial»*.

Este modelo laboral de Diamantis ha encontrado, sin embargo, críticos dentro de nuestro sistema. Así, Del Rosal [322] señala que en realidad la propues-

---

[319] Principalmente critica el modelo propuesto por Abbott y Sarch. Proponen estos autores prever algún tipo de responsabilidad penal para la persona que se designe como responsable de la IA, al que habrá que asignar unos deberes de control. Esta «*Responsible Person*» estaría sujeta a deberes de tipo objetivo, principalmente vinculados a un registro o licencia, y esos deberes de control se traducirían, en algunos casos, en algún tipo de responsabilidad objetiva. La «*responsible person*» podría ser tanto una persona física como jurídica, pero parece que por el tipo de IA comerciales en que piensan y el tipo de actividades de las que estaría a cargo esta persona –hablan de fabricantes, («*manufacturers*»), proveedores («*suppliers*») o propietarios («*owners*»)–, lo más normal será que este esquema se aplique en el ámbito empresarial. Sin embargo, cabe destacar que no especifican demasiado estos autores cómo debería concretarse esta responsabilidad, y que en otro de los pasajes reconocen que la doctrina del «*respondeat superior*» puede no ser óptima para resolver los problemas relacionados con los «*hard AI crimes*», por lo que el fundamento concreto de la responsabilidad queda sin concretar. *Vid.*, Abbott, R., /Sarch, A. F, *op cit.,* N.85, p. 378 y ss.

[320] Diamantis, M. E., «The extended corporate mind: When corporations use AI to break the law», NCL *Rev*., 98, 2019, p. 893.

[321] *Ibid.* p., 907. A partir de la p. 912 aporta la visión de los teóricos de la mente de tipo funcionalista que apuestan por la teoría de la mente extendida («*extended mind*»), afirmando que lo importante es que los deseos vayan unidos a funciones que se quieren realizar, y que no importa de dónde se extraiga el conocimiento o los medios que permitan convertir ese deseo o intención en la satisfacción del acto. Esta propuesta bebe de las teorías de Chalmers y Clarks. Clark, A./Chalmers, D., «The extended mind», *Analysis*, vol. 58(1), 1998, pp. 7-19.

[322] Del Rosal Blasco, B., *op cit*. N 35, p. 28.

ta del *«labor model»* de Diamantis acaba haciendo lo que trata de evitar, pues establece un sistema de responsabilidad objetiva, en el que la empresa responderá por los daños del sistema de IA simplemente porque lo controla (o más bien, podríamos añadir, lo posee), sacando provecho de esa posesión [323].

Dentro de las propuestas de corte objetivo, quizá la propuesta más polémica, al prescindir abiertamente del elemento subjetivo, es la de Osmani [324]. Esta autora, partiendo de la existencia de un espacio de impunidad debido a la dificultad de atribuir responsabilidad individual como consecuencia de los comportamientos dañinos de los robots, se aleja del modelo vicarial del *respondeat superior,* afirmando que no es posible vincular la acción individual con la acción del sistema de IA. Apuesta, por su parte, por el modelo de autorresponsabilidad, poniendo el foco en la culpabilidad organizacional, proponiendo usar el modelo de las *public welfare offences* [325], es decir, delitos en los que se prescinde del elemento mental. Este tipo de infracciones estaban orientadas, de hecho, a controlar las industrias y las actividades que afectaban al bienestar y seguridad colectiva, de una forma análoga con los riesgos que ha supuesto el avance tecnológico.

El problema, según señala Del Rosal y compartimos, es que esta doctrina contraviene claramente el principio de culpabilidad, pues el elemento subjetivo se sustituye por la mera asunción de riesgo [326], no pudiendo tener cabida en un sistema como el nuestro la responsabilidad objetiva. En todo caso, y más allá de las críticas a los delitos de *strict liability,* falta en esta propuesta la explicación acerca de la conexión del hecho de la IA con la organización, pues parece difícil afirmar, sin más, que en un contexto de uso en que, en aplicación de la normativa vigente se ha utilizado un sistema de IA, la existencia de un defecto de organización. En este sentido, creemos que la crítica es también aplicable no solo a los modelos de autorresponsabilidad que son incapaces de explicar de qué modo un acto de un sistema de IA se vincula objetiva y subje-

---

[323] *Ibid.,* p. 43.

[324] Osmani, N.: «The Complexity of Criminal Liability Systems», *Masaryk University Journal of Law and Technology*, vol. 14, 2020, pp. 53 y ss.

[325] *Ibid*, pp. 73-74. Esta autora considera que este modelo tiene dos ventajas, por una parte, surge como una respuesta a los riesgos generados por la revolución industrial, trasladables ahora al escenario de la revolución digital. Por otra, permite imputar responsabilidad penal a corporaciones por los daños causados por la IA, sin necesidad de probar intención o culpabilidad, basándose simplemente en la asunción del riesgo.

[326] De hecho, la autora reconoce de forma expresa esta renuncia a la culpabilidad: *«this doctrine [public welfare offences doctrine] omits the criteria of blameworthiness; a penalty can be imposed regardless of the actor's intent, so the plaintiff does not have to prove that the defendant acted purposely. Intent or blameworthiness is replaced by the assumption of the risk that the actor bears when engaging in a certain activity», Ibid*, p. 73.

tivamente a una entidad (simplemente se pasan de largo los problemas de causalidad, imputación objetiva y subjetiva hablando de una cultura de cumplimiento sin aclarar a quién compete implementarla [327]), sino también al modelo propuesto por Diamantis, que implica trasladar la doctrina del *respondeat superior,* siempre y cuando los trabajadores no humanos cumplan los dos requisitos exigibles a los humanos (que se trabaje en beneficio de la corporación y dentro del alcance del empleo). Es esta una idea que ha sido también apuntada en el ámbito español, si bien para su aplicación en el ámbito civil [328]. En definitiva, la propuesta de Diamantis torna una vez más en una responsabilidad objetiva de la empresa, que acabará respondiendo por los daños del sistema simplemente porque saca un rédito económico de él [329].

La respuesta que ofrece Del Rosal resulta, empero, insuficiente o al menos no del todo satisfactoria. Como no le convence la posibilidad de reconocer personalidad jurídica al sistema y atribuir responsabilidad directa al mismo (estamos de acuerdo), y tampoco le parece que haya un buen fundamento para imputar la responsabilidad penal a la persona jurídica que posee la herramienta (también lo podemos convenir) concluye el autor, derrotado, afirmando que los daños ocasionados por los sistemas de IA deben reconducirse extramuros del derecho penal, siendo solo el derecho civil el encargado de abordar esta problemática. Este es el punto en el que, necesariamente, debemos discrepar y que, en efecto, dota de sentido a esta investigación.

Resulta obvio que los sistemas de IA se desarrollarán en entornos normalmente empresariales, y que el contexto comercial será determinante para la individualización de la responsabilidad, pero parece que hay una tarea previa por determinar.

En este punto, nos adscribimos a lo señalado por Benítez Ortuzar [330]. La opción de la Unión Europea por liderar el desarrollo de la IA desde la perspec-

---

[327] En este sentido, recientemente Navarro Cardoso, poniendo de manifiesto la dificultad que encontrarán los partidarios de un modelo de autorresponsabilidad de la responsabilidad penal de la persona jurídica para negar que su esquema pueda ser trasladable a una IA; de forma que si el injusto de la empresa lo forma un defecto de organización y la culpabilidad una ausencia de cultura de cumplimiento, es posible considerar que el defecto de programación del sistema supla al defecto de organización y la culpabilidad siga siendo esa ausencia de cultura de cumplimiento. Navarro Cardoso, F., «Anotaciones sobre la responsabilidad penal de las «personas artificiales», *Revista Penal,* 56, 2025**,** p. 204.

[328] Así, Navas Navarro. Esta autora afirma la responsabilidad civil por hecho ajeno basada en la mera relación humano-sistema, asimilándola a la relación empresario-dependiente y proponiendo la aplicación del artículo 1903.4 CC, es decir, aplicación del régimen de responsabilidad del empresario por los hechos de su dependiente. Navas Navarro, S., *Daños ocasionados por sistemas de inteligencia articial: Especial atención a su futura regulación*, Granada, Comares, 2022, pp. 60-61. Una vez más, estamos ante la propuesta de aplicación de un modelo de tipo laboral.

[329] Crítico con ello: Del Rosal Blasco, B., *op cit*. N 35, p. 43.

[330] Benítez Ortuzar, I. F., *op. cit.* N. 43 p. 503-515.

tiva del control humano significativo, atendiendo al riesgo admisible, permitirá la construcción paralela a la ya vigente respecto de la persona jurídica, de previsión de modelos de cumplimiento normativo que permita eximir de responsabilidad a la persona (física o jurídica) que fabrique o comercialice, pero ello no resuelve por completo el problema. En efecto, en los casos en que el robot está programado para desarrollar un proceso de aprendizaje, ello dificultará el grado de predictibilidad de la respuesta. Es decir, *«cuando la máquina está configurada para tomar sus propias decisiones, la respuesta jurídica tradicional presenta muchas fisuras»* [331]. El hecho de que la UE prevea un sistema de gestión de riesgos basado en el control humano significativo no significa más que eso, que habrá que encontrar un humano capaz de controlar los riesgos de la IA, ya sea que esta se implante dentro o fuera de la empresa. Optar por el sistema de autorresponsabilidad en este caso puede ser un camino fácil, pero no por ello más satisfactorio. Parece que las propuestas aquí analizadas (Januario, De la Mata, Osmani, o el propio Diamantis), parecen huir de la problemática de fondo, que va a seguir existiendo.

Resulta ilustrativa la conclusión alcanzada por Aires de Sousa [332], que reconoce que indudablemente las conductas lesivas de los sistemas de IA se producirán en su mayoría en entornos empresariales, pero también admite que ninguno de los dos grandes modelos puede resolver los problemas propios de esta tecnología. Apunta que bajo el modelo de la autorresponsabilidad no habrá manera de resolver los casos difíciles de autonomía operativa del sistema, pues la imprevisibilidad derivada de esta autonomía introduce un elemento que será muy difícil incluir en los modelos de prevención o cumplimiento por parte de las organizaciones, dando lugar al dilema de cómo prevenir aquello que es imprevisible.

Hemos de señalar, además, que el artículo 31 bis de nuestro Código penal exige, como es sabido, la comisión de un hecho delictivo por una persona física, pues, a pesar de la afirmación de un pretendido «injusto propio» de la persona jurídica, solo un hecho debidamente imputable a un agente con capacidad criminal (en los términos arriba señalados), por mucho que no esté identificado, dará lugar a una transferencia legítima de responsabilidad. Solo con una ficción sin demasiado sentido po-

---

[331] *Ibid.,* p. 514.

[332] Aires de Sousa, S., *op cit*., N 264, p. 73: *«aqui se evidenciam dificultades na afericao dos pressupostos de que depende aquela responsabilidade do ente coletivo. Neste último modelo, em que a pessoa coletiva responde pelos actos practicados por aqueles que a representam ou nela ocupam uma posição de liderança –modelo essencialmente seguido na legislação portuguesa– será difícil enquadrar o dano na esfera de responsabilidade de quem representa o ente coletivo. Já no contexto de um modelo baseado na cultura corporativa ou num déficit de organização a autonomia da máquina pode introduzir uma variável de imprevisibilidade difícil de detetar ou de prevenir no contexto corporativo.»* Esta misma idea se reitera en un trabajo posterior: Aires de Sousa, S., «The importance of AI for financial crimes», *op. cit. N*. 245, p. 163-180.

dríamos concluir que un ente sin capacidad de consciencia, de culpabilidad, ni de pena, habría de ser quien, con su «acto» diera lugar a una responsabilidad de otro ente, sin necesidad de dirimir ningún tipo de problema propiamente penal (sería operativo, no cabe duda, pues, ante las evidentes dificultades de individualización y atribución del hecho cabría decir que estaría bajo el control de una persona jurídica). Sin embargo, nos parece que este modelo es, no por sencillo, más justo, pues trata de desrresponsabilizar a la/s persona/s físicas que, aunque en otro papel, estarán siempre detrás de la creación y supervisión de estos sistemas.

Por su parte, el modelo de heteroresponsabilidad o vicarial presenta sus propias dificultades, pues el daño producido por la IA ha de estar vinculado a la decisión tomada por una persona física que represente a la persona jurídica. En estos entornos será altamente complejo, mucho más de lo que lo era en el ámbito humano, encontrar o individualizar a esa persona humana, precisamente por el problema del *many hands* [333]. Hemos de huir, sin embargo, de aquellas soluciones que, aunque desde la óptica del hecho de la persona física, tratan de establecer una relación casi automática entre el daño y la consecuencia jurídica, con esquemas más propios del ordenamiento civil o, si se quiere, de los intentos más anglosajones de responsabilidad objetiva.

Sin conocer previamente cuáles son los nuevos deberes jurídico-penales que se derivan (tanto para las personas físicas como para las jurídicas) en el desarrollo de tareas por medio de la IA, es imposible determinar la individualización de la responsabilidad en estructuras complejas [334].

Es necesario, pues, ir de lo más general a lo más específico en esta construcción. Los verdaderos y primarios retos se encuentran, creemos, en un paso previo, como es el de determinar los deberes de cuidado en sectores novedosos, con escasa regulación y donde por el momento no sabemos ni cómo atribuir la responsabilidad a nivel individual. Comenzar una investigación como la que aquí se pretende con la elaboración del modelo de responsabilidad en el seno de la empresa, dejando de lado los deberes de las personas que vinculan a esa misma empresa y que

[333] De hecho, algunos autores han ido mucho más allá de la teoría de las «*many hands*», afirmando que lo que ocurre en los entornos de interacción de IA más compleja se asemeja más a una especie de inteligencia de enjambre (Zhao, S., *op cit*. N. 160., p. 3), hablando de un riesgo específico de estas tecnologías interconectadas y superinteligentes, generando un riesgo de interconectividad. Muy desarrollado está este riesgo en la obra de Teubner y Beckers, que se refieren a la naturaleza compleja y opaca de las interacciones entre los sistemas de IA, que operan de forma interdependiente. Dichas interacciones son automatizadas, surgiendo en muchas ocasiones de una colaboración coordinada de múltiples algoritmos no reconducibles a un solo agente, con una alta interdependencia técnica entre los sistemas que forman parte de esta red, dando lugar a una opacidad estructural. *Vid.*, Beckers, A., /Teubner, G., *op cit.* N. 111, capítulo 5, pp. 111 y ss.

[334] Aires de Sousa, S., *op. cit*. N 264, p. 73. Pagallo, U., *The laws of robots*, *op. cit*. N. 143., p. 70.

estarán, necesariamente, detrás de cada uno de los sistemas, de una forma más o menos intensa, sería tanto como construir la casa por el tejado.

## II. DELIMITACIÓN POSITIVA: OBJETO Y FOCO PRINCIPAL DE ANÁLISIS

### 1. LAS PROPUESTAS DOGMÁTICAS

Si el modelo de la autoría directa del sistema no funciona y aún no estamos preparados para individualizar la responsabilidad en el seno de las corporaciones, no quedará más remedio que explorar modelos de imputación individuales.

La hipótesis de trabajo principal de esta investigación es que en los casos en que se cometan delitos que aquí se han calificado de *hard AI crimes* (es decir, hechos lesivos de bienes jurídico-penales en los que, debido a las características de autonomía e imprevisibilidad derivadas de la complejidad del sistema no sea posible*, prima facie,* encontrar criterios de imputación individual del delito), el modelo que ofrece las mejores perspectivas de solución es el de los delitos de peligro, inserto dentro de una discusión que comparte toda una serie de elementos con la temática que aquí nos ocupa.

La apuesta por un modelo como el propuesto se sitúa en un escenario de desarrollo tecnológico y social apegado a la realidad. Se trata de abordar el problema que planten las lagunas de responsabilidad provocadas por la IA haciendo uso en primer lugar de categorías bien conocidas y estudiadas, para valorar su capacidad de rendimiento. Sólo después, si fueran insuficientes, sería el momento de buscar otras alternativas más transgresoras.

Merece la pena estar atentos a la nota distintiva de la actual criminalidad, presidida más que nunca por la ubicuidad y la globalidad. El fenómeno de las regulaciones exclusivamente nacionales encuentra sus límites naturales cuando el delito ya no solo no se comete en un solo lugar, sino que, quizá, el concepto de «lugar» ha cambiado por completo. Debe tenerse en mente este fenómeno, si no es para homogeneizar ordenamientos, al menos sí para considerar el acercamiento y el diálogo como un valor esencialmente necesario. Si la disciplina de la inteligencia artificial es, por definición, multidisciplinar, con más razón habrá de ser al menos multiculturalmente influido el panorama regulatorio a tener en cuenta en la solución de los problemas penales.

Se hace necesario explorar, por ello, cómo son los caracteres de los *hard AI crimes* en sus diferentes formulaciones, huyendo, por el momento, del

modelo nacional de delito de peligro, para poder extraer, con el análisis de las propuestas, algunos problemas de partida.

En el ámbito continental, autores como Salvadori proponen esta solución, reconociendo que, en casos en que las IAs hayan alcanzado un estadio de mayor autonomía, lo adecuado será crear delitos que contemplen «posibles y genéricos eventos de daño en abstracto» [335]. También así lo afirma Busato [336], que, a pesar de mostrarse crítico con las bases que legitiman este tipo de delitos y sus límites, considera que debe ser esta la solución a la incertidumbre generada por estos eventos dañosos, tal y como se ha enfrentado en otras ocasiones el problema de la causación de riesgos desconocidos [337]. Los italianos Lagioia y Sartor [338] realizan una propuesta similar, apostando por la creación de una responsabilidad penal específica para los creadores o desarrolladores de sistemas de IA, incluso cuando el hecho dañoso no llegue a ocurrir, simplemente por haber emprendido una actividad peligrosa que dé lugar a un riesgo significativo de daño.

También resulta esta una propuesta frecuente entre los autores anglosajones, pues, precisamente al tratar las dificultades propias de los *«hard AI crimes»,* señalan Abbot y Sarch [339] que quizá la mejor forma de afrontarlas es creando una *«constructive offence»* para todos los casos en que, a pesar de haberse realizado un primer acto de alguna forma culpable (puede ser, por ejemplo, una pequeña imprudencia del programador), sin embargo, y sin que fuera previsible realmente, se produce un daño de una magnitud muy superior. Los autores ponen el ejemplo de un algoritmo que se usa para generar un fraude bancario de pequeña entidad pero que, finalmente, da lugar a un cortocircui-

---

[335] Salvadori, I., *op. cit*. N. 101, p. 163.

[336] Busato, C., *op. cit*. N. 124 p. 359.

[337] Pone el ejemplo Busato de los delitos de manipulación genética del CP español y, concretamente, del artículo 160.3, que castiga la clonación con fines eugenésicos, pero lo cierto es que no es el único, pues la legislación penal española castiga, asimismo, delitos de peligro de «alta incertidumbre» amparado en la idea de los riesgos catastróficos que pueden suponer. Así, *vid.* Capítulo I del Título XVII del CP español, donde se castigan conductas relativas a la energía nuclear y radiaciones ionizantes (castigadas con pena de hasta veinte años de prisión), previendo también modalidades de imprudencia grave, o delitos de estragos, castigados con penas especialmente altas de prisión. Especialmente llamativo, en el Capítulo II, el delito de incendio provocado con peligro para la vida o integridad física de las personas, que se castiga con penas de prisión de entre diez y veinte años, previéndose también el castigo (atenuado) en caso de imprudencia grave. Baste dejar apenas apuntado este modelo, que se estudiará en profundidad *infra*, en capítulo sexto.

[338] Lagioia, F., & Sartor, G., «Ai systems under criminal law: a legal analysis and a regulatory perspective», *Philosophy & Technology*, 33(3), 2020, p. 459. La propuesta de estos autores pasa por ampliar el alcance de la *recklessness,* equivalente al dolo eventual del sistema continental para permitir castigar los casos en que el autor sabe que hay algún riesgo, pero no conoce con exactitud su naturaleza.

[339] Abbott, R.,/ Sarch, A., *op. Cit.* N 85, p. 371 y ss.

to general, provocando daños físicos en personas [340]. De esta forma, proponiendo la creación del delito de *«Causing Harm Through Criminal Uses of AI»* [341], podría afrontarse la totalidad del daño causado, de forma parecida a como ocurre con el delito de *«felony murder»* en EEUU. Partiendo de una conducta culpable pero imprevisiblemente dañosa para el bien jurídico finalmente dañado, se imputa todo el resultado a la misma conducta inicial. Calo [342] propone algo similar. Reconociendo que, debido a la incertidumbre generada por los posibles resultados dañosos de los robots, la tendencia actual es hacia la objetivación de la responsabilidad, la solución óptima para el derecho penal sería la creación de algo así como un delito menor *(«parasitic misdemeanor»),* que castigue a un acusado por poner en juego una tecnología físicamente capaz de causar un daño específico cuando ese daño se produzca realmente. Diríamos, entonces, que la idea es una vez más similar: la sola iniciación de una conducta que, *ex ante*, pueda ser peligrosa para un bien jurídico, aunque sea de forma remota, permite imputar el daño finalmente producido, con el objetivo de *«controlar moral y pragmáticamente el uso excesivo de una tecnología potencialmente peligrosa»* [343].

Todas estas propuestas, aunque similares en líneas generales, difieren en dos aspectos principales: el grado de *mens rea* o estado mental exigido en la acción inicial –a veces más estricto y otras más flexible–, y la severidad de la pena impuesta. Así, en ordenamientos continentales las penas son más graves solo si van acompañadas de estados mentales más rigurosos, mientras que en propuestas como la de Calo se admiten estados mentales menos exigentes y, en consecuencia, penas menores.

Estas diferencias revelan tendencias y problemas que vale la pena destacar. El denominador común de estas figuras diversas (el delito de peligro abstracto, en

---

[340] El ejemplo análogo que ofrecen con lo que podría ser la misma estructura de conducta en el mundo analógico es la de unos niños que, sin tener autorización para ello, juegan con fuegos artificiales sin licencia, lo cual da lugar a un incendio con daños catastróficos para personas y cosas. Hablamos, en todo caso, de un elemento común en estos escenarios: una fuente de peligro inicialmente pequeña se descontrola, produciendo un daño elevado.

[341] La denominación que finalmente se adoptase para este tipo de delito es una cuestión que requeriría una discusión propia, por sus complejidades y consecuencias particulares. Si bien el problema del *«nomen iuris»* no ha sido especialmente explorado en la tradición continental, el principio del *«fair labelling»* resulta de importancia en el contexto anglosajón, y esta cuestión debiera tratarse en la eventual introducción del delito. Para ver las implicaciones que puede implicar la correcta elección de la denominación jurídica del delito, *vid.*: Chalmers, J., / Leverick, F., «Fair labelling in criminal law», *The Modern Law Review*, 71(2), 2008, pp. 217-246.

[342] Calo, R.*, op. cit.* N. 62, p. 554.

[343] *Ibid.,* p. 554: *«providing a moral and pragmatic check on overuse of a potentially dangerous technology without justification».*

terminología continental, o implícito, usando la terminología inglesa [344]: la *«constructive offence»,* la *«parasitic misdemeanor»* o un posible nuevo delito de peligro abstracto genérico [345]) muestran el intento de solucionar un fenómeno ya conocido, que ambos ordenamientos han tenido que enfrentar. Con la proliferación de algunas actividades peligrosas, en las que hay una gran incertidumbre en cuanto a sus resultados potencialmente dañosos y una imposibilidad de dejar sin castigo los posibles daños, el derecho ha suavizado algunos estándares de imputación de responsabilidad, sea al abordar la incontrolable capacidad destructora del fuego, la capacidad disruptiva de la tecnología genética o la utilización de herramientas tecnológicas de alta imprevisibilidad y escaso margen de control.

## 2. LAS PROPUESTAS DE LOS PAÍSES

Las propuestas relativas al modelo de delito de peligro no solo provienen de la doctrina, sino que un vistazo atento a la evolución del derecho comparado parece devolvernos este modelo como uno de los prioritarios. Como muestra de la actualidad de la problemática, es oportuno examinar las frecuentes menciones al modelo de peligro del informe *«Criminalisation of AI-related offences»* [346], fruto del ya mencionado trabajo de la AIDP (XXI Congreso In-

[344] Por utilizar la terminología de Duff. *Vid.*: Duff, R. A., «Criminalizing endangerment», *La. L. Rev.*, *65*, 2004, p. 959.

[345] A este respecto, resulta de mucha utilidad atender a la normalización de este tipo de delitos en entornos anglosajones. Especialmente interesante es el caso del delito de *common law* del derecho escocés denominado *«recklness endangerment offence»,* donde, partiendo solo de un riesgo genérico a cualquier bien jurídico, se castigan conductas que, con evidente desprecio por estos bienes, los pongan en peligro. Algunos ejemplos de casos en que se ha aplicado son lanzamiento de objetos desde un puente a la vía pública, lanzamiento de ladrillos o incluso la puesta en peligro propia de montañeros que solicitan un rescate de los servicios de emergencias. *Vid.* caso de HMA v. Margaret Ferrier; «*It is a crime to endanger others by reckless conduct. This crime can be committed in many different ways. It involves exposing an individual, or particular individuals, or the public generally, to a significant risk to life or health. A high degree of recklessness is needed, more than carelessness or negligence.»* La acusada es condenada por este delito por haber realizado vida social con normalidad después de haber dado positivo en un test COVID. La pena impuesta es la de «*Community Paybak Order*», similar a los trabajos en beneficio de la comunidad contemplados en el CP español. Fuente: Judiciary of Scotland, disponible en: <https://judiciary.scot/home/sentences-judgments/sentences-and-opinions/2022/09/13/hma-v-margaret-ferrier#:~: text=It%20is%20a%20crime%20to, more%20than%20carelessness%20or%20negligence>. [Accedido por última vez el 22/01/2026].

[346] Association International de Droit Pénal (AIDP), Miró-Llinares, F., *General Report – XXI International Congress of Penal Law on 'Artificial Intelligence and Criminal Justice', International Colloquium of Section II, «Old and New Criminal Offences: AI Systems as Instruments and Victims»*, 2023. Publicado más adelante en: Miró-Llinares, F., Duvac, C., Toader, T., /Santisteban Galarza, M., (Eds.) «Criminalisation of AI-related offences», *Revue Internationale de Droit Pénal*, Vol. 95 issue 1, 2024 (*op. cit*. N. 160).

ternacional de Derecho Penal [347]). En el mismo, teniendo en cuenta que el momento clave en términos de riesgo es el del diseño del sistema y su implementación, se considera como una opción la tipificación de delitos que busquen disuadir precisamente las conductas que se producen en esos momentos [348]. Este propósito, según apunta el relator general, Fernando Miró, puede lograrse mediante la anticipación de la protección con delitos de peligro, que sancionen el incumplimiento de ciertos deberes en relación con intereses específicos dignos de protección. Estos delitos de peligro relacionados con el diseño e implementación de sistemas de IA que generen riesgos deberán ser tipificados cuando las acciones sancionadas representen una amenaza considerable para los intereses protegidos, haciendo una alusión implícita al principio de intervención mínima o última ratio [349]. Destaca en todo caso Miró que estos delitos no deben diseñarse sin tener en consideración la normativa disponible en el momento, que debe servir para identificar cuáles son los riesgos relevantes.

Se afirma que, en la mayoría de los casos, las fases más relevantes para el derecho penal serán las primeras etapas del ciclo de vida de la IA (diseño del algoritmo o selección de datos), por lo que tiene sentido que la tipificación se adelante precisamente a estos momentos o etapas [350]. En aquellos sistemas penales diseñados aún para motivar el comportamiento humano, la responsabilidad debe trasladarse a esos momentos clave en los que la IA es diseñada y puede ser controlada, adelantando a momentos alejados de la manifestación del daño, pero cercanos a la acción humana responsable, explorando la posibilidad de tipificar delitos que sancionen el diseño deficiente o la falta de control de ciertos sistemas de IA [351]. Así, la propuesta implica que los delitos abarquen la fase de *diseño* o producción de todos aquellos sistemas que permitan una actuación autónoma externa a los objetivos predefinidos [352], criminalizando, por ejemplo, como delitos autónomos, los clásicos actos preparatorios u omisiones del deber de cuidado exigido.

Relata Miró que varios países han considerado este modelo, especialmente en áreas consideradas de alto riesgo. Así, los informes nacionales de

---

[347] Fruto del trabajo anual realizado por los países parte, así como sus relatores generales, se celebró el Congreso Internacional en París, en junio de 2024 (AIDP, XXIst International Congress of Penal Law).

[348] Miró-Llinares, F., *op. cit*. N. 160, p. 14.

[349] Se hace de esta forma una referencia implícita al principio de proporcionalidad, afirmando que la consecuencia legal asociada a estos delitos deberá ser proporcional al nivel de riesgo causado y al interés en juego.

[350] Miró-Llinares, F., *op. cit*. N. 160, p. 47.

[351] *Idem.*

[352] Es decir, no es un modelo que esté pensando para responder a los sistemas que no puedan desarrollar esta «autonomía operativa», por lo que nos remitimos aquí a la importancia mayúscula del elemento de autonomía, y del concepto de IA que se ha adoptado en el capítulo primero.

Finlandia y Chile destacan que podría ser necesario incorporar nuevos delitos relacionados con el diseño y control de ciertos sistemas de IA [353]. República Checa, por su parte, considera muy probable que, debido a la complejidad de los sistemas, habrá evidentes dificultades para atribuir responsabilidad a un sujeto concreto, por lo que sugiere la introducción de «nuevos delitos dentro de las categorías de delitos formales o de peligro» (por ejemplo, delitos que penalicen el mero hecho de no cumplir con los niveles de seguridad requeridos en actividades como pruebas, fabricación, actualización o monitoreo de la IA, incluso cuando no haya habido daño derivado de dicha infracción [354]). Argentina, por su parte, propone verificar *ex post* las condiciones en las que se produjeron los eventos dañosos, para que el derecho penal sancione conductas que impliquen la omisión de medidas de seguridad o actos que dificulten el acceso al tipo de programación utilizada por el sistema [355].

El informe italiano reconoce que un enfoque basado en el riesgo parece ser necesario, y que el legislador probablemente tendrá que recurrir a una anticipación de la protección penal mediante la creación de nuevos delitos que impliquen la contravención, por ejemplo, de obligaciones de información o mandatos a organizaciones sancionados penalmente. Croacia considera que podrían incorporarse nuevos delitos que sancionen el diseño deficiente o la falta de control para algunos sectores críticos [356].

El informe chino propone un nuevo delito basado en la negligencia «en relación con incidentes de seguridad graves relacionados con productos de IA», sancionando a diseñadores, productores o usuarios que no ejerzan el debido cuidado en el diseño, programación, producción y uso, y que como

---

[353] Miró-Llinares, F., *op. cit*. N. 160, p. 48, p. 11. Se refiere el relator general a los siguientes trabajos: Lahti, R, Koistinen, J., Lahti, L., « *Finnish national* report on criminalisation of AI-related offences», *RIDP*, 95, 2024: *«It may be necessary to incorporate new offences related to the design and control of certain AI systems if harm is caused for certain protected interests»* y Krause, S.: «*Chilean national report on criminalisation of AI-related offences*».

[354] Provaznick, J.: «Czech Republic national report on criminalisation of AI-related offences», *RIDP,* 95, 2024. *«The complexity of AI and non-singularity of the legal regimes of ownership, production, maintenance, controlling etc. of the devices or systems that use it will almost certainly lead to a grave difficulty in establishing the individual criminal liability of particular subject when something goes wrong. I would thus predict that new crimes falling to the categories of formal or endangering ones will be introduced (e. g. crimes which would criminalize mere not keeping up with the necessary levels of safety standards of activities like test-ing, manufacturing, updating, monitoring etc. of AI, even though no harm came from such a breach). The same is imaginable on the part of the users of AI (e. g. criminalization of endangering others by operating an AI-device without due maintenance or actualization, continuance of using it despite a warning on an occurrence of a fatal flaw in the particular model and call to maintenance service)»,* p. 257.

[355] Cherñavsky, N. A; Riquert, M. A, Sueiro, C. C.: *«Argentinian National report on criminalisation of AI-related offences», RIDP*, mayo de 2024, p. 368.

[356] Se trata de los delitos relacionados con armas autónomas, vehículos de conducción totalmente autónoma, robots médicos o sistemas financieros autónomos.

consecuencia de ello causen daños graves. También se propone por China la criminalización de comportamientos preparatorios, elevándolos a delitos específicos, proponiendo la creación de un nuevo delito de abuso de productos de IA, que incluya el uso indebido de capacidades de aprendizaje profundo para entrenar productos de IA que originalmente fueron diseñados con fines lícitos para cometer otros delitos.

El relator francés propone introducir un delito que penalice a las personas que operen con una IA y tengan conocimiento (o hubieran debido tenerlo) de que su funcionamiento puede generar un desorden sancionable penalmente en caso de que no se actúe para detener este hecho. De hecho, y tal y como señala el informe holandés, algunos sistemas ya pueden sancionar los comportamientos riesgosos mediante delitos relacionados con la creación de productos defectuosos. En el Código Penal de los Países Bajos hay dos artículos sobre la venta de artículos peligrosos para las personas: el artículo 174 tipifica como delito vender o distribuir bienes peligrosos, y el artículo 175 sanciona a quien sea responsable de la venta de bienes perjudiciales para la salud en su modalidad imprudente [357].

Finaliza Miró advirtiendo que lo esencial será aclarar el papel del derecho penal en casos de incumplimiento de obligaciones en fases de diseño o control de herramientas de IA, incluso en casos en que no haya conocimiento sobre el riesgo específico o el daño potencial de las herramientas, pero sí lo haya sobre la violación de un requisito que está pensado, precisamente, para reducir los riesgos asociados a su uso [358].

En conclusión, tanto las propuestas de la doctrina (continental y anglosajona), como las propuestas regulatorias más recientes de los países, apuntan hacia el modelo de adelantamiento en la protección, comenzando a explorar, aun tentativamente y de forma dispar, el modelo del delito de peligro, con una variedad de conductas en la base de los mismos (desde la mera generación de peligros genéricos, hasta deficiencias de diseño genéricas, incumplimiento de pruebas en fase de fabricación, omisión de medidas de seguridad, control insuficiente, o abuso de técnicas de aprendizaje complejas, entre otras).

---

[357] Smit, A. M.: *«The Netherlands national Report on criminalisation of AI-related offences», RIDP*, 95, 2024, p. 437, argumentando que, sobre estos delitos: *«Just as criminal proceedings can be initiated against producers of goods that cause damage that is not caused by the use of AI, this can also take place when AI has been used.»*

[358] Miró-Llinares, F., «General report», *op. cit.* N. 160, p. 50.

## 3. LAS CRÍTICAS. LOS PROBLEMAS NUCLEARES COMO PUNTO DE PARTIDA

Explorar un modelo como el de los delitos de peligro tiene, desde luego, ventajas e inconvenientes. Por una parte, nos sitúa en un campo familiar, que ha sido largamente estudiado y, por ello, objeto de críticas. No nos movemos en un escenario absolutamente novedoso, por mucho que la materia lo pareciera en un principio, sino que la adopción de un modelo familiar para el derecho penal nos obliga a traer a colación todo aquello que sobre él ya se ha dicho, para tratar de convenir si es capaz de superar esas críticas y ser aplicado en este contexto tecnológico de la manera más respetuosa posible con nuestro sistema de derechos.

Tanto la doctrina española como la anglosajona han tenido que lidiar con las críticas que se han vertido sobre los delitos de peligro, en los que el daño es cada vez más remoto y donde el nexo de imputación subjetiva (los estados mentales exigidos por el requisito de *mens rea)* se debilita cada vez más. Los retos planteados por estas figuras son comunes a ambos ordenamientos, como lo son las dificultades al trasladarlos a los *«hard AI crimes»,* donde las críticas tradicionales se tornan, aún, más acuciantes.

Aunque un desarrollo más pormenorizado se reserva para otros momentos de esta investigación, puede ya apuntarse que dichos problemas pueden dividirse en tres grandes grupos.

### 3.1 **La pérdida de importancia del principio de lesividad o la crisis del principio del daño**

El problema que tradicionalmente ha surgido en el ámbito de los delitos de peligro, en concreto, y los delitos de daño remoto (por utilizar la denominación acertada que Von Hirsh propone [359]), en general, tiene que ver con la pérdida de importancia sucesiva del principio de lesividad en sus diferentes manifestaciones. Lo que en España es, a grandes rasgos, representado por el principio de le-

[359] Von Hirsch, A., «Extending the harm principle: 'remote' harms and fair imputation», en Simester, A. P.,/ Smith, a.t.h, (eds.), Harm and Culpability*,* Oxford University Press, 1996. El autor agrupa bajo esta denominación problemas diversos con un denominador común, que es la mayor lejanía del «daño» en una serie de delitos, que no solo abarcan los delitos de peligro, con los delitos de peligro abstracto como ejemplo más paradigmático, sino también los casos de daños cumulativos o incluso casos de participación psíquica. En todo caso, cuando de aquí en adelante se utilice la denominación genérica de «delitos de daño remoto» se entenderá que se adscribe solo a los casos que aquí se analizan, más parecidos al paradigma del delito de peligro abstracto continental.

sividad o antijuridicidad material puede reinterpretarse, en términos anglosajones, como la crisis del principio del daño, o la necesidad de adaptar este principio cuando el «daño» deja de ser tal y se comienza a diluir, alejándose y haciéndose cada vez menos tangible respecto de la conducta inicial.

Todos los problemas referidos al «daño» o «lesividad» que se asocian a la expansión del derecho penal, se ven profundamente confirmados con la proliferación de delitos de «daño remoto» tecnológicos.

En España, autoras como Mendoza Buergo o Corcoy Bidasolo se han mostrado especialmente preocupadas por la afectación a este principio, considerándolo como uno de los puntos clave para poder solucionar los desajustes que generó, en su momento, la proliferación de delitos de peligro abstracto. Así, Mendoza advierte sobre la *«difuminación de la lesividad de las conductas incriminadas», así como sobre la «insuficiente carga de lesividad de la conducta individual desde parámetros estrictamente penales y no administrativos»* [360], proponiendo una reinterpretación de las garantías del derecho penal y las funciones que le son propias. Directamente reticente con la proliferación de delitos de peligro abstracto o estadístico, Corcoy asegura que principios como el de lesividad o antijuridicidad material son clave «*para excluir la relevancia penal*» de este tipo de delitos [361].

También en España, precisamente al hilo de la necesidad de modificar la dogmática para su adaptación al crimen tecnológico, Agustina [362] se ha mostrado reticente sobre la posibilidad de que la distinción entre delitos de peligro concreto y abstracto [363] tenga algún sentido en una época como la actual, donde la completa confusión entre la apariencia y la verdad ha hecho que los bienes tradicionalmente protegidos en nuestro ordenamiento se vayan diluyendo y confundiendo. Apuesta este autor por superar la férrea atención continental a la germana teoría del bien jurídico para explorar caminos más apegados al *«harm principle»* anglosajón. Y así, concluye: *«Para acercarse a los nuevos problemas con un criterio adecuado de dañosidad o peligrosidad, en ese sen-*

---

[360] Mendoza Buergo, B., «Exigencias de la moderna política criminal y principios limitadores del derecho penal», *Anuario de derecho penal y ciencias penales*, *52*(1), 1999, p. 300.

[361] Corcoy Bidasolo, M., «Principio de culpabilidad ¿responsabilidad objetiva en determinados delitos?», en AA. VV., *Estudios político criminales, jurídicos penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, Tirant Lo Blanch, Valencia, 2023, 2.ª ed., 2023, p. 576.

[362] Agustina, J. R., *op. cit*. N.62, p. 748 y ss.

[363] Las categorías tradicionalmente manejadas entre la doctrina española respecto de los delitos de peligro (la distinción fundamental entre los delitos de peligro concretos y abstractos, siendo estos últimos lo que sin duda mayores problemas dogmáticos han generado) encuentra un paralelismo en la distinción que Anthony Duff realiza entre los delitos de peligro implícitos o explícitos. Me remito, por ello, a esta categoría. *Vid.*: Duff, R. A., «Criminalizing endangerment», *op. cit*. N. 344.

*tido, sería necesario acudir al hombre medio y a la teoría de la conmoción en un mundo en el que la apariencia posee un significado distinto»* [364].

Lo cierto es que ya Feinberg reconoció en su primigenia obra la necesidad de castigar, sin saltar fuera de los márgenes del principio del daño, los casos de actividades peligrosas donde había que entrar a evaluar otros factores además del daño efectivamente causado. Con su pensamiento adelantaba de alguna forma los problemas presentes. Y así, justificaba la punición de daños remotos atendiendo a tres parámetros: la conjunción de la gravedad del posible daño y la probabilidad de su concurrencia (lo que él denomina, sencillamente, riesgo), el valor social asignado a la conducta examinada y la consideración de algunos principios que puedan actuar como limitaciones que hagan desaconsejable la criminalización, por tratarse de conductas extremadamente valiosas (refiriéndose, principalmente, al ejercicio de derechos fundamentales en colisión [365]). Resulta también adelantado a los problemas que más adelante surgirían cuando habla de la importancia de excluir de la punición los daños «triviales», al hilo del principio *«de minimis non curat lex»* [366], o cuando expresa los límites que deben imponerse en los casos de lo que él denomina «daños acumulativos», que no son más que una de las manifestaciones del daño remoto que requieren una respuesta penal tras la proliferación de actividades riesgosas.

Así, puede apreciarse ya lo que a nuestro juicio es un precedente del principio de precaución en su discurso (punto de apoyo fundamental para la justificación de los delitos de peligro en el ordenamiento español, por herencia del *«vorsorgeprinzip»* de la tradición alemana [367]), cuando señala que en aquellos supuestos de delitos ya no clásicos o del núcleo duro, sino más relacionados con bienes supraindividuales (pone el ejemplo del delito medioambiental), el legislador requiere de unos conocimientos técnicos para fijar umbrales de permisión o límites *(«thresholds»)* de daño, pero para ello no solo tiene que remitirse a la autoridad científica competente, sino luego crear una regulación con-

---

[364] Agustina, J. R., *op. cit*. N.62, p. 756.

[365] Feinberg, J., *Harm to others*, 1984. Vol. I, Oxford University Press, p. 216.

[366] *Ibid*. p. 215. Más adelante, y en un intento por desarrollar una teoría de criminalización que suponga verdaderos límites dogmáticos al legislador, tratando de poner freno a la excesiva criminalización en el contexto estadounidense, y con una inspiración básica en el pensamiento de Feinberg: Husak, D., *Overcriminalization: The limits of the criminal law*, Oxford University Press, 2008.

[367] Sobre la importancia de este principio precisamente para resolver los problemas de los delitos tecnológicos: Galán Muñoz, A. «La problemática utilización del principio de precaución como referente de la política criminal del moderno derecho penal: ¿hacia un derecho penal del miedo a lo desconocido o hacia uno realmente preventivo?» *Revista General de Derecho penal*, 23, 2015, pág. 1 y ss. Crítico, sin embargo, con la aplicación de este principio en la resolución de problemas de adscripción de responsabilidad en el ámbito de la responsabilidad robótica: Quintero Olivares, G., *op. cit.* N. 126, p. 20. En todo caso, se desarrolla la importancia que el principio de precaución tiene en una época como la actual *infra* (capítulo cuarto).

creta, externa al campo penal, que sea la encargada de supervisar y dar forma legal a los permisos y usos legítimos. Solo cuando se rompa esta legislación podrá acudirse al derecho penal, con algunas limitaciones o requisitos extra si se quiere dar cumplimiento en este sector al principio del daño.

Con estas sólidas bases desde las que partir, sin embargo, no se ha podido frenar el desarrollo y proliferación de cada vez más delitos en los que el principio del daño ha podido verse comprometido. En este sentido, los esfuerzos de la doctrina inglesa por mantener férreos los límites derivados de este principio merecen todo el reconocimiento. Especialmente valiosos, a estos fines, los trabajos de Von Hirsh y Simester[368]. Tratan ambos autores de encontrar algún tipo de justificación a la creación, en genérico, de delitos que denominan *«non-constructive»,* es decir, aquellos en cuya formulación no hay referencia específica al daño concreto que se puede ocasionar (rasgo que sería el propio de un supuesto delito futuro de «generación de daños por el uso de IA» arriba propuesto). Pues bien, señalando diferentes manifestaciones de este fenómeno, destacan la de los delitos de peligro, y muy señaladamente, los de peligro abstracto (en contraposición a los de peligro concreto que, según los autores, sí pueden adecuarse con facilidad al principio del daño). En ellos el problema fundamental es que los delitos de puesta en peligro parecen prescindir del requisito del daño, ya que prohíben una conducta independientemente de las contingencias que hacen que esa conducta sea peligrosa. Es decir, la conducta genera un peligro solo si se dan ciertas contingencias, pero la norma se aplica incluso aunque esas contingencias estén ausentes y el autor sea consciente de ello. Ello obliga a encontrar alguna justificación posible a la punición de estas conductas, y la única de las alternativas a la que los autores dan algo de crédito[369] tiene que ver con la idea de asumir obligaciones de cooperación entre participantes de actividades colectivas, concibiendo este tipo de normas como un esquema de protección recíproca; es decir, si cada uno de nosotros cumple

---

[368] Se hará referencia a los razonamientos contenidos en los siguientes trabajos: Simester, A. P., / Von Hirsch, A., «Remote harms and non-constitutive crimes», *Criminal Justice Ethics*, 28(1), 2009, pp. 89-107; y Von Hirsch, A., *op. cit*. N. 359.

[369] Concretamente, con referencia a Nils Jareborg, aunque en opinión de la autora otros autores como Duff han dejado apuntada una idea similar en alguna ocasión, al hilo de la necesidad de actuar como ciudadanos en condiciones de igualdad. Así, Duff, R. A., «Criminalizing endangerment», *op. cit*. N. 344, p. 962 «*a driver who claims to know that he can safely ignore such rules claims a certain superiority over his fellows: they must obey these rules, because they cannot be trusted to decide for themselves, but I need not. What is wrong with such a claim is not that it is false (though it often will be), but that it is a denial of fellowship with my fellow citizens: a recognition of fellow citizenship (and of the dangers involved in allowing exemptions to the law's demands) should motivate me to accept such laws even if I believe (truly) that they are unnecessary in my case. At least so long as the demands the law makes on me are not that onerous, this is a modest burden that I ought to accept as an implication (and expression) of citizenship».*

las normas (por ejemplo, las relativas al tráfico rodado), esto no solo protegerá al propio afectado por la norma, sino a todos los ciudadanos [370]. Con esta idea de aceptación comunitaria de protección recíproca se ofrece, por tanto, una de las justificaciones más convincentes de la siempre polémica figura de los delitos de peligro abstracto, permitiendo así afirmar la existencia del injusto («*wrong*») y, por tanto, la vigencia del *«harm principle».*

No es esta, sin embargo, una justificación pacífica y aceptada genéricamente, pues autores como Husak, al analizar la legitimidad de este tipo de delitos, y examinando el argumento de la «arrogancia cívica» que propone Duff, niega legitimidad a estos delitos, en tanto no encajan en su *«wrongful constraint»,* es decir, no tienen en esencial, una lesividad suficiente [371].

Por su parte, Von Hirsh [372], en un artículo anterior en el tiempo, se encarga de lo que él considera el trabajo pendiente en la obra de Feinberg, pues debe darse el salto de afirmar no solo que el resultado castigado constituye un genuino daño, sino que ese mismo daño es el que cabe «adscribir» o «imputar» a la persona de detrás, no en un sentido causal ni de conexión subjetiva, sino más bien en cuanto a su incumbencia (lo que, para la teoría dogmática continental podríamos situar aún en el plano de la imputación objetiva). Él bautiza a este problema con el concepto *«fair imputation»* [373], y al menos algunas de sus soluciones pueden resultar interesantes. La idea básica de Von Hirsh es que hay que atender a los roles asumidos por cada persona, habida cuenta del papel que desempeña en cada momento, destacando que los deberes específicos de personas especialistas o técnicas que se dedican a una determinada parcela de conocimiento pueden tener asignados mayores deberes y, por ello, justifica los delitos de peligro abstracto solo en casos de actividades altamente especializadas y reguladas. Atendiendo a este razonamiento, que el autor reconoce no ser

---

[370] *«the fact that the scheme provides reciprocal benefits to all participants suggest a prima facie reason why all should comply, even those of superior ability. All participants have an interest grounds responsibility in each of us for its compliance. The regime is a package deal».* Simester, A. P./Von Hirsh, A., *op. cit*. N. 368, p. 96.

[371] Para ver el desarrollo de este argumento de Husak, así como la crítica a los intentos de justificación de Duff o Green, *vid.* Husak D., *op. cit*. N.366., p. 108 y ss. Para profundizar en los argumentos de Duff, vid: Duff, A. R., «Crime, Prohibition, and Punishment» *19 Journal of Applied Philosophy,* 2002, pp. 97-102, 2002. Una vez más ocurre que Husak analiza la adecuación de los delitos de peligro como parte de un problema mayor, y las categorías que en su trabajo resultan problemáticas son tan amplias como las «*offences of risk creation*» o lo que él denomina «*overlapping crimes*» o «*ancilliaty offences*». Todos ellos tienen como denominador común su desatención al principio del daño y por ello él los agrupa bajo el mismo fenómeno, tratando de buscar soluciones comunes.

[372] Von Hirsch, A., *op. cit.,* N. 359.

[373] *Ibid,* p. 265.

mayoritario ni aceptado por otros teóricos de los delitos de peligro [374], se podría dar cabida y justificación a los delitos de peligro abstracto como delitos «especiales», de tal forma que no se podría justificar, bajo el criterio de Von Hirsh, la creación de delitos de peligro abstracto dirigidos a usuarios de tecnologías altamente autónomas, pero sí, en cambio, a los «profesionales» cuya actividad estuviera relacionada con dichas tecnologías. Por trasladarlo a nuestro escenario, un hipotético «delito de conductas peligrosas generadas por IA» solo sería aplicable a programadores y, en su caso, a propietarios con conocimientos en el sector, pero nunca a un usuario ajeno a las especificidades del sector tecnológico.

En definitiva, como los argumentos señalados ponen de manifiesto, la discusión en torno al principio de lesividad será esencial en la búsqueda de un modelo de imputación para los *hard AI crimes*, haciendo necesario revisitar los argumentos que los autores han puesto sobre la mesa en el pasado. De esta forma, las reflexiones aquí expuestas no deben interpretarse como un punto final a la discusión, sino como una manera de trasladar las preguntas clásicas a las nuevas realidades [375].

## 3.2 **Tendencia a la objetivación de la culpabilidad**

Otra de las constantes y razonables críticas a las que se someten este tipo de delitos de daño diluido o remoto es la acusación de ser cada vez más tendentes a exigir menos vinculación subjetiva. El problema de la objetivación de la responsabilidad –según la terminología continental– o, en términos del *common law*, la reducción de los estados mentales *(mens rea)* exigidos en los delitos hasta su mínima expresión, plantea dificultades en ambos sistemas jurídicos.

---

[374] Pone él mismo como ejemplo a Hassemer, que propugna directamente la eliminación de los delitos de peligro abstracto. *Vid.*: Hassemer, W., «Kennzeichen und Krisen des modernen Strafrechts», *Zeitschrift für Rechtspolitik*, 25, 1992, pp. 378-383. De la traslación de todos estos problemas en el ámbito continental, en profundidad, nos encargaremos en el capítulo sexto.

[375] Así, resultará básico, por ejemplo, establecer si la lógica de la actividad en la que intervienen actores variados necesita de una regulación previa y coordinada (explorando, para ello, la normativa que actualmente tenemos, que se estudiará *infra* en capítulo séptimo), que dará lugar a normas de conducta y fijará los contornos de la confianza y ese derecho a arrogarse potestad de autorregulación. La cuestión sobre a qué actores confiamos la regulación o fijación de los contornos de estos nuevos estándares de cuidado y esferas de confianza es, en este sentido, una de las más importantes, pues entronca directamente con el principio democrático (también en profundidad en capítulo séptimo). A este respecto, Beck, sobre la conformación de los estándares y los actores que deben entrar a crear este concepto: Beck, S.: «Robotics and Criminal Law», op. cit. N. 259, pp. 41-56.

Quizá no sea sorprendente encontrar, entre los autores anglosajones, un reconocimiento más naturalizado de la rebaja en el contexto de crímenes de IA, e, incluso, de la posibilidad de aceptar supuestos de responsabilidad objetiva. Así, es fácil constatar, con Calo [376], que la tendencia hacia la objetivación de la culpabilidad es evidente, y que hace tiempo los estándares de *mens rea* se reducen hasta casi diluir la separación entre lo objetivo y lo subjetivo, especialmente en el campo de la robótica, donde las cualidades de personificación y emergencia arriba señaladas pueden conllevar un rol más significativo de la responsabilidad objetiva en el derecho penal. Una idea similar transmite Pagallo [377] cuando afirma que la mayor impredecibilidad del riesgo en este sector conducirá a un mayor protagonismo de la responsabilidad objetiva.

Lo cierto es que algunos autores del ámbito continental y, más concretamente, español, han reconocido abiertamente que esta posibilidad no está tan alejada cuando se trata de crímenes cometidos por robots. Merece la pena, por ello, reproducir sus términos, pues, tras afirmar que la opción más viable por el momento sería acercarnos a un régimen parecido a la responsabilidad objetiva, propone Lledó Benito lo siguiente: *«tendríamos una presunción iuris tantum de responsabilidad criminal del titular del robot, que podría ser desvirtuada probando que el fallo del robot no se debió a su uso doloso o imprudente, sino a fallos en su diseño o montaje»* [378].

A pesar de que, como acaba de examinarse, no faltan las propuestas más tendentes a la objetivación, la inclinación mayoritaria que se intuye al repasar la literatura sobre IA es la negativa frontal a aceptar la responsabilidad objetiva

---

[376] Calo, R., *op. cit.* N 62, p. 554.

[377] Pagallo, U., «When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law», *Minds&Machines*, 27, 2017, pp. 625-638. En este trabajo Pagallo señala que la concepción del riesgo cambiará significativamente, y que la idea que acabará imperando será la de que la gente debe pensárselo dos veces antes de interactuar con robots. Todo el texto es extremadamente interesante pues en el mismo propone algunas soluciones imaginativas como la idea de que sea el propio robot el que llegue a instrumentalizar al humano o la posibilidad de crear una futura causa de exculpación o exención de la responsabilidad (una *«defence»)* del tipo *«robot loss of self-control»*.

[378] Lledó Benito, I., *El derecho penal, robots, IA y cibercriminalidad: desafíos éticos y jurídicos: ¿hacia una distopía?,* Dykinson, Madrid, 2022, pp. 94-95. El autor cita en este punto a otro autor español, al que se remite. Álvaro Pascual, D, «Inteligencia artificial: un panorama de algunos de sus desafíos éticos y jurídicos», Universitat de Girona, disponible en: <http://hdl.handle.net/10256/14950> [Accedido por última vez el 22/01/2026], pero parece que se realiza, en este punto, una mala interpretación, pues el último autor solo se refiere al régimen de responsabilidad objetiva contemplado para los robots en el ámbito civil. En todo caso, la figura que plantea introducir recuerda, en parte, a la «*due diligence defence*» del derecho anglosajón, sin que quepa, a nuestro juicio, hacer una traslación, sin mayores explicaciones o justificaciones de esta lógica.

como opción, pues, en palabras de Votjus *et al*[379], supondría que se tambalearan las bases de los sistemas continentales.

Si bien es indudable la relevancia central del principio de culpabilidad en los sistemas continentales, resulta pertinente señalar que, en tiempos recientes, se ha producido una cierta matización en su aplicación y alcance. Aunque no de forma tan abierta como en los ordenamientos de influencia angloamericana, la doctrina española reconoce, desde un tono eminentemente crítico, que la tendencia a la objetivación de la responsabilidad está presente en ordenamientos como el español y que sus manifestaciones son, de hecho, muy variadas. Corcoy Bidasolo[380] lo ejemplifica magistralmente en un reciente trabajo, donde afirma que, entre otros fenómenos, la proliferación de delitos de peligro ha llevado a objetivar el dolo, minando el principio de culpabilidad de una forma insoportable.

Llegamos, pues, a un punto común: la proliferación de delitos de peligro o daño remoto supone, para ambos sistemas, un reto que genera incomodidades dogmáticas. La doctrina anglosajona, que convive desde hace tiempo con las *«strict liability offences»* y que, además, han proliferado en los últimos años al hilo del desarrollo de actividades altamente peligrosas, ha desarrollado un valioso trabajo a la hora de conceptualizar y limitar esta expansión. Un análisis de esta cuestión puede aportar claridad sobre problemas comunes y, en el ámbito continental, obligarnos a considerar si esta tendencia nos es realmente ajena.

Una de las defensas comunes a este respecto pasa por afirmar la singularidad del ámbito propio de este tipo de delitos, donde, de alguna forma, puede llegar a justificarse la excepción a principios como el de culpabilidad. Así, Simester[381], al analizar los límites que cabe imponer a los delitos de este tipo, acaba abogando por una solución llamativa, que pasa por distinguir dos tipos de delitos y, por ende, dos regímenes de responsabilidad: en primer lugar, aquellos que forman parte del núcleo duro del derecho penal, cuyo estigma aparejado es

---

[379] Vojtuš, F., Kordík, M./Dražová, P., «Artificial Intelligence and the criminal responsibility-challenges, obstacles and possible solutions», en *2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA),* 2022, p. 666. «*This would therefore be the introduction of the concept of strict criminal liability for the unlawful conduct of artificial intelligence, which would be imposed on its creator or the entity that has placed it on the market or put it into use, or uses it. In this case, we believe that the adoption of such an approach, or the application of other models of criminal liability that have recently emerged in the Anglo-American legal environment, which do not rely on the examination of the subject's fault, 16 would mean a paradigm shift in the way criminal liability is viewed, especially in the continental (European) legal environment».*

[380] Corcoy Bidasolo, M., «Principio de culpabilidad …*op cit.* N 361, p. 568.

[381] Simester, A. P., «Is strict liability always wrong?», en Simester, A. (ed.), Appraising Strict Liability, Oxford University Press, 2005, pp. 21-50.

el tradicionalmente asociado a la censura en sentido estricto del derecho penal. En segundo lugar estarían los delitos menos estigmatizantes (él los identifica con las *«regulatory offences»,* aunque las categorías o denominaciones usadas por la doctrina son variadas [382]), que se corresponden a su vez con hechos menos estigmatizantes unidos a conductas que no han formado típicamente parte del derecho penal, en los que es posible rebajar el estándar de culpabilidad, pues en algunos casos los beneficios pueden superar a los perjuicios. Simester realiza un recorrido por los argumentos a favor y en contra de este tipo de responsabilidad, para finalmente mostrar que quizá sí haya que reconocer un lugar legítimo a este tipo de delitos en un ámbito muy concreto (relacionado con el desarrollo de actividades que supongan la creación de riesgos, es decir, en contextos típicamente empresariales, donde ya el ordenamiento da cabida a inversiones de la carga de la prueba), pero solo acudiendo a castigos que no supongan, en ningún caso, un estigma en sentido fuerte ni tampoco un trato desproporcionado, situando la multa como la mejor candidata a estos propósitos expresivos y disuasivos, donde se pone el foco en la disuasión o como parte de un régimen que sirve para distribuir el coste de una actividad, emprendida voluntariamente, que beneficia a una persona y que normalmente impone riesgos a otras que no se han expuesto voluntariamente a dicha actividad arriesgada.

Esta distinción entre dos grupos de delitos, con menor exigencia legitimadora en uno de ellos, no resulta extraña del todo en el ámbito del derecho penal español, donde autores como Silva Sánchez [383] han señalado, al hilo de la expansión del derecho penal en la sociedad del riesgo, que los delitos de peligro deben ser tratados como una especie de «grupo diferenciado», de forma que el derecho penal contemporáneo se configure como un «derecho penal de dos velocidades», donde la necesaria regularización de sectores peligrosos en los que los principios clásicos se ven comprometidos, haya una relajación de los principios, siempre y cuando la pena a imponer no cause un efecto estigmatizante fuerte (excluyendo, desde luego, la pena de prisión). Quizá nuestros ordenamientos, sus problemas, soluciones y justificaciones no estén, en este punto, tan alejados. Cabe señalar que estas propuestas justificadoras no han estado, sin embargo, exentas de críticas [384].

---

[382] *Ibid*, pp. 23-24. Habla de otras categorías como «*public welfare offences*», «*mala prohibita*» o, su denominación predilecta, «*cuasi-criminal regulations*», como oposición a lo que denomina «*stigmatic crimes*».

[383] Silva Sánchez, J. M., *La expansión del Derecho penal. Aspectos de la política criminal en las sociedades postindustriales*, 3.ª ed. Edisofer, Madrid, 2011.

[384] Todas ellas se desarrollan *infra*. Por ahora, y para una visión general: en el ámbito alemán, por excelencia, Hassemer. *Vid.:* Hassemer, W., «Viejo y nuevo derecho penal», en Hassemer, W., *Persona, mundo y responsabilidad*, Tirant lo Blanch, Valencia, 1999. En el ámbito español, sin ánimo de exhausti-

Otros autores han entendido justificada una parcela restringida a estos delitos. Así, Duff [385] se ha mostrado partidario de aceptar este tipo de delitos solo en supuestos muy restringidos, siendo precisamente uno de ellos las actividades que crean un peligro serio y que son beneficiosas socialmente, pero opcionales. Pone el ejemplo de la persona que vende comida en mal estado o el empleador que responde por la seguridad de sus trabajadores en el lugar de trabajo. En este sentido alude a un riguroso deber de cuidado.

Resulta útil tener en mente que cuando se habla de los problemas asociados a la responsabilidad objetiva, un concepto amplio puede llevarnos a problemas que en absoluto son ajenos a nuestros ordenamientos. Green [386] expone de forma magistral los diferentes escenarios posibles, lo que resulta de enorme utilidad a efectos comparativos En esencia, los inconvenientes que se señalan para algunos de los tipos de «responsabilidad objetiva» son los mismos problemas con los que el ordenamiento español viene lidiando.

Dentro de esta tipología de responsabilidad objetiva en sentido amplio, Green habla de:

– Delitos donde al menos un elemento material de la definición del delito no requiere el correspondiente requisito de *mens rea* (las que, según este autor serían las únicas en caer en esta categoría estricta).

– Delitos donde se niega la posibilidad de alegar una «*defence*» (lo que nuestro sistema continental sería una causa de justificación o exoneración de la culpabilidad). Resulta este problema familiar al caso español, específicamente respecto de un problema ya apuntado por Corcoy [387]. Abusando de la aplicación de otras instituciones (señaladamente la institución de la *«actio libera in causa»),* se niega sistemáticamente la posibilidad de aplicación de una causa de exclusión de la responsabilidad (siendo la eximente de estado de intoxicación el ejemplo más claro) para afirmar la presencia de los elementos subjetivos en un momento anterior al de la acción.

– La tendencia a presumir ciertas intenciones en el acusado (una vez más, situación familiar que se produce con la interpretación del dolo cada vez más objetivado en el ordenamiento español).

---

vidad, Díez Ripollés, J. L., «De la sociedad del riesgo a la seguridad ciudadana: Un debate desenfocado», *Nuevo Foro penal*, n.º 69, junio 2006, pp. 199-248; Mendoza Buergo, B., *El derecho penal en la sociedad del riesgo*, Civitas, Madrid, 2001.

[385] Duff, A. R., «Strict Responsibility, Moral and Criminal», *Journal of Value Inquiry*, 43, 2009, pp. 311 y ss.

[386] Green, S. P., «Six senses of strict liability: a plea for formalism», en Simester, A. (ed.), Appraising Strict Liability*,* Oxford University Press, 2005.

[387] Corcoy Bidasolo, M., «Principio de culpabilidad… *op. cit*. N. 361, p. 576.

– La importancia cada vez mayor de una tipología de *mens rea* mucho más flexible o menos exigente: paradigmáticamente, la negligencia en el ámbito inglés, reflejo de nuestra imprudencia, que se extiende de forma innegable en todo el catálogo delictivo.

– La evidente pérdida de importancia del principio del daño en la tipificación de delitos *(«harmfulness»);* lo cual se une directamente con la primera de las críticas vertidas en el apartado 3.1.

En definitiva, los problemas con los que lidia la doctrina anglosajona más crítica con estas figuras no se encuentran, en absoluto, alejados de nuestras preocupaciones y, en esencia, las soluciones planteadas son bastante parecidas.

Para concluir con este necesario enfoque comparativo resulta clarificador leer a Spencer y Pedain [388], pues tratan el problema de la responsabilidad objetiva desde un plano dogmático más amplio, con la intención de situarlo debidamente en los sistemas continentales. Así, reconocen ambos autores, los problemas de la responsabilidad objetiva tienen un reflejo (específicamente hablando del sistema alemán, pero con argumentos extrapolables al español) en varias instituciones: en el esquema de las *«Ordungswidrigkeiten»,* que asimilan a las *«regulatory offences»* inglesas*,* que abarcan estándares de culpabilidad muy debilitados, pero que conllevan penas con estigma social moderado; en los delitos cualificados por el resultado *(«erfolgsqualifizierte Delikte»),* donde, si el resultado se produce, aunque el elemento subjetivo sea débil, la pena será considerable; o en las llamadas condiciones objetivas de punibilidad, sobre las que no es necesario que concurra elemento subjetivo alguno. Otros sistemas continentales, señalan, como el francés, prevén, de hecho, un cierto régimen de responsabilidad objetiva en el que se afirma, incluso, la inversión de la carga de la prueba (las llamadas *«contraventions»).* Los autores afirman que tanto en los sistemas continentales como en los de *common law* se viene aceptando la creación de delitos no estigmatizantes con pequeñas penas, pero la cuestión a determinar es si estos hechos son los suficientemente «poco importantes» como para justificar una excepción a la regla general [389].

Resulta útil, pues, unir las discusiones en torno al debilitamiento del principio de culpabilidad con la discusión sobre los límites de la criminalización, y esta es una lógica que debe aplicarse a todos los preceptos que, aunque en teoría sí exigen todos los elementos del *mens rea*, en realidad se sirven de ins-

[388] Spencer, J. R.,/ Pedain, A., «Approaches to strict and constructive liability in continental criminal law», en Simester, A. (ed.), Appraising Strict Liability*,* Oxford University Press, 2005, pp. 237-283.

[389] *Ibid.* p. 276.

trumentos propios de la responsabilidad objetiva. Terminan los autores con una reflexión interesante, y directamente extrapolable al panorama español:

*«A pesar de –o tal vez debido a– su inquebrantable compromiso teórico y práctico con el principio de culpabilidad, el Derecho alemán responde cada vez más a la tentación de allanar el camino a la condena, no suprimiendo el requisito de la culpa, sino tipificando como delito conductas que no son en sí mismas lesivas, pero que a menudo conducen al daño. De este modo, intenta eludir las cargas de la prueba de la culpa en relación con la causación de un resultado lesivo, manteniendo a primera vista la pureza dogmática»* [390]. Lo cierto es que dicha pureza dogmática no existe ya ni en el panorama alemán ni desde luego en el español, donde nuestros problemas no están en absoluto alejados de la realidad anglosajona. Quizá debiéramos ser mucho más conscientes de estas realidades cuando nos propongamos examinar los problemas y límites de la responsabilidad objetiva para tratar de buscar un modelo óptimo para regular los *hard AI crimes*.

### 3.3 **El problema de la pena y el llamado *«retribution gap»***

El último de los problemas que se deriva de la posibilidad de crear una ofensa de «pequeño calado», que capture un hecho base de «pequeña lesividad» acompañado de un «bajo estándar de culpabilidad» es el relativo a la concreción de la pena resultante a imponer.

Es este un fenómeno especialmente preocupante en el ámbito de la IA, pues cada vez más en este ámbito la sociedad demanda castigos ejemplarizantes, suficientes y capaces de mostrar el rechazo tajante que la sociedad expresa ante la posibilidad de que, por el uso de una herramienta tecnológica, y bajo la excusa de la opacidad y la falta de previsibilidad, hechos de gravedad significativa queden sin castigo o con un castigo menor. Este problema es descrito por Danaher [391] cuando trata el problema del *«retribution gap»* en el ámbito tecnológico. La idea de partida es que con el castigo penal se busca una sensación de alivio para la sociedad, y una función expresiva del reproche, pero que, precisamente con la interacción de agentes artificiales en las acciones, la necesidad de este tipo de castigo se presenta con algunas características especiales, precisamente por la entrada en el terreno de juego de estos entes extraños.

---

[390] *Ibid*, p. 282 (traducción propia).
[391] Danaher, J., *op. cit.* N.43, pp. 299-309.

El argumento de Danaher, que en su interpretación mayoritaria nos sitúa en el escenario de la laguna de punición, puede presentar una mayor capacidad de rendimiento si se explora en otro sentido: no es solo que la sociedad desconozca a quién imponer la pena, sino que debemos preocuparnos esencialmente sobre cómo ha de ser esa pena.

Así, el tercero de los problemas que plantean los *hard AI crimes* (común a los delitos de peligro) es conseguir una pena que no cause una distorsión. Danaher nos habla de la distorsión entre aquello que la sociedad demanda y lo que los estándares de imputación de responsabilidad normalmente utilizados requieren. Debemos estar atentos a que, con las nuevas formas de tipificar o reaccionar ante estos delitos, no acabemos señalando a «chivos expiatorios morales» [392] más que a verdaderos agentes responsables.

Una vez convenido que la responsabilidad debe asignarse en todo caso al agente humano, la ulterior cuestión a resolver será, necesariamente, en qué tipo de pena se traduce esa responsabilidad. Si no queremos dejar sin castigo conductas en las que el daño es remoto (y decidimos, por tanto, salvar esta tipificación acudiendo a la tipificación de delitos de peligro), el siguiente paso será decidir qué tipo de pena encaja, siendo necesaria, suficiente, justa y oportuna en un contexto tan singular como el presente. Pone de manifiesto este problema McCallister [393] cuando apunta que el problema es que, si la *«mens rea»* no es completamente abandonada, sino que se opta por rebajar su límite, entonces el castigo, por razones de equilibrio, será demasiado ligero para compensar a la víctima (aunque se podría incluir aquí tanto a la sociedad como a las necesidades de estabilización de la norma), y a su vez desproporcionado (ante la duda de si realmente hubo culpa por parte del acusado). Corremos el riesgo, por lo tanto, de que la pena sea a la vez injusta por exceso y por defecto.

Sin duda, el problema de la pena desproporcionadamente alta para un comportamiento comparativamente poco reprochable es, como se ha visto en el apartado anterior, un problema familiar en los delitos que hemos venido a llamar de «daño remoto», así como en la esfera de las *strict liability offences* (entendidas en sentido amplio). Se pueden señalar, por tanto, los argumentos que los autores han puesto de manifiesto a este respecto. Este problema puede unirse, además, con algunos de los problemas que ya han surgido al hilo de los casos de punición de daños remotos, como pueden ser los delitos de peligro o las ya nombradas *constructive liability offences*. Aborda este problema de forma muy lúcida Clark-

[392] *Ibid.,* p. 20.
[393] McAllister, A., *op. cit.* N. 43.

sonn [394], pues señala ese desajuste entre grado de culpabilidad y pena impuesta, que acaba dependiendo del resultado. Autores como el propio Danaher [395], Ashworth [396] o Husak [397] han advertido la conveniencia de acudir al principio de proporcionalidad como límite a las penas no solo desproporcionadas, sino que corren el riesgo de acabar resultando un verdadero trato injusto.

La idea a extraer sería la siguiente: no basta con que se supere la barrera de emprender una actividad de alguna forma indeterminada arriesgada o peligrosa, sino que es necesario que ese concreto peligro derive de forma natural de la actuación desarrollada. Sólo así podrá mantenerse la coherencia entre el injusto y la pena. De lo contrario, se obtendrá una pena desproporcionada y un estigma innecesario. Esta idea conecta directamente con la anterior: si entendemos que la solución pasa por tipificar conductas meramente imprudentes (con un bajo estándar de *mens rea)* por hechos que luego tienen unas consecuencias mucho mayores e imprevisibles, y aunque haya en la sociedad una tendencia innegable a buscar estos «chivos expiatorios» que respondan con una pena con significado expresivo potente, si nos atenemos a principios como el de proporcionalidad, la única pena posible sería la proporcionada a la culpabilidad y, por ello, una con una capacidad estigmatizante limitada.

Esto supone, ciertamente, poner de manifiesto una tensión que deja entrever la eterna discusión sobre cuáles deben ser los fines del castigo y cuál debe ser la finalidad específica de un nuevo delito como el que aquí se propone. No por redundante resulta esta discusión menos oportuna. En definitiva, como señala Farmer con cita a Williams, a la hora de determinar los estándares de conducta *«la clave para el análisis no es empezar por la pregunta de si está justificado utilizar la negligencia, sino empezar considerando la actividad que va a ser regulada y el tipo de regla que es apropiada en función del fin que se busca»* [398]. La pregunta por la finalidad perseguida con la regulación de la actividad y el sector concreto en cada caso es una cuestión previa y necesaria para abordar esta problemática de una forma completa.

---

[394] Clarkson, C., «Aggravated endangerment offences», *Current Legal Problems*, *60* (1), 2007, p. 278. De hecho, su exposición se basa en la criminalización bajo la legislación penal australiana de una ley que permitía imponer penas de prisión de hasta 20 años por un acto de conducción negligente con resultado de muerte, pudiendo ser el acto negligente la conducción con una licencia caducada (pp.291 y ss.).

[395] Danaher, J., *op. cit.* N.43, p. 14.

[396] Ashworth, A., «Should strict criminal liability be removed from all imprisonable offences?» *Irish Jurist,* 45, 2010*,* p. 18.

[397] Husak, D., *op. cit*. N.366, capítulo 1, pp. 3-54, *passim.*

[398] Farmer, L., *Making the modern criminal law: Criminalization and civil order*, Oxford University Press, 2016, p. 185. *Vid.:* Williams, G., *Criminal Law. The General Part,* London, Stevens & Sons, 1953.

En definitiva, los tres problemas aquí apuntados se acaban entremezclando, y en esencia nos conducen a preguntas que el derecho penal ha tenido que enfrentar en tiempos pasados, y en escenarios que tienen elementos en común con el que aquí nos ocupa.

Solo teniendo en cuenta los problemas y argumentos que hasta ahora se han puesto de manifiesto podremos comenzar a analizar de qué forma estos pueden reproducirse en los nuevos contextos de aplicación.

En definitiva, y cerrando esta primera parte de la investigación dedicada a estudiar el estado de la cuestión, ha quedado establecido cuál es el concepto de IA que se maneja (con la especial importancia como elemento diferencial de la bautizada como «autonomía operativa»), cuáles son las lagunas o necesidades del derecho penal ante esta nueva forma de comisión de delitos (tensión entre imprevisibilidad y control), y cuál es el modelo que se propone como adecuado, (el de los delitos de peligro), planteando a la vez los tres grandes problemas que deben resolverse para aplicarlo: preservar los principios de lesividad, culpabilidad y proporcionalidad de la sanción.

A la resolución de estos problemas y al desarrollo de este modelo se dedicarán los siguientes capítulos.

# CAPÍTULO CUARTO. UNA REVISIÓN DE LOS PARADIGMAS DEL DERECHO PENAL DEL RIESGO

*«Diziam os argonautas que navegar é preciso, mas que viver não é preciso. Argonautas, nós, da sensibilidade doentia, digamos que sentir é preciso, mas que não é preciso viver»* [399]

Fernando Pessoa tomó como base de sus versos la conocidísima máxima latina «*Navigare necesse est; vivere non est necesse*», a la que da un vuelco poético. Si los antiguos romanos ya usaban en época de Plutarco [400] esta arenga sobre la inevitabilidad de los riesgos inherentes a la vida, el poeta luso da un paso más, afirmando que lo necesario no es solo vivir, sino sentir, y que, sin esa sensibilidad tan nuestra, tal vez la vida no merezca la pena ser vivida. Así, una

[399] Cita célebre del poeta luso Fernando Pessoa. En español: «*Decían los argonautas que navegar es preciso, pero que no es preciso vivir. Argonautas, nosotros, de la sensibilidad enfermiza, digamos que sentir es preciso, pero que no es preciso vivir*». Pessoa, F., 450, «Capítulo sobre la indiferencia o algo parecido», *Libro del desasosiego*, Trad. Por Ángel Crespo, Seix Barral, Barcelona,1984, p. 252.

[400] Las referencias a la máxima latina se han asociado con diferentes orígenes, aunque quizá el más común se encuentra en los escritos de Plutarco, historiador que narra cómo el aforismo fue gritado por Pompeyo el Grande en el siglo I a sus hombres, instándoles a zarpar a pesar de una tormenta muy violenta en el mar. Pessoa la atribuye, sin embargo, a la leyenda de Jasón y los Argonautas, que vuelve a la vida cuando los exploradores portugueses de los siglos XV y XVI la toman como inspiración para navegar a pesar del peligro, buscar aventuras y enfrentarse a lo desconocido. Bartholomew, R., «'Navegar é preciso; viver não é preciso': The Impossible Journeys of Kierkegaard and Pessoa», en Falcato, A., / Cardiello, A. (Eds.), *Philosophy in the Condition of Modernism*, Palgrave Macmillan, Cham, 2018, pp. 385-414, p. 386. No obstante, también se han encontrado asociaciones de este lema con la llamada «Liga Hanseática», una confederación de gran éxito de ciudades mercantiles y gremios alemanes, que controlaba el báltico a finales de la Edad Media. Así, Hilgendorf, con cita a su vez a Binding, K., *Die Normen und ihre Übertretung, Eine Untersuchung über die rechtmäßige Handlung und die Arten des Delikts*, vol. 4: Die Fahrlässigkeit, Scientia, Aalen, 1965, 1919, p. 437. *Vid.*: Hilgendorf, E., «The dilemma of autonomous driving…, *op. cit.* N. 220, p. 80.

vida sin riesgos, o una vida sin sentimientos, ambos como componentes inherentes al proceso mismo de vivir, aparecen como carentes de significado.

Esta es la idea que vertebra las reflexiones de las que parte este capítulo, donde el riesgo aparece como inevitable en una sociedad que, alejada como está de los pilares que la sostenían en la época del imperio romano, sigue planteando esa tensión inevitable para la existencia humana. En este binomio (o trinomio, para quien se adhiera al concepto de Pessoa), riesgo, sentimiento y vida, son, en todo caso, ineludibles.

Se analizará cómo esta inexorabilidad del riesgo (cualquiera que sea su origen) es un elemento clave para abordar los peligros que las viejas y las nuevas técnicas generan en el conglomerado social, haciendo un repaso por el surgimiento y reformulación de los postulados del llamado derecho penal del riesgo, y sus manifestaciones más palmarias (respectivamente, el principio de precaución, la responsabilidad penal por el producto, y los delitos de peligro), para establecer si este discurso y estos métodos sirven a la sociedad argonauta de hoy.

## I. UNA ACTUALIZACIÓN DE LA MODERNA SOCIEDAD DEL RIESGO

### 1. LA TRADICIONAL DISCUSIÓN SOBRE LA MODERNIDAD Y LA SOCIEDAD DEL RIESGO

Aunque recorrer todo el camino histórico que dio lugar al discurso sobre la modernidad y el riesgo podría ser objeto de una investigación en su totalidad, nos parece adecuado traer aquí a colación la discusión que nace en la década de los 70 en Alemania y penetra en España en los 90, para trazar una breve línea temporal de continuidad con el discurso actual, relacionado con la expansión y la modernización, y explorar cuáles son esos elementos comunes que pueden servir como bases de entendimiento de una situación sociológica como la actual, y que son esenciales para comprender los debates político-criminales de nuestro tiempo.

Si los accidentes catastróficos de la producción en masa que marcaron un antes y un después en el momento de «la expansión» fueron los protagonistas de la discusión dogmática y político-criminal, hoy lo son los ejemplos de tecnologías sobre las que nuestro control disminuye y que amenazan con trastocar nuestras formas de concebir el delito (y su castigo).

A estos efectos, para situarnos, puede servir la introducción que el profesor Arroyo Zapatero realizaba al inicio de la obra colectiva *Crítica y justifica-*

*ción del derecho penal ante el cambio de siglo,* que recopila una fuente valiosa de trabajos de expertos en ese viraje del derecho penal, señalando: «*el momento del calendario coincide con un momento material, una situación del derecho penal cuya sostenibilidad se discute, sobre todo desde Frankfurt, lugar donde incluso se combate ferozmente el Derecho penal actual, al que se descalifica como «moderno». Efectivamente, la situación del Derecho penal es prerrevolucionaria respecto de un Derecho penal clásico, si por tal entendemos el Derecho penal acuñado durante el siglo XX, en especial tras la segunda guerra mundial y hasta finales de los 80*» [401].

No fue el cambio de siglo, sin embargo, el momento en que el discursó comenzó a cambiar, y aunque, como suele ocurrir con todos los procesos de evolución social y política, es difícil situarnos en un concreto momento temporal, sí es posible apuntar que hubo un caso primigenio que lo cambió todo, por lo catastrófico de sus efectos y por el novedoso debate jurídico al que dio lugar. Nos referimos al primero de los *leading cases* de la modernidad penal, el conocido caso *Contergan*. En palabras de Cancio Meliá comienza con este caso el Derecho penal del riesgo, alcanzándose «un punto de inflexión en el funcionamiento del sistema penal» [402]. Con él comenzamos, pues, nuestra línea temporal, haciendo un viaje en el tiempo que nos retrotrae a la década de los 60. Nos centraremos en este capítulo en el análisis de la perspectiva sociológica y sus efectos en el discurso político-criminal, (a los hechos), para adentrarnos más adelante (en el capítulo V), en los problemas jurídicos y dogmáticos que todos estos casos suponen.

## 1.1 **La década de los 60: un caso que lo cambió todo**

El medicamento *Contergan*, cuyo principio activo era la talidomida, desarrollado por la farmacéutica *Chemie-Grünenthal*, fue comercializado en la República Federal Alemana entre 1960 y 1962. Este medicamento, disponible sin receta médica, era recomendado por sus efectos sedantes, especialmente a mujeres embarazadas. Los ensayos clínicos pertinentes se habían realizado, pero no en este tipo de pacientes. Ya en 1959 un neurólogo informó a la mercantil sobre los posibles efectos neurotóxicos de la talidomida. Hasta 3000

[401] Arroyo Zapatero, L., «Presentación», en Arroyo Zapatero, L./Neumann, U., /Nieto Martín, A., (Coords.), *Crítica y justificación del derecho penal en el cambio de siglo. El análisis crítico de la Escuela de Frankfurt*, Ediciones de la Universidad de Castilla-La Mancha, Cuenca, 2003, p. 18.

[402] Cancio Meliá, M., «Caso Contergan», en Sánchez-Ostiz Gutiérrez, P. (Coord.), «Casos que hicieron doctrina en derecho penal», *La Ley,* Madrid, 2011, p. 210.

personas denunciaron sus efectos tóxicos durante la etapa de comercialización. La farmacéutica realiza en este tiempo algunos estudios clínicos adicionales en ratas, procede a introducir una advertencia en el prospecto sobre los posibles efectos secundarios y solicita que se suministre bajo previa prescripción médica. La comercialización continúa y los casos de malformaciones se siguen sucediendo, comenzando a hacerse eco en los medios de comunicación. Dos años más tarde, en 1961, varios informes médicos y científicos vinculan el uso de la talidomida con malformaciones congénitas graves tras su uso durante el embarazo. Pese a una reticencia inicial de la empresa, la retirada del mercado sobrevino en noviembre de 1961, tras las crecientes presiones y eco en los medios. En el mes de diciembre se iniciaron las investigaciones judiciales contra *Grünenthal.* Las malformaciones dentro de la República Federal de Alemania ascendieron a entre 4000 y 5000 casos, de los cuales llegaron a sobrevivir aproximadamente 2800 [403].

Al caso *Contergan* siguieron muchos otros, paradigmáticos también de los problemas que, apenas un par de décadas más adelante, ocuparían la discusión dogmática (así, el caso *Lederspray*, también en Alemania o el caso del aceite de colza, en España, analizados *infra),* pero lo que este primer caso supone es, a nuestro modo de ver, el inicio de un cambio de paradigma. Compartimos, en este sentido, el análisis que Cancio realiza al respecto, afirmando que el caso Contergan supuso todo un mar de dudas en aquellos sectores en los que la sociedad industrial actuaba sin dominar el fondo de las fuerzas que desataban su actividad, dando a entender que algunos de los principios fundamentales del ordenamiento jurídico penal vigente en aquel momento habían llegado a su límite (o, al menos, añadiríamos, estaban necesitados de alguna adaptación). Señala Cancio que a partir de entonces solo queda un camino, si es que entendemos que este sector de actividad de riesgo debe ser objeto del proceso penal (como nadie puso en duda): aprehender estos riesgos «en la fuente». Se inicia aquí, afirma Cancio, la era de los delitos de peligro abstracto [404]. La década de los 60, época de la contracultura y el rupturismo, toca a su fin, y con ella el discurso sobre el derecho penal cambiaría irremediablemente, iniciando su propia época de revolución y revisionismo.

---

[403] Se trata de una versión resumida de los hechos narrados por Cancio Meliá, M., *op cit*. N 402, pp. 208-209.

[404] *Idem,* p. 222. Toma Cancio la referencia a la necesidad de aprehender los riesgos en «la fuente» de Armin Kaufmann, que estuvo implicado en los hechos del caso, en calidad de perito de una de las defensas. Un resumen de la postura del alemán en Kaufmann, A., «Tatbestandsmäßigkeit und Verursachung im Contergan-Verfahren: Folgerungen für das geltende Recht und für die Gesetzgebung», *JZ*, 1971, pp. 569-576.

## 1.2 **La década de los 70: los orígenes de la Escuela de Frankfurt** [405]

Debemos hablar aquí de la figura de Winfried Hassemer y sus discípulos de la Escuela de Frankfurt, situando sus incipientes inicios en su obra *Theorie und Soziologie des Verbrechens* [406], de 1973, donde el profesor alemán expone por primera vez su postura, marcando el inicio de su reflexión crítica sobre el derecho penal desde una perspectiva sociológica [407].

En esencia, y aunque los postulados de esta escuela a la que se suman autores como Herzog o Prittwitz se van desarrollando en los años posteriores, ya podemos intuir una línea de pensamiento que se opone a lo que se comienza a denominar, de forma despectiva, el derecho penal moderno (por contraposición al clásico derecho penal del paradigma de la Ilustración). Con la creación de nuevos tipos penales destinados a proteger nuevos bienes jurídicos de contenido indeterminado, se está produciendo una crisis debida a la expansión incontrolada y a una sobrecarga del derecho penal, siendo necesario reconducir su expansión a un derecho penal nuclear mínimo [408].

Aunque las obras más significativas de Hassemer son algo posteriores [409], se dejarán apuntadas las principales ideas, que, siendo concebidas ya en los años 70, permean en el pensamiento de sus discípulos germanos y que acaban irrumpiendo, inevitablemente, en el panorama español. El derecho penal moderno es concebido por Hassemer como un derecho que viene a imponer una

---

[405] La bautizada como Escuela de Frankfurt tiene, en realidad, un origen anterior, pues surgió en los años 20 con la fundación del Instituto de Investigación Social en la Universidad de Frankfurt, gracias al apoyo de Felix Weil. Sin embargo, se debe hacer una matización sobre el nombre, que, aunque común ya en la literatura, reúne a personas con muy distintas concepciones. Así lo explica Feijoo en: Feijoo Sánchez, B., *Normativización del derecho penal y realidad social,* Univ. del Externado, Bogotá, 2007, nota al pie 100. El mismo contenido del capítulo II de este libro puede encontrarse en Feijoo Sánchez, B., «Sobre la «administrativización» del Derecho penal en la «sociedad de riesgo»: un apunte sobre la política criminal a principios del siglo xxi», en AA.VV., *Derecho y justicia penal en el siglo XXI: liber amicorum en homenaje al profesor Antonio González-Cuéllar García,* Colex, 2006, nota al pie 11. En adelante, las citas que se hagan a este trabajo se referirán a la primera de las versiones, es decir, la contenida en la monografía *Normativización del derecho penal...*

[406] La referencia completa de la obra es: Hassemer, W., *Theorie und Soziologie des Verbrechens: Ansätze zu einer praxisorientierten Rechtsgutslehre,* Scriptor-Verlag, Kronberg im Taunus, 1973.

[407] La corriente de Hassemer y sus discípulos ha sido calificada de «individualismo monista». Schünemann, B., «Consideraciones críticas sobre la situación espiritual de la ciencia jurídico-penal alemana», *ADPCP*, 49, 1996, p. 190.

[408] *Idem*., p. 191, resumiendo los postulados de la Escuela de Frankfurt.

[409] Principalmente, como paradigmática en este sentido: Hassemer, W., «Kennzeichen und Krisen des modernen Strafrechts», *op. cit.* N. 374 publicada más tarde en español en Hassemer, W., «Rasgos y crisis del Derecho penal moderno», *ADPCP*, 45 (1), pp. 235-250, y reeditada en su obra colaborativa con Muñoz Conde. Así, Hassemer, W./Muñoz Conde, F., *La responsabilidad por el producto en derecho penal*, Tirant Lo Blanch, Valencia, 1995, cuya primera parte («Viejo y nuevo derecho penal»), es, de hecho, una versión reelaborada por Muñoz Conde precisamente de esta obra.

lógica dialéctica frente al «clásico» derecho penal basado en la idea del contrato social. El bien jurídico pasa de ser un elemento negativo a un elemento que pide el castigo, pasando el principio de bien jurídico a obligar a recurrir a la amenaza penal, *«convirtiendo dolorosamente la «prohibición de exceso» en una «prohibición de defecto»* [410]. El segundo de los rasgos de esta modernidad es la exacerbación de la prevención y una orientación a las consecuencias, incluso pretendiendo una especie de pedagogía y sensibilización de la sociedad con determinados delitos, sin que importe tanto la adecuación de la intervención (usa como ejemplos los delitos medioambientales o la discriminación de la mujer). Considera los delitos de peligro abstracto como «la forma más representativa del moderno Derecho penal» [411], donde se amplía enormemente su ámbito de aplicación y donde, al prescindirse del perjuicio, se prescinde también de la necesidad de demostrar la causalidad. Basta con probar la realización de la acción incriminada, cuya peligrosidad no tiene que ser verificada por el juez. La solución concreta que proponen Hassemer y Muñoz Conde es diferenciar entre el derecho penal básico y un «Derecho de intervención» [412]. En cuanto al derecho penal básico, no impide tener en cuenta bienes jurídicos universales, pero deben ser precisados de la mejor manera posible y «funcionalizado» desde el punto de vista de los bienes jurídicos individuales. En cuanto al derecho de intervención, este *«estaría ubicado entre el Derecho penal y el Derecho sancionatorio administrativo, entre el Derecho civil y el Derecho público, con un nivel de garantías y formalidades procesales inferior al del Derecho penal, pero también con menos intensidad en las sanciones que pudieran imponerse a los individuos»* [413].

Nos parece sumamente interesante la crítica que Schünemann lanza a estos postulados, en lo que tiene que ver con la errónea interpretación sociológica que Hassemer realiza de la sociedad que él llama moderna, pues pone el énfasis en uno de los elementos que consideramos esenciales en el análisis político-criminal, que tiene que ver con el estudio de las estructuras de producción. Afirma que la evolución cultural se ha alejado de la idea de personalidad como felicidad suprema, siendo esta imagen más propia de los postulados culturales del siglo XIX, habiendo superado ya, en nuestra actual cultura de masas de la sociedad postmoderna postindustrial, este paradigma. En esta sociedad, continúa, todo está prefabricado y producido, generando un consumo barato, la moda se calcula en un sentido exclusivamente económico, donde todo se

[410] HASSEMER, W./MUÑOZ CONDE, F. *La responsabilidad por el producto..., op. cit.* N. 409, p. 23.
[411] *Idem*., p. 29.
[412] *Idem*., p. 41. y ss.
[413] *Idem*., p. 46.

transforma en entretenimiento y el desarrollo de la personalidad pierde su sentido. En definitiva, el individuo, que vive en una sociedad de masas, *«es incapaz de ver la génesis manipuladora e industrial de su forma de vida sólo aparentemente individual»* [414]*, afirmando con solemnidad que «la teoría personal del bien jurídico ha caído en la trampa de esta sociedad postmoderna»* [415].

Sin que sea necesario compartir este pensamiento pesimista que afirma un «pseudoindividualismo fabricado industrialmente» [416], lo que resulta claro es que los postulados de la Escuela de Frankfurt, que no hacen más que iniciarse en los 70, parecen no reflejar, a nuestro juicio, una imagen actualizada o realista del individuo, que no vive aislado en un contexto asocial o vacuo, sino cuyo modo de vida, y por tanto cuyas amenazas a su modo de vida, van a verse afectadas por la evolución de la sociedad en la que se inserta. La continuidad de los conceptos dogmáticos debe verse, sin duda, condicionada por la forma en que el individuo se relaciona con el mundo.

### 1.3 **La década de los 80: la verdadera revolución sociológica del riesgo**

En 1984, el sociólogo americano Charles Perrow publica su libro *Normal Accidents* [417], donde expone su teoría sobre los inevitables riesgos de la alta tecnología. La idea fuerza de la obra es que a medida que el tamaño de los sistemas va creciendo, y el número de funciones atribuidas a ellos se multiplica, los ambientes serán más hostiles. Cuanto más se ensamblen los sistemas entre sí, más inabarcables e inesperadas serán las interacciones que entre ellos surjan [418]. El autor va recorriendo diferentes sectores donde dichos accidentes se producirán de forma inevitable, partiendo del ejemplo del accidente de la central nuclear de *Three Mile Island* en 1979, y hablando de los riesgos que se derivarán de la energía nuclear, la industria química, la aviación, los accidentes de tráfico marítimo y terrestre, así como en lo que él denomina «*exotics*» [419]:

---

[414] Schünemann, B., «Consideraciones críticas...», *op cit*. N. 407, p. 195.

[415] *Idem*.

[416] *Idem*.

[417] Perrow, C., *Normal Accidents: Living with High-Risk Technologies*, Basic Books, Harper Collins Publisher, 1984.

[418] Esta es una idea recogida y reconocida, a su vez, por Herzog, con referencia explícita a la obra de Perrow. Así, Herzog, F., «Sociedad del riesgo, derecho penal del riesgo, regulación del riesgo», en Arroyo Zapatero, L./Neumann, U., /Nieto Martín, A., (Coords.), *Crítica y justificación del derecho penal en el cambio de siglo. El análisis crítico de la Escuela de Frankfurt*, Ediciones de la Universidad de Castilla-La Mancha, Cuenca, 2003, p. 251.

[419] Perrow, C., *op cit.,* N. 417, capítulo 8. Es habitual encontrar, de hecho, una conexión entre la energía nuclear y la inteligencia artificial (símil que se usará a lo largo de esta investigación en un aspecto

el espacio exterior, las armas nucleares y la tecnología genética (en absoluto exóticos en nuestra era). Las conclusiones a las que el sociólogo llega encuentran, desde luego, puntos comunes con el debate de nuestra actual sociedad altamente tecnologizada, y por eso merece la pena una revisión de sus conclusiones. El análisis de Perrow pretende trascender los errores individuales, resaltando la importancia del enfoque sistémico (las organizaciones suelen ser incapaces de operar sin errores y los organismos reguladores son con frecuencia ineficaces). La peligrosidad de estos accidentes, señala, se ubica en el sistema, y no en sus componentes [420]. También advierte sobre una realidad: estos sistemas son resistentes al cambio, pues los privilegios y ganancias privadas hacen que las construcciones planificadas sean inflexibles. Pero no dejan de ser construcciones humanas, lo que implica que son los humanos quienes pueden destruirlas o reconstruirlas. El autor señala que las catástrofes o accidentes nos mandan señales de emergencia que hemos de interpretar, quizá abandonando determinadas tecnologías, rediseñando otras, o regulando otras. El foco central está en la idea de que aún algo podemos hacer para luchar contra esta potencialidad catastrófica. El hecho de que algunos de los desastres que se narran a lo largo de la obra no se pudieran concebir no implica que debamos abandonar toda nuestra pretensión de controlarlos. No podemos resignarnos a aceptar que las ganancias bien merecen los riesgos. El mensaje final de la obra es, pues, una llamada al inconformismo y al cuestionamiento de los sistemas. Sí, los accidentes serán inevitables si seguimos en la lógica expansiva de análisis coste-beneficio operada por las entidades privadas, pero no tiene por qué ser así si otros factores de valoración entran en la ecuación. Los problemas, por tanto, no están en los motivos personales, los errores individuales o en las ideologías políticas. Las señales provienen de los sistemas, tecnológicos y económicos. Son sistemas que las élites han construido y que, por tanto, pueden cambiarse o abandonarse [421].

---

principalmente regulatorio), en importantes filósofos contemporáneos. Así, Coeckelbergh, al hilo de la reflexión sobre cómo la IA transforma el poder, señala que, al igual que con el poder nuclear, "mediante la IA el poder es ejercido de arriba abajo desde una posición central y de una forma antidemocrática. De igual forma que nunca nos han preguntado si queremos el poder nuclear, nunca se nos ha preguntado si queríamos la vigilancia y la biométrica de la IA […]. Coeckelbergh, M., *La filosofía política de la inteligencia artificial,* Ediciones Cátedra, Madrid, 2023, p. 127.

[420] Perrow, C. *op cit*., N. 417, p. 351: *«The main point of the book is to see these human constructions as systems, not as coleections of individuals or representatives of ideologies» […] «it is the way the parts fit together, interact, that is important».*

[421] *Idem*: *«These systems are human constructions, whether designed by engineers and corporate presidents, or the result of unplanned, unwitting, crescive, slowly evolving human attempts to cope. Either way they are very resistant to change. Private privileges and profits make the planned constructions resistant to change; layers open layers of accommodations and bargains that go by the name of tradition make the unplanned ones unyielding. But they are human constructions, and humans can destruct then or reconstruct them».*

Aunque la lógica del riesgo en la sociedad industrializada y tecnológica estaba ya sobre la mesa, la verdadera obra de referencia que marcó un antes y un después en el discurso del riesgo, afectando directamente a la discusión político-criminal y dogmática de la época, la encontramos en los escritos del sociólogo alemán Ulrick Beck [422].

El surgimiento del término *Risokogesellschaft* (sociedad del riesgo), debido al citado sociólogo, ha tenido un impacto innegable en la literatura penal no solo de lengua germana, llenado a influenciar muchos otros ordenamientos [423]. Un repaso esencial de su obra lo realiza Bertolesi, analizando el *topos* de la sociedad del riesgo [424], precisamente poniéndolo en relación con los daños causados por IA. En el proceso de definición de la sociedad del riesgo la sociedad industrial es fundamental, pero esta no responde ya a la lógica tradicional, sino que, a la creación de riqueza se suma la inevitable producción social de riesgos.

En la sociedad del riesgo que contornea Beck, la distribución de riqueza está asociada a problemas y conflictos que surgen de la producción, de la definición y de la distribución de riesgos producidos tecnocientíficamente [425]. Los riesgos no se presentan como un invento de la modernidad, pero sí presentan unos caracteres diferentes. En primer lugar, los riesgos pasados eran personales, los actuales (tales como los riesgos nucleares), tienen dimensión global, sin poder ubicarse en un contexto espacial o temporal [426]. Otra de las características modernas del riesgo es que escapa a la percepción humana, por originarse en entornos altamente científi-

---

[422] Nos referiremos principalmente a la obra *Risikogesellschaft. Aufdem Weg in eine andere Moderne*, publicada en 1986, con independencia de que algunos de los aspectos relevantes de su teoría sociológica en lo que aquí nos interesa pueden extraerse de una obra posterior: *Gegengifte: Die organisierte Unverantwortlichkeit»,* Suhrkamp Verlag, Frankfurt am Maine*,* 1988. Cuando no nos refiramos a esta última explícitamente, se entenderá que nos referimos a la primera, en cuyo caso haremos referencia a la edición publicada en Paidós, Barcelona, 1998, en español, o bien a la edición en lengua inglesa publicada por Sage en 1992 (Beck, U., *Risk Society: Towards a New Modernity*, Londres, Sage, 1992).

[423] Es cierto, tal y como señala Gracia, que no es la única de las concepciones sociológicas del riesgo que han influido en la dogmática penal, si no que estas se han ido enriqueciendo con otras aportaciones (que se mencionarán *infra).* Gracia Martín, L., «Concepto categorial teleológico y sistema dogmático del moderno Derecho penal económico y empresarial de los poderosos», *InDret*, 3/2016, nota 33. Lo cierto es que también en Italia dicha influencia ha sido notable. Así lo afirma Bertolesi, a pesar de reconocer la influencia de autores italianos como Perini, *(Il concetto di rischio nel diritto penale moderno,* Milan, 2010). Así, Bertolesi, R., *Intelligenza artificiale e responsabilità penale per danno da prodotto*, 2020, tesis doctoral en abierto, disponible en: https://air.unimi.it/handle/2434/711618 [Accedido por última vez el 22/01/2026], p. 118 y ss.

[424] Un análisis en español, que describe de forma resumida la postura de Beck en Beck, U. «De la sociedad industrial a la sociedad del riesgo» (trad. Del Río Herrmann), *Revista de Occidente*, n.º 150, noviembre de 1993, pp. 19-40.

[425] Beck, U., *Risk society… op cit. N 422,* p. 19., traducción de la autora.

[426] *Ibid.,* p. 21. Bertolesi aclara que los riesgos globales en los que estaba pensando Beck tenían que ver con problemas actuales como la fisión nuclear o los depósitos de residuos radioactivos. Bertolesi, R., *op cit*., N. 423, p 119.

cos, y ello repercute en la manera de medirlos (conlleva, necesariamente, la medición por medio de teorías, experimentos o parámetros científicos)[427].

Lo decisivo no es tanto el aspecto cuantitativo del riesgo como el cualitativo, pues las causas son propias de la modernización, y no ajenas a ella[428]. Señala Bertolesi, al revisar este análisis del sociólogo, que, en este paradigma de sociedad, los peligros ya no se atribuyen a los dioses o a la naturaleza, pues la propia modernidad prometió superar estas amenazas, sino que ahora las amenazas provienen de este progreso, de forma que la cultura del miedo deriva «del hecho paradójico de que las instituciones encargadas del control han producido lo incontrolable»[429]. Precisamente, como se ha insistido ya en los primeros capítulos de esta obra, la creación de lo incontrolable por definición es el rasgo más significativo de la tecnología de la industria tecnológica actual.

Es crítico el autor (y en este sentido puede encontrarse quizá una conexión con Perrow) con la falta de reflexión en la evolución hacia la modernización, señalando la idea de que la transición de una época a otra (de la sociedad industrial al periodo de la modernidad) se produce de una forma indeseada, inadvertida y compulsiva[430]. El último de los rasgos que Bertolesi trae a colación en su análisis de la obra de Beck tiene que ver con que la práctica del conocimiento científico parece estar marcada por un proceso de feudalización por parte de los intereses económicos y políticos, y esto acaba por socavar la tradicional pretensión de neutralidad que se le atribuía[431]. En efecto, podemos convenir que si hay un postulado que caracteriza la sociología del riesgo actual, con conexión directa con los postulados de Beck, es el cuestionamiento de la neutralidad tecnológica, patente desde los primeros discursos sobre los riesgos asociados a la inteligencia artificial.

En definitiva, algunos de los rasgos ya apuntados por Beck (causas propias de riesgo, progreso irreflexivo y aparente neutralidad que, sin embargo, oculta intereses económicos privados), son algunos de los elementos más im-

---

[427] *Ibid.,* p. 27: «The focus is more and more on hazards which are neither visible nor perceptible to the victims; hazards that in some cases may not even take effect within the lifespans of those affected, but instead during those of their children; hazards in any case that require the "sensory organs" of science *–theories, experiments, measuring instruments–* in order to become visible or interpretable as hazards at all» (la cursiva es del original).

[428] Son, en palabras del propio Beck, producto de la industrialización. Beck, U., *Risk society… op cit. N 422,* p. 21.

[429] Bertolesi, R., *op cit*., N. 423, p. 120: *«Le istituzioni che sono preposte al controllo hanno prodotto incontrollabilità».*

[430] En este sentido, Beck, en *The Reinvention of Politics: Towards a Theory of Reflexive Modernization*, en Beck., U., /Giddens, A.,/Lash, S., *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*, Stanford University Press, 1995, p. 1 y ss.

[431] Beck, U., *Risk society… op cit. N 422*., p. 167 y ss., hablando del fenómeno de *«feudalization of cognitive practice».*

portantes en el análisis de la sociedad tecnológica de hoy, como se tratará de mostrar *infra*.

Finalmente, y para cerrar el estudio de la productiva década de los 80, merece la pena traer a colación otro trabajo de Hassemer, *«Symbolisches Strafrecht und Rechtsguterschutz»* [432], publicado ya en 1989, pues da lugar a que, inmediatamente después, la década de los 90 se convierta en el momento en que el concepto de expansión del derecho penal explota, y donde el derecho penal del riesgo es ya un tópico recurrente.

### 1.4 **La década de los 90: la expansión definitiva del derecho penal**

A principios de los 90 se publica otra de las obras de referencia en la sociología del riesgo de la mano de Niklas Luhmann. En su obra *Sociología del riesgo* [433] expone el sociólogo alemán su concepto de riesgo, diferente del presentado por Beck, pero igualmente influyente en todo el discurso político-criminal en adelante. Mientras Beck partía de una concepción crítica y normativa del riesgo, con el foco puesto en la denuncia de la creación de riesgos globales por parte del progreso, la construcción de Luhmann es sistémica y funcionalista, es decir, no critica, sino observa cómo los diferentes sistemas (económico, político o científico) gestionan complejidades. Para Luhmann, el riesgo no se presenta como una amenaza objetiva, sino como una selección comunicativa, una forma en la que el sistema anticipa las consecuencias de sus propias decisiones. Por tanto, mientras en Beck el riesgo se trata desde una perspectiva política y social, caracterizado como una consecuencia real y política del progreso técnico, en Luhmann no es más que una construcción comunicativa interna a cada sistema, pues parte de una visión del sistema social cerrada, autopoiética [434].

---

[432] El propio Schünemann, B., «Consideraciones críticas...», *op cit*. N.407, p. 198 y ss., sitúa dicha obra de Hassemer como la obra en que vincula sus reflexiones con la concepción de la sociedad del riesgo, y es precisamente este trabajo el que luego da pie a que su discípulo (Felix Herzog), en su célebre escrito de habilitación de 1991, «*Gesellschaftliche Unsicherheit and strafrechtliche Daseinsvorsorge»,* desarrolle toda su crítica del derecho penal del riesgo.

[433] Publicada en alemán como *Soziologie des Risikos,* Walter de Grurter, 1991. En español: Luhmann, N., *Sociología del riesgo*, (trad. Por Silvia Pappe, Brunhilde Erker, Luis Felipe Segura, Javier Torres Nafarrete), Universidad Iberoamericana, Jalisco, 1992.

[434] Para una interpretación de las tesis luhmannianas y una explicación de su teoría de sistemas, *vid.*: Martínez García, J. I., «Pensar el riesgo. En diálogo con Luhmann», en Pérez Alonso, E./Arena García, E./Mercado Pacheco, P/ Serrano Moreno, J. L. (Coords.), *Derecho, globalización, riesgo y medio ambiente*, Tirant Lo Blanch, Valencia, 2012, pp. 323-354.

Aunque parte del discurso basado en la interconexión de sistemas autopoiéticos ha traspasado directamente al discurso más actual del riesgo tecnológico [435], otros muchos autores contribuyeron en esta década al discurso del riesgo y su expansión. Así, siguiendo directamente la estela de Beck, uno de sus discípulos más señalados, Herzog, desarrolla en 1991 su pensamiento [436] con una crítica directa a la expansión del derecho penal y a la lógica de los delitos de peligro, determinando que a la expansión de los riesgos sociales le corresponde una política que no se plantea nuevas perspectivas de desarrollo social. Los riesgos solo se tienen en cuenta cuando son catastróficos y entonces las exigencias de reacción se ven colmadas con el recurso al derecho penal. Este uso del derecho penal para paliar estos riesgos catastróficos el autor lo tilda de «dramatización», precisando que la ampliación del derecho penal sirve al debate político como coartada para demostrar que es consciente de un problema, utilizándose de forma instrumentalizada [437]. En algún escrito más reciente, Herzog caracteriza al Derecho penal del nuevo siglo como decadente [438], en el que reina la «irresponsabilidad organizada» [439].

Herzog parte de un contexto de capitalismo mundial, economía de mercado y liberalización de muchos espacios, tales como como Internet, sin que paralelamente se cumpla el ideal de Adam Smith de que haya un equilibrio que logre el bienestar de la mayoría. Y, aun así, siendo consciente de estas deficiencias producidas en determinados sectores, reconociendo que «al progreso de la globalización le es inherente el injusto reparto de cargas y riesgos» [440], no puede afirmar que el derecho penal deba encargarse de corregir estas injusticias, pues este no surge como instrumento de la política estructural proactiva, sino que es reactivo y orientado a la responsabilidad individual. Así, alude al título de su tesis doctoral para afirmar que «el Derecho Penal no puede lograr la prevención del injusto estructural, sino sólo la manifestación del derecho intersubjetivo» [441].

---

[435] Un reflejo muy claro de la influencia luhmanniana en la explicación de la interconexión de sistemas de IA lo vemos en la obra Teubner y Beckers (*op. cit*. N. 111), sobre la que se profundiza en este mismo capítulo, *infra.*

[436] Herzog, F., *Gesellschaftliche Unsicherheit and strafrechtliche Daseinsvorsorge. Studien zur Vorlegung des Strafrechctschutzes in den GefÄhrdungsbereich*, Heidelberg, R. V. Deckers, 1991. Una versión resumida y en español de esta obra la encontramos en Herzog, F., «Límites al control penal de los riesgos sociales (Una perspectiva crítica ante el derecho penal del peligro)», trad. Larrauri Pijoan y Pérez Álvarez, *Nuevo Foro Penal,* 53, 1991, pp. 303-312.

[437] Herzog, F., «Límites al control…», *ibid,* p. 326.

[438] Herzog, F., «Sociedad del riesgo…» *op cit*, N.418, p. 249 y ss.

[439] Toma el término de Beck, en su obra, *Gegengifte.*

[440] Herzog, F., «Sociedad del riesgo…», *op. cit., N. 418,* p. 254.

[441] *Idem.*

Explica, entonces, su concepción sobre la sociedad de la información, donde se hace necesario indagar sobre la interdisciplinariedad o la codependencia de los humanos en un mundo globalizado [442], pues, en realidad, frente a todas las promesas de la sociedad de la información lo cierto es que cada día sabemos menos sobre los procesos de decisión, sobre los actores y sobre su motivación. Denuncia Herzog que esta sociedad del riesgo, donde el control social debe basarse en la libertad de información y en la transparencia, necesita de restricciones internas, pero las respuestas que ofrece el derecho penal no pueden lograr esta necesaria autorregulación social [443].

Otro de los teóricos del riesgo apegados a la crítica del expansionismo es Prittwitz, si bien se separa algo de la concepción de Beck, al que tilda de catastrofista [444] y cuyo concepto de sociedad de riesgo no acepta. Su caracterización del derecho penal del riesgo parte de la crítica de los intentos de preservación del futuro por medio del derecho penal [445] o, en general, del derecho penal de una sociedad donde la definición de riesgos se ha convertido en uno de los aspectos centrales del conflicto social [446]. Pone el énfasis este autor en que la idea de fragmentariedad no es suficiente, pues, por mucho que exista un ataque grave a un interés digno de protección, la verdadera pregunta será si este ataque merece ser protegido por el ordenamiento penal [447]. Esta idea de que el derecho penal debe tener un rol modesto en la sociedad globalizada del riesgo es reiterada en otros trabajos [448].

Otros representantes de la crítica a la expansión los encontramos en Italia, que, de la mano de Sgubbi y su concepción del derecho penal como riesgo social, reproduce un discurso similar al de la Escuela de Frankfurt, con un carácter marcadamen-

---

[442] *Idem.,* p. 257.

[443] *Idem*., pp. 257 y 258.

[444] Prittwitz, C., «Sociedad de riesgo y derecho penal», en Arroyo Zapatero, L./Neumann, U., / Nieto Martín, A., (Coords.), *Crítica y justificación del derecho penal en el cambio de siglo. El análisis crítico de la Escuela de Frankfurt*, Ediciones de la Universidad de Castilla-La Mancha, Cuenca, 2003, p. 261.

[445] Entrando aquí en abierta disputa con Stratenwerth, G., «Zukunftssicherung mit den Mitteln des Strafrechts?», *ZStW*, 105, 1993, pues lo que éste último defendía, esto es, que el derecho penal ha de encargarse de la preservación del futuro, a juicio de Prittwitz, pasa por alto que la seriedad de la situación no conduce de modo irremediable a la intervención del Derecho penal, sino más bien determina a su renuncia, pues se revela como un tratamiento inadecuado frente a un mal conocido, de cuya existencia nadie discute. Prittwitz, C.*, op cit. «Sociedad de riesgo…»,* N. 444*,* p. 269.

[446] *Ibid.,* p. 265.

[447] *Idem*., p. 279, *passim*.

[448] Así, más recientemente, en Prittwitz, C., «La función del Derecho Penal en la sociedad globalizada del riesgo: defensa de un rol necesariamente modesto» en Pérez Alonso, E./ Arana García, E./ Serrano Moreno, J. L./ Mercado Pacheco, P. (Coords.), *Derecho, globalización, riesgo y medio ambiente*, Tirant Lo Blanch, Valencia, 2012, pp. 415-428.

te liberal. En su monografía de 1990[449], el autor aborda la evolución del derecho penal hacia un sistema basado en el riesgo social, donde el delito ya no depende exclusivamente de la culpabilidad o daño causado, sino de factores externos como la posición económica y social del individuo. El derecho penal moderno se centra en la transgresión normativa, priorizando el daño a la normatividad sobre el daño a bienes jurídicos. El delito, que para Sgubbi pasa a ser un hecho amoral socialmente neutro, es un riesgo social en la medida en que depende de la colocación económico-social del individuo, que además será más elevado según más socialmente significativa sea la actividad que este desarrolla[450] (esto lo denomina el autor la primera ley del riesgo). Esta estructura genera una responsabilidad penal más artificial y menos vinculada a principios naturales, incentivando la negociación colectiva sobre lo lícito e ilícito. Su contextualización del fenómeno parte de la crítica al Estado intervencionista que regula administrativamente todas las parcelas de libertad, creando áreas de legalidad condicionada. Estas áreas de legalidad condicionada no lo son con respecto a la libertad, sino a la norma[451]. El núcleo del delito pasa a ser la mera transgresión de la norma, y el Estado expropia a la sociedad civil los bienes jurídicos[452], creando unos nuevos, artificiales, para asegurar su propio fin (con la estructura propia de un monopolio). En este contexto, las responsabilidades penales se distribuyen de forma desigual según el rol o posición social del individuo. Las clases altas, antes inmunes al sistema penal, ahora enfrentan mayores riesgos debido a la regulación de actividades técnicas y simbólicas. Por otro lado, los sectores marginados quedan fuera del radar penal o son tratados como sujetos al margen del sistema. En este contexto, la ejecución efectiva de penas es limitada, y el derecho penal se convierte en un instrumento simbólico y negociable que incentiva la ilegalidad total[453].

---

[449] Nos referimos a la obra: Sgubbi, F. *Il diritto penale come rischio sociale. Ricerche sulle scelte di allocazione dell'illegalità penale*, Il Mulino, Bolonia, 1990. Una visión más actualizada de este análisis crítico de la sociedad actual y el reflejo penal se puede encontrar en el mismo autor, en Sgubbi, F., *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Il Mulino, Bolonia, 2019, donde el autor critica el recurso al derecho penal como herramienta política y mediática para abordar problemas sociales, sacrificando principios fundamentales como la garantía y la certeza jurídica. Este modelo castiga más por lo que el individuo es que por lo que hace (personalidad), reemplazando el binomio inocente/culpable por puro/impuro.

[450] Sgubbi, F., *Il diritto penale come rischio..., op. cit*. N. 449, p. 9.

[451] *Idem*. p. 14.

[452] *Idem*. p. 18.

[453] En el contexto italiano, pese a que algunos autores se muestran críticos con la expansión y partidarios de un derecho penal mínimo (así, Stortoni, aunque de forma más matizada, pues reconoce que, pese a las distorsiones que se producen en el derecho penal clásico, cuya capacidad de rendimiento va a ser mínima, acepta la posibilidad de adoptar la perspectiva de los delitos de peligro de mera conducta de forma muy cautelosa, con especial atención a la esencia del injusto y el principio de legalidad), otros han señalado que adoptar un derecho penal mínimo resultaría clasista y discriminatorio. Ejemplo de estos últimos: Marinucci y Dolcini. Del primer grupo, Stortoni. Stortoni, L., «El riesgo procedente del desconocimiento tecnológico», en AA. VV, *Responsa iurisperitorum digesta,* Ediciones Universidad Salamanca, vol. 3*,* 2000*,* pp. 83-104. También en Stortoni, L., «Angoscia tecnologica ed

Sin embargo, las réplicas no tardan demasiado en llegar, y lo hacen, una vez más, en el contexto alemán, de la mano de autores como Kuhlen [454] o Schünemann [455], que son capaces de replicar los argumentos de la Escuela de Frankfurt con críticas que son calificadas por Gracia como demoledoras y concluyentes [456].

Kuhlen [457] acepta que la creciente distancia personal y social propias de la producción industrial, unidas al desnivel técnico-científico entre fabricante y usuario limitan las posibilidades de que el consumidor pueda ejercer un control [458], pero de aquí deriva el autor una necesidad de afrontar la situación mediante los delitos de peligro abstracto, y no extramuros del derecho penal.

Enmarca Gracia en los siguientes términos el pensamiento de ambos autores: cuando las relaciones sociales tenían lugar entre sujetos claramente identificados e individualizados, que se podían comunicar de una forma directa y con grandes posibilidades de autoprotección frente a todos los posibles peligros, la intervención penal podía retrasarse, pero en la sociedad actual el modo de relación se ha sustituido, de forma que las interacciones son anónimas y en ámbitos donde confluyen multitud de cursos causales diversos, con víctimas potenciales que tienen sus capacidades de autoprotección esencialmente mermadas [459]. Es solo en este contexto donde se puede (y se debe, según Schünemann, de una manera muy obvia [460]), anticipar la intervención penal sin que se tambaleen las bases del sistema.

---

esorcismo penale», *Rivista italiana di diritto e procedura penale*, vol. 47, 2004, pp.71-89. También Baratta o Ferrajoli (Baratta, A., «Principi di diritto penale mínimo. Per una teoría dei diritti umani come oggetto e limiti della legge penale» en español en: Baratta, A., «Principios de Derecho Penal Mínimo (Para una teoría de los derechos humanos como objeto y límite de la ley penal)», en *Criminología y Sistema Penal (Compilación in memoriam*), B de F, Buenos Aires, 2004, pp. 299-333; Ferrajoli, L., *Diritto e ragione. Teoria del garantismo penale*, en español en: *Derecho y razón: teoría del garantismo penal*, Trotta, Madrid, 1995). Del segundo grupo, Marinucci y Dolcini. Por todos: Marinucci, G./Dolcini, E., «Derecho penal 'mínimo' y nuevas formas de criminalidad», *Revista de Derecho Penal y Criminología*, 9, 2002, pp. 147-167.

[454] Kuhlen, L., «Zum Strafrecht der Risikogesellschaft», *GA*, 1994, pp. 347 y ss.

[455] Por todas, Schünemann, B., «Kritische Anmerkungen zur geistigen Situation der deutschen Strafrechtswissenschaft», *GA* (5), 1995. En español en *ADPCP*, (49), 1996, *op cit*. N.407.

[456] Gracia Martín, L., «Concepto categorial teleológico…», *op cit., N. 423,* p. 16.

[457] Aunque su obra fundamental en cuanto contrarréplica es de 1994 (en forma de recensión a la obra de Prittwitz y Hilgendorf), a saber: Kuhlen, L., «Zum Strafrecht der Risikogesellschaft», *op cit. N. 454;* una exposición de su pensamiento crítico podemos encontrarla en Kuhlen, L., «*Grundfragen der strafrechtlichen Produkthaftung», JZ, 49*(23), 1994, pp. 1142-114. En español en: Kuhlen., L., «Cuestiones fundamentales de la responsabilidad penal por el producto», en Mir Puig, S., /Luzón Peña, D. (Coords.), *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto,* J. M. Bosch Editor, 1996, pp. 231-246. También en Kuhlen., L., «Necesidad y límites de la responsabilidad penal por el producto», *ADPCP*, 55 (1), 2002, pp. 67-90.

[458] Kuhlen, L., «Cuestiones fundamentales…», *ibid*, p. 232.

[459] Gracia Martín, L., «Concepto categorial teleológico …», *op cit., N. 423,* p. 17.

[460] Schünemann, B., «Kritische Anmerkungen…», *op cit.* n. 455, p. 212 (igual en *ADPCP*, *op cit.* N.407, p. 199), pues dicha forma de intervención anticipada «deriva de forma natural de la naturaleza de las

Es rotundo Schünemann al recalcar que *«cuando la escuela de Frankfurt postula la reconducción, por principio, del Derecho penal a los delitos de resultado, está abogando por una negativa a la modernización del Derecho penal, negativa que necesariamente ha de fracasar en la finalidad del Derecho penal de proteger bienes jurídicos, al ignorar las condiciones de actuación de la sociedad moderna»* [461].

Estas condiciones de actuación de la sociedad moderna son, sin duda, la base sobre la que, ya en el inicio de década y de siglo, vuelve a suscitar la reflexión esta vez de la doctrina española, dando lugar a toda una oleada de reacciones.

### 1.5 **Los años 2000: la discusión española** [462]

Cuando la recepción en España del discurso de la expansión y la modernización del derecho penal del riesgo es inevitable, se suceden los trabajos de investigación que, en muy diferentes sentidos, se proponen analizar el fenómeno. Se destacarán esencialmente tres posturas, en aras a la sistematización y mejor entendimiento. Podemos distinguir al menos tres grupos:

– Los defensores moderados de la expansión. Dentro de este grupo se deben incluir las dos monografías de obligada referencia en España, las de Mendoza Buergo y Silva Sánchez [463], que reaccionan a la crítica alemana, proponiendo su propio modelo para afrontar la situación.

– Los favorables a la expansión, como Gracia Martín o Pérez Alonso que aportan, en este sentido, argumentos definitivos.

---

cosas».

[461] *Idem*., p. 200 (versión española).

[462] No es casualidad que sea precisamente con la llegada del siglo y, más concretamente, con la aprobación del nuevo Código penal español en 1995, cuando, tras apenas unos años, las discusiones alemanas se trasladan al contexto español, donde la introducción de algunas nuevas modalidades delictivas trae a colación los debates que se tenían en Alemania. En este sentido, Gracia, hablando específicamente de la introducción de los delitos económicos, en Gracia Martín, L., «Concepto categorial teleológico …», *op cit*., N. 423, p. 23.

[463] Mendoza Buergo, B., *El derecho penal en la sociedad del riesgo*, *op. cit.* N. 384; y Silva Sánchez, J. M., *La expansión …, op. cit.* N. 383. Aunque somos conscientes de que los postulados de Mendoza y de Silva no son idénticos, parece adecuado agruparlos bajo el mismo paraguas, pues ambos, partiendo de la inevitabilidad de la confrontación de los riesgos, que afirman como una realidad, se pueden considerar defensores del modelo garantista. Esta última afirmación es, en cambio, muy rebatida por Gracia, que cree que ningún penalista aboga por dejar de lado los principios constitucionales. En todo caso, y a efectos de simplificar su discurso, se deja apuntado que se consideran parte de la misma rama porque ambos aceptan como inevitables algunos postulados de la expansión.

– Los críticos de la expansión, que, más allá de criticar los puntos que ya destacaba la Escuela de Frankfurt, extienden sus críticas a las nuevas posturas acogidas en España. Destacan aquí Paredes Castañón [464] o Díez Ripollés.

a) Los moderados

En su obra monográfica *El derecho penal en la sociedad del riesgo* [465], Mendoza analiza el impacto que la concepción contemporánea de la denominada sociedad del riesgo está teniendo tanto sobre el derecho penal como sobre la política criminal. Caracteriza a la sociedad como impersonal, con riesgos atomizados, complejidad organizativa y creciente percepción de inseguridad, caracteres que han venido demandando una creciente intervención punitiva. La autora parte de una perspectiva crítica para abordar las transformaciones en el Derecho penal europeo (tomando como base, especialmente, Alemania y España), donde identifica dos corrientes contrarias: aquella crítica con la evolución del derecho penal del riesgo, y aquella que defiende la adaptación del Derecho penal a las nuevas realidades y necesidades de protección y seguridad [466]. Mendoza describe los pilares de la ya asentada teoría sociológica del riesgo, donde señala como características los nuevos peligros, la complejidad organizativa y la sensación de inseguridad subjetiva [467], analizando cómo la función preventiva del derecho penal se ha visto reforzada, con la consiguiente funcionalización del derecho penal, que busca asegurar funciones sociales y el mantenimiento del sistema. Sin embargo, una vez presentadas todas las posturas en liza, reconoce que la intervención del Derecho penal en estos nuevos escenarios resultará admisible solo en tanto y en cuanto se respete su función garantista; es decir, los fundamentos del Estado liberal. La crítica iría solo dirigida, pues, a aquellas propuestas de expansión que pretendan usar el derecho penal con el simple objeto de mantener un orden normativo sin efectiva protección de derechos individuales. En este sentido, la obra de Mendoza supone un punto de partida intermedio, que más adelante desarrollará en su concepción sobre los delitos de peligro abstracto como ejemplo

[464] Paredes Castañón, J. M., «¿Al otro lado del discurso jurídico penal? Sobre las bases político-criminales de la "modernización" del Derecho penal. Reseña crítica de Luis Gracia Martín» *RDPC*, 17, 2006, pp. 437-472.

[465] Mendoza Buergo, B., *El derecho penal en la sociedad del riesgo*, *op. cit.* N. 384.

[466] *Ibid.,* p. 118.

[467] Algunas referencias adicionales a la concepción sociológica de las que parte Mendoza Buergo, que beben de Beck, Luhmann, Prittwitz o Giddens, en: Mendoza Buergo, B., «Principio de precaución, derecho penal del riesgo y delitos de peligro», en Romeo Casabona, C. M., (Coord.), *Principio de precaución, biotecnología y derecho,* Comares, Bilbao-Granada, 2004, pp. 436-442.

paradigmático de este nuevo modelo[468]. Ya apuntaba esta autora en algún escrito previo sobre la moderna política criminal que la alternativa no está entre dos únicas posturas claramente opuestas, sin que sea defendible que las únicas alternativas sean o bien *«la total funcionalización del Derecho penal que modifique o incluso renuncie a principios garantistas o, por el contrario, en el refugio inconmovible en el llamado Derecho penal nuclear más básico, sino que se hacen necesarias las posturas intermedias que intenten alcanzar un punto de posible solución»*[469].

Por su parte, Silva escribe una de las monografías[470] de referencia en torno a la modernización del derecho penal del riesgo, dando lugar a un vocablo propio: la expansión[471]. El punto de partida de Silva es, desde luego, la obra de Beck, que caracteriza como «lugar común» de la sociedad postindustrial[472], donde los riesgos se refieren a daños no delimitables, globales y, con frecuencia, irreparables: que afectan a los ciudadanos, y que surgen de decisiones humanas[473]. Resalta este autor otra característica: las consecuencias lesivas del fallo técnico aparecen como problema central en este modelo, cuyo punto de partida es que cierto porcentaje de accidentes graves es inevitable, teniendo en cuenta la complejidad del sistema, por lo que habrá que decidir cuál es el criterio de ubicación de dichos fallos[474]. Otro de los rasgos que señala Silva es la interdependencia entre individuos, resaltando que la indemnidad de los bienes de un individuo depende de la realización de con-

---

[468] *Vid. infra*, capítulo VI.

[469] Mendoza Buergo, B., «Exigencias de la moderna política …, *op. cit.* N. 360, p. 292.

[470] La primera de las ediciones de *La expansión* es de 1999, aunque las referencias que se harán de aquí en adelante se referirán a la última de las ediciones, puesta al día y ampliada por el autor, es decir, a la 3.ª edición, tal y como aparece *op. cit.* N. 383. Aunque, para ser justos, la más reciente revisión por el autor de sus postulados la encontramos en el editorial publicado en *InDret* 2/2025, «Expansión 2.0: los nuevos riesgos», donde Silva analiza la intensificación de la expansión del Derecho penal en respuesta a los nuevos riesgos surgidos desde finales del siglo XX, como el terrorismo global, la crisis financiera de 2008, los desafíos tecnológicos y ambientales, la erosión del Estado del bienestar, la inmigración masiva, la guerra en Europa y, más recientemente, los riesgos derivados del uso acelerado e intensivo de la inteligencia artificial y la biotecnología sobre el cuerpo y la mente humana. Sostiene que estos riesgos, de carácter estructural y sistémico, han propiciado un incremento de la intervención penal en ámbitos tradicionalmente ajenos a ella, pese a que esta herramienta resulta en muchos casos inadecuada. En este contexto, advierte sobre la creciente tendencia a canalizar el miedo social hacia figuras criminalizadas como forma de control simbólico. La continuación de esta reflexión la podemos encontrar en su monumental tratado de derecho penal, que se estudiará, en lo relativo a estos postulados, *infra.*

[471] Alabando el término, por lo acertado de su denominación, aséptica en cuanto a las referencias temporales (huyendo de la terminología hasta entonces empleada que se refería a la modernidad o modernización); Gracia Martín, L., «Concepto categorial teleológico …», *op cit., N. 423,* p. 27, que también señala lo novedoso de la concepción de Silva.

[472] Silva Sánchez, J. M., *La expansión …, op. cit.* N. 383, p. 13.

[473] *Idem.,* nota al pie 25

[474] *Idem.,* p. 15, con referencias a Perrow en nota 28.

ductas de control por terceros [475]. Señala, asimismo, como rasgo cultural de la sociedad contemporánea, la tendencia a expandir la imputación de responsabilidades *(Zurechnungsexpansion),* pues se tiende a considerar al accidente fortuito *(Unglück)* como injusto *(Unrecht)* [476]. Reconoce Silva, entonces, cómo estas tendencias globales se han venido a materializar en una administrativización del Derecho penal [477], y que esto ha supuesto la progresiva transformación del derecho penal en las modernas sociedades a un elemento más del derecho de gestión (donde triunfa la lógica de las acumulaciones o repeticiones propias de lo administrativo), dejando de lado la dimensión de la lesividad del comportamiento individual [478].

Expone finalmente Silva cuáles deben ser, a la vista de esta caracterización sociológica y penal, las características del derecho penal del presente, aceptando que no puede suceder que el derecho penal actual comparta las mismas características del derecho penal liberal del siglo pasado. El sistema que propone pasa por «*admitir la gradualidad de la vigencia de las reglas de imputación y de los principios de garantía en el propio seno del Derecho penal, en función del concreto modelo sancionatorio que éste acabe asumiendo*» [479], proponiendo un derecho penal de dos velocidades, donde el núcleo duro del derecho penal (necesariamente integrado por las sanciones privativas de libertad), mantuviera todas las garantías clásicas, y las nuevas tipificaciones, con sanciones no privativas de libertad, admitieran flexibilización [480], aceptando, y solo para algunos casos (conductas alejadas de la creación de un peligro real para bienes individuales), con resignación, la expansión del derecho penal [481].

---

[475] *Idem.,* p. 17. Cita Silva en este punto a Luhmann (*Die Gesellshaft der Gesellshaft*), por ser uno de los autores que vaticina el advenimiento de la sociedad de la información, donde esta interdependencia resulta fundamental. Silva Sánchez, J. M., *La expansión …, op. cit.* N. 383, nota al pie 45.

[476] En este punto cita a Prittwitz. La obra a la que se refiere Silva es: Prittwitz, C., *Strafrecht und Risiko: untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft*, Vittorio Klostermann, Frankfurt am Main, 1993. Versión en español: Prittwitz, C., *Derecho penal y riesgo: Investigaciones sobre la crisis del Derecho penal y la política criminal en la sociedad del riesgo,* Marcial Pons, Madrid, 2021.

[477] A partir de p. 136 y ss., donde defiende su postura de diferenciación cuantitativa, que ha sido desarrollada posteriormente en otros de sus trabajos. En *La expansión* basaba esta distinción en la forma de entender la lesividad de una y otra clase de infracciones (penales y administrativas, se entiende), y en los criterios utilizados para su imputación.

[478] *Ibid.,* p. 142.

[479] *Ibid.,* p. 171, reconociendo que su modelo bebe de aportes como el realizado por Hassemer y otros, de construir un *Interventionsrecht* (Derecho de intervención), entre el Derecho penal nuclear y el *Recht* del *Ordungswidrigkeiten*, entre el Derecho civil y el público, para ilícitos en materia de drogas, económicos, ecológicos, etc.

[480] En p. 181 y 182 explica en qué concretos casos podría darse esta flexibilización.

[481] *Ibid.,* p. 182.

b) Los favorables

A la visión cautelosa de Mendoza y resignada de Silva, hemos de añadir las versiones más animadamente expansionistas de los autores arriba categorizados como «favorables a la expansión».

Así, Gracia Martín ha sido uno de los defensores más convencidos de la misma. En su obra *Prolegómenos* [482] caracteriza el fenómeno, más allá de sus tintes sociológicos, como una cuestión de conflicto social y político, perfilando a los críticos de la expansión como portadores de un discurso que califica de resistencia, y que tilda de irracional, reaccionario y anacrónico [483]. Gracia da una vuelta de tuerca al discurso, desmantelando las resistencias intelectuales al Derecho penal moderno y partiendo de otra base discursiva. Para Gracia la modernización se presenta no como un elemento negativo, sino como una exigencia histórica derivada de los cambios sociales, económicos y tecnológicos, por lo que es deseable que el derecho penal se adapte a nuevos fenómenos como la globalización, los riesgos tecnológicos o el crimen organizado [484]. Su crítica se centra en considerar al modelo liberal como decimonónico, superado, por ser portador de un sesgo clasista. Que el nuevo derecho penal incluya, por primera vez, el castigo de los poderosos es, para Gracia, algo a celebrar [485]. Quienes critican esta expansión (los que él denomina partidarios del discurso de resistencia) son tildados de políticamente reaccionarios, pues sus argumentos perpetúan un modelo jurídico desfasado que obstruye la creación de un derecho penal acorde a las necesidades contemporáneas [486]. Llega Gracia a criticar incluso algunos de los principios limitadores tradicionales del Derecho penal liberal, pues han sido en ocasiones aplicados en detrimento de ciertas clases sociales, por lo que la solución es reinterpretarlos según las exigencias no del Estado liberal, sino del Estado social y democrático de derecho [487].

---

[482] Gracia Martín, L., *Prolegómenos para la lucha por la modernización del Derecho penal Y para la crítica del discurso de resistencia. A la vez una hipótesis de trabajo sobre el concepto de Derecho penal moderno en el materialismo histórico del orden del discurso de criminalidad*, Valencia, Tirant lo Blanch, 2003.

[483] Ibid. Así, en p. 166, estableciendo que la modernización ha de definirse como un proceso de constitución de un discurso de criminalidad completamente nuevo, o en p. 199, afirmando que ya no se puede partir del postulado unilateral del «Estado liberal de Derecho», pues este debe considerarse completamente superado.

[484] *Ibid.,* pp. 89 y ss.

[485] *Ibid.,* pp. 190 y 191: «entiendo por Derecho penal moderno en sentido *material,* al «sistema» de regulaciones jurídico-penales que tienen por objeto la criminalización formal, de conformidad con el sistema de garantías políticas del Estado social y democrático de Derecho, del *sistema de acción ético-socialmente reprochable de las clases sociales poderosas*» (cursiva en el original).

[486] *Ibid.,* pp. 213-216.

[487] *Ibid., passim*. Aunque de forma marcadamente incisiva en pp. 208-209, afirmando que el Estado social no sólo es un auténtico Estado de Derecho, sino que en la medida en que lo es de forma material, es

Pérez Alonso es otro de los autores que se muestra abiertamente favorable a la expansión o modernización, y lo hace con una confrontación abierta tanto con la Escuela de Frankfurt como con el discurso de Silva, al que tilda de liberal o clásico [488]. Para este autor el derecho penal del riesgo es de un corte claramente social, pues protege derechos fundamentales de segunda generación, transmitiendo la importancia o trascendencia social de estos [489], mientras el discurso de la resistencia defiende de manera encubierta solo los intereses individuales clásicos, calificando la vuelta al Derecho penal liberal de la Ilustración como un anacronismo histórico, que sin duda conduciría a un derecho disfuncional a nuestro tiempo, que caracteriza como sociedad posmoderna [490]. Pérez Alonso se muestra partidario, por tanto, de un derecho penal mínimo, pero siempre teniendo en cuenta el marco constitucional actual, pues el Estado debe responder las necesidades y demandas sociales *«de hoy, y no a las de otra época»* [491].

En definitiva, concluye con cita a Feijoo, no puede interpretarse cuál debe ser el derecho penal mínimo de cada momento con abstracción de las características de la sociedad y las funciones que se esperan del derecho penal, que no permanecen, desde luego, inamovibles y ajenas al cambio de cada orden social temporal [492].

Finalmente, Feijoo, con su estudio exhaustivo sobre el paradigma de la sociedad del riesgo [493], donde uno de sus mayores méritos consiste en destacar

---

mucho más Estado de derecho que el liberal.

[488] Pérez Alonso, E., «Expansión, riesgo y principio de precaución en el Derecho penal», en Pérez Alonso, E./ Arana García, E./ Serrano Moreno, J. L./ Mercado Pacheco, P. (Coords.), *Derecho, globalización, riesgo y medio ambiente*, Tirant Lo Blanch, Valencia, p. 399. Otros de los autores que se han encargado de criticar la propuesta del derecho de dos velocidades de Silva, considerando que el derecho penal debe interpretarse como un todo con garantías comunes: Zugaldía Espinar, J. M., «¿Qué queda en pie del Derecho penal del principio mínima intervención, máximas garantías?», *Cuadernos de Política Criminal*, 79, 2003, p. 109-124. Señala, incluso, Zugaldía, el carácter social de la administrativización del derecho penal y su contrario (la penalización del derecho administrativo).

[489] Pérez Alonso, E., *op. cit.* N. 488, p. 397.

[490] *Ibid.,* p. 399. En lo relativo al término, a lo largo de las discusiones referentes a la expansión del derecho penal, los términos moderno y posmoderno aparecen entremezclados. Así lo explica Miró Llinares, refiriéndose a una discusión «tácita, de tipo terminológico», reconociendo que en la mayoría de los casos ambos comparten características político criminales idénticas o similares. Profundizando en la discusión terminológica: Del Rosal Blasco, B., «¿Hacia el Derecho penal de la Postmodernidad?», *RECPC*, 11-08, 2009, pp. 1-64. Se utilizan, aquí, indistintamente, respetando la terminología propia de cada autor. *Vid.*: Miró Llinares, F., «El «moderno» Derecho penal vial y la penalización de la conducción sin permiso», *InDret*, 3/2009, nota al pie 3.

[491] Pérez Alonso, E., *op. cit. N*. 488, p. 397.

[492] En este punto, Pérez Alonso acera su postura a los argumentos empleados por Feijoo Sánchez en dos de sus trabajos más reseñados en la materia (Feijoo Sánchez, B., «Sobre la administrativización …», *op. cit.* N. 405 y, como parte de un trabajo monográfico completo, Feijoo Sánchez, B., *Normativización del derecho penal y realidad social… op. cit*. N. 405.) El pensamiento de Feijoo se resume inmediatamente *infra.*

[493] Feijoo Sánchez, B., *Normativización del derecho penal y realidad social…* op. cit. N. 405.

la verdadera naturaleza de la crítica al derecho penal moderno, no como construcción verdaderamente histórica, sino como «modelo conceptual ideal» [494], deja claro que lo que se discute esencialmente en la contienda entre expansionistas y resistencia es, en definitiva, la implantación de un modelo teórico político, basado en el rechazo frontal a la evolución de las modernas sociedades occidentales (que interpretan como de tendencia anti-liberal y anti-ilustrada [495]).

Para Feijoo, sin embargo, el discurso de resistencia no tiene sentido, pues *«El derecho penal no puede obviar las nuevas características de las sociedades contemporáneas si quiere cumplir sus funciones sociales»* [496]. La idea que expresa Feijoo, partiendo de los postulados de Beck, es que ante determinados descubrimientos sobre los sistemas sociales la política criminal no puede permanecer inamovible, aceptando que, en una sociedad donde los riesgos son inevitables, no es viable delegar completamente en la tecnocracia o en la administración la definición de niveles aceptables de riesgo. La sociedad del riesgo surge precisamente cuando los marcos normativos y las instituciones protectoras del modelo industrial fracasan en su promesa de garantizar seguridad frente a amenazas derivadas de decisiones, conscientes o no, que generan nuevos peligros [497].

Merece la pena citarlo textualmente, por lo contundente de su idea:

> *«Si el paradigma social ha cambiado, tal y como están intentando hacerlo ver las ciencias sociales, y la sociedad vigente ya no se corresponde con el modelo de sociedad industrial, el derecho penal tendrá que acompañar ese cambio de paradigma. Si las características de la sociedad actual como sociedad del riesgo obligan a cambiar las estrategias preventivas públicas o estatales modificando el perfil de los Estados modernos, el derecho penal no se puede mantener impasible ante esa dinámica evolutiva, sino que tiene que procesarla internamente de acuerdo con sus fines y funciones. No se trata de decidir sencillamente si la pena debe resolver las nuevas necesidades que plantea la configuración de la sociedad como sociedad del riesgo, sino de determinar el papel que debe cumplir la pena dentro del conjunto de medidas jurídicas (preventivas, sancionadoras, etc.) que debe adoptar el Estado para resolver los nuevos conflictos sociales de las sociedades actuales.»* [498]

---

[494] *Ibid.,* p. 195.
[495] Ibid. Nota al pie 105, con fragmentos donde los propios Hassemer, Herzoz o Prittwitz así lo expresan.
[496] *Ibid.,* p. 199.
[497] *Idem*.
[498] *Ibid.,* p. 200.

La verdadera crítica, que compartimos plenamente, es sencilla: los postulados de la Escuela de Frankfurt propugnan un derecho penal del todo disfuncional por estar completamente desvinculado del contexto histórico-social [499]. Por ello, la llamada de Feijoo es a la reflexión pausada, determinando que no parece pertinente realizar una crítica en bloque a todas las manifestaciones de esta expansión, sino que la tarea de la dogmática y la política criminal es realizar un análisis sosegado de cada una de sus manifestaciones [500] que permita deslindar la intervención legítima de la ilegítima [501].

### c) Los críticos

El discurso de rechazo abierto a la modernidad y a la expansión llega a España de forma limitada, y son pocos los autores que, abiertamente, aceptan el discurso de la escuela de Frankfurt sin mayores cuestionamientos [502]. Sin embargo, sí merece la pena resaltar que algunos académicos han puesto de manifiesto críticas a algunas de las vertientes acogedoras de la expansión que hemos mencionado.

---

[499] *Ibid.,* 201. En este sentido, también claro Silva Sánchez, J. M., *La expansión …, op. cit.* N. 383. *«Parece razonable pensar que un derecho penal mínimo sólo pueda generarse en el marco de un Estado mínimo»,* p. 56, nota al pie 118. La verdadera pregunta es qué esperamos, pues, de nuestro Estado a la vista de los problemas sociales que en la actualidad son acuciantes y requieren de soluciones.

[500] Dichos estudios pormenorizados, atendiendo a los diferentes contextos de criminalización, se han ido sucediendo, siendo pertinente que en cada uno de los ámbitos se realicen reflexiones no solo del problema genérico, sino de los concretos contextos. Así, y a título de ejemplo, al hilo de los delitos contra la ordenación del territorio: Pozuelo Pérez, L., «Notas sobre la denominada «expansión del Derecho penal»: un análisis al hilo de los delitos contra la ordenación del territorio», *RDPP,* 9, 2003, pp. 13-47. De hecho, se muestra esta autora crítica con la tendencia a no individualizar los problemas vinculados genéricamente a la expansión, siendo más conveniente atender a las necesidades específicas de cada uno de los ámbitos. Así, en Pozuelo Pérez, L., «De nuevo sobre la denominada «expansión» del derecho penal: una relectura de los planteamientos críticos», en *El funcionamiento en derecho penal. Libro Homenaje al profesor Günther Jakob*s, Universidad Externado de Colombia, Tomo II, 2003, p. 109. En lo relativo a otro de los topos más comunes (tráfico viario): Miró Llinares, F., *op cit. N*. 490. Lo mismo en lo relativo al desarrollo de los delitos medioambientales. Por todos, en un texto iniciático: Heine, G., «Accesoriedad administrativa en el Derecho Penal del Medio Ambiente», *ADPCP*, 46, 1, 1993.

[501] Feijoo Sánchez, B., *Normativización del derecho penal y realidad social… op. cit*. N. 405, p. 204: «*volver a un análisis dogmático y una crítica político-criminal con mayor detalle y concreción que permita ir delimitando las expansiones razonables del derecho penal de las irrazonables*».

[502] Podría citarse, a lo sumo, a Muñoz Conde, partidario y suscriptor directo de las opiniones de Hassemer. Hassemer, W./Muñoz Conde, F. *La responsabilidad por el producto…, op. cit.* N. 409. Aceptando, con cautelas, la inevitable modernización, Demetrio Crespo: Demetrio Crespo, E., «Del «Derecho penal liberal» al «Derecho penal del enemigo»*, RDPC*, 14, 2004, pp. 114 y ss., entendiendo que la modernización del Derecho penal es «necesaria, y por otra parte, imparable», pero que ésta debe llevarse a cabo «con escrupuloso respeto a las garantías del Estado de Derecho» (p. 114). El autor es partidario de lo que llama un «garantismo» partiendo de las ideas matizadas del «Derecho penal mínimo» de Ferrajoli. Pp. 105 y ss. Basándose fundamentalmente en los postulados de: Ferrajoli, L., *op. cit.* N. 453.

Así, Paredes Castañón se muestra especialmente crítico con la visión de Gracia [503]. Se confiesa, sin embargo, simpatizante de su punto de partida: el derecho penal necesita de una modernización para su adaptación a las nuevas realidades sociales, políticas y económicas propias de sociedades contemporáneas plurales, multiculturales y complejas [504]. Se cuestiona, sin embargo, si las respuestas a estos problemas contemporáneos pueden encontrarse en el denominado «derecho penal moderno», apuntando que los «viejos problemas» asociados a este derecho no encuentran en el trabajo de Gracia respuesta alguna. Reconoce Paredes echar en falta una concepción «menos acorde con el discurso político predominante en Europa» (al que califica de «bienestarista»), sobre lo que es y debe ser el Derecho penal moderno [505], pues sería necesario, para este autor, poner de manifiesto que este discurso de la modernización no es el único posible [506]. Pone en cuestión, finalmente, el materialismo histórico de Gracia, que da por hecho esta vinculación inescindible entre el derecho penal liberal y las características de la sociedad burguesa [507]. A Paredes la parece discutible, en primer lugar, la vinculación funcional entre derecho y sociedad, (el derecho es producto de la estructura de la sociedad), y, en segundo lugar, que de aceptar que esa vinculación existe, este concepto sociológico serviría, a lo sumo, para describir el derecho, pero no para prescribir cómo debe ser [508]. Finalmente, cuestiona Paredes la concepción evolucionista del derecho, considerando que tal vez, más que pasar a etapas donde todo lo anterior queda atrás, asistimos a un «eterno retorno» [509]. Y así cierra su crítica mostrándose reacio a la afirmación simplificada del «progreso», dando, en cierto modo, la razón al discurso de la resistencia, que sostienen (como él mismo hace) «*que el supues-*

---

[503] Paredes Castañón, J. M., «¿Al otro lado…», *op cit.,* N. 464. Las respuestas a las críticas aquí vertidas por Paredes se encuentran en: Gracia Martín, L., «Concepto categorial teleológico …», *op. cit., N. 423, passim*, y especialmente en extensa nota al pie 148.

[504] *Ibid.,* p. 447.

[505] *Ibid.,* p. 455.

[506] Alude este autor a un trabajo previo suyo: Paredes Castañón, J. L., «El riesgo como construcción conceptual: sobre el uso y el abuso de las ciencias sociales en el discurso político-criminal del «Derecho Penal del riesgo»», *Revista Catalana de Seguretat Pública*, 13, 2003, pp. 11 y ss., marcando el camino metodológico, dando por hecho que, cuando se trate de discutir la conveniencia político-criminal de castigar un determinado tipo de comportamientos (su trabajo reflexiona sobre el medioambiente, pero alude explícitamente a otros como la propiedad intelectual o industrial), el punto de partida debe ser que no es necesario vincular a un determinado género de discurso político y económico único. Paredes Castañón, J. M., «¿Al otro lado…», *op cit.,* N 464., p. 455.

[507] Ibid, p. 456.

[508] Ibid.., p. 458, *passim.*

[509] *Ibid.,* p. 459. Con más detalle en p. 460: *«Dicho con más propiedad: no existe el avance ni el retroceso desde el punto de vista histórico, sino que en cada momento la política y los ordenamientos jurídicos constituyen una combinación inextricable –pero explicable– de elementos antiguos y nuevos (en el sentido estrictamente descriptivo de ambos términos). No hay, pues, «progreso» ni «reacción»; lo que sí que hay son cambios (término neutro)».*

*to progreso que habría tenido lugar en el «Derecho Penal moderno» no es tal, sino franco retroceso o, al menos, cuestionable desviación por caminos no deseables. Cuestionan, pues, de un modo ciertamente muy post-moderno, que exista tal progreso lineal y que haya que renunciar a principios y valores «del pasado» en aras del mismo.»* [510]

Paredes se muestra reticente a afirmar la unión entre derecho penal moderno y estados sociales y democráticos [511], pues *«si alguna conexión instrumental hay que atribuir a las figuras delictivas que suelen englobarse en el "Derecho Penal moderno", parece que la misma debe ser más bien con el desarrollo del capitalismo, con la creciente multiculturalidad, con el desarrollo de las tecnologías y con la globalización. Es decir, el «Derecho Penal moderno» parece más bien el Derecho Penal de una sociedad capitalista desarrollada, abierta al mundo y multicultural»* [512].

En este punto coincidimos plenamente con Paredes cuando señala que las características esenciales de aquello que ha venido a denominarse como «derecho penal moderno» se basa mucho más en la evolución social que en la política. Se ha de atender, entonces, a factores como «el uso creciente de las tecnologías, con el desarrollo del consumo, con el predominio de la gran empresa, con la inmigración, con el pluralismo cultural e ideológico de los ciudadanos y grupos sociales, con la integración supraestatal, con la mayor facilidad de las comunicaciones, etc.» [513], y que este modelo, de hecho, es el propio no del estado de bienestar, sino de sociedades capitalistas tan paradigmáticas como la estadounidense, sin que, siquiera, se corresponda con regímenes democráticos, como son muestra las grandes potencias asiáticas [514]. La propuesta que de aquí deriva Paredes es que, entonces, el criterio para valorar lo adecuado de una determinada política criminal será si logra los objetivos propios de la sociedad. Señala, Paredes, como ejemplo, la preservación de determinadas condiciones de funcionamiento de la economía capitalista [515].

Para Paredes el verdadero debate sobre el derecho penal moderno no es resuelto por Gracia (ni por ninguno de los teóricos), pues las preguntas a las que hay que dar respuesta (las relativas a qué criterios de imputación de responsabilidad pueden aplicarse sin cuestionar garantías básicas, qué técnicas de tipificación son más adecuadas, qué relación debe tener el derecho penal con

---

[510] *Ibid.,* p. 461.
[511] *Ibid.,* p. 464
[512] *Ibid.,* p. 464.
[513] *Ibid.,* p. 465.
[514] *Idem*.
[515] *Ibid.,* p. 466.

el administrativo y qué penas son eficaces, entre otras [516]), quedan, aún pendientes, y son la verdadera tarea de este tiempo.

No podemos más que estar de acuerdo con este punto del discurso crítico de Paredes, que no nos parece tan alejado del punto de partida de quienes afirman una separación respecto del modelo de Estado liberal. Sin negar que el Estado ha de responder a un mayor auge de la protección social y de derechos más amplios, insertado como está en el paradigma del Estado social y del bienestar, al menos en las sociedades occidentales de lo que hoy se denomina norte global, no por ello estos Estados sociales (cuya crisis, por cierto, parece ser innegable), son ajenos a los postulados del capitalismo. Se puede afirmar que el Estado en que se inserta el derecho penal no es Estado liberal, pero aceptando, a su vez, que los rasgos del Estado social de bienestar y del modelo capitalista conviven en una tensión constante. Que este análisis político sea el que enmarque el discurso político-criminal sirve a los fines de entender el contexto situacional, y seguramente también los riesgos propios de cada época, amén de los bienes que interesa proteger, pero, como Paredes acertadamente destaca, deja aún muchos interrogantes abiertos sobre cómo reaccionar frente a esos ataques.

Por su parte, Díez-Ripollés toma el discurso de la modernización como antecedente necesario para lo que él considera el mal de la política criminal actual, que denomina securitaria [517]. Díez Ripollés conecta el discurso sobre la expansión del derecho penal moderno con el derecho penal de la seguridad ciudadana al señalar que este último ha parasitado las propuestas modernizadoras para justificar medidas autoritarias. Según el autor, mientras que el derecho penal de la sociedad del riesgo busca adaptarse a los desafíos de los riesgos tecnológicos y proteger bienes jurídicos colectivos, el modelo de seguridad ciudadana redirige estas herramientas hacia la criminalización de la delincuencia clásica y marginada. Esto se logra mediante la anticipación de la intervención penal, la flexibilización de garantías y un enfoque en la inocuización de ciertos grupos sociales. Sin embargo, Díez Ripollés critica esta instrumentalización, argumentando que ambos modelos (el de la expansión y el securitario, respectivamente) responden a realidades sociales e ideológicas diferentes y que, por tanto, debemos estar atentos a esta extrapolación.

Señala que el modelo penal de seguridad ciudadana prioriza la eficacia sobre principios fundamentales como la proporcionalidad y la humanidad, consolidando desigualdades sociales y abandonando objetivos resocializado-

[516] *Ibid.,* p. 470.
[517] Díez Ripollés, J. L., *op. cit*. N. 384, *passim.*

res. Además, cuestiona su efectividad a largo plazo y advierte sobre los riesgos de legitimar políticas punitivas que refuerzan un derecho penal autoritario y excluyente. Para Díez Ripollés, es fundamental distinguir entre las demandas legítimas de modernización del derecho penal y las estrategias securitarias que buscan instrumentalizarlas para justificar un control social más rígido [518].

## 2. CONCLUSIONES SOBRE LA DISCUSIÓN «CLÁSICA»

En primer lugar, aunque el tópico de discusión relativo a la ampliación o expansión del derecho penal moderno (o posmoderno) ha encontrado críticos y defensores, parece inevitable encontrar, hasta en los autores inicialmente más críticos, un cierto reconocimiento de la inevitabilidad de la creación de nuevas formas de tipificación [519]. La primigenia discusión en torno a la modernización y expansión en genérico ha ido tornando en discusiones centradas en los diferentes ámbitos de aplicación. Algunas de las discusiones que estaban presentes desde un inicio se han ido, sin embargo, reproduciendo. Es el caso, precisamente, del cuestionamiento de la legitimidad de los delitos de peligro, y concretamente de los delitos de peligro abstracto, sobre los que se discutirá *infra* [520].

En segundo lugar, todos los análisis del fenómeno político-criminal parten de una determinada forma de concebir la sociedad. Los rasgos del sistema social y de las estructuras de producción no son en absoluto ajenos a la discusión política criminal, por lo que el análisis de los postulados sociológicos

---

[518] *Idem*, *passim*. Las reflexiones en torno al modelo de derecho penal de la seguridad ciudadana deben quedar, pues, extramuros de nuestras consideraciones, en la medida en que, si bien nacen de las reflexiones de la modernidad, acaban por alejarse de sus postulados, como así también reconoce Gracia Martín, L., «Concepto categorial teleológico …», *op. cit., N. 423,* p. 44. Para Gracia no hay ningún tipo de similitud entre el derecho penal económico que castiga a los poderosos y el de la seguridad ciudadana que va dirigido a reprimir a las clases más oprimidas. «*Entre la criminalidad económica de los poderosos y la callejera de sectores de población marginada no cabe ver ningún punto específico de conexión empírica más allá de su mera coexistencia espaciotemporal en el acontecer social*», p. 44. Por ello la asociación que hace Díez Ripollés de estos dos fenómenos le parece a Gracia infundada, porque responden a modelos de Estado opuestos (el Estado de policía y el Estado social de Derecho, respectivamente), dejando claro que el derecho penal del riesgo comprende exclusivamente a los tipos penales relativos a los denominados «riesgos tecnológicos» (atómicos, nucleares, científicos, por la fabricación y comercialización de productos, por la producción laboral, etc.). *Ibid.,* p. 91. Diferencia, asimismo, Demetrio Crespo, algunos de los postulados de la modernidad como «corriente regresiva», situando como paradigma el derecho penal del enemigo que construye Jakobs. Demetrio Crespo, E., *op cit.* N.502, p. 109.

[519] Así lo reconoce Gracia cuando dice que hasta el propio Hassemer ha acabado cediendo en su discurso y que autores partidarios de un derecho de corte más naturalista como Greco han acabado por ceder a las tendencias modernizadoras. Gracia Martín, L., «Concepto categorial teleológico …», *op. cit., N. 423*, p. 17.

[520] Capítulo sexto.

acaba por impregnar la discusión penal. Si desde los años 80 la discusión de la modernidad estuvo marcada por la construcción teórica del riesgo, caracterizada como una sociedad tecnológica, moderna y arriesgada, sin duda será interesante explorar cuáles son los postulados de una sociedad como la actual, aunque sea, siquiera, para tratar de entender cuáles son las características (origen, extensión y dimensiones) de los riesgos que el derecho penal debe (o puede) afrontar.

Un cuarto de siglo nos separa de las discusiones que, allá por el inicio de siglo XXI, ocupaban la reflexión político-criminal. Resulta esencial, pues, realizar una revisión de estos postulados, con el objeto de entender cómo las nuevas estructuras de producción, de interconexión social, y de política, pueden venir a moldear el derecho penal del futuro (que, en realidad, no es más que el derecho penal del presente). Merece la pena realizar la tarea casi adivinatoria de explorar la base sociológica de la sociedad ultramoderna y ultratecnologizada y preguntarnos cuáles serán los nuevos postulados que guíen el discurso. ¿Todo puede ser tan diferente en apenas un cuarto de siglo?

## II. LA NOVEDOSA DISCUSIÓN SOBRE LA ULTRAMODERNA SOCIEDAD ALGORÍTMICA

Resulta útil comenzar el análisis de los postulados actuales de la mano de una de las voces más autorizadas en el análisis del fenómeno expansivo. Nos referimos a Silva Sánchez que, en su monumental tratado de Parte General [521], publicado en 2025, llama la atención sobre los nuevos caracteres de la expansión. Así, para Silva, el miedo ya característico de la sociedad del siglo XXI [522], se expande hacia nuevos sectores, que son los que comienzan a caracterizar a la sociedad contemporánea. Para este autor, la crisis económico-financiera de 2007 fue capaz de poner en primer plano la importancia de los riesgos patrimoniales como esenciales para la vida [523], condicionando que la discusión en las dos últimas décadas haya estado dominada por otros intereses, alejados de

[521] SILVA SÁNCHEZ, J. M., *Derecho Penal. Parte General*, Civitas, Madrid, 2025.

[522] Una reflexión teórica interesante sobre este miedo en BAUMAN, Z., *Miedo Líquido: La sociedad contemporánea y sus temores*, Paidós, Barcelona, 2007. También en SUNSTEIN, C., *Leyes de miedo: más allá del principio de precaución*, Katz Editores, Buenos Aires-Madrid, 2009.

[523] SILVA SÁNCHEZ, J. M., *Derecho Penal. Parte General, op cit*. N. 521, 5/88. En adelante, las referencias a la monografía se citarán siguiendo este estilo, recomendado por el propio autor, por número de capítulo y número marginal.

los clásicos peligros relacionados con la vida y la salud [524]. Reconoce Silva que la expansión no se debe entender como un problema de populismo o derivado del sensacionalismo, sino que responde a la *«conformación actual de las comunidades políticas»* [525]. Sin examinar, por tanto, cómo se conforman estas comunidades, poca información podremos extraer. Reconoce también Silva que nos encontramos en los albores del debate sobre el riesgo más específico de la evolución tecnológica de nuestro siglo, que no es otro que el derivado de la inteligencia artificial [526]. Integra, entonces, Silva, su concepción antropológica de la sociedad contemporánea, donde resalta el triunfo del individualismo absoluto, *«que tiñe de provisionalidad cualquier relación interpersonal»* [527]. A la ideología de la emancipación que denomina radical se une, además, el auge de las tecnologías de la información y comunicación, generando la sinergia de ambos (individualismo y aceleración), *«un marco propicio para la emoción instantánea»* [528].

Aunque Silva habla en general de la sociedad contemporánea en su reciente obra, nos será útil explorar, tal y como se ha hecho con los postulados propios de la modernidad, en un orden cronológico, los distintos postulados políticos, filosóficos, sociológicos y político-criminales que han ido vertebrando lo que aquí se ha bautizado como «sociedad algorítmica». Otros autores autorizados hablan, simplemente, de sociedad de la innovación *(Innovationsgesellschaft),* para hacer referencia a esta realidad acelerada y tecnologizada, resaltando la necesidad de despojar de la connotación exclusivamente positiva asociada a la innovación, y partiendo de la necesidad de explorar sus postulados más allá de la economía y la antropología, para realizar un acercamiento a otras esferas como la política, la social o el propio estudio de la tecnología [529]. Este es, esencialmente, este enfoque interdisciplinar que en este apartado se adopta como propio.

Resulta de obligada referencia la obra del historiador Noah Yuval Harari [530], aunque sea simplemente por la trascendencia de sus escritos, citados en nume-

---

[524] Reconoce Silva que estos tienen que ver tanto con la importancia y efectos de las crisis bancarias como las preocupaciones sobre el endeudamiento público y la sostenibilidad de las prestaciones públicas. *Ibid.,* 5/89.

[525] *Ibid.,* 5/90

[526] *Ibid.,* 5/89; el autor equipara este fenómeno con el del *big data*, caracterizando a la IA, de forma simplificada, como el «procesamiento artificial de macrodatos».

[527] *Ibid.,* 5/91.

[528] *Idem.*

[529] En este sentido, y con esta terminología: Ibold, V., *op. cit*. N. 104, p. 40 y ss.

[530] Harari es calificado por Pérez Luño como «uno de los más afamados e informados exponentes del posthumanismo». Pérez Luño, A. E., *op. cit*. N 36, p. 35 (Una versión ampliada de este artículo se encuentra en: Pérez Luño, A. E., «El posthumanismo no es un humanismo» en *Doxa,* 44, 2021, pp. 291-312)

rosas ocasiones como explicación (revestida a veces de vaticinio) de los problemas que configurarán el siglo XXI. En *Homo Deus: Breve historia del mañana* [531], Harari analiza el pasado, el presente y el futuro de la humanidad, proponiendo una visión especulativa sobre los próximos pasos evolutivos de nuestra especie.

Harari plantea cómo, a lo largo de la historia, la humanidad ha enfrentado grandes desafíos existenciales (principalmente hambruna, epidemias y guerras). Sin embargo, en la era contemporánea, el autor sostiene que dichos enemigos han sido, si no eliminados, al menos controlados en gran medida gracias a los avances en la tecnología, la medicina, la organización social y la cooperación global. En este nuevo panorama, el ser humano ya no se dedica únicamente a sobrevivir, sino que empieza a aspirar a un horizonte distinto. El proyecto del siglo XXI, según Harari, consiste en superar las limitaciones biológicas del *Homo sapiens* y avanzar hacia la condición de *Homo Deus*, una especie que persigue activamente la inmortalidad, la felicidad plena y capacidades que tradicionalmente se atribuían a los dioses.

Este nuevo impulso está motivado por el rápido desarrollo de tecnologías como la inteligencia artificial, la ingeniería genética y la neurociencia, que abren la posibilidad de modificar no solo nuestro entorno, sino también nuestra naturaleza interna. En este contexto emerge una nueva cosmovisión: el *dataísmo*, una ideología que considera el universo como un flujo de datos y entiende a los organismos como algoritmos [532]. El dataísmo privilegia el procesamiento y circulación de información por encima de valores tradicionales como la libertad individual o la dignidad humana, y sugiere que los algoritmos –biológicos o digitales– son capaces de tomar decisiones más eficientes que los seres humanos. De esta manera, los pilares sobre los que se ha construido el pensamiento moderno, como el libre albedrío o la autonomía moral, comienzan a tambalearse ante la posibilidad de que nuestras decisiones sean, en última ins-

El pensamiento de Harari es aquí desarrollado pues se encuentra a caballo entre el transhumanismo de corte más futurista (que queda extramuros de esta investigación), y los postulados sociológicos más apegados a la realidad social.

[531] HARARI, Y. N., *op. cit. N. 56*.

[532] El analista Antoni Gutiérrez-Rubí bautiza a este modelo como «datacracia», donde «el flujo de información convertido en conocimiento» constituye «el valor supremo para la toma de decisiones». GUTIÉRREZ-RUBÍ, A., «Tecnopolítica y los algoritmos», en SABARIEGO, J./ JOBIM DO AMARAL, A./ CARVALHO SALLES, E. B., (Orgs.), *Algoritarismos,* Tirant lo Blanch, Sao Paulo, 2020*,* p. 109. En este escrito, Gutiérrez Rubí analiza esta era tecnopolítica como aquella en la que la tecnología queda al servicio de la política y la democracia, donde caracteriza a la sociedad actual como «perezosa», pues el entorno algorítmico, confortable y donde las burbujas informativas nos aíslan, nuestros deseos son cada vez más previsibles y homogéneos. A su vez, la aceleración se abre paso, en una sociedad donde la interacción social hace que todo torne en un estímulo permanente, sin tiempo para la reflexión o la pausa (se habla de *fast food* conceptual). *Ibid.,* pp.105-106.

tancia, el producto de procesos bioquímicos predeterminados. Sin embargo, este cambio de paradigma no está exento de riesgos. Harari advierte que el desarrollo de superhumanos o élites mejoradas tecnológicamente podría provocar una brecha irreversible entre quienes tienen acceso a estas mejoras y quienes no, dando lugar a una nueva era de desigualdad extrema. En efecto, la combinación entre poder computacional, control de datos y capacidad biotecnológica plantea un futuro en el que los humanos podrían perder el control sobre su destino, cediéndolo a sistemas que no comparten sus valores ni su conciencia ética.

Desde luego, esta concepción casi nietzscheana de la humanidad [533] ha encontrado críticos autorizados y, sin ir más lejos, Pérez Luño, uno de los filósofos más autorizados en el campo de la tecnología en España, se ha manifestado radicalmente en contra de esta interpretación posthumanista [534]. Considera el pensamiento de autores como Harari como «tecnocientífico», y aunque reconoce que el inicio del nuevo milenio se caracteriza por la presencia innegable y avasalladora de las nuevas tecnologías y las tecnologías de la información y comunicación, es necesario cuestionar planteamientos que sugieren la superación del *homo sapiens* [535]. Le parece discutible a Pérez Luño el planteamiento de que la IA permitirá un conocimiento sin trabas, y la idea derivada de que el futuro de la libertad política deba ser fruto de decisiones asépticas e inteligentes que serán tomadas por programas inteligentes [536].

La reflexión que realiza Pérez Luño tiene en cuenta los riesgos de estos planteamientos, que comparte con Rodotá, afirmando que los vaticinios posthumanistas no le dan importancia al bienestar humano, sino a la lógica del beneficio económico. Esta revolución, que caracterizan como elitista (dirigida por una minoría de científicos y tecnólogos), daría lugar a una evidente brecha tecnológica, concluyendo que la robótica, sin valores humanistas y democráticos detrás, *«corre el riesgo de condenar a ingentes masas de población a convertirse en una muchedumbre inútil para los procesos productivos y la evolución de nuestro sistema»* [537].

La crítica central que Pérez Luño dirige a Harari es su actitud derrotista o conformista con este avance tecnológico. Bajo un manto de aparente neutrali-

---

[533] Basada de forma indudable en el modelo del superhombre sugerido por Nietzsche. Destacadamente en: Nietzsche, F., *Así habló Zaratustra*, (trad. Sánchez Pascual), Alianza Editorial, Madrid, 1972.

[534] Así, en Pérez Luño, A. E., *op. cit*. N. 36.

[535] *Ibid.,* p. 18.

[536] *Ibid.* p. 27.

[537] *Ibid.,* p. 29, con citas al trabajo de Rodotá, en el este autor se basa. Rodotà, S., «Del ser humano al posthumano», en de la Quadra-Salcedo y Fernández del Castillo, T. (Dirs.), *Sociedad digital y derecho*, BOE, 2018, pp. 87-94.

dad, Harari presenta un análisis del futuro tecnocientífico. Aunque el autor afirma no hacer profecías, sino describir tendencias actuales, en realidad asume de forma implícita las tesis del posthumanismo como verdades incuestionables. A pesar de reconocer que la inteligencia artificial aún no posee conciencia ni emociones, sostiene que la inteligencia, desligada de la conciencia, ya supera a la humana en muchas tareas complejas. Además, su obra considera a las religiones, ideologías y al humanismo como obsoletos ante el avance tecnológico, sin proponer alternativas éticas que defiendan el valor de la conciencia frente a la inteligencia artificial. En consecuencia, Pérez Luño acusa a Harari de aceptar, aunque sea de forma implícita, que la conciencia ha perdido su capacidad de guiar la inteligencia y que el ser humano está dejando de ser el agente que dirige el rumbo del mundo, sin ofrecer otras alternativas éticas [538].

Incidiendo en la lógica de la era de la información y el auge de las tecnologías, resulta fundamental examinar las obras de, al menos, dos autores del entorno angloamericano, que empiezan a poner de manifiesto ya no las bondades, sino los riesgos del masivo procesamiento de datos. Así, Frank Pasquale y Virgina Eubanks, respectivamente, comienzan a cuestionar las ventajas del progreso tecnológico y a reflexionar sobre lo sistémico de sus riesgos.

En *The Black Box Society* [539] (2015), Pasquale examina cómo los algoritmos y las prácticas opacas de las grandes empresas tecnológicas y financieras afectan a la privacidad y a la justicia en la sociedad. Su construcción parte de la premisa de que vivimos en una sociedad de caja negra *(black box),* donde tanto las empresas como los gobiernos operan con opacidad, usando tecnologías capaces de tomar decisiones cruciales que afectan a nuestras vidas, cada vez más expuestas. Pasquale cuestiona el desequilibrio entre privacidad y transparencia corporativa. Se pone de manifiesto que vivimos en la era de los datos descontrolados, donde lo que está en juego es la reputación, pues los algoritmos se usan cada vez más para tomar decisiones que tienen que ver con parcelas esenciales de la vida humana (calificaciones crediticias, decisiones laborales, segmentación de consumidores). Se pone de manifiesto también el poder de las redes en cuanto a la manipulación del pensamiento. Pone como ejemplo la crisis financiera de 2008, dando a entender que también en el sector financiero la adopción de decisiones en contextos de opacidad y complejidad

[538] Pérez Luño, A. E., *op. cit*. N 36, pp. 36-37.

[539] Pasquale, F., *The black box society: the secret algorithms that control Moner and information*, Harvard University Press, Cambridge, 2015.

ya de por sí peligrosa, ahora se ve mediado por algoritmos que perpetúan desigualdades estructurales y riesgos sistémicos [540].

También advirtiendo sobre los riesgos estructurales que esta nueva tecnología comporta, Eubanks escribe su célebre ensayo *Automating Inequality. How High-Tech Tools Profile, Police, and Punish The Poor* [541], donde realiza un análisis sobre cómo los sistemas algorítmicos se están usando para gestionar servicios públicos, advirtiendo cómo estas tecnologías no solo reproducen, sino que profundizan las desigualdades sociales existentes, especialmente hacia los sectores más vulnerables. Tras analizar algunos de los sistemas aplicados en EEUU [542] y de dar cuenta de sus fallos, concluye afirmando que la desigualdad se automatiza en tres sentidos: primero, porque perpetúa estereotipos y sesgos preexistentes; segundo, porque controla a los pobres de una forma más severa; y tercero, porque reduce la asistencia social, amparados en un manto de eficiencia tecnológica. Eubanks formula, por tanto, toda una censura al uso acrítico y descontextualizado de la tecnología, advirtiendo sobre los riesgos estructurales de estas nuevas formas de contacto de la sociedad con el poder [543], en una revisión actual de la crítica a la neutralidad tecnológica presente, como se ha visto, desde la época de Beck.

Las reflexiones sobre el impacto de la tecnología no tardaron en volverse globales, pese a la implantación algo más pausada del paradigma algorítmico en Europa. Barona Vilar reflexiona, también, en varios de sus trabajos, sobre lo que supone este paradigma en lo que ella denomina la «neomodernidad» [544]. Partiendo de una concepción explícitamente influenciada por el pensamiento

---

[540] Califica Pasquale a estos eventos con consecuencias macro y de perfiles catastróficos como «*never events*», es decir, eventos que no deberían volver a pasar nunca más. Pero, a diferencia de lo que ocurre con otros «*never events*» en situaciones de bienes jurídicos que podríamos interpretar como clásicos (él pone el ejemplo de errores obvios y graves en el ámbito de la salud), parece que en el mercado financiero este problema no acaba de tomarse en serio. Pasquale, F. *op cit*. N. 539, p. 102. Esta idea se conecta, de hecho, con los postulados de Silva, *supra,* cuando afirma que comenzamos ahora a ser conscientes de la fragilidad de otros sistemas y bienes más allá de los clásicos.

[541] Eubanks, V., *Automating Inequality: how high-tech tools profile police and punish the poor,* St. Martin's Press, Nueva York, 2018.

[542] Así, el denominado sistema automatizado de bienestar en Indiana (para gestionar el acceso a beneficios sociales), el sistema de predicción de personas sin hogar en Los Ángeles, o el sistema de evaluación de riesgo infantil de Pittsburg (destinado a predecir casos de abuso infantil).

[543] El discurso de Eubanks, así como el de Pasquale, encuentran puntos en común con el de otros autores contemporáneos del ámbito angloamericano. Es pertinente citar, a título de ejemplo, a Cathy O'Neil que, más allá de poner el foco, como hace Eubanks, en la generalización de los algoritmos en las políticas públicas, amplía la perspectiva y denuncia también los daños que estas «armas de destrucción matemática» generan en sectores públicos y privados (finanzas, educación o trabajo), haciendo una reivindicación sobre su transparencia. Así, O'Neill, C., *Weapons of math destruction: how big data increases inequality and threatens democracy,* Crown, Nueva York, 2016.

[544] Barona Vilar, S., «Justicia penal desde la globalización y la postmoderidad hasta la neomodernidad», *Revista boliviana de Derecho*, 27, 2019, pp. 20-53.

de Bauman [545], Barona considera que en el momento actual se está realizando una liquidación del marco institucional sobre el que se asentó la modernidad, propiciando un minimalismo estatal que cabalga, por exigencias del mercado, hacia la desregulación [546]. Caracteriza la autora la justicia de este momento como una justicia líquida, donde el derecho penal se enmarca en una sociedad «*sin asideros, en la que las relaciones personales, el amor, la vida, y las relaciones sociales –jurídicas– responden a unos parámetros de inconsistencia, falta de compromiso, y en las que predominan los cambios, las mutaciones metamorfósicas, la inestabilidad, la fugacidad, la falta de respeto hacia las normas, y un espíritu egoísta de supervivencia*» [547]. Esta sociedad, para Barona, que comparte en cierto modo el pesimismo existencial que veíamos en Silva, genera conformistas, pero también destructores.

Sin embargo, se muestra mucho más esperanzada la autora con la nueva etapa que ella viene a denominar neomodernidad, y que se sitúa justo después del escenario catastrofista que presenta como postmodernidad [548]. Así, Barona recorre el escenario desde el miedo líquido a la esperanza renovadora que llega con algunos cambios de paradigma sociales, reconociendo que la sociedad globalizada, aún bajo el paraguas de un pensamiento esencialmente economicista, ha ido evolucionando. Reconoce la irrupción de las redes sociales como el rasgo que impera en su construcción de sociedad neomoderna, que será aquella sociedad en la que se rescate algún componente social para introducirlo en la «agresiva sociedad posmoderna neoliberal» [549]. Algunos de los rasgos sociológicos que destaca Barona son: la consolidación de la transformación de la producción capitalista (instantáneos intercambios en cantidades ingentes sin fronteras), el consumo de masas, el culto al hedonismo, la incorporación de las redes sociales (con el consiguiente descontrol para la privacidad y de descomposición del componente identitario), y una dependencia absoluta respecto de las tecnologías (tanto en el campo industrial como personal) [550]. Para Barona, todos estos rasgos permiten reconocer el resurgimiento de los valores humanistas, pero sin dejar de

---

[545] Barona Vilar, S., «De la sociedad líquida a la justicia penal líquida. Una lectura desde la obra de Zygmunt Bauman», *Ius Puniendi. Política Criminal y Criminología,* 2017, pp. 371-397. En mayor profundidad en: Barona Vilar, S., «Justicia Penal líquida (desde la mirada de Bauman)», *Teoría y Derecho*, 22, 2017, pp. 64-91.

[546] *Ibid.,* p. 373.

[547] *Ibid.,* p. 374.

[548] En su trabajo Barona Vilar, S., «Justicia penal desde la globalización …», *op cit. N*. 544, parte de una crítica frontal a los postulados de la expansión del Derecho penal, reproduciendo las críticas que se han explicado *supra* asociadas al derecho penal expansivo moderno.

[549] *Ibid.,* p. 40.

[550] *Ibid.,* pp.40- 42. Este es el rasgo que precisamente Barona ha venido a denominar la transformación 4.0 o cuarta revolución industrial.

lado las nuevas características de una sociedad global. En ella ve Barona atisbos de cambio, en tanto el ser humano, consciente de su vulnerabilidad y de las consecuencias del mercantilismo más extremo, toma conciencia de la fragilidad del planeta, reconociendo la necesidad de solidaridad [551]. Ve Barona esperanza también en el papel de la sociedad civil frente al Estado, en una sociedad neomoderna que tiene como reto principal «*deconstruir la sociedad global desde una mirada más humana*» [552]. No niega el papel del Estado en este proceso, pero habrá de ser un Estado protagonista no solo en el ejercicio de control, sino en su papel tuitivo y garante de los intereses sociales.

Las reflexiones de corte económico más influyentes, en torno al sistema productivo capitalista, no son, sin embargo, tan alentadoras como las expectativas de Barona. La socióloga estadounidense Shosana Zuboff se convierte en una de las primeras personas en descifrar la «fórmula mágica» de las nuevas tecnologías, cuyo valor añadido y elemento diferencial es magistralmente reflejado por esta autora estadounidense en su influyente obra *La era del capitalismo de la vigilancia* [553]. La tecnología llevaba ya muchos años entre los factores que caracterizaban las sociedades, pero los entresijos de las grandes tecnológicas de Sillicon Valley que vienen a marcar el rumbo político y económico de nuestro tiempo estaban, aún, por descifrarse. ¿Qué hace a esta concreta tecnología, al nuevo paradigma de la inteligencia artificial y el procesamiento masivo de datos, tan especial?

En esta extensa monografía la autora explica cómo el fenómeno por ella bautizado como «capitalismo de la vigilancia» constituye una nueva forma de mercado, que utiliza la propia experiencia humana como materia prima, cambiando el paradigma de la creación de valor. Este modelo económico, esencialmente novedoso en cuanto instrumentaliza al ser humano hasta límites hasta ahora insólitos, amenaza valores fundamentales (entre los que se encuentran la privacidad, pero también la propia democracia o la autodeterminación). Zuboff analiza en la primera parte del libro el surgimiento de este modelo, adentrándose en los entresijos de las grandes empresas tecnológicas, detallando el proceso por

---

[551] *Ibid.,* p. 43. Continúa en p. 44: «*Rasgos de esta nueva etapa son la conciencia ecológica, la nueva búsqueda de los espiritual en corrientes orientales o cósmicas, el respeto a la diversidad, la valorización de minorías étnicas, la defensa de la igualdad real entre hombres y mujeres, y todo ello sin perder el dato de que la persona es ciudadana del mundo, un mundo que sigue interconectado a través de los medios de comunicación y las redes sociales, que han cambiado la manera de concebir las relaciones jurídicas personales, profesionales, de consumo, etc*.»

[552] *Ídem.*

[553] Las citas y referencias en adelante se harán respecto de la edición en inglés: Zuboff, S., *The age of surveillance capitalism. The fight for a human future at the new frontier of power*, Profile Books, Londres, 2019. Zuboff, licenciada originalmente en filosofía, es catedrática emérita de la *Harvard Business School*, y realizó su doctorado en psicología social en la Universidad de Harvard.

el cual empresas como Google o Facebook iniciaron este modelo aprovechando el vacío legal reinante en un contexto neoliberal que favorecía la libertad individual frente a la regulación estatal. El éxito de este negocio se basa en la transformación de datos personales en productos que sirven a la predicción, y que se comercializan en mercados de futuros conductuales [554]. Este modelo ha ido transitando del entorno digital al físico, haciendo que toda la sociedad sea objeto de control, no solo mediante la predicción de comportamientos, sino mediante su modelación, sirviendo este sistema de poder al servicio del capital para transformar espacios íntimos en zonas de vigilancia masiva.

Al menos cuatro características son extraídas por Zuboff de esta sociedad capitalista: en primer lugar, y como base de todo el sistema, la extracción y análisis masivo de datos como presupuesto; en segundo lugar, la automatización constante; en tercer lugar, la personalización de los productos y los servicios; y, por último, la continua experimentación del modelo sobre el usuario [555]. Este modelo económico aquí bosquejado presenta serias amenazas, y, más allá de la pérdida evidente de privacidad y el riesgo de manipulación y socavamiento de la democracia, Zuboff destaca la deshumanización de este modelo económico, sosteniendo que la humanidad está siendo explotada y transformada en datos comerciables, de una forma similar a como el capitalismo de base industrial explotó los recursos naturales en el siglo xx [556].

Las reflexiones sobre el impacto de la sociedad algorítmica y su relación con el Derecho penal se suceden, también, en el entorno germano. Burchard [557] llama la atención sobre lo singular de la inteligencia artificial en comparación con otros avances científicos anteriores, reclamando que no se banalice su im-

---

[554] Capítulos 1 a 3, y, especialmente, explicando lo que significa el «*behavioral surplus*» o excedente conductual.

[555] *Ibid.,* p. 8, dando una definición de lo que entiende por capitalismo de la vigilancia, con esos cuatro rasgos.

[556] «*If industrial civilization flourished at the expense of nature and now threatens to cost us the Earth, an information civilization shaped by surveillance capitalism will thrive at the expense of human nature and threatens to cost us our humanity.»,* p. 346. En p. 416 y siguientes desarrolla la idea de la instrumentalización del humano. Los pasajes del capítulo 15 son, de hecho, claves para entender la idea del capitalismo de la vigilancia en su totalidad, no solo en cuanto poder transformador de la economía, sino como clave para generar un nuevo orden social y político basado en el «poder instrumental» *(instrumentarian power),* donde lo que se busca no es manipular las creencias sino monetizar el comportamiento humano en sí mismo. Esta modernidad, descrita por Zuboff como «tercera modernidad», implica una sociedad colmena (*hive*) donde la autonomía es sustituida por la previsibilidad.

[557] Burchard, C., «Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft», *Normative Orders Working Paper*, 02/2019, pp. 3-31. También disponible en: Joerden, J. C./ Schuhr, J. C., (Coords.), *Jahrbuch für Recht und Ethik. Annual Review of Law and Ethics,* vol. 27, Duncker & Humblot, Berlin, 2019, pp. 527-555. Las referencias se entenderán hechas a la primera de las versiones.

portancia [558]. Señala este autor como una de las características esenciales de esta sociedad la desconfianza general de los sujetos, que se ven entre ellos como meros portadores de riesgo. La desconfianza se extiende también a los jueces, poniendo en duda que sus decisiones no estén sesgadas. Resulta valiosa la reflexión que realiza el autor alemán sobre los cambios que la IA puede suponer en el proceso penal, y que van mucho más allá de lo que hasta ahora se ha experimentado (alude a la anticipación de la penalización o la protección de bienes jurídicos colectivos) [559]. Corremos el riesgo de que, al banalizar la importancia de la IA, creemos una espiral de desarrollo (*Entwicklungsspirale)* en la que generemos hechos sociales sin ser conscientes de sus fundamentos normativos [560]. Uno de los ejemplos que ofrece es la confianza cada vez mayor depositada en el algoritmo de toma de decisiones (perpetuando así el efecto anclaje), y favoreciendo una desresponsabilización individual en la toma de decisiones, sobre las que se acaba culpando a la máquina [561]. Llama el autor a una reflexión previa a la implantación de esta tecnología [562], que tiene que ver con la pregunta acerca del derecho penal deseable y debe ir unida a una reflexión sobre qué tipo de sociedad queremos (¿qué derecho penal para qué sociedad? [563]). Enfatiza Burchard la idea de que se está instaurando una pérdida fundamental de confianza interpersonal, que el autor denomina generalización de la desconfianza automatizada (*Generalisierung automatisierten Misstrauens* [564]). La conclusión que alcanza Burchard es que debemos buscar justo el ideal contrario; es decir, es necesario instaurar el ideal de confianza como filosofía social para contrarrestar los riesgos de la IA.

Resulta interesante explorar, en este punto, por las similitudes que presentan respecto del pensamiento de Burchard, las ideas del antropólogo y filó-

---

[558] Es consciente el autor de que las técnicas de vigilancia predictiva no nacen al hilo de la IA, y reconoce que el paradigma del control unido al *big data policing* nace en EEUU antes, pero sí señala que la vigilancia basada en *big data* supone un cambio fundamental, principalmente porque estas nuevas técnicas apelan directamente a la ciudadanía, haciendo que, a través de predicciones conductuales cada vez más intensas, se internalice la inhibición de comportamientos. *«Je (daten )intensiver Verhaltensvorhersagen zur Kriminalitätsinhibition ausfallen, desto mehr wird der Internalisierung des Kriminalitätsinhibitionsprojekts». Ibid.,* p. 14.

[559] *Ibid.,* p. 18.

[560] *Ibid.* p. 19.

[561] Este mecanismo de desconexión o separación moral (*«moral disengagement»)* ha sido explicado por Bandura, refiriéndose al conjunto de mecanismos sociocognitivos por los cuales los individuos suspendemos nuestros estándares morales internos para participar en conductas sin experimentar culpa o autorreproche. Bandura, A., *op. cit. N.* 178, pp. 193-209.

[562] El autor se refiere más a su aplicación como mecanismo de predicción y vigilancia, aunque lo cierto es que puede trasladarse su argumento a cualquiera de los usos de estas tecnologías.

[563] *«Welches Strafrecht für welche Gesellschaft?»,* Burchard, C., *op cit*. N. 557, p. 28.

[564] *Idem,* reconociendo que en el plano procesal se convierte en una sospecha inicial automatizada general.

sofo francés Jean Lassègue y el jurista, también francés, Antoine Garapon, realizando toda una radiografía antropológica sobre la justicia digital [565]. En su primigenia obra en coautoría, los autores reflexionan sobre cómo la informática determina toda una serie de mutaciones en lo antropológico, particularmente en el derecho, pues la justicia pasa a estar dominada por el paradigma del *big data* [566]. Los autores analizan el impacto de la revolución digital como una transformación triple (simbólica, sociológica y cognitiva), donde se reconfiguran algunas estructuras profundas del derecho. Destacan al menos tres ideas: primero, la ruptura del espacio (donde lo digital elimina el encuentro físico y ritualizado, y donde no hay espacio compartido, sino mera superposición de espacios digitales [567]), segundo, se produce una aceleración de la justicia, donde el tiempo se compacta y la deliberación se reduce, pues se favorece la respuesta inmediata algorítmica [568], y, finalmente, una dimensión política, donde la descentralización es evidente [569], dando lugar a un universo que se califica de horizontal, donde la autoridad se diluye. Resulta interesante cómo Garapon explora este camino ya iniciado en su anterior monografía en su obra de 2021, *La despazializzazione della giustizia* [570], culminando su reflexión sobre la pérdida de espacio en todos los aspectos del derecho y la justicia actual [571]. Aunque a lo largo de la obra presenta hasta cuatro dimensiones (las normas, el

---

[565] Analizaremos dos de sus obras; Garapon, A./Lassègue, J., *Justice Digitale: révolution graphique et rupture anthropologique*, Presses Universitaires de France, Paris, 2018**,** y Garapon, A., *La despazializzacione della giustizia*, Mimesis, Milán, 2021.

[566] Fronza, E., «Justice digitale: révolution graphique et rupture anthropologique» De Antoine Garapon y Jean Lassègue», *En letra: Derecho penal*, 9, 2019, p. 180. Reseña publicada originalmente en Fronza, E., «Code is Law'. Note a margine del volume di Antoine Garapon e Jean Lassègue, justice digitale. Révolution graphique et rupture anthropologique, Puf, Paris, 2018», *Diritto penale contemporaneo*, dic. 2018, pp. 1-5.

[567] Resaltando Fronza en su reseña que se produce una «congelación del proceso de simbolización y del proceso político», privando al juicio de su función comunicativa y representativa» Fronza, E., «Justice digitale…», *op cit.* N. 566, p. 187.

[568] *Idem.* Se une esta idea a la señalada por muchos autores de que se da una alteración de la verificación de la verdad con prácticas que buscan la correlación, pero no la verdadera interpretación.

[569] Es una idea reiterada por los autores. En palabras de Fronza, se tiende a «*una especie de mundo pospolítico de-simbolizado y desespecializado, desprovisto de un poder que gobierne*». *Ibid.,* p. 186.

[570] Garapon, A. *La despazializzacione…*, *op. cit.* N. 565. Aunque el término «desespacialización» no está reconocido oficialmente en el castellano, se mantiene aquí una traducción literal, por lo gráfico del término, que es usado en algunos contextos académicos. Así, Rodríguez-Iglesias, I., «Desespacialización y privilegio en el discurso glotopolítico en la prensa sobre mujeres andaluzas. Implicaciones para la comunicación intercultural y el multilingüismo», en Guerrero Salazar, S., (Ed.), Lenguaje *y discurso sobre las mujeres en la prensa española*, Círculo de Lingüística Aplicada a la Comunicación 91, 2022, pp. 97-112.

[571] Es esta una obra que pretende una mejor comprensión de los «desafíos de la contemporaneidad» *(comprendere le sfide del contemporaneo),* integrando una visión basada en la reflexión filosófica, histórica, antropológica y sociológica. Así se explica en el prólogo de la obra, a cargo de Emanuela Fronza y Carlo Guarnieri.

proceso, el juicio y la valoración) que han sido víctimas de esta «*despazializzazione*» (disolución del espacio), aclara el autor que el objeto de la obra es, precisamente, encontrar remedio a esta situación, analizando la realidad jurídica en su conjunto para intentar recuperar los espacios que la tecnología está haciendo desaparecer [572]. En lo relativo a la «desespacialización» de las normas, Garapon señala cómo se ha pasado del paradigma del derecho nacional al del derecho global de carácter técnico, señalando una superposición de planos (el espacio físico y digital se superponen), produciendo una «revolución cuántica de las relaciones sociales» [573], pues el espacio se empieza a explotar para que produzca valor (alude Garapon a las técnicas de trazabilidad, vigilancia e incluso auto-geolocalización), destruyendo la distinción entre espacio público y privado. El gran triunfo de la economía es, precisamente, conseguir crear un espacio sin territorio [574]. Señala Garapon que el paradigma del tiempo también ha cambiado, pues pasa a ser un continuo. En lo relativo al proceso cívico (*la procedura*), Garapon destaca la deshumanización que implica la privación del espacio común, que priva de ser influenciado por «la percepción carnal del otro» [575], pues el anonimato nos empuja a ver en los otros lo perverso que no reconocemos en nosotros mismos, con la facilidad de establecer a otros como enemigos gracias a esa abstracción física. En lo relativo al proceso judicial, Garapon resalta (en desarrollo de la idea ya apuntada en la obra coescrita con Lassègue) el punto de no retorno que supuso la pandemia en la aceptación definitiva de la delegación en la máquina [576]. Realiza una defensa del espacio jurídico concreto, que está desapareciendo, en cuanto a su función pedagógica, pues la presencia en los rituales (judiciales) como mero consumidor y no partícipe del espectáculo, favorece una presencia cada vez más pasiva, donde se pierden las bondades de habitar colectivamente un espacio en el que más allá de mirar y escuchar, hay lugar para la asamblea [577]. Finalmente trata el espacio de la decisión (del juicio como valoración), donde, ahondando en la idea de la

---

[572] Así, en la última parte de la obra. Garapon, A. *La despazializzacione…*, *op. Cit.* N. 565**.**, p. 172.

[573] *Ibid.,* p. 50.

[574] Se basa el autor en los postulados de Marx en este punto, señalando que ya este filósofo entendió que el capital tiende a abolir todo límite espacial que se oponga al tráfico y a conquistar toda la tierra como mercado propio. *Ibid.,* capítulo 1, sección II «*La deterritorializzacione come conseguenza dell'economia neoliberale*». De estos postulados deriva el autor la idea de una «horizontalización» de la legitimidad, pues aquello que es técnicamente posible pasa a ser legítimo.

[575] *Ibid.,* p. 74.

[576] *Ibid.,* p 103.

[577] Nos parece acertadísima, en este punto, la reflexión que Garapon presenta, haciendo de la metáfora de salir de casa (necesaria en tiempos previos al *teletrabajo* y los *telejuicios*), la experiencia de salir de uno mismo y enfrentarse a la socialización que requiere el papel de ciudadano activo. El autor lo resume magistralmente con la frase «Salir de casa es salir de uno mismo» («*Uscide dalla propria casa e quindi anche uscire da sé*». *Ibid.,* p. 111.

sustitución del espacio lingüístico por el discurso del cálculo, advierte sobre las consecuencias de la aceleración de las valoraciones. Todo requiere una valoración por parte de la masa, (en forma de reseñas o *likes*), y esta masa se opone al pueblo, pues mientras el ciudadano tiene un componente propio, la masa es uniforme, donde la capacidad crítica se pierde, falta el diálogo y hay un auge de la productividad [578]. La reflexión de Garapon invita, en definitiva, a interpretar toda esta revolución espaciotemporal como una advertencia, y termina su obra con una propuesta de «re-espacialización» de la sociedad, recuperando el tiempo y el espacio compartido, como único camino para sostener lo político, lo simbólico y lo común.

Continuando el estudio de los postulados antropológicos, localizamos también autores que defienden la continuación del legado de Luhmann, en un intento de explicar la lógica de la sociedad (y, concretamente, de la delincuencia) en el tiempo de los algoritmos. Así, Teubner y Beckers, en su monografía *Three liability regimes* [579], ahondan en esta concepción al señalar los riesgos propios y característicos de la IA. Los autores introducen el concepto de *instituciones socio-digitales* para explicar cómo los algoritmos adquieren roles sociales y capacidades comunicativas [580]. Argumentan que los algoritmos no tienen por sí mismos cualidades ontológicas de actores, pero pueden ser personificados dentro de instituciones que les atribuyen capacidad de acción según sus premisas normativas. Este enfoque subraya la importancia de las prácticas institucionalizadas en la atribución de autonomía y responsabilidad a los algoritmos, destacando que estas instituciones actúan como acoplamientos estructurales entre sistemas técnicos y sociales. Una de las ideas base de la obra (la necesidad de personificación de los algoritmos), bebe principalmente de las ideas sociológicas de Luhmann [581], argumentando los autores que los algoritmos pueden ser tratados como actantes *(actants)* en lugar de actores completos, y que esto dependerá de su mayor o menor integración en las instituciones socio-digitales, aceptando que puede darse ya una comunicación entre algoritmos y humanos (si bien asimétrica [582]), donde los algoritmos pueden

[578] Ve Garapon, de hecho, la gobernabilidad algorítmica como la actualización del panóptico de Bentham, en un sistema de control de la población que ha conseguido fusionar el ideal comunista y la economía capitalista, llegando incluso a «*gamificar*» la obediencia. *Ibid.,* p. 161.

[579] Beckers, A., /Teubner, G., *op cit.* N. 111.

[580] *Ibid.,* pp. 14 y ss.

[581] También se ve influenciada claramente por la teoría del actor red de Latour. Así, en p. 25 y ss. Los autores son, sin embargo, precavidos, dando a entender que no se trata de aceptar lo que llaman la «*android fallacy*», es decir, «*the mistaken confation of the concept of personality tout court with "humanity" as such, one has to understand the peculiarity of the digital capacity for action*». *Ibid.,* p. 23.

[582] La idea es que «Los agentes de software, al igual que las empresas y otras organizaciones formales, no son más que flujos de información. Éstos se transformarán en personas en sentido estricto cuando

generar eventos comunicativos a pesar de que sus capacidades no sean idénticas a las humanas. A lo largo de la obra advierten de que, como consecuencia de esta construcción social, existen tres riesgos asociados a los algoritmos: el riesgo de autonomía; el riesgo de asociación (resaltando las crecientes interacciones híbridas humano-máquina [583]); y el riesgo de interconectividad [584] (incluso entre las propias redes algorítmicas). Para Beckers y Teubner, solo comprendiendo las atribuciones que las instituciones sociales hacen a los algoritmos se podrán enfrentar los riesgos emergentes de la IA, integrada ya como un sistema (una institución socio-digital), más.

En definitiva, podemos tomar como anclaje temporal las palabras de De Asís Pulido [585], que, aludiendo una vez más al dataísmo [586] y la algocracia como conceptos centrales, obliga a despertar del *«gran sueño analógico»,* afirmando que nos encontramos inmersos ya en los albores de una Edad Media dentro de la Era Digital [587]. Era, que, según este autor, caracterizada por la celeridad y la radicalidad de sus cambios y evoluciones, nos sitúa ya fuera del umbral de la Edad Antigua (la Digital), «*para alcanzar su medioevo: la Edad del algoritmo*» [588]. Reflexiona este filósofo acerca del dato como petróleo del siglo XXI, y en torno al dataísmo como punto de inflexión en las relaciones sociales, mediadas ya irremediablemente por el algoritmo y las redes.

Finalmente, merece la pena señalar los postulados filosóficos de uno de los autores más influyentes en el marco de la sociedad tecnológica: el filósofo

---

construyen una identidad social en el proceso de comunicación y cuando la institución sociodigital emergente crea expectativas que les atribuyan efectivamente la capacidad de actuar». *Ibid.,* p. 30, traducción propia.

[583] Ponen el ejemplo real del «*computational journalism*», o de algoritmos formando parte de juntas de dirección de empresas. *Ibid.,* p. 90. Sobre este último escenario, explorando la viabilidad de que una IA sea nombrada miembro de un consejo de administración, y con ejemplos de países donde así ha ocurrido (principalmente Hong Kong y Noruega): MOSCO, G. D., «AI and the Board Within Italian Corporate Law: Preliminary Notes», *European Company Law Journal*, 3, 2020, pp. 87-96. Ejemplos en p. 89.

[584] Identifican esta interconectividad con lo que los sociólogos cognitivistas llaman «cognición distribuida», pues se trata de procesos cognitivos sin un sujeto identificable, distribuido a través de la estructura del grupo social. Las «multitudes» (*«crowd»),* no son como un grupo, pues no desarrollan un propósito común ni una estructura organizativa. Son un «fenómeno colectivo de acciones individuales interconectadas y mutuamente responsables que se dan por una infraestructura social y material que produce resultados colectivos espontáneamente». BECKERS, A., /TEUBNER, G., *op cit.* N. 111, p. 115.

[585] DE ASÍS PULIDO, M., «La justicia predictiva: tres posibles usos en la práctica jurídica», en LLAMO ALONSO, F. H., (Dir.), *Inteligencia artificial y filosofía del derecho*, Laborum ediciones, Murcia, 2022, p. 287.

[586] Sitúa este autor este término ya en 2013. Así, en: BROOKS, D., «The Philosophy of Data», *The New York Times,* 4 de febrero de 2013*,* disponible en: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html> [Accedido por última vez el 22/01/2026].

[587] *Ibid.,* p. 285.

[588] *Ibid.,* p. 286.

surcoreano Byun-Chul Han. A lo largo de su prolija obra [589], el autor reflexiona sobre los efectos del capitalismo globalizado y la tecnología en nuestras vidas. Su enfoque crítico combina tradiciones filosóficas orientales y occidentales, abordando temas como el neoliberalismo, la tecnología, el consumismo y la autoexplotación. Algunos de las ideas más destacados de su pensamiento son:

– La idea de autoexplotación y la sociedad del rendimiento. Han sostiene que hemos pasado de una sociedad disciplinaria, donde el poder se ejercía externamente, a una sociedad del rendimiento, donde los individuos se autoexplotan bajo la ilusión de libertad. En *La sociedad del cansancio* [590], describe cómo el sujeto contemporáneo se convierte en «emprendedor de sí mismo», sometido a una presión constante por optimizar su rendimiento. Esto genera agotamiento crónico, depresión y pérdida de sentido.

– La transparencia como imperativo moral: así, en su obra *La sociedad de la transparencia* [591], Han critica cómo la obsesión por la transparencia elimina la intimidad y la negatividad. La vigilancia constante y el intercambio superficial de información deshumanizan las relaciones y convierten al individuo en un objeto expuesto.

– Una fuerte crítica al capitalismo y al neoliberalismo. Han denuncia el impacto del capitalismo global y el neoliberalismo en todos los aspectos de la vida humana, argumentando que estos sistemas fomentan un conformismo radical que reduce al individuo a un cliente o productor, anulando su capacidad crítica y su autonomía.

– Una reflexión en cuanto al tiempo. En obras como *Vida contemplativa* [592], Han aboga por recuperar el ocio y la inactividad como fuentes de creatividad y equilibrio interior. Frente a la hiperactividad impuesta por la sociedad del rendimiento, propone una vida contemplativa que permita reconectar con lo esencial.

---

[589] La obra de Han puede calificarse de prolija, aunque sus publicaciones se caracterizan por ser cortas y de estilo desenfadado, dirigidas al público general. Podemos destacar, de las publicaciones donde desarrolla su crítica contra la sociedad contemporánea: Han, B.-C., *Müdigkeitsgesellschaft*. Matthes & Seitz, 2020; Han, B.-C., *Transparenzgesellschaft*. Matthes & Seitz, 2012; Han, B.-C., *Im Schwarm: Ansichten des Digitalen*. Matthes & Seitz., 2013; Han, B.-C., *Psychopolitik: Neoliberalismus und die neuen Machttechniken*, Matthes & Seitz, 2014; Han, B.-C., *Undinge: Umbrüche der Lebenswelt*, Ullstein, 2021; Han, B.-C., *Infokratie: Digitalisierung und die Krise der Demokratie*, Matthes & Seitz, 2021.

[590] Edición en español: Han, B-C., *La sociedad del cansancio*, (trad. A. Saratxaga), Herder, Barcelona, 2018, 2.ª ed.

[591] Han, B.-C., *La sociedad de la transparencia*, (trad. R. P. Gabás), Herder, Barcelona, 2013.

[592] Nos referimos a una de sus últimas obras, Han, B.-C., *Vida contemplativa: Elogio de la inactividad* (trad. M. Alberti), Taurus, Madrid, 2023.

– El concepto de «no-cosas» (*Undinge*). En *No-cosas* [593], Han analiza cómo hemos pasado de un mundo lleno de objetos tangibles a uno dominado por información efímera. Este cambio desmaterializa nuestra experiencia y nos desconecta de lo concreto. Encontramos aquí puntos comunes con la filosofía de la desespacialización explicada inmediatamente *supra.*

Finalmente, es preciso traer a colación la visión del filósofo y sociólogo Hartmut Rosa [594]. Su «teoría crítica de la modernidad tardía» [595] se articula en torno a dos ideas: la alienación y la aceleración *(Beschleunigung und Entfremdung).* El pensamiento del autor se basa en la idea de que vivimos en una sociedad en constante aceleración, tanto tecnológica (los avances tecnológicos cada vez son más veloces), como social y vital (con un sentimiento de que nunca tenemos tiempo para nosotros [596]), que trae causa en la lógica de la competencia. Este ritmo vertiginoso provoca una alienación, que implica una pérdida de nuestra capacidad de establecer una relación significativa con el mundo [597]. Analiza, asimismo, cómo las estructuras políticas, económicas y educativas favorecen esta aceleración, denunciando que incluso los espacios que debían servir a la reflexión se ven dominados por una lógica de eficiencia, rendimiento e innovación constante [598]. Pone en cuestionamiento, por tanto, la pretendida libertad que prometen las sociedades modernas capitalistas, que, más que por la libertad, están dominadas por la temporalidad [599].

---

[593] Han, B.-C, *No-cosas: Quiebras del mundo de hoy* (trad. J. Chamorro), Taurus, Madrid, 2021.

[594] Podemos enmarcar a Rosa dentro de la conocida como «tercera generación» de la Escuela de Frankfurt, por lo que tiene pleno sentido analizar en un texto como este sus postulados, que parten de una teoría crítica influenciada por los postulados weberianos y marxistas.

[595] Así titula su obra de 2016, Rosa, H., *Alienación y aceleración: hacia una teoría crítica de la temporalidad en la modernidad tardía*, Katz Editores, Madrid, 2016.

[596] Destaca esta paradoja de la sociedad Tula Molina, señalando que «por un lado, el incremento de tecnología debería liberar el tiempo de las personas y, por otro, resulta evidente que masivamente el tiempo se vuelve cada vez más escaso», Tula Molina, F. «Hartmut Rosa (2016), Alienación y aceleración: hacia una teoría crítica de la temporalidad en la modernidad tardía, Buenos Aires, Katz, 192 pp.», Redes: Revista de estudios sociales de la ciencia y la tecnología*,* 46, 2018 p. 156.

[597] De hecho, y como destaca Tula, Rosa toma su idea de alienación de Marx, haciéndola valer respecto de cinco aspectos: «*respecto de las acciones (trabajo), de sus productos (cosas), de la naturaleza, de los otros seres humanos (mundo social) y de si-mismo*», para acabar convergiendo en una idea general: «*los sujetos persiguen fines y prácticas que les son impuestos, pero que no tienen ningún deseo real de apoyar*», Tula Molina, F. *op cit.* N. 596, p. 159.

[598] La denuncia de fondo es que la vida individual y social deja de tener un sentido, sino que se presenta como «*cambio frenético, sin dirección*». Rosa, H., *Alienacion*, *op. cit*. N. 595, p. 69.

[599] Señala el autor la paradoja de la Modernidad; «*cuanto más tiempo ahorramos […] menos tiempo parece que tenemos*», Bialakowsky, A. «Alienación, aceleración, resonancia y buena vida. Entrevista a Hartmut Rosa», *Revista Colombiana de Sociología*, Volumen 41, número 2, 2018, p. 249-259. Ante esta aceleración y esta alienación que conlleva, Rosa propone como alternativa lo que él denomina «resonancia» (título de una de sus obras más recientes), que constituye una nueva forma de vincularnos con nuestro entorno, manteniendo una relación viva, vibrante y recíproca del mundo, exponiendo su concepción de la

Finalmente, merece la pena terminar este apartado de estudio con una obra esencialmente económica, de reciente publicación, que viene de la mano de uno de los economistas que más protagonismo tuvo en la política europea en los últimos años. Varoufakis [600] señala algunas de las características de la sociedad tecnológica actual, estableciendo un puente entre teoría económica y digital. El autor utiliza un lenguaje efectista para afirmar que el capitalismo ha muerto, y que sus dinámicas se han visto sustituidas por lo que denomina «tecnofeudalismo», una nueva forma de sistema económico dominado por las plataformas digitales y las grandes empresas tecnológicas. La paradoja central de su tesis es que el capitalismo ha sido destruido por una nueva forma de capital, más poderosa, surgida en las últimas dos décadas, explicando que esta transformación se debe, principalmente, a tres factores: la privatización de internet (por parte de los gigantes tecnológicos, que él personaliza en EEUU y China), la respuesta de los gobiernos y bancos centrales tras la crisis financiera de 2008, y la pandemia del COVID-19, que generalizó el trabajo domiciliario, generando estos tres factores una nueva forma de capital [601]. Según el autor, los mercados y el beneficio, elementos centrales del capitalismo, han sido desplazados. En su lugar, aparecen las plataformas digitales, que funcionan como feudos, y una economía basada en la renta, especialmente la «renta en la nube», que es el precio a pagar por acceder a estos nuevos espacios digitales. Esta nueva estructura depende del trabajo no remunerado de lo que él llama los «siervos de la nube» (en clara contraposición con los medievales siervos de la gleba), evocando estructuras más propias del feudalismo que del capitalismo. La solución que propone Varoufakis para derrocar este nuevo tecnofeudalismo es formar amplias coaliciones que incluyan al proletariado tradicional, a los trabajadores digitales y también a ciertos sectores del propio capital subordinado. La reseña que Amiune realiza de la obra nos sitúa, ya, en un escenario de presente absoluto, pues se cuestiona, a la luz de la obra del griego, cómo «*En este nuevo modo de extracción de valor, la Inteligencia Artificial acelera y ensancha el poder de los propietarios del capital de la nube. Ellos, los Señores Tecnofeu-*

buena vida, reclamando un acercamiento al mundo más afectivo, donde los vínculos afectivo-personales llenen el vacío que la aceleración ha supuesto. Desarrollado en: Rosa, H., *Resonancia: Una Sociología de la Relación con el Mundo*, Katz Editores, Bogotá, 2019.

[600] Varoufakis, Y., *Tecnofeudalismo: El sigiloso sucesor del capitalismo*, Ediciones Deusto, Barcelona, 2024.

[601] Así lo señala Amiune, J. M., «Varoufakis, Yanis. Tecno-feudalismo. El sigiloso sucesor del capitalismo. Ciudad Autónoma de Buenos Aires», Ciclos en la historia, la economía y la sociedad, 62, 2024, p. 194.

*dales de la Era Digital, o la nueva clase dominante, decidirán el sentido y alcance de la Cuarta Revolución Industrial y la globalización del futuro*» [602].

En un mundo acelerado dominado por estas premisas, donde parece que, a golpe de parpadeo, el orden mundial geopolítico puede verse amenazado de un momento a otro, tiene sentido enmarcar estas reflexiones, y poner en contraposición esta larga línea temporal que nos ha permitido viajar desde los inicios de la modernidad hasta el presente. ¿Y todo esto, para qué?

## III. UN DISCURSO COMÚN

### 1. LOS RASGOS PROPIOS Y GENUINOS DE LA ULTRAMODERNA SOCIEDAD ALGORÍTMICA

Mediante un repaso de las ideas de los autores más influyentes de la actualidad, se ha tratado de caracterizar a la denominada «sociedad algorítmica», expresión contemporánea de la ultramodernidad, para tratar de explicar qué rasgos introducen transformaciones relevantes en las estructuras políticas, económicas, sociales y antropológicas que sustentan la comunidad jurídica. A diferencia de la sociedad postindustrial o del riesgo, esta nueva etapa se define por cuatro rasgos propios: una lógica económica diferente, un cambio antropológico, una resignificación del derecho y un cuestionamiento del poder político.

#### 1.1 **Lógicas estructurales: dataísmo, vigilancia y tecnofeudalismo**

Una primera convergencia se observa en la crítica al nuevo modelo de poder económico fundado en la extracción masiva de datos. Autores como Zuboff, con su concepto de *capitalismo de la vigilancia,* y Varoufakis, al acuñar el término *tecnofeudalismo*, coinciden en señalar que el nuevo poder económico ya no reside en el capital productivo, sino en la captura de la experiencia humana como recurso explotable. Pasquale y Eubanks amplían esta visión al advertir que la

[602] Ibid, p. 195. Amiune nos sitúa directamente en el momento presente, donde se suceden eventos como: «*la decisión de Elon Musk de comprar Twitter y convertirla en X, hasta la confrontación hegemónica entre Estados Unidos y China. Cómo la guerra de Ucrania amenaza el reinado del dólar; la muerte del sujeto liberal; las dificultades de la socialdemocracia y la urgente necesidad de recuperar la autonomía para pensar y tomar el timón de nuestras vidas, así como para recuperar y ensanchar nuestra libertad*».

opacidad algorítmica perpetúa desigualdades estructurales y profundiza la exclusión social mediante la automatización de la discriminación.

El dataísmo, tal como lo plantea Harari, actúa como ideología rectora de esta etapa, promoviendo una cosmovisión en la que los organismos son algoritmos y las decisiones humanas se conciben como procesos previsibles y replicables por máquinas.

En definitiva, como consecuencia de un nuevo modelo económico basado en la extracción de valor directamente del núcleo de mayor intimidad del individuo, se generará una nueva dinámica de ataque a los bienes más esenciales de este nuevo sujeto (a la vez consumidor y consumido) de la sociedad algorítmica.

### 1.2 Antropología del sujeto algorítmico: individualismo, rendimiento y deshumanización

La sociedad ultramoderna construye un nuevo sujeto antropológico caracterizado por el individualismo radical, la autoexplotación emocional y la fragilidad en sus vínculos. Silva Sánchez, Han y Barona coinciden en advertir que el ser humano se convierte en una unidad emocional autónoma, que transita por vínculos sociales y emocionales precarios, mediado por pantallas y estructuras algorítmicas que exigen rendimiento, transparencia y exposición constante. Han, particularmente, diagnostica una transición de la sociedad disciplinaria a la sociedad del rendimiento, donde el sujeto es emprendedor de sí mismo y víctima de un agotamiento estructural. Por su parte, Rosa aporta el concepto de *aceleración* como núcleo de la alienación contemporánea, denunciando que la velocidad impuesta por la competencia impide una relación significativa con el mundo, lo cual pretende contrarrestar con su ideal de *resonancia*.

La innegable aceleración de las interacciones sociales instituye, pues, una lógica diferente, con repercusiones en todos los ámbitos de lo humano, dando pie a postulados que, de hecho, pretenden superar lo humano para trascender la falibilidad que nos aleja de la perfección, asociada en ocasiones con el ideal del cálculo como respuesta a todos los interrogantes de la vida.

### 1.3 Desestructuración del espacio jurídico y crisis de deliberación

Un rasgo común identificado por Garapon, Lassègue y Burchard es la desestructuración de las formas clásicas del derecho. La justicia, en la era digital, se deslocaliza, se acelera y se deshumaniza. El proceso jurídico pierde su

dimensión simbólica y pedagógica, sustituido por decisiones automatizadas. Ello, a su vez, provoca una generalización de la desconfianza, no solo interpersonal, sino también institucional. El algoritmo se presenta como neutral, pero en realidad desplaza al sujeto y erosiona la autoridad tradicional. No es necesario compartir los postulados antropológicos o económicos anteriormente descritos, creemos, para afirmar como realidad que el juicio humano se está viendo sustituido, en la toma de decisiones sobre las más diversas parcelas de la vida, por el juicio algorítmico. La deliberación, el razonamiento, y la explicación se ven sustituidas, una vez más, por el cálculo.

### 1.4 **Emergencia de nuevas formas de poder y redefinición de lo político**

Autores como Barona, De Asís Pulido y Garapon alertan sobre la desaparición del espacio político tradicional y la emergencia de un poder algocrático, regido por lógicas privadas y deslocalizadas. Este nuevo orden se caracteriza por la disolución del Estado como garante, por la delegación del juicio en sistemas automatizados, y por una creciente horizontalidad aparente que diluye las jerarquías sin ofrecer verdaderas garantías deliberativas.

## 2. COMPARATIVA ENTRE LA SOCIEDAD MODERNA Y LA SOCIEDAD ALGORÍTMICA

### 2.1 **Elementos comunes**

Podemos afirmar que tanto la moderna sociedad del riesgo como la ultramoderna sociedad algorítmica parten del riesgo como categoría fundacional y central.

Si en la modernidad (Beck, Perrow) el riesgo derivaba de los sistemas industriales complejos, en la ultramodernidad es el riesgo digital y estructural (algorítmico, automatizado, inmaterial) el que se impone. En ambos modelos, el derecho penal debe responder a formas de peligrosidad no tradicionales, anticipadas y despersonalizadas.

Asimismo, ambas etapas suponen una crisis de los postulados clásicos del derecho penal liberal, lo que motiva respuestas que oscilan entre la expansión punitiva (Silva Sánchez, Gracia) y la crítica garantista (Escuela de Frankfurt, Barona, Burchard).

## 2.2 **Diferencias esenciales**

a) El origen del riesgo. En la sociedad postmoderna, el riesgo es un subproducto de la industrialización y la complejidad tecnológica (energía nuclear, biotecnología, redes industriales). En cambio, la sociedad ultramoderna lo genera desde estructuras informacionales, algorítmicas y globales, donde los sistemas de decisión automatizada crean nuevos tipos de amenaza invisibilizada y desconocida (de caja negra).

b) La configuración del sujeto. Mientras que la sociedad postmoderna denuncia la erosión de la personalidad autónoma frente a las masas, la ultramodernidad plantea un sujeto hiperconectado, fragmentado, emocionalmente inestable y sometido a un imperativo de rendimiento en una sociedad acelerada (Han, Silva, Rosa). Si la renuncia a la autonomía era, tal vez, la preocupación de la modernidad, en una sociedad en masa, la renuncia a la privacidad (hasta en el sentido más íntimo del término, con las amenazas al núcleo más íntimo de la intimidad mental), y la discriminación sistémicas son las preocupaciones más patentes de la sociedad algorítmica.

c) Los modos de control social. En la etapa anterior, el derecho penal del riesgo emerge como forma de gestión preventiva basada en el peligro abstracto. En la ultramodernidad, el control se traslada a estructuras algorítmicas que operan en la sombra: predictivas, no deliberativas y autorreferenciales. Si las decisiones sociales, directamente implicadas en la consecución del bienestar social y en la realización del ideal de justicia e igualdad, se delegan en el algoritmo, lo que ocurre es que, como vaticinaba Lessig, el algoritmo se convierte en ley.

d) Las relaciones sociales. La sociedad postmoderna se estructura en torno a la masificación y la cultura de consumo, mientras que la algorítmica responde a una lógica de vínculos efímeros e hiperindividualismo, con ausencia de una comunidad real (Barona, Rosa). Podemos aventurarnos a afirmar que, si en la discusión en torno al individualismo de la modernidad una de las preocupaciones esenciales era qué bienes sociales eran merecedores de protección penal (pues, desde postulados exageradamente individualistas, el bien jurídico había de ser únicamente personal), puede parecer que el hiperindividualismo y la desconexión social de hoy dan lugar a cuestionarnos la necesidad de proteger otros bienes.

e) La naturaleza del poder político. El poder moderno mantiene un rol estatal, aun en crisis. En la sociedad ultramoderna, se da un desplazamiento del poder hacia plataformas privadas que asumen funciones de decisión normativa (Zuboff, Varoufakis). La discusión sobre la tecnocracia, si bien no novedosa, se intensificará, favorecida por la circunstancia de que, a mayor complejidad de la

técnica, menos democratizado estará el discurso público, con el riesgo de que la regulación se torne cada vez más inaccesible al público general.

f) Representación del tiempo y el espacio. La modernidad tardía aún conserva cierta linealidad temporal y espacial. La era algorítmica impone la aceleración constante, la simultaneidad, y la desaparición de los espacios rituales de la justicia (Garapon, Rosa).

Si las propuestas de los autores estudiados pasaban por dar sentido a la vida actual afirmando la importancia de la resonancia, la recuperación de la comunidad y la vida contemplativa, nuestro objetivo aquí será dotar de sentido al derecho penal actual, tomando en consideración lo que tiene de peculiar la sociedad en la que se inserta.

Frente a la amenaza invisibilizada y desconocida de los nuevos riesgos, conocimiento y acercamiento a los mismos; frente al imperativo de rendimiento y la hiperproductividad (también normativa), garantías; frente al control trasladado a las propias estructuras algorítmicas, reivindicación del control humano; frente a los vínculos efímeros y el hiperindividualismo, reflexión conjunta y democrática; frente al desplazamiento del poder a las grandes plataformas tecnológicas, regulación; y frente a la aceleración, parsimonia [603].

Trataremos de desarrollar cómo estos antídotos se pueden expresar en las nuevas estructuras de imputación de la sociedad algorítmica.

## 3. REFLEXIÓN FINAL: ¿QUÉ LE IMPORTA ESTO AL DERECHO PENAL QUE SURJA EN LA SOCIEDAD ALGORÍTMICA?

El análisis presentado ha evitado de forma deliberada adentrarse en el discurso del control, entendido como la dimensión securitaria y preventiva del derecho penal (desarrollada por autores como Garland o Brandariz [604]), centra-

[603] Se usa aquí el término parsimonia con una connotación amplia y positiva. Del latín *parsus,* (probablemente del verbo *parcere;* ahorrar, economizar), se usa sin embargo aquí como «circunspección, templanza» (3.ª acepción del Real Diccionario de la Lengua).

[604] Aludimos aquí a todo el discurso que se ha generado en torno a la lógica del control, el cálculo y el adelantamiento de la intervención penal. El discurso, muy explorado especialmente al hilo de las técnicas de predicción de la violencia, la peligrosidad o la reincidencia, si bien encuentra puntos en común con lo aquí tratado (pues se enmarca, como es evidente, dentro de la misma época y sociedad), pone el foco en la vertiente preventiva, que aquí no se pretende explorar. Nos remitimos a los escritos más influyentes en este sentido, tanto del ámbito norteamericano (Harcourt, Simon y Feley, o Garland), como español (Brandariz, Miró o Martínez Garay). Así, y sin ánimo de exhaustividad: Harcourt, B., *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age*, University of Chicago Press, Chicago, 2007; Feeley, M./ Simon, J., «Actuarial Justice: The Emerging New Criminal Law» en Nelken, D., (Ed.), *The*

da en el castigo anticipado, la vigilancia o la tecnificación del enemigo. Si bien esa vertiente ha sido ampliamente explorada en la literatura político-criminal contemporánea, el enfoque aquí adoptado responde a otra preocupación: cómo las transformaciones sociales, económicas y tecnológicas acaban por permear, de forma más sutil pero profunda, en las estructuras de imputación penal.

En este sentido, se parte de la premisa de que el Derecho Penal no solo es instrumento de control, sino también expresión reactiva de un determinado modelo de sociedad. La pregunta fundamental que articula esta reflexión no es solo cómo se penaliza, sino a quién se puede atribuir responsabilidad, cómo se estructura la imputación, y qué lugar queda para la culpabilidad, la autoría o la autonomía decisional en una sociedad donde las decisiones están mediadas, fragmentadas o directamente sustituidas por lógicas automatizadas.

La expansión algorítmica no solo introduce nuevos delitos o modifica el riesgo: cuestiona la epistemología del juicio penal, reconfigura la narrativa de los hechos, problematiza el vínculo causal, y difumina la responsabilidad. El sujeto ya no actúa solo: actúa con, sobre, o a través de un sistema que puede ser autónomo o semi-autónomo, opaco o interpretativo, pero nunca plenamente humano.

Por ello, el foco de este trabajo será la reconstrucción de los marcos de imputación partiendo de un determinado contexto estructural, dejando de lado la denuncia sobre el hipervigilantismo y el control, que, aunque permeará necesariamente en el discurso, no será central en él. Comprender qué derecho penal puede ser legítimo en esta sociedad algorítmica exige, por supuesto, redefinir las garantías que rijan la imputación, pero consideramos valioso hacerlo partiendo de una caracterización contextual y subjetiva, conociendo cuál es el sujeto al que, en una era acelerada, despersonalizada y con muy pocas certidumbres, imputamos, juzgamos y penamos.

## IV. EL PAPEL DEL PRINCIPIO DE PRECAUCIÓN

Uno de los principios más comúnmente invocados cuando se ha tratado de afrontar el riesgo propio de la incertidumbre tecnológica ha sido el principio de precaución.

---

*Futures of Criminology*, Sage Publisher, London, 1994, pp. 173-201; GARLAND, D., *The culture of control. Cime and Social Order in Contemporary Society*, University of Chicago Press, 2001; BRANDARIZ, J. A., *El modelo gerencial-actuarial de penalidad: Eficiencia, riesgo y Sistema penal*, Dykinson, Madrid, 2016. Se puede considerar, incluso, que esta lógica de la hipervigilancia e hiperprevención encuentra una continuidad con las denuncias sobre la securitización y el paradigma del derecho penal del enemigo que, aun encontrando puntos en común con la expansión del modelo de imputación, encuentra sus propias particularidades. Consideramos ambos fenómenos dos caras de la misma moneda; control y reacción, respectivamente, ambos necesarios para comprender la política criminal en su conjunto.

La lógica precautoria parece tener, al menos intuitivamente, especial sentido cuando se trata de tomar decisiones sobre la realización de actividades de las que pueden derivarse riesgos. Sin embargo, veremos cómo ha sido objeto de polémica en cuanto a su aplicación al derecho penal de la modernidad. La discusión se ha replicado, como era de esperar, cuando se trata de afrontar la regulación penal de los riesgos de las nuevas tecnologías, incluyendo, desde luego, a la inteligencia artificial.

Estudiaremos, brevemente, el recorrido tradicional de este principio, para analizar su posible capacidad de rendimiento en un momento como el presente y si la intuitiva idea del «más vale prevenir que curar», tan asentada en el imaginario colectivo, puede ser, sin más, trasladada a la regulación penal.

## 1. ORIGEN

El principio de precaución surge originalmente en Alemania *(Vorsorgeprinzip)* como una respuesta a los desafíos ambientales y sociales derivados de la industrialización y el desarrollo tecnológico [605]. Su enfoque radica en actuar preventivamente frente a riesgos potenciales, incluso cuando no exista certeza científica completa sobre los daños. Este principio aboga por la protección del medio ambiente mediante medidas anticipatorias, considerando la falibilidad del conocimiento humano y la necesidad de salvaguardar los sistemas naturales para las generaciones futuras.

No es, sin embargo, en este país donde acaba desarrollándose, sino en la práctica europea e internacional, aunque sí es el contexto medioambiental, en torno al cual giran las primeras aproximaciones. Ha sido, sin embargo, de difícil aprehensión, pues al no existir un acuerdo general sobre su definición, no se ha llegado a integrar como principio del derecho internacional consuetudinario, pero sí ha sido reconocido en reiteradas ocasiones en resoluciones internacionales [606].

---

[605] Este principio fue introducido formalmente en 1971 como parte del *Umweltprogramm der Bundesregierung* (Programa Ambiental del Gobierno Federal) y luego consolidado en 1974 mediante la Ley de protección contra emisiones (*Bundesimmissionsschutzgesetz*). Hay una parte de la doctrina alemana que considera que, de hecho, tiene reconocimiento expreso en la Constitución (concretamente en el artículo 20.a. de la *Grundgesetz*: «Protección de los fundamentos naturales de la vida y de los animales»). Asimismo, en la sentencia del Tribunal Constitucional Federal (BVerfG NJW 1979, caso Kalkar), estableciendo la obligación del legislador de actuar para proteger de los riesgos derivados de la energía nuclear. En detalle: Ibold, V., *op. cit*. N. 104, p. 104 y ss. Sobre su génesis, en detalle, a título ejemplificativo: Martínez Pérez, E. J., «El principio de cautela en la práctica internacional y europea: concepto, naturaleza jurídica y contenido», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 17-49.

[606] Así, en el Convenio sobre la protección del medio marino de la zona del Mar Báltico (Helsinki, de 24 de marzo de 1992), el Convenio para la protección del medio marino del Atlántico nordeste (París, de 22 de septiembre de 1992), o el Convenio para la protección del Mar Mediterráneo contra la contaminación, hoy de-

También en el ámbito de la UE se ha desarrollado. De hecho, la Comisión Europea, en su declaración sobre este principio del año 2000, reconoce su entidad y establece un procedimiento para su aplicación [607], que implica una identificación de riesgos, una caracterización del peligro, una evolución de la exposición al riesgo y una caracterización completa del mismo (siempre basada en los mejores datos científicos disponibles) [608]. Solo una vez realizada esta evaluación por los expertos científicos, cabe una decisión por parte de las autoridades, tendente a establecer cuál es el nivel de riesgo aceptable.

Una de las ideas presentes en este postulado es que, con el avance tecnológico, no es posible afirmar el riesgo cero, pues esto implicaría frenar el desarrollo de la tecnología, y que lo que se busca con este principio es la adopción de medidas proporcionadas. Este principio se formalizó en el artículo 191 del Tratado de Funcionamiento de la Unión Europea (TFUE) y se aplica en la actualidad, de forma indiscutida, en áreas como el medio ambiente, la salud pública y la seguridad alimentaria [609].

## 2. CONCEPTO Y FUNDAMENTOS FILOSÓFICOS

En un intento por definir este principio de una forma más concreta, Cortina [610] habla de algunos de sus rasgos característicos. Destaca que este principio no se aplica a cualquier situación de riesgo, sino solo a las que puedan dar lugar a daños eventuales graves e irreversibles en contextos de incertidumbre científica. La magnitud de estos daños debe ser elevada, pues permite poner en entredicho la libertad de comercio. Para Cortina, tiende a ser una forma global de gestionar riesgos, caracterizándolo como una «nueva forma de prudencia» [611].

Cortina sostiene una fundamentación filosófica de este principio que parte de la lógica ciudadana, entendiendo que, aunque son los Estados quienes hacen efectivas las políticas en que se plasma este principio, son los ciudadanos los que, por medio de la opinión pública y la participación, lo configuran.

---

nominado Convenio para la protección del medio marino y de la región costera del Mediterráneo. Más referencias en: Martínez Pérez, E. J., *ibid.*

[607] *Comunicación de la Comisión sobre el recurso al principio de precaución,* Comisión de las Comunidades Europeas. (2000, 1 de febrero). (COM(2000) 1 final).

[608] Martínez Pérez, E. J., *op cit*. N. 605, p. 43.

[609] Reconoce el artículo 191 del TFUE que la política medioambiental de la Unión «se basará en los principios de cautela y de acción preventiva».

[610] Cortina, A., «Fundamentos filosóficos del principio de precaución» en Romeo Casabona, C. M. (Ed.), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, pp. 5-6.

[611] *Ibid.,* p. 6.

Y esta tarea ciudadana se relaciona, para esta autora, con la ética de la responsabilidad y las tecnologías, que considera que la responsabilidad debe entenderse necesariamente de forma relacional [612].

Por su parte, Andorno también lo relaciona con la prudencia, fundamentando su existencia en la idea de apreciar adecuadamente las circunstancias del caso concreto, evitando una de las dos situaciones extremas indeseables: o bien el temor irracional ante lo novedoso o bien la pasividad irresponsable [613]. Al igual que Cortina, y también basándose en la obra de Hans Jonas, apela a la democracia y a la responsabilidad, no solo individual, sino de la humanidad en su conjunto. Caracteriza al principio de precaución como un estándar jurídico que es más bien una guía amplia, una «regla de contenido indeterminado», pero que otorga un amplio margen de flexibilidad en cuanto a las medidas específicas a tomar, que además dependerán del contexto cultural en que se implanten.

Este autor enumera seis condiciones para que este principio opere:

1) Situación de incertidumbre acerca del riesgo: que ocurre cuando aún no se ha demostrado científicamente de forma concluyente la relación causal entre una determinada tecnología y el posible daño, pero existe una sospecha razonable de que dicha relación podría existir [614].

2) Evaluación científica del riesgo: se diferencian aquí dos etapas, la evaluación (a cargo de expertos, sobre el riesgo potencial y las consecuencias eventuales de la inacción), y la etapa de gestión del riesgo, que es de decisión política con intervención de todos aquellos que puedan tener interés en el asunto.

3) Perspectiva de un daño grave o irreversible: la evaluación de la gravedad es, en cierta parte, algo arbitraria, pero puede haber algunos criterios que orienten, por ejemplo, el carácter renovable de los bienes afectados.

4) Proporcionalidad en las medidas: quien soporta estas medidas debe ser la sociedad en su conjunto, por lo que no pueden suponer una carga excesiva.

5) Transparencia en las medidas.

6) Inversión de la carga de la prueba: no en un sentido literal, es decir, no se debe exigir a quien introduce el producto o desarrolla la actividad que

---

[612] *Ibid.,* p. 10 y ss. Bebe Cortina de al menos cuatro influencias: el imperativo categórico kantiano, la ética de las capacidades de Sen, la heurística del temor de Hans Jonas y la ética del discurso. En todo caso, otros pensadores han dado especial importancia a la responsabilidad relacional cuando se trate de encontrar respuestas a las políticas que surjan en la era de la IA. Así, Coeckelbergh, M., *op. cit.* N. 85, *passim.*

[613] Andorno, R., «Validez del principio de precaución como instrumento jurídico para la prevención y la gestión de riesgos» en Romeo Casabona, C. M. (Ed.), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, p. 18.

[614] *Ibid.,* p. 28.

elimine cualquier duda científica sobre su peligrosidad precisamente en un campo donde lo que hay es incertidumbre. Lo que exige es promover un rol más activo del introductor del riesgo en el esfuerzo de determinar el grado de probabilidad y magnitud del riesgo [615].

La reflexión más empírica (desde el punto de vista de la economía conductual), la podemos encontrar en el trabajo de Sunstein. El autor comienza reconociendo que, cuando se alude a este principio, las acepciones son muy variadas [616]. Sunstein está de acuerdo en adoptar una versión «débil» del principio, pero en este sentido dice que el reto está en hacer coincidir el nivel de evidencia con el nivel de respuesta; una evidencia muy baja del daño puede obligar a investigar más, una evidencia más fuerte a que el riesgo se haga público y una aún más grande, a que se deba regular jurídicamente.

Sin embargo, examina el fenómeno más complejo, es decir, aquel que, ante la mínima evidencia de daño, propone férreos controles. Pone Sunstein de manifiesto la realidad psicológica de que la población, en general, prefiere asumir un riesgo seguro que una probabilidad muy baja de riesgo catastrófico. Y esta tendencia de la aversión al riesgo puede hacer común aceptar una versión débil del principio de precaución, pero no una fuerte. El autor afirma que el principio de precaución en su versión fuerte llevaría a la inacción, pues es inútil como guía de actuación [617]. Sunstein atribuye la popularidad del principio a una serie de sesgos psicológicos compartidos por la mayoría de las personas. Entre ellos destaca el sesgo de aversión al cambio o *status quo bias*, la creencia en la benevolencia de la naturaleza frente a la intervención humana, el sesgo de disponibilidad, la negación de la probabilidad influida por las emociones, y la desatención de los sistemas *(systems neglect)* [618], que impide percibir los efectos colaterales de una decisión en

---

[615] *Ibid.,* p. 32.

[616] Sunstein, C., «Beyond the Precautionary Principle», *University of Pennsylvania Law Review*, 3, 2003, pp. 1003-1058. Señala el autor al menos cuatro, desde las más débiles hasta las más fuertes: el principio de no exclusión (la regulación no debe cerrarse por la ausencia de certeza científica); margen de seguridad (las actividades deben limitarse por debajo del nivel donde los efectos adversos aún no han sido identificados); mejor tecnología disponible (exigir el uso de la mejor tecnología avanzadas en actividades con riesgos inciertos, salvo que se demuestre que no existe un riesgo apreciable); y prohibición preventiva (deben prohibirse actividades con potencial incierto de causar daño sustancial, salvo que se pruebe la ausencia de riesgos significativos).

[617] Utiliza varios ejemplos sobre riesgos cuya magnitud no se conoce con certeza (arsénico en el agua potable, modificación genética de la alimentación, energía nuclear, etc.), poniendo de manifiesto que en ocasiones el principio de precaución actúa, más que como pauta de acción, como pauta de inacción, pues no es posible saber qué alternativa genera más riesgos. Así, se pregunta Sunstein si sería más acorde con el principio de precaución dejar de experimentar con nuevos fármacos por el posible daño que causen o no experimentar con ellos y dejar sin tratamiento ciertas dolencias. *Ibid.,* p. 1023.

[618] *Ibid.,* p. 1049.

el conjunto del sistema. En conclusión, Sunstein no aboga por desechar el principio de precaución, sino por superarlo mediante una evaluación más comprehensiva y racional de los riesgos. A nivel individual, esta perspectiva puede contribuir a una mejor comprensión de los problemas; sin embargo, a nivel gubernamental, el principio de precaución resulta, según el autor, inoperante como guía de acción. Por tanto, sostiene que una regulación racional de los riesgos debe incorporar medidas precautorias, pero no depender del principio de precaución en sí mismo. Propone superar los sesgos psicológicos y adoptar un enfoque racional para regular los riesgos, basado en una visión más amplia y en un análisis coste-beneficio, sin adherirse estrictamente al principio en su versión fuerte.

Finalmente, aún fuera del derecho penal, Clarke muestra una visión actualizada del pensamiento de Sunstein, donde trata la problemática del principio de precaución en futuros escenarios tecnológicos y distópicos, hablando de tecnologías entonces emergentes como la IA [619]. La idea es muy similar pues, además de aludir a cómo los sesgos cognitivos influyen inevitablemente en nuestra percepción del riesgo, señala que evitar riesgos puede generar nuevos riesgos, dando lugar a paradojas fuertes del tipo «parálisis por análisis» [620], por lo que descarta, con Sunstein, su aplicación en su versión fuerte, proponiendo que se use para fomentar una mayor sensibilidad hacia los riesgos y consecuencias sociales y ambientales, en lugar de servir como un marco normativo rígido.

## 3. LOS PRIMEROS DEBATES SOBRE EL PRINCIPIO DE PRECAUCIÓN EN EL DERECHO PENAL

Con el auge de la expansión del derecho penal, no es de extrañar que un principio que nace para la gestión de riesgos científicos termine por trasladarse a las discusiones sobre el derecho penal del riesgo. En este sentido, Mendoza señala que el principio de precaución es particularmente afín a la «*filosofía propia de la sociedad del riesgo, puesto que se trata de un mecanismo de de-*

[619] Clarke, S., «Future technologies, dystopic futures and the precautionary principle», *Ethics and Information technology*, 7(3), 2005, pp. 121-126.

[620] Pone de manifiesto Clarke lo que denomina «*the precautionary paradox*», ejemplificada con el «grey goo problem» (p. 123), un escenario hipotético en el que máquinas microscópicas autorreplicantes (ensambladores moleculares) se descontrolan y convierten toda la materia en copias de sí mismas, destruyendo el entorno. La paradoja surge porque el principio de precaución podría recomendar detener la investigación en nanotecnología para evitar este riesgo, pero al hacerlo también impediríamos desarrollar soluciones que podrían prevenirlo. Esto genera una contradicción: el principio exige tanto investigar como no investigar, lo que lo hace inviable como herramienta de política.

*fensa frente a riesgos eventuales todavía no presentes, que pretende anticiparse a todo peligro imprevisto no conocido y, por tanto, no dominable*» [621].

El principio de precaución ha sido analizado por detractores y defensores al hilo de contextos delictivos variados: paradigmáticamente el fraude de alimentos, el delito medioambiental y los delitos genéticos. Veamos, brevemente, la discusión del papel que en la dogmática penal ha tenido en un momento muy concreto, para analizar finalmente cómo estos postulados se encuentran, una vez más, presentes en la discusión actual acerca de cómo regular penalmente el riesgo tecnológico derivado de la IA.

El concepto que toma como punto de partida la discusión penal bebe de las definiciones arriba propuestas, siendo los denominadores comunes de este principio la actuación ante riesgos inciertos (o, al menos, no demostrables con certeza científica) donde falla la construcción clásica de la previsibilidad (es decir, se trata de contextos de insuficiencia epistémica). Cabe matizar que la mayoría de los autores señalan las diferencias con una institución dogmática similar, pero que no comparte los mismos rasgos: el riesgo permitido. La lógica de la precaución se diferencia, en esencia, del riesgo permitido, en que este último se basa en un paradigma de riesgos conocidos, mensurables y previsibles [622].

Cuando se analiza la discusión clásica sobre este principio dentro del derecho penal, al menos se pueden identificar tres funciones: la prohibitiva (el principio de precaución debe prohibir ciertas conductas en ausencia de certeza sobre su inocuidad); la proactiva/positiva (el principio debe actuar como impulsor de medidas de cautela); y dogmática (este principio apoya la creación o interpretación de tipos penales de cierto tipo). Las tres funciones tienen implicaciones diferentes y se entremezclan en las críticas y defensas de los autores.

---

[621] Mendoza Buergo, B., «Principio de precaución… *op. cit*. N. 467, p. 440.

[622] Así, por ejemplo, Romeo Casabona, aunque señalando que ambas instituciones tienen un objetivo común, que es la prevención del daño temido. Romeo Casabona, C. M., «Aportaciones del principio de precaución al Derecho penal» en Romeo Casabona, C. M. (Ed.), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, p. 388. También Alonso Álamo, M., «¿Riesgos no permitidos? Observaciones sobre la incidencia del principio de precaución en el Derecho penal», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estad*o, Tirant Lo Blanch, Valencia, 2014, diferenciando el ámbito de uno y otro en función de que el riesgo sea cierto o no, pp. 51 y ss. Consideramos, en este punto, más acertadas aquellas posturas que asimilan, en parte, el principio de precaución y el riesgo permitido, con el siguiente matiz: dado que el objetivo preventivo del derecho penal es evitar el riesgo (cuya estimación siempre implica que, a partir de los datos conocidos estimamos lo desconocido), la diferencia entre ambos estriba, simplemente, en el nivel de conocimiento con el que contamos en un determinado momento. Es decir, en el escenario del riesgo permitido normalmente nos basamos, para emitir el juicio de riesgo aceptable, en una repetición de situaciones previas que han dado lugar a la materialización de un riesgo, lo que nos permite afirmar que lo «conocemos», aunque nunca lo conozcamos por completo, porque contamos con más datos. En el caso de la precaución, por el contrario, los datos y nuestro conocimiento son algo más incompletos, pero se trata, en todo caso, de una cuestión de grado.

### 3.1 El encuadre dogmático del principio de precaución dentro del derecho penal

Se analizan, en esta sección, los argumentos relativos a la tercera de las funciones posibles del principio de precaución, que se ha calificado de dogmática. La discusión sobre cuál es el papel que puede tener el principio de precaución dentro de las diferentes categorías dogmáticas de la teoría del delito se ha reflejado en tres aspectos:

#### a) Delitos de peligro

Según una parte de la literatura, el principio de precaución puede ayudar a justificar la creación o interpretación de los delitos de peligro pues, ante riesgos que no son previsibles *ex ante,* cuya materialización no puede descartarse, se hace necesario tipificar esta hipótesis de peligrosidad. Con esta propuesta, que ha encontrado fuertes críticas, se pretenden demostrar las bondades de la tipificación de delitos simplemente por la peligrosidad que suponen, sin que sea necesario mostrar una relación de causalidad entre conducta y resultado que será difícil [623]. Baste dejar apuntado este problema por el momento, pues se estudiará en profundidad *infra,* en el capítulo relativo a los delitos de peligro, al examinar las críticas a los delitos de peligro presunto o hipotético.

#### b) Delitos imprudentes

La idea es que el principio de precaución puede tener algún papel a la hora de determinar la previsibilidad objetiva de la conducta. La propuesta, en este caso, implica reformular el deber de cuidado, de tal forma que este se vea conculcado simplemente con la mera sospecha científica fundada [624]. El peligro evidente de esta propuesta es que puede acabar por sustituir el juicio de previsibilidad por la mera sospecha.

---

[623] Así, Romeo Casabona, C. M., *ibid: «El principio de precaución en relación con los comúnmente llamados delitos de peligro, - fundamentalmente los delitos de peligrosidad o de acción peligrosa - podrían contribuir a la elaboración de un procedimiento de imputación para algunos de estos delitos, basándose para ello en la innecesaridad de imputar un determinado resultado, y pudiendo prescindir entonces y en consecuencia de la constatación de un nexo causal». Ibid.,* p. 406.

[624] *Ibid.,* p. 409: «*el principio de precaución podría ofrecer un paradigma diverso, al menos en ciertos ámbitos, consistente en que la no previsibilidad no sería ya necesariamente equivalente a la negación de una conducta imprudente*».

c) Causalidad

Aunque el principio de precaución no exige suprimir la prueba del nexo causal, sí ayuda a reconocer que en ciertos contextos esta causalidad no es fácilmente demostrable. El principio de precaución vendría, en este caso, a favorecer una valoración basada en indicios o conocimiento disponible más que en certezas científicas.

Las posturas más críticas se han centrado en resaltar que existe un riesgo evidente de que este principio, al colarse entre las categorías dogmáticas arriba descritas, acabe por sustituir también otros principios esenciales como la lesividad, la culpabilidad o incluso la presunción de inocencia. Es una crítica común la que tiene que ver con la tensión con el *principio in dubio pro reo* [625].

Por el contrario, las posturas más favorables lo ven como necesario para afrontar los riesgos sistémicos, catastróficos o irreversibles (tales como el medio ambiente o la salud global), pudiendo integrarse como un criterio interpretativo de la peligrosidad, no como fundamento exclusivo de punición.

### 3.2 **Las posturas críticas**

El principio de precaución ha suscitado un intenso debate doctrinal, especialmente en su traslación al ámbito penal. A continuación, se analizan las posturas más escépticas, que alertan sobre los riesgos que supone una incorporación acrítica de este principio en un sistema penal fundado en garantías.

Massaro señala que la función constitucional del derecho penal exige que cualquier restricción de derechos se base en juicios de peligrosidad empíricamente fundamentados, algo que resulta imposible en los contextos donde opera la lógica de la precaución. Critica, además, una de las posibles consecuencias de trasladar el principio de precaución de la infracción administrativa a la infracción penal: que la infracción de simples formalidades administrativas (como la falta de autorización) sea elevada al rango de conducta penalmente punible, sin una relación próxima y demostrable con el bien jurídico protegido [626].

---

[625] En este sentido, Corcoy: «El principio de precaución es exactamente el principio contrario al *in dubio pro reo* puesto que supone la obligación de intervenir en caso de duda sobre la peligrosidad, mientras que el *in dubio pro reo* supone que en caso de duda sobre la nocividad hay que absolver, es decir, no intervenir», Corcoy Bidasolo, M., «La lección de la experiencia española en materia de responsabilidad por el producto del caso de la colza a nuestros días», en Foffani, L., (Coord.), *La sicurezza agroalimentare nella prospettiva europea. Precauzione, prevenzione, repressione*. Ed. Giufre, Milán, 2014, p. 442.

[626] Massaro, A., «Principio di precauzione e diritto penale: nihil novi sub sole?», *Diritto Penale Contemporaneo*, 9 mayo 2011, *passim.*

Alonso Álamo, por su parte, se muestra abiertamente escéptica sobre la validez penal del principio de precaución. Reconoce que este principio puede tener una función legítima en sede administrativa o política, pero no como criterio estructurador de la tipicidad penal. Además, denuncia una «intromisión del derecho administrativo en el derecho penal» al penalizar conductas que en realidad constituyen meras desobediencias normativas, sin prueba suficiente del peligro material [627].

En el caso de Aires de Sousa, podemos encontrar una evolución en su forma de aceptar este principio. En sus primeros textos sobre responsabilidad por el producto se mostraba especialmente recelosa, defendiendo que la sanción administrativa es el cauce más adecuado para hacer operativo el principio de precaución. Consideraba la autora que este principio es insuficiente como fundamento inmediato de una incriminación, ya que se basa en un juicio de plausibilidad más que de previsibilidad, lo que desnaturaliza el modelo clásico de imputación penal, fundado en la necesidad de establecer un nexo entre conducta y resultado [628].

Por su parte, Sola Reche, partiendo de una visión clásica del injusto penal, donde la peligrosidad de la conducta debe derivarse del conocimiento de los procesos causales, se pregunta: ¿Es legítimo que una conducta pueda ser constitutiva de delito cuando el nivel de conocimiento de los procesos causales es tan limitado que no permite ni siquiera afirmar su peligrosidad? [629]. Aunque admite que el principio puede cumplir funciones distintas (funciones limitativas, prohibiendo conductas, y garantistas, exigiendo medidas de prevención de riesgos) alerta de los peligros de trasladar al derecho penal una lógica de pro-

---

[627] Alonso Álamo, M., *op cit*. N. 622. La autora se refiere de forma muy crítica a lo que llama «delitos de riesgo» (que no son solo los específicamente tipificados como tal, sino también los relacionados con la salud pública, la seguridad e higiene de los trabajadores y los relativos a energía nuclear y radiaciones ionizantes). Los caracteriza como aquellos en que se ha administrativizado el sector y también hay delito previsto para cuando se incumple la normativa administrativa, por lo que la autora señala que son delitos donde no solo se castiga un riesgo incierto sino también la mera desobediencia, a lo que se suma que en algunos casos no hay un bien jurídico identificado a proteger, por lo que se protege en última instancia el miedo. El único papel del principio de precaución que considera aceptable es el de criterio interpretativo para excluir la tipicidad (función negativa o de exclusión de la tipicidad, p. 64).

[628] Así, en: Aires de Sousa, S./Moniz, H., «Manifestações do princípio da precaução no direito português», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estad*o, Tirant Lo Blanch, Valencia, 2014, p. 372. La razón de Aires de Sousa parte de una idea previa, según la cual se altera el paradigma penal, pues no hay peligro de lesión de bienes jurídicos, sino un delito de desobediencia formal, concluyendo esta autora que hay una incompatibilidad ente el modelo de precaución y el modelo de previsibilidad necesaria para la intervención penal. Más desarrollada su postura inicial en su monografía: Aires de Sousa, S., *Responsabilidade Criminal pelo Produto e o Topos Causal em Direito Penal. Contributo para uma protecção de interesses do consumidor,* Coimbra Editora, Coimbra, 2014.

[629] Sola Reche, E., «Principio de precaución y tipicidad penal», en Romeo Casabona, C. M., (Coord.), *Principio de precaución, biotecnología y derecho,* Comares, Bilbao-Granada, 2004, p. 479.

hibición por ignorancia, esto es, castigar conductas simplemente por no haber demostrado su inocuidad, pero acaba reconociendo que sí puede ser conforme con el principio de intervención mínima cuando el desvalor de la acción penal se derive de la infracción de una disposición (debe entenderse administrativa), inspirada en el deber de precaución [630].

Es crítico a este respecto Feijoo, señalando la tibieza del argumento de Sola Reche, al negar ninguna cualidad a este principio como fundamento de los delitos de peligro abstracto, pues, bajo su concepción *«El derecho penal sólo puede prevenir riesgos futuros sobre la base de retribución de efectivos riesgos pasados de los que la sociedad tiene que aprender, mientras que el derecho administrativo es el encargado de la prevención de situaciones de riesgo o de inseguridad. Si quiere ser legítima, la pena estatal tiene vetada la posibilidad de punir un hecho que individualmente considerado carezca de lesividad o peligrosidad para otros ámbitos de organización»* [631].

Se muestra crítico, también, Díaz-Maroto, en este caso con la interpretación que del principio ha venido realizando la jurisprudencia [632]. Las ideas que se derivan de este análisis jurisprudencial es que se opta por reconocer que, incluso en presencia de incertidumbre sobre el riesgo concreto que se puede derivar de una determinada sustancia, puede ser legítimo adoptar medidas penales cuando hay una sospecha razonable de riesgo, y de esa presunción derivar la imputación penal. Para Díaz-Maroto esta forma de interpretar el papel del principio de precaución (al aplicar delitos de peligro abstracto) es radicalmente incorrecta, pues «*hace que se identifique el injusto con la mera infracción de un deber, calificación que se atreve a dar el TS cuando dice de forma categórica que "Se trata de un delito que puede ser considerado de infracción de deber y que, como tal, se comete por quien está obligado a su cumplimiento"»* [633]. En definitiva, concluye el autor, el principio de precaución, que puede resultar válido en el ámbito

---

[630] *Ibid,* p. 486.

[631] Feijoo Sánchez, B., «Normativización del derecho penal…» *op. cit* N. 405, p. 206. Finalmente, en p. 214: «*Mediante el principio de precaución se incide en una política criminal de la seguridad (de bienes) como idea directriz y de esa manera la expansión del derecho penal resulta imposible de frenar, siendo cada vez más evidente que la pena queda desligada de la culpabilidad por el hecho y se convierte en un instrumento de política social, es decir, en una sanción administrativa cualificada*».

[632] Díaz-Maroto y Villarejo, J., «Sociedad del riesgo y fraudes alimentarios», en AA. VV., *Un modelo integral de Derecho penal. Libro homenaje a la profesora Mirentxu Corcoy Bidasolo*, BOE, Madrid, 2022, pp. 1099 y ss., analizando la interpretación que, principalmente en las figuras de los así llamados «fraudes alimentarios» ha hecho la jurisprudencia española sobre el principio de precaución, tomando como referencia la primigenia sentencia del Tribunal de Justicia de las Comunidades Europeas de las «vacas locas» (STJCE 1998/80, de 5 de mayo de 1998). Una referencia detallada y actualizada a las sentencias donde se aplica este principio, tanto del TS como de Audiencias Provinciales, en nota al pie 15.

[633] *Ibid.,* p. 1102.

del derecho civil y administrativo, no puede trasladarse al derecho penal. Mendoza, también crítica con la interpretación jurisprudencial del principio afirma que estas sentencias: *«Señalan el relieve que la jurisprudencia del Tribunal de Justicia de la Comunidad Europea ha dado al principio de precaución para garantizar un alto nivel de protección de la salud pública, que justifica que se tomen medidas legislativas protectoras aun cuando exista incertidumbre respecto a la existencia o importancia de esos riesgos»* [634].

Finalmente, Pérez Alonso se muestra rotundo en cuanto a la compatibilidad de los postulados del principio de precaución con las garantías penales, *«por cuanto que supone la anticipación del Derecho Penal como medio de cautela frente a futuros e inciertos riesgos y el abandono de los criterios penales de la imputación objetiva y subjetiva del comportamiento típico»* [635], afirmando que solo podría ser admisible ante un cambio radical de paradigma científico. Le reconoce, sin embargo, un papel al principio. Se rechaza la existencia de «delitos de precaución», entendidos como la prohibición general de conductas arriesgadas sin evidencia de daño, ya que implicarían una anticipación excesiva del derecho penal. No obstante, el principio de precaución sí se reconoce como útil para interpretar las conductas peligrosas en los delitos de peligro abstracto [636].

### 3.3 **Las posturas favorables**

Romeo Casabona propone una lectura integradora del principio de precaución, sin contraponerlo a la prevención, sino como una prolongación necesaria de esta en contextos de incertidumbre. Reconoce que, en los casos donde no hay previsibilidad objetiva, el principio puede servir como fundamento auxiliar de imputación, especialmente en delitos de peligro. Sostiene: «*Podría ayudar a construir una acción peligrosa al margen del criterio de partida de la previsibilidad, pero con semejantes, si no superiores, presupuestos de seguridad jurídica basados en las prescripciones de conducta que tal principio impone»* [637]. Para este autor, la precaución no erosiona las garantías, sino que

---

[634] Mendoza Buergo, B., *El derecho penal en la sociedad del riesgo*, *op. cit.* N. 384, nota al pie 73.

[635] Pérez Alonso, E., *op. cit.* N. 488, p. 408.

[636] *Idem*. Y en p. 412, respecto de su papel limitado: «*el principio de precaución no puede ser el sustituto del riesgo permitido para evaluar la peligrosidad de la conducta propia de un delito de peligro abstracto, pues en tal caso como hemos señalado, se estaría sustituyendo el delito de peligro abstracto por un "delito de precaución"».*

[637] Romeo Casabona, C. M., «Aportaciones del principio de precaución …», *op cit*. N. 622, p. 406.

traslada el centro de gravedad del análisis desde la constatación del daño hacia la adopción de conductas normativamente seguras en presencia de riesgo no descartable.

Mendoza es quizá la autora más autorizada en la defensa de una función proactiva del principio dentro del derecho penal, al cual caracteriza como «principio orientador en las políticas de gestión de riesgos» [638]. Su tesis es que, si bien no puede servir para suplantar los requisitos clásicos de la imputación penal (señalando una tensión inherente con principios como el de intervención mínima y última ratio), sí puede contribuir a definir el contenido del deber de cuidado, en particular en ámbitos donde el conocimiento técnico es fragmentario o está en evolución. Distingue esta autora entre la función prohibitoria y la función proactiva o positiva del principio, inclinándose claramente por esta última. Lo que es jurídicamente razonable, por tanto, es establecer medidas de control en la realización de determinadas actividades, y el papel del derecho penal puede empezar «a través de los correspondientes tipos penales en blanco», cuando se produzcan *«graves infracciones a las prohibiciones o controles que, en atención al principio de precaución, se establezcan en el orden administrativo»* [639]. Sin embargo, Mendoza va algo más allá de esta función proactiva (o de integración o interpretación adecuada del desvalor de los tipos) y desarrolla, además, una propuesta concreta para integrar el principio de precaución en la dogmática penal, en los delitos imprudentes, admitiendo que la peligrosidad de la acción puede ser redefinida en términos de seria posibilidad de riesgo, aun sin certeza *ex ante*. Reconoce la autora que para la imprudencia resulta irrenunciable el elemento de previsibilidad y evitabilidad, pero la previsibilidad *«se puede construir sobre la base de una sólida conjetura o hipótesis científica todavía no suficientemente contrastada desde el punto de vista científico, pero cognoscible por aquel a quien se quiere exigir responsabilidad, porque se integra en una norma de cuidado; y ello tanto cuando se trate del deber de omitir conductas, como de actuación cumpliendo con ciertos controles o medidas de precaución establecidas normativamente»* [640].

Así, la autora defiende un modelo de responsabilidad que no renuncia a la lesividad ni a la exigencia de imputación objetiva, pero que amplía sus márgenes operativos ante la complejidad de los riesgos contemporáneos [641].

---

[638] Mendoza Buergo, B., *El derecho penal en la sociedad del riesgo*, *op. cit.* N. 384, p. 442, pero solo puede operar, según esta autora, cuando se trate da daños graves y/o irreversibles contra bienes jurídicos de la mayor importancia y no renovables.

[639] *Ibid.,* p. 451, con referencia a la STS 1442/2002, de 14 de septiembre, FJ2.

[640] *Ibid.,* p. 455.

[641] Su postura respecto del papel de precaución en los delitos de peligro, a partir de p. 457, se examinará *infra*. La amplitud que Mendoza acaba otorgando a este principio, a pesar de afirmar que debe

Gómez Tomillo [642] adopta una posición firmemente favorable al principio, considerando que su utilización es imprescindible para abordar ciertas amenazas emergentes, especialmente aquellas que comprometen la vida colectiva, la salud pública o la supervivencia humana, por medio de la técnica del delito de peligro. Defiende la existencia de los que denomina «delitos de peligro hipotético» o «delitos de riesgo normativo», definidos como aquellos en los que el legislador prohíbe una conducta sobre la base de una sospecha científica razonable, aun sin prueba empírica concluyente. Para Gómez Tomillo, hay una exigencia de proporcionalidad inversa: cuanto mayor es la magnitud del daño potencial, menor debe ser la exigencia de certidumbre para activar la respuesta punitiva. Desde esta lógica, sostiene que el derecho penal no puede permanecer inerte ante riesgos que, aunque no estén plenamente demostrados, podrían tener consecuencias catastróficas [643].

Finalmente, Corn aporta una defensa abiertamente política del principio de precaución, al afirmar que su aplicación no debe subordinarse a la dogmática penal tradicional, sino que debe responder a un imperativo de protección efectiva. Reivindica el carácter legítimamente político del principio, y afirma que negarle entrada al ámbito penal por su falta de adecuación a la teoría clásica sería una forma de hipocresía jurídico-política. [644] Así lo expresa: *«¿Es o no lógico parar y castigar a los que podrían causar un daño catastrófico e irreversible para una multitud de personas y sin embargo quieren seguir con sus actividades de lucro particular justificándose en la falta de unanimidad de la comunidad científica?»* [645]. Reconoce, eso sí, que la aplicación del principio exige la concurrencia de daño potencial de gran magnitud e irreversible (donde el derecho administrativo no pueda ser el único llamado a intervenir), y exige, asimismo, una aplicación esencialmente democrática de este principio. El modelo que defiende Corn para plasmar de la mejor manera este principio es el de «*un tipo penal de*

operar en estricto respeto de las garantías penales, le vale la crítica de Feijoo Sánchez, B., «Normativización del derecho penal…» *op. cit* N. 405, pp. 213 y ss.

[642] Gómez Tomillo, M., «El principio de precaución en el Derecho penal. Contribución político-criminal a los delitos de riesgo normativo», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estad*o, Tirant Lo Blanch, Valencia, 2014, pp. 81-122.

[643] Defiende Gómez Tomillo, de manera explícita, la necesidad de que sea el derecho penal, y no el administrativo, el que aborde esta penalización, considerando que es una cuestión de diferenciación gradual, pues el administrativo es una manifestación más del *ius puniendi*, si bien con menos garantías, y menos herramientas para que el Estado investigue, señalando también que los principales protagonistas son las personas jurídicas y que la sanción administrativa sería insuficiente. *Ibid.,* pp. 112 y ss.

[644] Reconoce, sin embargo, que su postura es minoritaria entre la doctrina italiana.

[645] Corn, E., «El tímido presente y el futuro incierto del principio de precaución en la parte especial del derecho penal italiano: organismos genéticamente modificados y otras aplicaciones», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estad*o, Tirant Lo Blanch, Valencia, 2014, p. 294.

*ofensividad intermedia entre delitos de resultado e ilícitos de simple relevancia administrativa, en que el contenido específico de la violación es determinado concretamente por la autoridad respecto de la específica conducta de riesgo»* [646].

En definitiva, la postura de los autores favorables no es uniforme, pero converge en algunos puntos clave:

– No proponen eliminar el principio de lesividad ni de culpabilidad, sino adaptarlos a contextos de incertidumbre.

– Considera que la falta de certeza científica no puede paralizar la acción del derecho penal ante peligros graves e irreversibles.

– Apelan a la función orientadora y estructuradora del principio, más que a su uso como fundamento autónomo de punición, rechazando, en general, la existencia de presunciones de peligrosidad como fundamento de los tipos penales.

– Admiten que la incorporación del principio debe realizarse bajo límites normativos precisos, como el deber de fundamentar la sospecha en conocimientos científicos disponibles, la relevancia del bien jurídico y la adecuación de las medidas adoptadas.

Estas posturas reflejan un cambio de paradigma en la forma de concebir la imputación penal en un mundo en el que los riesgos ya no se definen exclusivamente por su manifestación empírica, sino por su potencialidad estructural y sistémica.

## 4. LA VUELTA AL DEBATE DE LA PRECAUCIÓN AL HILO DE LA IA

Con la irrupción de tecnologías disruptivas como la inteligencia artificial, la controversia sobre el principio de precaución, que había quedado confinada a los supuestos arriba analizados, vuelve a reabrirse, reproduciendo de alguna forma el debate ya iniciado, planteando la posibilidad de que, trasladado al ámbito penal, este principio pudiese operar en estos novedosos contextos, donde los daños son paradigmáticamente inciertos, debido esencialmente a la fal-

[646] Sigue, en este sentido, las propuestas de Pongiluppi, C. «Riflesione sulle funzione della pena nei contesti precauzionali» en Foffani, L., (Coord.), *La sicurezza agroalimentare nella prospettiva europea. Precauzione, prevenzione, repressione*. Ed. Giufre, Milán, 2014, pp. 549-558, partiendo de una idea básica de no distinción cualitativa entre derecho administrativo sancionador y derecho penal, propugnando un modelo «*ingiunzionale*», más similar a nuestro esquema de normas penales en blanco que de delitos de peligro abstracto, aludiendo a la mayor inseguridad jurídica que supone la interpretación de estos últimos.

ta de conocimiento no solo sobre los efectos, sino sobre el propio funcionamiento de los sistemas, pero potencialmente catastróficos, en un sentido algo particular.

El concepto que se maneja es exactamente el mismo, por lo que aplican las mismas consideraciones arriba señaladas tanto respecto del concepto como de sus fundamentos filosóficos [647]. Tomemos la definición que ofrece Galán Muñoz, que, partiendo de la base de que podemos distinguir entre concepciones fuerte y débiles del principio, considera como punto básico que «*el Derecho podrá intervenir limitando la libertad de los ciudadanos, estableciendo restricciones, prohibiciones e incluso previendo sanciones, en relación a aquellas conductas respecto de las cuales exista sospecha o indicios de que pueden dar lugar a la producción de resultados catastróficos y/o irreversibles para el medio ambiente o la salud de las personas, por más que exista incertidumbre científica sobre la realidad de dicho riesgo y su extensión*» [648], reconociendo la incidencia de este principio en tecnologías de todo tipo: *«[…] la nuclear, las biotecnológicas, las informáticas, las relativas a la nanotecnología o incluso la utilización conjunta de varias de ellas*» [649].

Autoras como Aires de Sousa han examinado su incidencia como herramienta de regulación de las tecnologías de inteligencia artificial, reconociendo que dicho principio sí tiene un lugar en este campo como herramienta de regulación en la concepción, desarrollo y uso de productos inteligentes, y que sus medidas deben ser reversibles y flexibles a medida que los riesgos se controlan y gestionan [650]. En un trabajo anterior ya justificaba esta autora la lógica de la precaución en el campo de la IA, pero estructurado como un principio normativo. Así, considera que en los usos de IA fuerte (paradigma de *strong AI*), existe una verdadera incertidumbre sobre su funcionamiento, que se asocia a una cuasi presunción genérica de peligrosidad. Localiza Aires De Sousa hasta tres situaciones en las que, a su juicio, se puede encontrar un reflejo de la lógica de la precaución en decisiones penales donde, a la vista de que se podía producir con algún grado de probabilidad o previsibilidad ciertos daños, sin un conocimiento demasiado detallado sobre las características del riesgo, se ha producido o propuesto una prohibición penal. Estos tres ámbitos son: la prohibición de armas autónomas, prohibiciones de algunos países de la Unión Eu-

---

[647] Así, Aires de Sousa se remite directamente a la concepción de Adela Cortina, *op cit*. N. 610, resaltando el fundamento ético de la responsabilidad. Aires de Sousa, S., «The importance of AI for financial crimes», *op. cit. N*. 245. También en su monografía *Responsabilidade pelo*…, *op cit*. N. 628, p. 65.

[648] Galán Muñoz, A., «La problemática utilización …», *op. cit*. N. 367, p. 8.

[649] *Ibid.,* p. 2.

[650] Aires de Sousa, S., «The importance...», *op. cit*. N. 647, p. 175.

ropea sobre al acceso a determinados datos (con el ejemplo paradigmático de Francia prohibiendo penalmente el análisis de datos sobre magistrados y decisiones judiciales para prever sus próximas decisiones [651]), y la cláusula *anti-spoofing* introducida en el mercado de valores de EEUU tras un supuesto de negociación algorítmica de alta frecuencia [652]. Para De Sousa, por tanto, parece evidente que, en el caso de la IA, donde la imprevisibilidad y la incontrolabilidad son rasgos inherentes, la lógica precaucional ha de tener un papel [653].

Por su parte, Piergallini [654] ha planteado recientemente la cuestión de la precaución desde otra perspectiva, afirmando que, *«frente a máquinas capaces de descifrar el entorno que las rodea y de remodelarlo de manera imprevisible, ¿no sería quizá necesario plantearnos, incluso, antes de un problema de atribución (y por tanto de responsabilidad), una penetrante cuestión ética, relativa a la relación entre riesgo y precaución? En esta última perspectiva, surge con fuerza la necesidad de interrogarnos sobre el «margen de seguridad»,* es decir, el grado de riesgo que estamos dispuestos a tolerar: se trata de una elección política, proactiva y no reactiva, es decir, no se orienta en una perspectiva consecuencial, en cuanto a la identificación de los centros de responsabilidad, sino que, ante todo, intenta incidir sobre cómo gobernar, preventivamente, estos riesgos, cuestionando el nivel de tolerabilidad social» [655].

Sin embargo, aunque la invocación de este principio sigue siendo frecuente en el contexto de la expansión de la IA, las cautelas son similares a las realizadas en su aplicación a los anteriores riesgos tecnológicos. Una vez más, en este campo, un sector significativo de la doctrina subraya las tensiones entre el principio de precaución y los principios estructurales del Derecho penal, como la legalidad, la tipicidad y la culpabilidad.

---

[651] Nos referimos a la ley francesa *Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*, que en su artículo 33 establece: *«Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées. La violation de cette interdiction est punie des peines prévues aux articles 226-18,226-24 et 226-31 du code penal…»* (Los datos de identidad de los magistrados y del personal judicial no pueden reutilizarse con el fin o efecto de evaluar, analizar, comparar o predecir sus prácticas profesionales reales o supuestas**.** La infracción de esta prohibición se castiga con las penas previstas en los artículos 226-18, 226-24 y 226-31 del Código Penal).

[652] Se refiere Aires al famoso caso *United States v. Michael Coscia*. 866 F.3d 782, 786 (7th Cir. 2017) que da lugar a la tipificación expresa de esta práctica por medio de la *Dodd-Frank Act.* Reconoce la propia autora que, aunque en este caso el HFT no era propiamente IA, se puede extrapolar la lógica. En el mercado bursátil actual resulta innegable la presencia de *traders* algorítmicos dotados de IA que negocian como otros operadores más. Aires de Sousa, S., «Não fui eu...», *op. cit*. N. 236, p. 88.

[653] *Ibid.,* p. 89.

[654] Piergallini, C., «Intelligenza Artificiale: De "mezzo" ad "autore" del reato?», Rivista italiana di diritto e procedura penale, 4, 2020, pp. 1745. También publicado en Acquaroli, R/Fronza, E./Gamberrini, A. (Coords.), *La giustizia penale tra ragione e prevaricazione*, Aracne, Roma, 2021, pp. 219-244.

[655] *Ibid.,* p. 1750.

Quintero Olivares sostiene, de forma rotunda, la vertiente que aquí se ha llamado prohibitiva, que *«la vía de prohibir el uso de la máquina [...] supondría incluir en las normas de cuidado el principio de precaución, lo cual no es posible»* [656]. Añade que las legislaciones penales no pueden crear delitos en los que no haya ninguna agresión a un bien jurídico, defendiendo que las normas de cuidado no se inspiran en la precaución sino en la experiencia.

Por su parte, Bertolesi, en su intento por encontrar un modelo de imputación correcto para los daños causados por sistemas de IA, trae a colación la postura de la doctrina italiana, diferenciando entre dos enfoques del principio (su implicación para el legislador y su implicación para el ciudadano). En lo que atañe al legislador, se pregunta si el principio puede fundamentar la sanción penal de una conducta, descartando la posibilidad en el caso de los delitos de peligro concreto, pero aceptando que en el caso de los delitos de peligro abstracto puede haber un lugar [657], mostrándose más partidario del modelo *«ingiunzionale»* [658], que garantiza la valoración de la administración competente o de la agencia de control, a los cuales confiar la valoración del equilibrio de intereses, a la luz de las indicaciones provenientes de los sujetos interesados. Se habla de un mecanismo de gestión de riesgos compartido, entre la autoridad administrativa, las agencias de control, y el legislador.

En cuanto al papel del principio de precaución en su relación con el ciudadano, se reproduce el debate clásico de qué papel juega la precaución como parámetro de juicio para configurar la responsabilidad, que se puede reconducir al debate sobre la configuración del deber de cuidado en la imprudencia. Reconoce que la jurisprudencia italiana sí ha aplicado la regla de precaución tradicionalmente en los casos que denomina de toxicidad (lo que aquí podríamos identificar con fraude alimentario), reconociendo también la doctrina una incidencia interpretativa en el concepto normativo de «*colpa per imprudenza*». Acoge Bertolesi la propuesta de Ruga Riva [659], afirmando que cuando existan serias conjeturas sobre la peligrosidad de determinados productos bajo determinadas condiciones, la regla de precaución debe ser la de recopilar toda la información

---

[656] Quintero Olivares, G., *op. cit.* N. 126, p. 19.

[657] Aunque Bertolesi nombra a diversos autores (Pulitanò, Forti, Giunta, o Pongiluppi), que han sido partidarios de la legitimidad del delito de peligro hipotético legitimado con base en el principio de precaución, que se explicará *infra*, reconoce que la opinión dominante es la contraria, según la cual los delitos de peligro abstracto necesitan de una base etiológica «nomológicamente fundada», de forma que, en caso de incerteza, la protección sea extrapenal. Bertolesi, R., *op cit*., N. 423, p. 97.

[658] El mismo modelo que adopta Corn, E., *op. cit. N. 645.*

[659] Formulada en Ruga Riva, C., «*Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*», en Dolcini, E/ Paliero, C. E. (Coords.), *Studi in onore di Giorgio Marinucci*, II, Giufre, Milán, 2006, pp. 1743-1777.

posible y actuar sobre la base de las mejores tecnologías disponibles, pero sin que esto implique una obligación de abstención, sino más bien obligaciones de actualización o conocimiento [660]. Señala, sin embargo, que la doctrina dominante (entre ellos, Piergallini), aún es reacia a la posibilidad de situar el principio de precaución como base de una regla cautelar, prefiriendo, como arriba se ha indicado, la «regulación prudencial codificada», que es la única que permitiría la función de garantía de cognoscibilidad explícita de la codificación de las reglas prudenciales cautelares. Pone como ejemplo Bertolesi la posibilidad de aplicar esa lógica precaucional en la creación de un delito de peligro que establezca la prohibición de producción en el caso de las armas autónomas letales, pero aquí lo diferencial para Bertolesi es que sí hay certeza sobre los riesgos que puede causar, considerando problemática la prohibición cuando hay incertidumbre sobre la peligrosidad de determinadas propiedades. Solo ve posible esta opción el autor italiano en áreas donde hay un gran consenso social, sin que se vea compensada la aplicación o uso tecnológico con los beneficios para la sociedad [661].

Cappellini, por su parte, advierte contra una «ultraresponsabilidad» penal fundada en los sentimientos de inseguridad y miedo, especialmente en contextos tecnológicos donde el control humano, además de poco realista, no garantiza mayor seguridad [662].

No obstante, debemos señalar que existe todo un conjunto de posturas que, si bien critican su uso directo como base de tipificación penal, aceptan que el principio pueda proyectarse de manera indirecta sobre el Derecho Penal a través de la configuración de deberes de cuidado o en la interpretación de la culpa.

Aires de Sousa plantea que la naturaleza de ciertos productos de IA puede exigir, además de un deber intensificado y ampliado de supervisión y control del producto tras su comercialización, una regulación dinámica (*responsi-*

---

[660] Así lo explica Bertolesi, R. *op cit.,* N. 423, p 99, aceptando que en las hipótesis de extrema gravedad en que las conjeturas impliquen una peligrosidad muy grave, no podrá exigirse el deber de abstención.

[661] *Ibid.,* p. 223. De hecho, el mismo autor reconoce que en el caso de las armas autónomas letales, si bien ha sido propuesta su prohibición, no es uno de los campos con mayor consenso, sino uno muy debatido. Salvadori también reconoce esta posibilidad de prohibir las armas autónomas letales como una de las razonables soluciones en este campo, en aplicación, precisamente del principio de precaución. *Vid.*, Salvadori, I., *op. cit*. N. 101, p. 171: «*la comunidad no podrá asumir los riesgos inherentes al desarrollo de a.a. [agentes artificiales] que, por su ámbito de aplicación, el contexto en el que operan o los fines para los que fueron creados, carecen de toda utilidad social o cuyo funcionamiento puede escapar a cualquier control por parte del hombre. Piense, en este sentido, en robots asesinos, armas autónomas inteligentes o los llamados Sexbots creados con fines de pornografía infantil*».

[662] Cappellini, A. «Reati colposi e tecnologie dell'intelligenza artificiale»., en Balbi, G./ De Simone, F./ Esposito, A./Manacorda, S. (Coords.), *Diritto penale e inteligenza artificiale. «Nuovi scenari»,* Giappichelli Editore, Turin, 2022, p. 30-32.

*ve*), y que en este sentido el principio de precaución puede ser un criterio orientador para la regulación de los productos de IA, con reglas reversibles a medida que los riesgos se van haciendo cada vez más controlables [663]. Con el objeto de poder establecer esta lógica, la autora se muestra partidaria del *sandbox approach* [664] (espacio seguro de experimentación), que permita obtener un conocimiento cada vez mayor de los sistemas de IA y superar la incertidumbre sobre sus riesgos.

De la Cuesta Aguado apunta, también, alguna idea de reconocimiento velado de este principio [665], al afirmar que el mismo puede colarse en nuestra legislación en los ámbitos tecnológicos más novedosos en los casos en que el derecho penal es accesorio del derecho administrativo (es decir, en los casos de delitos de peligro con bienes jurídicos abstractos). Admitiendo que la intervención penal en ámbitos relacionados con tecnologías altamente peligrosas y novedosas, entre las que se encuentra la IA, se asienta en el reconocimiento de un riesgo permitido que debe delimitarse en el caso concreto con el respecto de ciertas normas de cuidado (que no son otra cosa que medidas de precaución [666]) administrativamente impuestas. El hecho de afirmar que se ha superado el riesgo permitido implica afirmar, necesariamente, la eficacia del principio de precaución.

En definitiva, es posible afirmar que existen también algunos puntos de convergencia entre la doctrina más actual en este punto: podemos encontrar un rechazo mayoritario a la aplicación directa del principio como fundamento autónomo de responsabilidad penal pero, a su vez, gran consenso a la hora de considerarlo en el diseño de normas de prevención o como criterio subsidiario para configurar deberes de cuidado en contextos técnicos complejos.

Sí existe disenso, no obstante, en cuanto al grado de exigibilidad que deriva de su aplicación. Quizá lo que se puede entrever detrás de esta discordancia no es más que una diferente concepción en cuanto al papel que debe jugar el derecho penal, que oscila entre una visión más garantista y otra más

---

[663] Aires de Sousa, S., *op cit*. N. 652, p. 32.

[664] Dicho enfoque es, en efecto, el que prevé el Reglamento de Inteligencia Artificial ya aprobado, en sus artículos 57 y ss. (se examinará en detalle en el capítulo séptimo).

[665] De la Cuesta Aguado, M. P., «Reflexiones a propósito de «La robótica ante el derecho penal: el vacío de respuestas jurídica(s) a las desviaciones incontroladas», de Gonzalo Quintero Olivares», *Revista Electrónica de Estudios Penales y de la Seguridad*, 1, 2017, pp. 4 y 5. Se califica este reconocimiento como «velado» pues la autora no deja del todo claro cuál es su postura sobre dicho principio («*sin que ello signifique que aceptar la eficacia directa del principio de precaución»),* pero admitiendo su posible papel (quizá, entonces, no de eficacia directa), en el reconocimiento del riesgo permitido de actividades reguladas administrativamente.

[666] En este punto, la autora vuelve a ser ambigua, al situar esta afirmación entre interrogantes, sin que llegue a quedar claro si realmente afirma o cuestiona que las normas de cuidado que configuran el riesgo permitido no sean más que una concreción del principio de precaución.

apegada a la lógica del derecho penal como mecanismo para la gestión del riesgo social.

En todo caso, parece pacífico que el principio de precaución parece encontrar su lugar más legítimo en el derecho administrativo y en la política criminal preventiva, sirviendo como criterio para delinear deberes normativos en contextos de incertidumbre, pero sin erosionar los pilares del derecho penal garantista. Cualquier incorporación más directa deberá transitar por la codificación explícita de normas prudenciales y el respeto a los principios de culpabilidad y legalidad.

## 5. TOMA DE POSTURA: CONCLUSIONES SOBRE EL PRINCIPIO DE PRECAUCIÓN Y SU PAPEL

A la luz del análisis doctrinal, puede afirmarse que el principio de precaución, en su formulación contemporánea, constituye un criterio normativo relevante para el diseño de políticas de prevención, especialmente en el ámbito de las nuevas tecnologías y la inteligencia artificial. Su funcionalidad principal, no obstante, debe entenderse dentro de un marco de intervención pre-penal, con capacidad para informar la creación de prohibiciones administrativas o generales, pero sin que ello implique su recepción como fundamento autónomo de responsabilidad penal.

En este sentido, resulta particularmente ilustrativo el reciente Reglamento de Inteligencia Artificial de la Unión Europea, que contempla una serie de usos prohibidos de sistemas de IA considerados de riesgo inaceptable. Entre ellos se incluyen: la manipulación cognitivo-conductual de personas mediante técnicas subliminales, la explotación de vulnerabilidades de grupos específicos (como niños o personas con discapacidad) los sistemas de calificación social por parte de autoridades públicas y el uso de reconocimiento biométrico remoto en tiempo real en espacios públicos con fines policiales [667].

Estos supuestos reflejan riesgos que pueden ser considerados graves, irreversibles y cargados de incertidumbre, ya que afectan directamente a bienes jurídicos como la dignidad humana, la libertad individual, la no discrimina-

[667] Plasmados en artículo 5.1 del Reglamento, analizado en profundidad por sus implicaciones penales en el capítulo séptimo. Reconoce Miró como evidente que la lógica que inspira el Reglamento es absolutamente precaucional en: Miró-Llinares, F., *General Report – XXI International Congress of Penal Law on «Artificial Intelligence and Criminal Justice», International Colloquium of Section II, «Old and New Criminal Offences: AI Systems as Instruments and Victims»*, 2023. Publicado más adelante en: Miró-Llinares, F., Duvac, C., Toader, T., /Santisteban Galarza, M., *op. cit.* N. 160, p. 34, «*The AI Act is grounded on a risk-based approach with «a strong emphasis in the precautionary principle»*.

ción y la intimidad. Tales bienes, por su importancia estructural en los ordenamientos democráticos, son plenamente susceptibles de protección penal, especialmente cuando los mecanismos preventivos y administrativos resultan insuficientes.

Así, resulta concebible que, partiendo de la lógica precautoria que inspira estas prohibiciones, se estructuren normas penales en blanco o delitos de peligro abstracto que sancionen el desarrollo, comercialización o uso de estas tecnologías en contextos prohibidos, siempre que se identifique un injusto penal propio y se respeten los principios de legalidad, tipicidad y culpabilidad, así como el resto de requisitos de imputación de la teoría del delito. Para que estas prohibiciones administrativas se trasladen al campo penal habrán de cumplirse los principios propios del derecho penal. Cuando el problema tiene que ver con la traslación de una prohibición administrativa al derecho penal (el ya citado problema de la administrativización), más que negar la incidencia del principio de precaución en la fundamentación de la norma de cuidado que da lugar a la prohibición, cuestión que nos parece incontrovertida, hemos de ser capaces de encontrar un injusto penal propio, que vaya más allá del mero ilícito administrativo, constituyendo esta la verdadera tarea del derecho penal. Si, de partida, los bienes que pueden ponerse en peligro son de entidad suficiente (así parece en el caso de los usos prohibidos), no resultará difícil fundamentar que es incumbencia del derecho penal asegurar su integridad. Sí a la precaución como fundamento de delitos de peligro abstracto o de delitos en forma de normas penales en blanco, pero siempre que encontremos un injusto jurídico penal propio (nos encargaremos de esto *infra).*

Donde sí debe trazarse un límite claro (y aquí la doctrina es prácticamente unánime) es en el papel dogmático estricto de este principio. No es posible sustituir el paradigma de la previsibilidad, en ninguna de sus manifestaciones. No es posible afirmar la tipificación de meros peligros no descartables (el problema de los delitos meramente presuntos o estadísticos, que se estudiará *infra),* no será posible sustituir el juicio de peligrosidad *ex ante* propio de la imputación objetiva de los delitos dolosos o imprudentes por un juicio de no descartabilidad del daño, y no será posible omitir el juicio de causalidad por una no descartable conexión en casos de absoluta incertidumbre.

Remitiéndonos, por tanto, a la triple función arriba descrita, la precaución debe operar, no en un sentido dogmático fuerte, sino solo en sus facetas prohibitiva y proactiva (fundamentando prohibiciones y delimitando deberes de cautela, con los matices que se han apuntado en su traslación al campo penal). En definitiva, el principio de precaución puede y debe jugar un papel en

la construcción normativa y política del riesgo en sociedades tecnológicamente avanzadas, pero su proyección sobre el derecho penal debe ser mediada, prudente y garantista. Su verdadera utilidad no reside en reformular la teoría del delito, sino en ofrecer criterios para la delimitación de riesgos inaceptables y para la configuración de deberes de cuidado que, en ciertos casos, pueden devenir en responsabilidades penales solo si se cumplen el resto de los requisitos que la dogmática penal exige.

# CAPÍTULO QUINTO. LOS PROBLEMAS DEL MODELO DE RESPONSABILIDAD PENAL POR EL PRODUCTO

*NADA SUCEDE DOS VECES*

*Ni va a suceder, por eso*
*sin experiencia nacemos,*
*sin rutina moriremos.*
*En esta escuela del mundo*
*ni siendo malos alumnos*
*repetiremos un año,*
*un invierno, un verano.*
*No es el mismo ningún día,*
*no hay dos noches parecidas,*
*igual mirada en los ojos,*
*dos besos que se repitan.*
*Ayer mientras que tu nombre*
*en voz alta pronunciaban*
*sentí como si una rosa*
*cayera por la ventana.*
*Ahora que estamos juntos,*
*vuelvo la cara hacia el muro.*
*¿Rosa? ¿Cómo es la rosa?*
*¿Como una flor o una piedra?*
*Dime por qué, mala hora,*
*con miedo inútil te mezclas.*
*Eres y por eso pasas.*
*Pasas, por eso eres bella.*

*Medio abrazados, sonrientes,*
*buscaremos la cordura,*
*aun siendo tan diferentes*
*cual dos gotas de agua pura* [668].

La galardonada poeta polaca nos deleita con estos versos que ponen de manifiesto una realidad por todos conocida y, sin embargo, a menudo olvidada. El tópico clásico de la impermanencia, de la absoluta variabilidad, resulta inapelable. Y, sin embargo, nos empeñamos en encontrar puntos comunes ante situaciones diversas que nos recuerdan a lo entonces sucedido. Si entonces reaccionamos así, ¿por qué hoy iba a ser diferente? El tópico que ocupa este capítulo central pretende, de alguna forma, sacudirnos del espejismo de lo conocido (aunque sea, solo, momentáneamente).

Los postulados que sirvieron en un determinado momento, ante ciertos casos y ante determinados escenarios, no son ya, o no son en todo, directamente aplicables ante escenarios que nos recuerdan a lo que pasó, y en este sentido se presentan como las dos gotas de agua de Szymborska: tan parecidos y, a la vez, inevitablemente diferentes.

## I. ORÍGENES TEÓRICOS Y PRÁCTICOS

Si hasta ahora nos hemos encargado de examinar los antecedentes sociológicos y político-criminales, así como el contexto económico que fue el caldo de cultivo de un determinado modelo de derecho penal, se hace necesario adentrarse, por fin, en ese concreto modelo, explicitando de qué manera se reflejó en el derecho penal toda la tendencia expansiva y modernizadora.

Cuando nos referimos, en este capítulo, a la responsabilidad penal por el producto lo hacemos respecto de los intentos de castigar, por la vía penal, la producción de daños a bienes jurídicos individuales (bienes jurídico-penales) por la existencia de algún tipo de defecto o fallo en alguno de los productos generados por las industrias de la sociedad moderna del riesgo [669]. En el siguiente capítulo

[668] Szymborska, W., (Polonia,1923-2012), «Nada dos veces», *Llamando al Yeti*, 1957, publicado en español en: Szymborska, W., Beltrán, G., Murcia, A., Poniatowska, E., Beltrán, G., & Murcia, A. (2013). *Poesía no completa*, Fondo de Cultura Económica.

[669] Podemos tomar como referencia, si se quiere, la conceptualización que realiza Rodríguez Montañés, definiéndola como *«responsabilidad penal de los productores o distribuidores de bienes de consumo respecto de las lesiones o puestas en peligro de bienes jurídicos esenciales de los consumidores (fundamentalmente la vida y la integridad) derivadas de la normal utilización del producto».* «Incidencia dogmática de la jurisprudencia del caso de la colza y otros casos en materia de productos defectuosos»,

examinaremos el segundo de los modelos vinculados al *topos* de la responsabilidad penal por el producto, a saber, los delitos de peligro, entendiendo que ambos presentan dos «hijos de su tiempo», o dos formas de abordar un mismo problema, pero desde perspectivas diferentes y con criterios esencialmente distintos [670].

Esta forma de responsabilidad penal, cuyo término *(Produkthaftung)* debemos a la doctrina alemana [671], tiene al menos dos particularidades: en primer lugar, bebe del derecho civil (condicionando, entendemos, algunos de sus rasgos, como se intentará explicar), y, en segundo lugar, su construcción teórica se produce a partir de los casos jurisprudenciales que, tras la sucesión de eventos catastróficos y con numerosas víctimas, obliga a los tribunales penales a pronunciarse sobre cuestiones que no se habían tratado hasta el momento [672].

Se examinarán, de forma breve (al no ser este el objeto más importante de nuestra investigación, sino tan solo un pretexto para intentar trasladar su esquema a los problemas que generan los «productos» tecnológicos aquí investigados), los más importantes *leading cases* de la jurisprudencia alemana y española, así como las reacciones de la doctrina penal.

---

en Boix Reig, F. J./ Campos Cristóbal, R./ Bernardi, A. (Coords.), *Responsabilidad penal por defectos en productos destinados a los consumidores*, Iustel, Madrid, pp. 115-116.

[670] Siguiendo a Aires de Sousa, «La responsabilidad penal por el producto, entendida como responsabilidad por la colocación en el mercado de productos peligrosos para bienes jurídicamente valiosos, plantea distintos problemas jurídico-penales, dependiendo de la referencia temporal que se considere, distinguiendo entre, un primer momento relacionado con la colocación del producto en el mercado, haciéndolo accesible a la comunidad de potenciales consumidores, así como un ulterior momento referente al uso del bien por un consumidor concreto, materializado en una ofensa a bienes jurídicos individuales como la vida o la integridad física. Las cuestiones –y su respectivo encuadre jurídico-penal– difieren según el momento al que se atienda. Así, en su base está una interrogante fundamental: ¿debe la intervención penal reservarse al momento de lesión de bienes jurídicos individuales o, de manera diversa, y concediendo primacía al primero de los momentos enunciados, debe esa intervención ser admitida en el marco de la protección de bienes jurídicos colectivos como es el caso de la salud pública?». Aires de Sousa, S., *op. cit. N. 628*, p. 170 (traducción propia). La referencia relativa a la distinción entre los dos momentos se debe a Corcoy Bidasolo, M., «Responsabilidad penal derivada del producto: En particular la regulación legal en el código penal español: delitos de peligro», en Mir Puig, S./ Luzón Peña, D. M. (Coords.), *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, J. M. Bosch, Barcelona, 1996., p. 248.

[671] Aires de Sousa plantea que ya Vogel utiliza esta expresión *(strafrechtliche Produkthaftung)* en 1990 para referirse a «los daños (o riesgos) para la integridad física, la vida o, en algunos casos, el patrimonio, causados mediante la producción y comercialización de diversos productos peligrosos». Aires de Sousa, S., *ibid,* p. 147. Se refiere a Vogel, J., «Verbraucherschutz durch strafrechtliche Produkthaftung – Kriminologische und funktionale Aspekte», *GA,* 1990, p. 246. Es, sin embargo, a Hilgendorf a quien debemos el trabajo de verdadero rastreo de los orígenes del término, empezando ya a usarse en Maurach, R./ Schröder, F-C., *Strafrecht. Besonderer Teil*, 6.ª edición, 1981, p. 57. Así, en Hilgendorf, E., *Strafrechtliche Produzentenhaftung in der «Risikogesellschaft»*, Duncker & Humblot, Berlin, 1993, p. 89. Este autor, sin embargo, sostiene una concepción original respecto de los casos paradigmáticos resueltos por los tribunales alemanes, desarrollados *infra,* al considerar que no son verdaderamente una manifestación del derecho penal del riesgo, sino más bien una flexibilización de las estructuras dogmáticas clásicas, donde, más que otra cosa, se pone de manifiesto la compleja relación entre el derecho penal económico y la parte general. Hilgendorf, E., *ibid.,* p. 57.

[672] Una explicación detallada de sus orígenes y de su vinculación con el derecho civil en Aires de Sousa, S., *op. cit*. N.628, pp. 145 y ss.

## II. LOS CASOS QUE HICIERON NACER LA DISCUSIÓN

En el ya descrito caso *Contergan* se deja ver uno de los problemas centrales que surgirán en adelante en toda la discusión relativa a daños causados por productos: la relación de causalidad. Explica Cancio la resolución del Tribunal Regional de Aquisgrán [673], aclarando que el problema aquí tiene que ver con *«la existencia de múltiples riesgos nuevos, producidos por la generalización de productos químicos […] poco conocidos»,* por lo que aquí la complejidad deriva de la necesidad de determinar qué requisitos debe exigir el juez para afirmar la responsabilidad por un delito de resultado en los casos en que el conocimiento sobre los efectos de una sustancia es incompleto y no pacífico entre la disciplina científica [674] (concretamente, si se podía afirmar la causalidad entre la ingesta del medicamento por las gestantes y las malformaciones de sus hijos).

Señala asimismo, con cita a Puppe [675], que el problema causal pivota, en realidad, en torno a tres subproblemas: cuál debe ser el contenido que debe exigir el juez respecto de la ley causal (en este sentido parece que el tribunal afirma que, una vez probada la ley causal, es posible afirmar la causalidad concreta, aunque no de forma explícita); cómo debe introducirse esa ley causal en el proceso (quedando convencido el tribunal por medio de las declaraciones periciales); y qué relación debe haber entre la afirmación de la relación causal en el caso concreto y la correspondiente ley causal general (afirma su competencia el tribunal para dar por probada una de las posiciones sostenidas por los expertos).

El problema del caso concreto estaba, precisamente, relacionado con el primer elemento, pues no quedaba clara la existencia de una ley causal general según la cual la ingesta de ese medicamento, con esos concretos componentes, explicase la producción de tales malformaciones. Este es el argumento que, desarrollado por Kaufmann [676], la doctrina penal critica en adelante [677], frente a

---

[673] El auto que resuelve el famoso caso, si bien reconoce tanto la tipicidad de la conducta, así como la causalidad y la imprudencia de los acusados respecto de las lesiones causadas a los fetos, no condena, en uso de la facultad que le otorgaba el código penal del momento (concretamente, el § 153.3 StPO). Cancio Meliá, M., *op. cit*. N. 402, p. 211, nota al pie 6. Para una traducción al español de los razonamientos jurídicos relativos a la causalidad, Gómez Benítez, J. M., *Causalidad, imputación y cualificación por el resultado*, Ministerio de Justicia Centro de publicaciones, Madrid, 1988, pp. 123 y ss.

[674] Cancio Meliá, M., *ibid.*, p. 218.

[675] Puppe, I., «"Naturgesetze" vor Gericht. Die sogenannte generelle Kausalität und ihr Beweis, dargestellt an Fällen strafrechtlicher Produkthaftung», *JZ,* 1994, p. 1147.

[676] Kaufmann, A., *op. cit.* N. 404. De hecho, el propio Kaufmann actuó en la causa como perito de una de las defensas.

[677] Una de las voces más críticas con esta forma de interpretar la causalidad es Corcoy, afirmando (respecto del caso de la colza, donde el problema causal pivota esencialmente sobre este postulado), que se ha sustituido la causalidad por la imputación objetiva y que, por razones de «alarma social» se han

la posición que, por el contrario, adoptan los tribunales en casos posteriores. La crítica, en esencia, viene a sostener que, si no se conoce de una forma general una ley causal, afirmar la causalidad sobre la base de varios factores de correlación llevaría a una condena ilegítima.

Unos años más adelante, el caso *Erdal* o *Lederspray* se convierte en el segundo de los *leading cases*[678]. Resumidamente, en este supuesto, un *spray* que se vendía para el tratamiento de los productos de cuero, tras 20 años de comercialización, empieza a mostrar algunos efectos nocivos en ciertos clientes (principalmente lesiones relacionadas con problemas respiratorios). Tras la realización de toda una serie de investigaciones internas por parte de la empresa, se informó de que no era posible afirmar la toxicidad del producto, decidiendo realizar una advertencia en los envases, pero sin proceder a su retirada del mercado. En este supuesto, a diferencia del caso *Contergan*, sí hubo una condena por parte de los tribunales, que condenaron a los administradores (de la productora y de todas las filiales distribuidoras), como coautores de delitos de lesiones en comisión por omisión por la no retirada y como coautores de lesiones dolosas activas respecto de los productos vendidos con posterioridad[679].

Finalmente, el caso español que traslada la discusión a nuestro contexto es el ya célebre del aceite de colza[680]. Los hechos, resumidos, son los que siguen. En mayo de 1981 se detectó un brote epidémico en diversas zonas de España, inicialmente atribuido a una posible neumonía transmisible por vía respiratoria. No obstante, desde un primer momento también se consideró la hipótesis de una intoxicación alimentaria. La enfermedad, desconocida hasta entonces y con una sintomatología compleja fue posteriormente denominada «síndrome tóxico». A comienzos de junio de ese mismo año se comenzó a vincular el origen del brote con la ingesta de aceite en mal estado, y en junio las autoridades sanitarias alertaron públicamente sobre esta posibilidad. Tras años de investigaciones, los órganos judiciales competentes concluyeron que los acusados habían manipulado aceites destinados al consumo humano, incorporándoles aceite de colza desnaturalizado junto con otras sustancias quí-

---

modificado, llegando a anularse, los requisitos de prueba de los elementos esenciales del delito. Corcoy Bidasolo, M., «La lección de la experiencia española en materia de responsabilidad por el producto. *Op. cit.* N. 625, p. 455.

[678] El caso fue resuelto definitivamente por medio de sentencia del Tribunal Supremo alemán, BGHSt 37, 106, de 6 de julio de 1990, que confirma el pronunciamiento del tribunal de instancia.

[679] Lo explica con más detalle Rodríguez Montañés, T., «Incidencia dogmática…», *op. cit*. N. 69, pp. 117-118. Una exposición completa del caso Leadrespray y sus implicaciones dogmáticas en Gimbernat Ordeig, E., *Ensayos Penales*, Tecnos, Madrid, 1999, pp. 304-325.

[680] Resuelto por sentencia del Tribunal Supremo de 23 de abril de 1992 (ponente Bacigalupo Zapater), que resolvió los recursos de casación interpuestos frente a la sentencia de la Sala de lo penal de la Audiencia Nacional de 20 de mayo de 1989.

micas, entre ellas la anilina, un compuesto venenoso que no alteraba perceptiblemente las características sensoriales del aceite. Esta sustancia reaccionó con los ácidos grasos formando anilidas, cuya presencia fue detectada en los aceites implicados. La enfermedad presentaba síntomas como lesión de los vasos sanguíneos, edema pulmonar, incremento de eosinófilos, esclerodermatosis, atrofia muscular, daños pulmonares, pérdida de peso, desnutrición y afectación generalizada de órganos. Sin embargo, persistieron en el proceso probatorio importantes incertidumbres científicas, como la imposibilidad de identificar con exactitud el agente tóxico causante, la insuficiencia de pruebas para confirmar si las anilidas eran la causa directa de la enfermedad, la falta de correspondencia entre el síndrome observado y la intoxicación por anilina, y el desconocimiento sobre el mecanismo patológico que explicara la sintomatología descrita [681].

El Tribunal Supremo consideró que la relación de causalidad entre las acciones de los acusados y los resultados lesivos podía tenerse por probada, recurriendo a un criterio jurídico basado en la exclusión razonable de otras posibles causas, dado el número significativo de casos y datos coincidentes. El fallo de la Audiencia Nacional había condenado a los responsables por delitos de homicidio y lesiones imprudentes (imprudencia con resultado de muerte y lesiones), pero el Tribunal Supremo elevó la calificación jurídica a dolo eventual.

El problema en cuanto a la relación de causalidad fue, una vez más, determinante, pero la condena se sucedió pese a la existencia de importantes lagunas científicas, pues no se logró identificar con certeza el agente tóxico específico responsable del síndrome. Aunque la presencia de anilidas se constató en los aceites implicados, no se pudo determinar si eran suficientes por sí mismas para causar la enfermedad o si su aparición se debía a manipulaciones posteriores. Se demostró, además, que la sintomatología del síndrome tóxico no coincidía con la típica de una intoxicación por anilina, y se desconocía el mecanismo preciso por el cual el agente tóxico producía los efectos observados en los afectados.

La doctrina causal que el TS sostiene, y que posteriormente da lugar a toda una serie de reacciones críticas, viene a afirmar que, aunque en el caso del síndrome tóxico no existía una ley causal natural plenamente aceptada por la comunidad científica que explicara los hechos, ante esta carencia corresponde al órgano judicial valorar si los datos disponibles permiten, pese a esa incerti-

[681] Se presenta un resumen de los hechos tal y como son narrados en: Paredes Castañón, J. M., «Caso del aceite de colza», en Sánchez-Ostiz Gutiérrez, P. (Coord.), *Casos que hicieron doctrina en derecho penal*, La Ley, Madrid, 2011, pp. 425-427.

dumbre, afirmar una conexión causal. Así, fija como criterio que puede considerarse probada la causalidad cuando un hecho se repite en numerosos casos similares y es posible descartar otras causas alternativas. No se exige conocer todos los pasos del proceso causal ni identificar con precisión el agente tóxico, ni siquiera que el fenómeno sea reproducible experimentalmente.

Para Paredes, el caso de la colza supone un ejemplo paradigmático del *«esfuerzo por adaptar las categorías dogmáticas a las dificultades probatorias que surgen en los ámbitos del llamado "Derecho Penal del riesgo"»* [682]. Se muestra crítico con este intento del alto tribunal de reducir los niveles de exigencia para la imputación de los delitos dolosos de resultado.

Desde luego, no son estos los únicos casos de consecuencias masivas donde los problemas de la responsabilidad penal del productor respecto de los daños causados por sus productos se ponen de manifiesto, pero sí los más citados y comentados entre la literatura [683]. Sirven, sin embargo, para poner sobre la mesa al menos dos de los problemas que, aun hoy, se plantean cuando se trata de aplicar el esquema de la responsabilidad por el producto a las nuevas tecnologías, cuales son el relativo a la afirmación de la relación de causalidad entre la conducta (activa u omisiva) del productor, y los deberes que, tras la aparición de algunos signos de alarma o peligro, pueden fundamentar la generación de responsabilidad en comisión por omisión [684].

Tras la exposición breve de los casos paradigmáticos estamos en posición de constatar en qué sentido se sucedieron los pronunciamientos contrarios y favorables a este novedoso fenómeno, donde, por vez primera, los tribunales comienzan a condenar por delitos de resultado a productores en casos altamente controvertidos. Los tres problemas esenciales sobre los que pivotan estos pronunciamientos tienen que ver con: 1) la constatación de la causalidad; 2) la de-

---

[682] *Ibid.,* p. 428. Se deja de lado, aquí, la argumentación del Tribunal relativa a la delimitación entre dolo eventual e imprudencia, pues no consideramos que sea esencial a los fines que aquí interesan, pero debe tenerse en cuenta que constituye todo un hito en la interpretación de esta categoría de dolo de aquí en adelante.

[683] Así, se pueden nombrar también como supuestos que dieron lugar a discusiones dogmáticas similares el caso del spray de madera (*Holzschutzmittel,* resuelto por sentencia del Tribunal Supremo Federal alemán, BGHSt 41*),* el caso del alimento enteral en Chile (Contreras Chaimovic, L., «La responsabilidad penal del fabricante por la infracción de sus deberes de vigilancia, advertencia y retirada», *Polít. crim*., 1.º (9), 2015, p. 271 y ss.), o el caso del Benoliel en Francia tras la explosión de un envase de aerosol (Bertolesi, R., *op cit*., N. 423, p. 157 y ss.).

[684] Sobre el problema de la comisión por omisión, que no ha sido directamente explicitado en la explicación de los hechos, pero sí fue objeto de controversias dogmáticas evidentes, se discutirá *infra*. Sobre el resto de problemas que de forma patente estos casos ponen de manifiesto (la difícil individualización de la autoría en organizaciones complejas o la difícil delimitación del elemento subjetivo, entre otras), no nos haremos cargo, sin perjuicio de dejar apuntado que fueron, desde luego, objeto de innumerables controversias.

terminación de la infracción del deber de cuidado/creación de un riesgo jurídicamente desaprobado [685]; y 3) la justificación de una imputación en comisión por omisión. Cada uno de estos aspectos se examinará desde una doble perspectiva: tanto en su formulación original (apegada a los casos aquí descritos y afines), como en su formulación respecto de los sistemas de IA, pues compartimos con Fragasso que el daño provocado por los dispositivos inteligentes repite, aunque acentuándolos, los problemas que ya surgieron en la responsabilidad penal por el producto: la indebida superposición entre la estructura comisiva y la omisiva; la difícil identificación de los sujetos responsables personalmente en el entorno de organizaciones complejas; la individualización del nexo de causalidad; y la evaluación de la culpa en situaciones de incertidumbre científica [686].

Esta acentuación de los problemas que apunta la autora italiana nos parece, adelantamos, un obstáculo insalvable para argumentar que la imputación al productor del daño final a los bienes jurídicos individuales en forma de delitos de resultado sea una estructura adecuada para resolver los problemas de los sistemas autónomos de IA. Se intentará explicar, en las páginas siguientes, el por qué.

## III. EL PROBLEMA DEL ENCAJE DEL MODELO HOY

Muchos son los autores que, al intentar encontrar una solución a los problemas de imputación generados por las nuevas herramientas de IA, conciben la IA como un producto, para afirmar, sin demasiado problema, que es plenamente aplicable el régimen de responsabilidad penal por el producto. Parece que muchos de estos autores son, pues, partidarios de la imputación de los delitos de resultado ya existentes en nuestro Código (paradigmáticamente, cuando se trate de daños graves para bienes jurídico-penales de indudable importancia) [687].

---

[685] Quizá sea este el problema con contornos menos definidos, pues, por el momento en que se producen las primeras discusiones, no hay acuerdo entre los autores a la hora de hablar sobre los elementos que dan lugar a la imputación del delito imprudente o doloso, empleándose terminologías diversas, desde la infracción de deber o la infracción del cuidado objetivo hasta la creación de un riesgo jurídicamente desvalorado (más apegada a la teoría de la imputación objetiva, que aún estaba en ciernes de convertirse en una teoría mayoritaria). En todo caso, se agrupan aquí todas estas discusiones que giran en torno a la peligrosidad o previsibilidad de la conducta peligrosa *ex ante* bajo un solo epígrafe, pues se considera que todos ellos responden esencialmente a los mismos problemas, a pesar de su diversa denominación.

[686] Fragasso, B., «La responsabilità penale del produttore di sistemi di intelligenza artificiale», *Diritto Penale Contemporaneo*, 1/2023, p. 33.

[687] Las referencias a la IA como producto y a la aplicabilidad del modelo del *product liability* son ubicuas, aunque aquí nos referiremos al modelo en sentido amplio, no solo la responsabilidad del productor, sino a la responsabilidad por el producto, siempre y cuando se pretenda imputar un delito de resultado (ya sea en su modalidad dolosa o imprudente), sin atender a las particularidades de los «defectos» inherentes a la dinámica de la IA. Así, una justificación muy reciente de aplicar esta lógica del delito de resultado (lesiones

Así lo expresan Gless, Silverman y Weigend en su célebre trabajo *«If robots cause harm, who is to blame»* [688]. Parten de la base de que Alemania ha establecido reglas estrictas de responsabilidad civil para productos inseguros, trasladando estas normas a la legislación penal. Los productores deben asegurarse de que sus productos cumplen con los estándares de seguridad y deben monitorear continuamente las quejas y accidentes relacionados. Si no cumplen, pueden ser responsables penalmente por daños causados, ya sea por negligencia o intencionalmente. El concepto de responsabilidad penal por productos, explican, se basa en la creación de riesgos legales al poner productos en el mercado, y esto puede equipararse al daño causado por un agente inteligente cuya reacción no se puede predecir ni controlar. Un acto legal, como comercializar un coche autónomo, puede generar responsabilidad penal por omisión si causa daño. Este tipo de responsabilidad podría ser difícil de aceptar para la industria automotriz, pero las víctimas de accidentes de coches autónomos defectuosos podrían encontrar injusto que nadie sea responsable de los daños en ausencia de un conductor [689].

Hemos de afirmar, empero, que cualquier intento de imputar delitos de resultado sin realizar un análisis pormenorizado sobre los elementos propios de estas tecnologías correrá el riesgo de realizar una imputación inadecuada.

---

u homicidio imprudente en el ámbito médico) ante los resultados de un diagnóstico errado de un sistema de IA, en Rueda Martín, M. A., «La concreción del deber objetivo de cuidado en el desarrollo de la actividad médico-quirúrgica curativa a la luz de los avances de los sistemas de Inteligencia Artificial», *InDret* 1/2025. También propuestas de aplicar este modelo en Fragasso, B., *ibid.,* haciendo referencia a que quizá el instrumento más idóneo para tratar estos problemas es el que tiene que ver con la responsabilidad civil por producto defectuoso, y que de hecho la legislación europea ya se está adaptando para dar cabida a estos problemas (invirtiendo la carga de la prueba y presumiendo el daño en ocasiones con el objetivo de proteger al consumidor), pero que el derecho penal, por sus rasgos más rígidos, es más difícil que se pueda trasladar como modelo. Incluso una voz autorizada como Piergallini, caracterizando el problema de la imputación penal con los nuevos sistemas de IA como una «prolongación» de los problemas por de responsabilidad penal por el producto profusamente estudiados en su primigenia obra «*Danno da prodotto e responsabilitá penale»:* Piergallini, C., «Intelligenza Artificiale: ¿De "mezzo" ad "autore"…, *op. cit.* N. 654, premessa. Otros muchos autores son, desde luego, partidarios de la aplicación del régimen en su vertiente civil, pero respecto de la responsabilidad objetiva pura no se tratará en este trabajo, por motivos obvios. Empero, algunos autores han tratado de extrapolar este modelo de responsabilidad directamente al campo penal, solo para los delitos menores, combinando la responsabilidad objetiva con los requisitos de negligencia, en un enfoque modelado según la responsabilidad penal por productos defectuosos. Así: Lima, D., «Could AI agents be held criminally liable: artificial intelligence and the challenges for criminal law», *SCL Rev., 69*, 2018, p. 693: *«strict liability could either be reserved only for minor offenses when they fall within the margin of error on the part of the human agent, whether it is a programming or an operating error, or it could be discarded completely as a model for ascribing criminal liability. Perhaps the best way to consider strict liability is in a context where it is combined with negligence requirements, in an approach modeled after (criminal) liability for faulty products».*

[688] Gless, S./ Silverman, E./ Weigend, T., *op. cit*. N. 280.

[689] *Ibid.,* pp. 427-428.

## 1. ¿ES LA IA UN PRODUCTO?

Partiendo de la posibilidad de aplicar el correspondiente delito de daños causados por la IA (en su modalidad imprudente) al programador o al usuario, Del Rosal se plantea precisamente la posibilidad de que estemos ante una situación asimilable a la responsabilidad penal por el producto defectuoso [690]. Partiendo de la doctrina civil, de la que, como se ha señalado, bebe este tipo de responsabilidad, señala del Rosal el carácter normativo del concepto producto [691].

Una primera definición general la encontramos en la normativa española [692]. Así, en la Ley de defensa de consumidores y usuarios (RDL 1/2007) [693], el artículo 136 establece que *«A los efectos de este capítulo se considera producto cualquier bien mueble, aún cuando esté unido o incorporado a otro bien mueble o inmueble, así como el gas y la electricidad».* Por su parte, el 137, apartado primero, concreta lo que debe entenderse por producto defectuoso: *«Se entenderá por producto defectuoso aquél que no ofrezca la seguridad que cabría legítimamente esperar, teniendo en cuenta todas las circunstancias y, especialmente, su presentación, el uso razonablemente previsible del mismo y el momento de su puesta en circulación».*

Sin embargo, sobre esta definición parca que la normativa nacional recoge (cualquier bien mueble), la normativa europea más reciente ha venido marcando un camino de modernización necesario para superar el concepto actual (que bebe, de hecho, de la primigenia directiva de responsabilidad por productos defectuosos de 1985 [694]). La aprobación de la nueva Directiva 2024/2853, sobre responsabilidad por los daños causados por productos defectuosos [695] viene motivada, precisa-

---

[690] Del Rosal Blasco, B., *op cit*. N 35, p. 23.

[691] *Ibid.,* p. 24.

[692] La normativa sobre responsabilidad por productos se encuentra sectorializada y armonizada por la Unión Europea, de forma que, sin perjuicio de que en adelante aludiremos a la norma genérica porque el objeto de nuestro trabajo no es examinar los problemas de un solo sector, debe dejarse apuntado que, en función del contexto del producto, la normativa específica es variada. Así, a título de ejemplo, y como se verá en detalle en capítulo séptimo, el hecho de que un producto como un juguete lleve integrado un sistema de IA hará que se active, a su vez, la aplicación de su legislación armonizada específica (en este caso, la Directiva 2009/48/CE del Parlamento Europeo y del Consejo, de 18 de junio de 2009, sobre la seguridad de los juguetes), sin que ello signifique que la nueva regulación sobre IA quede desplazada. El sistema de regulación resulta, pues, extremadamente complejo.

[693] Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias.

[694] Directiva que, de hecho, estuvo en vigor hasta la reciente aprobación de la nueva Directiva el 18 de noviembre de 2024.

[695] Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos y por la que se deroga la Directiva 85/374/CEE del Consejo. Aunque su aprobación se produjo el pasado noviembre de 2024, será aplicable a los productos introducidos en el mercado o puestos en servicio después del 9 de diciembre de 2026,

mente, por el cambio tecnológico, pues la aparición de productos con inteligencia artificial, las nuevas formas de interconexión entre productos, la incorporación de nuevos actores en la cadena de producción y un entorno cada vez más global y complejo han hecho indispensable la actualización de la protección al consumidor[696].

A todos los efectos, y según la normativa ya en vigor y aplicable a partir de diciembre de 2026 (art. 4. 1) se entenderá por producto *«cualquier bien mueble, aun cuando esté incorporado a otro bien mueble o a un bien inmueble o interconectado con estos; incluye la electricidad, los archivos de fabricación digital, las materias primas y los programas informáticos».* No parece problemático afirmar, por tanto, que, si la IA es caracterizada como un programa informático (en caso de no estar imbuida en otro tipo de sistema o producto, como, en la mayoría de los entornos aplicativos, así ocurrirá), cabrá aplicar, *prima facie,* la nueva Directiva.

Podemos convenir, como hacía en su día Íñigo Corroza con referencia a la normativa anterior, que el concepto de producto que se acoge es esencialmente amplio (más aún si se tiene en cuenta la ampliación operada por la nueva Directiva), pero que se refiere, exclusivamente, a los productos que se van a introducir en el mercado para su posterior comercialización[697]. Efectivamente, y al estar centrada la normativa en el resarcimiento de daños generados al consumidor, solo los productos dirigidos a él son limitados por esta normativa. Esto nos puede poner ya sobre la pista de algunas dificultades de asumir un concepto demasiado apegado a la normativa civil[698].

También Turner señala las dificultades de partir de un concepto estricto establecido por la normativa de responsabilidad por el producto. Aunque reconoce que la nueva normativa pretende dar cabida a la IA, señala Turner dos

---

momento en que su trasposición es obligatoria para todos los estados miembros (art. 22 de las Disposiciones finales). Se estudia su contenido, en profundidad, en capítulo séptimo.

[696] Precisamente, ya en el considerando 3 aludiendo a que era necesaria una actualización de la normativa, que *«debería revisarse a la luz de los avances relacionados con las nuevas tecnologías, incluida la inteligencia artificial (IA), los nuevos modelos de negocio de la economía circular y las nuevas cadenas de suministro mundiales, que han dado lugar a incoherencias e inseguridad jurídica, en particular en lo que respecta al significado del término "producto"».*

[697] Íñigo Corroza, E. *La responsabilidad penal del fabricante por defectos de sus productos*, J. M. Bosch Editor, Barcelona, 2001, p. 63, basándose en la idea de que, para que se pueda hablar de puesta en circulación tiene que haber transacción económica de por medio.

[698] Pues si acogemos este concepto, de partida, dejaríamos fuera los sistemas de IA desarrollados por Administraciones Públicas no ofertados al mercado, así como aquellos desarrollados con fines no comerciales por entidades privadas. Piénsese, a modo de ejemplo, en el software de IA *Sweetie* que, creado por la ONG holandesa *Terre des Hommes,* fue cedido para su uso público a gobiernos como el holandés o el neozelandés (entre otros) para la detección y el castigo de depredadores sexuales de menores en internet. Toda la información relativa a este programa en el libro colectivo: Van der Hof, S. S./Georgieva, I./ Schermer, B./ Koops, B. J. (Eds.), *Sweetie 2.0: Using Artificial Intelligence to Fight Webcam Child Sex Tourism*, Springer, Berlin, 2019.

problemas: el primero, que la doctrina suele dar por hecho que la IA es un producto, cuando, según la definición (bien mueble), aquellas IA no incorporadas a un sistema físico deben ser considerados servicios y no productos [699]. Pero más allá de esta cuestión de matiz, lo que Turner pone sobre la mesa es que, en todo caso, lo que sí asume la normativa es un rasgo esencial a todos los productos: que permanezcan estáticos una vez se lanzan al mercado [700]. De hecho, para este autor una de las mejores pruebas de que los regímenes legales de responsabilidad por el producto asumen este estatismo es que en todos ellos (no solo en el europeo, sino también en el derecho estadounidense [701]) se prevén excepciones al régimen de responsabilidad cuando se producen lo que aquí se ha venido a denominar «riesgos de desarrollo» [702], es decir, cuando se produce algún cambio respecto del momento en que el producto se lanza al mercado.

Sobre el problema de los «riesgos de desarrollo» se profundizará *infra,* pero baste dejar aquí anotado como esencial que lo que en el régimen ordinario se prevé como una excepción a la aplicación de la normativa general, puede llegar a convertirse en regla si todo sistema de IA se convierte en producto, dando lugar a una posibilidad de exoneración masiva de responsabilidad de los productores [703].

---

[699] Turner, J., *op. cit.* N 5, pp. 95-96.

[700] Así, afirma: *«Product liability regimes operate on the assumption that the product does not continue to change in an unpredictable manner once it has left the production line. AI does not follow this paradigm». Ibid.,* p. 98.

[701] Para un estudio más detallado de los matices: Lord Griffiths/ De Val, P./ Dormer, R. J., «Developments in English Product Liability Law: A Comparison with the American System», *Tulane Law Review*, 353, 2021, p. 384 y ss., con referencias concretas a casos polémicos en EEUU en relación con la excepción de los riesgos de desarrollo, que los autores tildan de controvertida (el término usado en el ámbito anglosajón es el de «*state of the art*»).

[702] De hecho, la Directiva aclara en su artículo 7.3 que un producto no se considerará defectuoso solo porque más adelante se introduzca en el mercado otro que es mejor.

[703] Al efecto, en español, y tratando los riesgos de desarrollo desde la perspectiva del derecho civil: Salvador Coderch, P./Rubí Puig, A., «Riesgos de desarrollo y evaluación judicial del carácter científico de dictámenes periciales. Avance del Tratado de Responsabilidad Civil por Daños Causados por Productos Defectuosos», *InDret*, 1/2008. También en: Salvador Coderch, P./Rubí Puig, A., «Causas de exoneración de la responsabilidad: Excepción por riesgos de desarrollo» en Salvador Coderch, P./Gómez Pomar, F. (Coords.), *Tratado de responsabilidad civil del fabricante*, Civitas, Madrid, 2008, pp. 568-584. Realizan ambos autores un repaso por la historia de esta excepción (altamente influenciada por el caso de la talidomida que, de hecho, da lugar a una modificación en la legislación norteamericana), repasando cómo diversos países europeos han modificado su legislación para acogerla (así, además de España, Francia, Luxemburgo o Finlandia). *Ibid.,* pp. 10 y ss. Finalmente, el texto reflexiona sobre la importancia de que exista un criterio objetivo para determinar si un defecto era cognoscible, para lo cual es esencial la existencia de conocimientos científicos, dando criterios para afirmar si este conocimiento científico es asentado (pp. 20 y ss.), analizando la doctrina asentada en el caso Daubert como estándar (exigiéndose que los jueces, para aceptar un testimonio como experto, valoren al menos cuatro factores: la falsabilidad o reproducibildiad de las pruebas; el control por pares; la tasa de error conocida y la aceptación general de la comunidad científica).

En un sentido similar, Cappellini afirma que se ha producido un cambio de paradigma respecto de la tradicional máquina entendida como objeto o artilugio [704], pues, cuando se trata de imputar un delito culposo (imprudente) por un mal uso o defecto en el diseño, precisamente la tipología de este tipo de productos reviste estos casos de especial complejidad. Ante ciertas situaciones u órdenes, un producto reacciona siempre de la misma manera (es decir, el objeto-producto tradicional tiene un comportamiento esencialmente previsible), el hombre que lo diseña o usa siempre estará, por tanto, en condiciones de representarse los efectos de este producto. Sin embargo, resalta Cappellini que en el caso de los sujetos artificiales inteligentes la imprevisibilidad al menos parcial de su comportamiento frente a análogos estímulos forma parte de su funcionamiento estructural, de forma que la máquina no es ya un intermediario neutral o un puro medio-objeto *(mezzo-oggetto),* sino que se convierte en un producto subjetivizado *(prodotto soggettivizzato),* que ya no se limita a realizar la voluntad humana que lo respalda, sino que actúa en el mundo de una manera que ya no está completamente gobernada por la mano del hombre [705].

La conclusión que podemos alcanzar, hasta aquí, es la siguiente: no resulta forzado, si atendemos a un concepto algo más amplio y apegado a la normativa civil (no necesariamente vinculante para la imputación penal, pero sí relevante, pues de sus postulados bebe la lógica de la responsabilidad penal por el producto), afirmar la entidad de la IA como producto. Sin embargo, este producto presenta unos rasgos singulares que, quizá, escapan a la tradicional concepción de los mismos. Desde luego no todos los productos, una vez lanzados al mercado, permanecen absolutamente estáticos (aunque esto, en realidad, se concibe en la normativa civil como una excepción), pero sí es la primera vez que este dinamismo surge *ad intra*, como consecuencia no solo de las propias características de aprendizaje del sistema, que superan las instrucciones de su productor, sino como consecuencia de la interacción con su entorno (que incluye, en un contexto ultra conectado, no solamente la interacción humana, sino la interacción con otras redes de inteligencias artificiales).

Estos caracteres, absolutamente novedosos, nos permiten seguir hablando de un producto sin deformar demasiado el alcance del término, pero difícilmente podremos hablar de que, en todo caso, las mutaciones y actuaciones de este producto dinámico deban ser consideradas defectos.

---

[704] Así, Cappellini, hablando de la «macchine» como *«(il) più rudimentale degli utensili»* refiriéndose en este caso a la posibilidad de usarla como medio o instrumento del delito. Cappellini, A., *op. cit*. N. 662.

[705] *Ibid.,* pp. 24-24.

Veamos cómo este dinamismo particular da lugar a una reconsideración sobre el segundo elemento del binomio en juego; sí es producto, pero quizá no es defectuoso.

## 2. ¿SON LOS ERRORES DE LA IA UN DEFECTO?

En un sentido tradicional, podemos afirmar, con Íñigo Corroza, que cuando se hablaba de producto defectuoso el concepto de defecto estaba ligado a «las expectativas de seguridad de los potenciales consumidores o usuarios» [706]. Este concepto depende de un triple condicionante: la naturaleza del producto, las expectativas de los consumidores sobre él y las expectativas de los fabricantes sobre los consumidores. Acepta Íñigo que estos tres factores irán variando según el desarrollo de la sociedad, el momento histórico y, muy marcadamente, el estado de los conocimientos (tanto científicos como técnicos). Para esta autora, el defecto es solo aquel imputable al fabricante, pues solo de él se puede afirmar que defrauda unas expectativas normativas (es decir, las expectativas de seguridad que el ordenamiento penal le presupone). Sin embargo, si es el consumidor quien defrauda las expectativas en cuanto a su comportamiento esperado, no podrá decirse que haya un defecto, sino a lo sumo un fallo, pues esta defraudación de expectativas es cognitiva y no normativa. El argumento que utiliza Íñigo para excluir la calificación de «defecto» y, por tanto, de posibilidad de responsabilidad penal, por parte del consumidor, es que si se produce este fallo respecto de lo que se esperaba del producto y el uso que realmente se hace de él (por una falta de conocimiento sobre el mismo), sucediendo algún tipo de fallo en su uso o consumo, será el propio consumidor el que sufra las consecuencias, por lo que la *poena naturalis* que sufra será suficiente.

No consideramos, empero, que este binomio tan diferenciado sea aplicable a una situación como la que se examina aquí, pues, a diferencia de los supuestos más clásicos donde el producto estaba dirigido al consumo humano (supuestos en los que parece estar pensando Íñigo, a pesar de que utiliza la palabra usuario junto a la de consumidor en alguna ocasión), son perfectamente imaginables escenarios donde se produzcan daños por la actuación de un usuario de IA que no recaigan sobre él mismo, sino sobre terceras personas. Por tanto, y sin que se prejuzgue todavía la cuestión relativa al diferente fundamento de la eventual responsabilidad penal de fabricantes o usuarios no profesionales, no podemos aceptar que en la mayoría de los casos las consecuen-

[706] Íñigo Corroza, E., *op. cit.* N. 697, p. 65.

cias dañosas de los errores provocados por un sistema de IA (en su interacción con usuarios) recaigan sobre el mismo usuario en exclusiva.

Otra de las ideas que destaca Íñigo es algo más pacífica: los defectos (adscritos, recordemos, al fabricante), pueden sucederse en diferentes etapas del *iter* de producción, diferenciando esta autora entre una etapa 0 (fase anterior a la producción del bien, donde se incluyen la etapa de investigación y diseño del producto), la etapa 1 (fase propiamente de fabricación, donde pueden producirse fallos humanos y técnicos), etapa 2 (fase en que se elabora la información y las instrucciones sobre el uso recomendado del producto), y la etapa 3 (distribución) [707]. A lo largo de todas ellas reconoce que se pueden producir defectos (defraudación de expectativas normativas, en su terminología) que pueden dar lugar a responsabilidad penal. Ahora bien, no considera la autora lo mismo respecto de los posibles defectos de una fase ulterior en el tiempo: los denominados «defectos de desarrollo» [708]. Como se ha apuntado previamente, este tipo de defectos responden a supuestos en que, por el estado en que la ciencia se encontraba en un momento determinado, cuando se diseñó y comercializó el producto, no se podían conocer los riesgos que más tarde se evidencian. Una vez más, en su distinción entre defraudación de expectativa normativa y cognitiva, Íñigo considera que al ser este defecto de naturaleza cognitiva (falta de conocimiento sobre el estado de la ciencia), no es imputable al fabricante. Esta no es, empero, la opinión mayoritaria de la doctrina penal que fundamenta una responsabilidad por omisión de los fabricantes que no realicen una adecuada supervisión a lo largo de toda la vida del producto *(vid. infra).*

Hemos de aceptar, de lo hasta ahora afirmado que, en la medida en que los defectos o fallos derivados de un producto de IA se enmarquen en las etapas señaladas (producción, fabricación, información y distribución) en un sentido análogo al de los clásicos defectos de productos dedicados al consumo, ninguna particularidad arrojaría la IA. Pongamos por caso que un robot de asistencia doméstica inteligente que lleva incorporado un sistema de IA complejo es lanzado al mercado. En la medida en que el defecto del producto sea reconducible a una de esas fases porque, por ejemplo, en el momento de fabricación de una pieza, se ha producido una actuación gravemente im-

---

[707] *Ibid.,* p. 71. Un reconocimiento similar en Bertolesi, R., *op cit*., N. 423, p. 128.

[708] Se remite la autora, de hecho, a la definición que de estos defectos hace la doctrina en el campo civil, pues son una categoría propia de esta disciplina. Así, Salvador Coderch, P./ Ruiz García, J. A./ Seuba Torreblanca, J. C./ Solé i Feliu, J./ Luna Yerga, Á./ Carrasco Martín, J., «Los riesgos de desarrollo». *InDret,* 1/2001, p. 6: *«Son* ***riesgos de desarrollo*** *los causados por un* ***defecto*** *de un producto que* ***no*** *era* ***reconocible*** *a la luz del estado de los* ***conocimientos científicos y técnicos*** *existentes en el momento de la* ***comercialización*** *del producto de que se trate»* (negrita en el original).

prudente que ha llevado a un defecto sobre la cadena de producción, no habrá problema en reconducir el supuesto al clásico esquema de responsabilidad. Lo mismo puede decirse respecto de las otras etapas. Habrá defecto si se detecta una línea de código errónea en un programa o una vulnerabilidad de tipo informático, porque el software no sea lo suficientemente robusto. Lo mismo cabe decir respecto de los errores de información (si la información o instrucciones de uso de un algoritmo es insuficiente habrá responsabilidad por este defecto de advertencia) [709].

Sin embargo, el tipo de fallo o defecto de un sistema de IA que puede dar lugar a daños en bienes jurídico-penales va mucho más allá de este escenario. Los defectos que aquí preocupan, (aquellos que dan lugar a lo que se ha caracterizado como *hard AI crime)* no derivan de este escenario, sino del error inherente a las capacidades de desarrollo, adaptación y aprendizaje de la IA. La característica propia de las IA es su capacidad de tomar decisiones (o ayudar a la toma de decisiones) en muy diferentes campos, suplementando las capacidades humanas con la ayuda de sus potentes capacidades de cálculo y correlación.

Tendremos que encontrar, pues, alguna otra caracterización capaz de captar esta particularidad. Hay quienes afirman que, de hecho, el mecanismo de funcionamiento de los algoritmos de IA hace imposible hablar siquiera del término «defecto». Señalan Teubner y Beckers [710] que el requisito de defecto no encaja bien con la naturaleza de los agentes digitales y su comportamiento. Si bien los agentes electrónicos pueden considerarse defectuosos en casos de errores de fabricación o diseño (los arriba aludidos), reconocen estos autores que los daños suelen derivarse de comportamientos algorítmicos impredecibles. Sin embargo, sería difícil hablar de productos defectuosamente fabricados o diseñados cuando los sistemas toman decisiones basadas en reglas probabilísticas [711], pues cuando se trata de sistemas complejos resulta problemático hablar en términos de productos defectuosamente fabricados o diseñados, ya que estos sistemas no funcionan siguiendo reglas deterministas que conduzcan siempre a un único resultado correcto, sino que toman decisiones a partir de cálculos probabilísticos. En este marco, el error no es una señal

[709] Así, Borghetti, J-S., «Civil Liability for Artificial Intelligence: What Should its Basis Be?», *La Revue des Juristes de Sciences Po,* 17, junio 2019, p. 96.

[710] Beckers, A., /Teubner, G., *op cit.* N. 111, pp. 74 y ss.

[711] Chagal-Feferkorn, K. A., «Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers», *Stanford Law & Policy Review*, 30(1), 2019, p 84, señalando que los sistemas de toma de decisiones se basan, por su propia naturaleza, en predicciones probabilísticas y que estas predicciones inevitablemente fallan: *«sophisticated systems, in particular self-learning algorithms, rely on probability-based predictions, and probabilities by nature inevitably get it wrong some of the time.»*

de defecto en el diseño o la fabricación, sino una consecuencia inevitable de la naturaleza misma del razonamiento probabilístico. Dicho de otro modo, la probabilidad nunca garantiza certeza absoluta, sino que asigna grados de confianza a distintos resultados. Por ello, es inherente a la técnica que existan márgenes de error: incluso el sistema mejor construido puede equivocarse en ciertos casos, del mismo modo que ocurre con cualquier modelo estadístico o con cualquier tecnología que opera bajo incertidumbre. Así, conviene distinguir entre error y defecto: el primero es consustancial al funcionamiento de sistemas que predicen y deciden bajo condiciones de probabilidad, mientras que el segundo aludiría a una falla evitable en el diseño, la ejecución o la calidad de un producto.

Además, añaden estos autores que resulta muy controvertido evaluar el carácter defectuoso *(«defective»)* de una decisión que está programada para ser autónoma. En este sentido, plantean los siguientes interrogantes: «¿Cuál es el estándar de comparación? ¿Una decisión humana o una tomada por otro algoritmo? ¿Qué se consideraría defectuoso: el hecho de que se haya causado un daño, o el hecho de que se haya materializado una alta probabilidad de una decisión incorrecta?» [712] Señalan además, analizando el régimen de responsabilidad por el producto desde el punto de vista civil, que no cabe considerarlo en ningún caso un régimen de responsabilidad objetiva en sentido estricto, sino que (tanto en su vertiente europea como estadounidense), admite la exoneración de responsabilidad por falta de culpa en casos de defectos posteriores y defectos de desarrollo.

Es especialmente incisivo en este sentido Borguetti [713] al examinar la posible defectuosidad de los algoritmos de cara a establecer responsabilidad por daños causados por IA. Parte este autor de dos posibles métodos para convenir la defectuosidad de un algoritmo: o bien comparar su desempeño o actuación con la de un ser humano (es decir, examinar si un humano razonable hubiera actuado de la misma forma que el algoritmo), o bien compararlo con el desempeño de otros algoritmos similares. Sin embargo, ambas comparaciones presentan limitaciones importantes [714]. En primer lugar, sostiene que los algoritmos deberían ser más eficientes y seguros que los humanos, por lo que no tendría sentido comercializar un algoritmo que supere en peligrosidad las acciones humanas que busca reemplazar. Sin embargo, una comparación directa con lo que haría una persona razonable no es suficiente para determinar si un

[712] Beckers, A., /Teubner, G., *op cit.* N. 111, p. 74. (traducción propia).

[713] También desde el campo civil, pero ahondando en lo que significa un defecto cuando se trata de decisiones probabilísticas): Borghetti, J-S., *op. cit*. N. 709.

[714] *Ibid.,* pp. 98 y ss.

algoritmo es defectuoso, ya que los algoritmos no funcionan como los humanos y pueden generar resultados derivados de una racionalidad distinta. De hecho, dos algoritmos diseñados para la misma tarea pueden comportarse de manera diferente debido a sus lógicas internas, sin que eso signifique que ninguno de ellos sea defectuoso.

Por otro lado, comparar los resultados generales de dos algoritmos parece ser la forma más adecuada de evaluar si uno de ellos es defectuoso. Sin embargo, si un algoritmo produce resultados menos favorables que otro, esto no necesariamente implica que sea defectuoso. En este caso, el hecho de que un algoritmo cause más accidentes o más graves no implica que esté defectuoso, ya que podría ser simplemente que aún no ha alcanzado el nivel de desempeño de un algoritmo más avanzado. Para evaluar la defectuosidad, una opción sería establecer un umbral de rendimiento en relación con un algoritmo de referencia, pero esto plantea dificultades prácticas y teóricas, como la identificación del algoritmo de referencia y la determinación de un porcentaje adecuado de comparación [715].

A pesar de estos desafíos, Borguetti destaca la importancia de tener una metodología clara para evaluar la defectuosidad, especialmente considerando que los algoritmos de autoaprendizaje evolucionan con el tiempo y su rendimiento no siempre es predecible.

Desde luego, la centralidad del concepto de defecto no es tal en el derecho penal como puede serlo en el derecho civil pues, en la línea seguida por Íñigo, lo que se trata es de considerar si ese defecto, fallo o funcionamiento anormal del sistema de IA puede reconducirse a una defraudación de expectativas por parte del fabricante (o cualquier otra persona involucrada). Sin embargo, traer esta discusión relativa al defecto puede servir para poner sobre la mesa que estamos ante un contexto donde la lógica tradicionalmente considerada no parece funcionar del todo.

### 2.1 ¿Una nueva tipología de defectos?

Paredes pone sobre la mesa una distinción que puede ser de utilidad en nuestro intento de ahondar en la nueva tipología de defectos con relevancia

[715] En este sentido recuerdan los argumentos empleados por Borguetti a la problemática señalada respecto de los algoritmos de predicción de la peligrosidad y la falta de consenso sobre la medición de su eficacia, puesta de manifiesto entre nosotros de forma magistral en: Martínez Garay, L./ Montes Suay, F., *op. cit.* N. 17. Señala Borguetti que establecer estándares estrictos, como que un algoritmo debe ser tan bueno como el mejor en el mercado, podría generar otros problemas, como injusticias en los juicios sobre la defectuosidad, ya que un algoritmo ligeramente inferior podría ser considerado defectuoso sin que realmente lo sea, haciendo que solo se privilegie la pervivencia de unos pocos algoritmos en el mercado.

penal, cual es la diferenciación entre que el producto conlleve un riesgo en sí mismo y que el producto se presuponga completamente inocuo [716]. La idea que aquí nos interesa, por el momento, es la siguiente: si un producto presenta un riesgo implícito o inherente a su propia existencia, no pudiendo considerarse inocuo, entonces es necesario tener muchos más conocimientos (de tipo técnico y especializado) para saber detectar cuándo este peligro va más allá de lo implícito. Esto es, en esencia, lo que ocurre con los sistemas de IA que, por sus propias características (su previsibilidad implícita e inherente, y su posibilidad de tomar decisiones operativas), requieren un conocimiento experto que permita detectar el verdadero defecto o funcionamiento anormal [717].

Por su parte, Selbst examina cómo la inserción de la IA en muchos procesos de toma de decisiones puede suponer perturbaciones para imputar responsabilidad por la impredecibilidad del error. Desde la perspectiva del usuario, afirma que es posible que la posibilidad de usar una IA sin errores no

---

[716] Paredes Castañón, J. M., «Responsabilidad penal por productos defectuosos», en Camacho Vizcaíno, A., (Dir.), *Tratado de Derecho penal económico*, Tirant Lo Blanch, Valencia, 2019, p. 620, aclarando que *«conviene distinguir dos hipótesis: de una parte, aquellos casos en los que el producto (un alimento, por ejemplo), de estar en buen estado, no tiene por qué implicar ningún riesgo para el consumidor y, de otra, aquellos otros en los que el consumo o utilización del producto (es el caso, por ejemplo, de muchos medicamentos) conlleva siempre algún riesgo, aunque éste se vea incrementado de existir deficiencias. La razón para hacer esta distinción estriba en que el conocimiento que fundamenta el dolo eventual es, en un caso y en otro, diferente. Así, cuando consumir el producto en buen estado no es en principio peligroso, son dos cosas las que hay que demostrar para afirmar el dolo. Primero, que el autor o autores supieran que la probabilidad de que el consumo del producto diese lugar a una lesión de la vida, integridad física, etc. era relevante (no, pues, una mera posibilidad remota). Y segundo, que dicho conocimiento de la probabilidad de lesión revistiera un grado de detalle razonable. "Razonable" quiere decir aquí un conocimiento suficiente, teniendo en cuenta la capacidad técnica del sujeto».* Toma Paredes esta diferenciación como base para distinguir el dolo eventual de la imprudencia consciente en los casos más difíciles, determinando que, si el uso del producto ya conlleva un riesgo implícito, el nivel de conocimiento sobre el riesgo derivado del producto debe ser mucho más elevado, por lo que solo se podrá afirmar dolo si se tienen capacidades técnicas elevadas, es decir, solo serán dolosas las actuaciones del personal técnico. El personal no técnico, por el contrario (al carecer de este conocimiento elevado) tendrán muchas más dificultades para distinguir entre el riesgo normal inherente al producto y el anormal, por lo que será mucho más difícil afirmar su responsabilidad a título de dolo eventual.

[717] Incluso, como pone de manifiesto De la Cuesta Aguado, con la posibilidad de que, como consecuencia de la interacción humano-máquina, un error humano se vea potenciado por la capacidad de decisión del agente. De la Cuesta Aguado, M., P. «Inteligencia artificial y …» *op. cit*. N. 58*,* p. 58. Aunque, de hecho, a nuestro modo de ver, puede ocurrir también en el sentido contrario; es decir, que el error en la decisión algorítmica se vea potenciado por la interacción con el humano, en lo que se ha venido a denominar bucles de retroalimentación *(feedback loops),* que funcionan en una doble dirección. Así, basándonos en la definición y el ejemplo de Barocas *et al.,* cuando un motor de búsqueda muestra resultados, normalmente registra los enlaces en los que el usuario hace clic y cuánto tiempo permanece en esas páginas, y trata esta información como señales implícitas de qué resultados fueron considerados más relevantes. Cuando una plataforma de videos recomienda un video, usa el pulgar arriba o abajo como una señal explícita y esta retroalimentación se utiliza por el sistema para refinar el modelo, Barocas, S./ Hardt, M./ Narayanan, A., *Fairness and machine learning. Limitations and opportunities*, MIT Press, Cambridge, 2023, pp. 13-14.

dependa del nivel de cuidado o la diligencia desplegada por el usuario [718]. Y ello es así por las peculiares características de los sistemas de IA, que por un lado tratan de replicar las habilidades humanas pero, por otra parte, intentan superarlas (piénsese en lo inviable que sería crear, y ofertar, una IA de asistencia en la conducción que condujese, de hecho, mucho peor que un humano, o una herramienta de diagnóstico médico que fuera sensiblemente peor que un médico medio). Está dentro de la lógica de planteamiento y diseño de la herramienta que desempeñe sus tareas mejor que una persona media en su campo concreto de aplicación (o, al menos, que aspire a hacerlo es el objetivo indudable), por lo que se deriva un conflicto si se parte de la presunción de que un usuario medio será capaz de detectar los errores de un sistema que ha sido configurado para superar las capacidades de ese usuario.

Selbst desarrolla estas ideas, afirmando que muchos de los sistemas de IA parten de que o bien no existe una «verdad de base» (un resultado claramente óptimo) o bien sí existe, pero el sistema modifica el resultado con su actuación y no hay manera de verificar si, en caso de no haber intervenido, el daño se hubiera o no producido. Por tanto, y sin estos principios de base para contradecir las predicciones o decisiones de la IA, una vez que el sistema toma una decisión, el hecho de que el usuario decida o no confiar en ella no puede basarse en algo racional, sino simplemente en una cuestión de fe [719]. El único juicio que puede considerarse razonable o no (o, en lo que a nosotros nos interesa, diligente o imprudente), tiene que ver con la decisión sobre el uso de la herramienta en un inicio, en un determinado contexto.

Sí reconoce Selbst que puede tener algún papel en la labor de supervisión de los errores o aciertos de la IA, la posible explicación sobre los factores que llevaron a tomar una determinada decisión (lo que se ha venido a denominar *explainability),* pero estas explicaciones *ex post* solo servirán para detectar errores allá donde la lógica de la decisión no sea lo suficientemente compleja para el cálculo humano [720].

---

[718] Selbst, A. D., «Negligence and AI's Human Users», *Boston University Law Review*, 4, 2020, p. 1331.

[719] *Ibid.,* p. 1339.

[720] Pone Selbst un ejemplo interesante: un sistema da como resultado una correlación inversa entre tener asma y sufrir una muerte por neumonía (es menos probable que un paciente con asma acabe muriendo por neumonía). Dicha correlación inversa, explica Selbst, si bien no tiene por qué resultar evidente en un primer momento, muestra una relación que nuestros cerebros procesan con relativa facilidad, a saber: si un paciente con una determinada dolencia respiratoria acude más frecuentemente a revisiones médicas, no será de extrañar que en alguna de esas intervenciones se detecte cualquier otra enfermedad respiratoria. Sin embargo, señala el autor, a medida que esta correlación sea menos evidente para nuestras capacidades, será mucho más difícil que sepamos controlar en qué medida dichas correlaciones son espurias. *Ibid.,* p. 1340.

La reflexión valiosa que podemos extraer de los trabajos hasta ahora examinados tiene que ver con la necesidad de reformular la caracterización de los errores de los sistemas, y esto nos lleva necesariamente a hablar del error por excelencia de los sistemas de IA: el sesgo.

Más allá del ámbito de la responsabilidad por producto defectuoso, una de las principales preocupaciones sobre los algoritmos es el riesgo de sesgo [721] y discriminación, ya que los datos de los que se nutren los sistemas no son imparciales, siendo ya muy común que los resultados de los sistemas de IA generen algún tipo de discriminación. No en vano en los últimos años se ha creado todo un campo especializado en informática enfocado en la equidad y la no discriminación. Dicha discriminación en los modelos de IA suele aparecer como efecto no deseado debido a decisiones subjetivas tomadas durante todo el proceso de desarrollo y entrenamiento de los algoritmos [722].

Tal y como señala Selbst [723], los debates más comunes hasta el momento han girado en torno al sesgo discriminatorio en contextos críticos (principalmente, acceso al empleo y al crédito, y también de forma paradigmática en ámbitos sensibles penales como la policía y justicia penal [724]). Pero lo cierto es

---

[721] Aunque el término sesgo es polisémico, pudiendo hablar, al menos, de sesgos cognitivos propios de las limitaciones de la psicología humana; sesgos de tipo técnico o estadístico (según la RAE: «*Error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras»)* o sesgos de tipo social, nos referimos en este punto a la acepción más común del sesgo algorítmico, lugar ya común entre la literatura, y para cuya definición podemos acudir a Friedman, B./ Nissembaum, H., «Bias in computer systems», *ACM Transactions on Information Systems*, Vol. 14, 3, 1996. p. 332-333. Afirman las autoras que, aunque el término sesgo se refiere, sin más, a una inclinación o tendencia, en el contexto de sistemas informáticos se debe entender como discriminación sistemática e injusta que un sistema realiza contra ciertos individuos o grupos en favor de otros, de forma que solo cuando se den estos dos elementos (reiteración y resultado injusto), estaremos hablando de un sesgo con relevancia a los fines que aquí nos ocupan. Cuando, en otros puntos de la investigación, nos refiramos al sesgo, lo haremos siempre partiendo de esta conceptualización.

[722] La literatura sobre el particular resulta inabarcable. Basten, a título de ejemplo: Barocas, S./ Selbst, A., «Big Data 's Disparate Impact» *Cal. Law Review*, 104, 2016; Kim, P. T., «Data-Driven Discrimination at Work», *Wm. & Mary L. Rev,* 58, 2016*,* pp. 857-936; Zliobaite, I.: «Fairness-aware machine learning: a perspective», 2017, disponible en *arXiv:1708.00754;* Veale, M., /Binns, R., «Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data», *Big Data & Society,* July–December, 2017, pp. 1-17; Adams-Prassl, J./Binns, R.,/Kelly-Lyth, A., «Directly Discriminatory Algorithms, *The Modern Law Review,* 2022*,* pp. 1-32*.;* Zarsky, T. Z., « Understanding Discrimination in the Scored Society», 89 *Wah. L. Rev*.,89, 2014, pp. 1375- 1385. En España: Soriano Arnanz, A. «Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules», *Revista d'internet, dret y política*, 38, 2023, pp. 1-12; Ortiz de Zárate, L., «Sesgos de género en la inteligencia artificial», *Revista de occidente*, 2023, 502, pp. 5-20. Nuestra propia postura al respecto, en: Bravo Bolado, A. «Justicia algorítmica: un enfoque sociotécnico», *op. cit. N. 221*, pp. 1-42.

[723] Selbst, A., «Negligence …», *op. cit*. N. 718., p. 1355.

[724] El *leading case* indiscutible cuando se habla de discriminación algorítmica es el famoso Caso Compas, acrónimo de la herramienta de predicción de la peligrosidad *Correctional Offender Management Profiling for Alternative Sanctions*. El algoritmo se hizo célebre tras el caso Eric Loomis, (resuelto por sentencia State v. Loomis. 881 N. W.2d 749 (Wis. 2016)), donde el acusado denunció discriminación por razón de

que la IA, según se va incorporando a muy diferentes contextos de uso, corre el riesgo de tomar decisiones personalizadas, tomando en cuenta características personales (como consecuencia de su dependencia sobre los datos), haciendo que, en prácticamente cualquier contexto de aplicación, el sistema pueda tomar decisiones sesgadas que resulten en un peor resultado para las personas por su pertenencia a un colectivo [725].

Utilizaremos, para ilustrar las dificultades propias del sesgo, un caso hipotético [726]. Pensemos en una herramienta de diagnóstico médico que incorpora IA, que, en general, es más precisa que el diagnóstico humano. Sin embargo (y como suele ocurrir a menudo en los ensayos clínicos, realizados mayoritariamente sobre un determinado perfil de población), la herramienta resulta ser menos precisa para las mujeres que para los hombres al detectar ciertas condiciones. El resultado global es, por tanto, mejor (se detectan en mayor medida las enfermedades), pero no lo es con respecto a un grupo de población. Supongamos que, en un caso concreto, el médico (usuario de la herramienta), diagnostica mal a una mujer, que acaba falleciendo. El mal diagnóstico se debe, en todo caso, a que la IA ha reproducido un defecto que ya ocurría previamente; se diagnostica mejor a los hombres que a las mujeres.

Cabe preguntarse, como hace Selbst, en qué se traduce un escenario como este en lo relativo al deber del usuario. ¿El hecho de recurrir a una herramienta, que de base puede estar sesgada y de hecho lo está, mostrando mucho peores resultados en un grupo, y mucho mejores resultados en otro, implica una contravención del deber de cuidado exigible a un profesional como un médico? Aunque Selbst pone de manifiesto este problema como parte de un

---

género en la herramienta, siendo desestimada su petición (amparándose el tribunal en el derecho de la empresa a guardar secreto comercial sobre el algoritmo), y destapándose después todo un debate científico sobre la discriminación racial de la herramienta tras la publicación de un estudio en: Angwin, J./Larson, J./ Mattu, S./ Kirchner, L., «Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks» *ProPublica*, [en línea], 23 de mayo de 2016, disponible en: < https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing > [Accedido por última vez el 26/01/2026]. La difícil y polémica cuestión sobre la incorrección de discriminatorio del algoritmo en: Bravo Bolado, A. «Justicia algorítmica: un enfoque sociotécnico», *op. cit. N. 221*, pp. 32 y ss.

[725] En todo caso, la dinámica discriminatoria que se despliega en estos supuestos se aleja de la tradicional discriminación cometida por el humano, pues no responde a los tradicionales criterios centrados en la intención discriminatoria. Una explicación sobre esta dicotomía en: Bravo Bolado, A., «El sesgo discriminatorio y el derecho a la igualdad en los sistemas de IA», Gregoraci Fernández, B.,/Castillo Abela, J./ Velasco Caballero, F., (Dirs.), *Derecho de la neurotecnología, Libro Blanco para el desarrollo de la Carta de Derechos Digitales en el ámbito de la neurotecnología,* Marcial Pons, 2026, pp. 847-866. Una versión más corta puede consultarse en: Bravo Bolado, A., «Igualdad e inteligencia artificial: los modernos ataques a la igualdad en el sistema de justicia penal», en Doval Pais, A./ Gutiérrez P., E. (Dirs.), *Manifestaciones de desigualdad en el sistema de justicia penal*, Aranzadi, Navarra, 2023, vol. 2, (online).

[726] El caso que se presenta es, con algunas variantes propias, el que presenta Selbst, A., «Negligence …», *op. cit*. N. 718, p. 1355.

debate propio del derecho antidiscriminacion (cuál es el coste que es aceptable asumir a costa de un beneficio global mayor [727]), en lo que aquí nos interesa pone de manifiesto una realidad difícil: en muchas ocasiones los «errores» propios de la interacción humano-máquina son muy complejos de trazar, y, ante un resultado dañino evidente (la muerte de una paciente por un mal diagnóstico), afirmar que dicho resultado es consecuencia de un defecto atribuible a una persona individualizable es altamente complejo.

## 3. INTENTOS DE ENCAJAR EL MODELO TRADICIONAL EN EL ACTUAL SISTEMA PENAL

Hemos de señalar que, pese a ser conscientes de todas las particularidades que los sistemas de IA presentan, aún podemos encontrar autores que, desde el conocimiento de estas tecnologías, han tratado de adaptar las tradicionales categorías y conceptos para seguir aplicando el modelo de responsabilidad penal por el producto.

Así, destacadamente, una experta en la responsabilidad penal por el producto como Aires de Sousa, ofrece algunas claves para entender las particularidades del sistema de hoy. De hecho, esta es la pregunta que sirve como punto de partida a un reciente trabajo donde revisita algunos de los postulados de su tesis doctoral [728]. Plantea la autora que «*La IA es un sistema tecnológico que suele incorporarse a un producto –un coche, un ordenador, un dispositivo con chip, una máquina sanitaria, un robot–, pero el funcionamiento de este producto o sistema puede variar; los riesgos inherentes a los productos "inteligentes" pueden ser predecibles o impredecibles para el productor o programador. Pero, ¿puede el sistema de responsabilidad por productos defectuosos ser útil para proteger al mercado y al consumidor de los daños relacionados con la IA?»* [729].

Para la autora lusa, puede ser útil considerar los diferentes tipos de productos inteligentes, principalmente desde la perspectiva de su autonomía y de

[727] *Ibid.,* p.1360. No es este un problema menor, y ha ocupado una gran parte de la literatura sobre sesgo algorítmico y de la discusión sobre la aplicación de las herramientas de IA en el campo propiamente penal. Así, Selbst, A., «Disparate Impact in Big Data Policing*», Georgia. Law. Rev*., 52, 2017, pp. 109- 195. Específicamente sobre la necesidad de hacer un balance entre los distintos parámetros de equidad y precisión para obtener decisiones aceptables. Kleinberg, J./ Mullainathan, S./ Raghavan, M., «Inherent trade-offs in the fair determination of risk scores», *PROC. INNOVATIONS THEORETICAL COMPUTER SCI*., 2017, disponible en: <https://arxiv.org/abs/1609.05807 > [Accedido por última vez el 26/01/2026].

[728] Aires de Sousa, S., «The importance of AI for financial crimes», *op. cit. N*. 245, p. 163-180.

[729] *Ibid.,* p. 173.

los riesgos que plantean, dividiéndolos entre: productos deterministas (por ejemplo, robots deterministas programados para realizar tareas específicas), y productos cognitivos (pone el ejemplo de los coches autopilotados), programados para tomar decisiones autónomas en situaciones complejas. Tomando estas dos categorías (producto determinado versus producto con autonomía), hace uso de la distinción entre la responsabilidad por el tipo de producto y la responsabilidad por el proceso o método de producción [730].

Esta distinción, que bebe de forma directa de la categorización realizada por Bricola [731], se considera útil aquí. Explica Aires ya en su monografía sobre responsabilidad penal por el producto que la categoría de responsabilidad por el tipo de producto sirve especialmente en los casos de productos (o actividades de producción) inherentemente peligrosos (y este peligro inherente se manifiesta tanto en el propio proceso de producción como en su utilización posterior). Se pone como ejemplo de este tipo de actividades inherentemente peligrosas las relacionadas con la producción de energía nuclear o de bienes genéticamente modificados [732]. Señala que, en este tipo de actividades especialmente peligrosas, hay una relación inmediata con la idea del riesgo permitido, pues se trata de actividades que, si bien son inherentemente peligrosas, están aceptadas por el ordenamiento jurídico [733].

En contraste, la responsabilidad por el modo de producción tiene que ver no con actividades peligrosas en sí mismas (con un riesgo inherente a la propia actividad o sector), sino con la neutralización de riesgos comunes a todas las actividades. Por ejemplo, daños en la integridad física de los trabajadores provocados por accidentes de trabajo o problemas sobre la calidad de los productos alimenticios [734]. Aquí se abandona la lógica de la precaución (regular fuertemente una actividad que puede dar lugar a riesgos pero que estará dentro del riesgo permitido en tanto se mantenga dentro de ese riesgo permitido), y se aplica la lógica de la prevención, basada en reglas de experiencia y en nexos causales fundamentados en principios científicos conocidos [735].

---

[730] *Ibid.,* p. 174.

[731] Bricola, F., «Responsabilità penale per il tipo e il modo di produzione», en *La Responsabilità dell'impresa per i danni all'ambiente e ai consumatori,* Giuffrè, Milán, 1978, pp. 75 y ss.

[732] Aires de Sousa, S., *Responsabilidade..., op cit*. N.628, p. 106-107. Los postulados de la doctrina italiana beben de: Piergallini, C., *Danno da Prodotto e Responsabilità Penale,* Giuffrè, Milán, 2004.

[733] Así, Piergallini, C., *ibid.*, pp. 769 y ss.

[734] Aires de Sousa, S., *Responsabilidade..., op. cit*. N.628, p. 108.

[735] *Idem:* «*As regras destinadas a acautelar estes riscos são criadas essencialmente ao abrigo do princípio da prevenção e fundamentam-se no recurso a regras da experiência e a nexos causais alicerçados em princípios científicos conhecidos. As normas de conduta têm em vista eliminar o risco aparentemente dominável e idóneo à produção de resultados danosos».*

Pone en relación, así, la autora, la responsabilidad por el tipo de producción con los riesgos asociados a consecuencias desconocidas e imprevisibles y la responsabilidad por el modo de producción a los riesgos con efectos previsibles y prevenidos mediante la observancia de reglas de cuidado bien conocidas.

Señala, finalmente, Aires de Sousa, separándose en este punto de la opinión de Piergallini, que la lógica de la persecución penal es, por tanto, diferente en función del riesgo al que se enfrente (riesgo por el modo o riesgo por el tipo), y que no en ambos supuestos será igual de sencillo afirmar la imputación penal. Señala las dificultades para legitimar una responsabilidad penal por el tipo de producción en tanto y cuanto genere algunas presunciones incompatibles con los principios estructurales del derecho penal [736].

Apunta Aires de Sousa que los casos tradicionales de responsabilidad por producto defectuoso son por el método o proceso de producción, es decir, cuando hay problemas o bien durante la fabricación o bien durante la distribución del producto. No tiene problema la autora en admitir que este tipo de defectos pueden darse, también, en productos que incorporen IA (pone el ejemplo del coche Tesla con un *software* que funciona de forma incorrecta, lo que obliga a retirar el producto) [737]. Sin embargo, y atendiendo a las características propias de la IA, la otra forma de responsabilidad (por el tipo de producto) parece tener más sentido. Este tipo de responsabilidad tiene en cuenta la naturaleza, cualidades y riesgos del producto más allá del momento de fabricación, y por tanto ha de tener en cuenta muchos otros niveles o etapas (la autora hace referencia a la concepción o diseño del modelo y su uso).

En definitiva, parece que Aires apuesta por un mejor conocimiento sobre los productos, pues, si lo que se persigue es minimizar los daños imprevisibles, es esencial poner el foco en la fase de diseño o prueba [738].

---

[736] *Idem*, contraponiendo su postura (remitiéndose a los argumentos por los cuales limita el recurso al principio de precaución en algunos aspectos), con la de Piergallini, que considera toda la responsabilidad por el producto transversal.

[737] Se refiere, como ejemplo, al caso por el cual Tesla tuvo que retirar caso un millón de coches eléctricos del mercado por problemas de seguridad: Jin, H., «Tesla recalls almost half a million electric cars over safety issues», 31 de diciembre de 2021, disponible en: <https://www.reuters.com/business/autos-transportation/tesla-recalls-over-475000-electric-vehicles-2021-12-30/> [Accedido por última vez el 26/01/2026].

[738] Es común la alusión a la importancia de la fase del diseño de la IA en cuanto a todas las posibles implicaciones éticas que puede conllevar, hablándose así de aproximaciones como «*legal compliance by desing*». En este sentido, Pompeu Casanovas, concluyendo que, para abordar adecuadamente el problema del cumplimiento legal, es necesario un enfoque interdisciplinar que combine el diseño técnico con perspectivas legales, éticas y sociales implantadas desde el momento del diseño. Este concepto ha sido más adelante trasladado a la IA. Hashmi, M./Koker, L. D./Casanovas, P., «Legal compliance through design:

Aunque es habitual encontrar un discurso muy favorable a la atención a las fases de diseño de los sistemas de IA (ideas como el *«privacy by desing»* o incluso *«safety by design»* o *«equity by design»),* parece que no somos conscientes de la importancia primordial que esta primera fase implica también de cara a la depuración de responsabilidades. Es de celebrar, por tanto, un enfoque que sitúa el énfasis de la responsabilidad en un momento previo, que es donde verdaderamente existe una mayor capacidad de dominio sobre las características que conformarán la herramienta (así, sobre el modelo concreto que se va a utilizar, sobre las bases de datos que van a nutrirlo, o sobre el diseño de los parámetros y objetivos).

También Bertolesi trae a colación esta distinción, aclarando la utilidad de recurrir a la normativa sobre responsabilidad por el producto, señalando que caben dos alternativas: o bien se acoge una perspectiva instrumental y se tutelan bienes jurídicos intermedios, buscando la lógica del adelantamiento de la tutela, o bien, si solo nos aferramos a la tutela desde la perspectiva de los bienes jurídicos penales (vida y salud tradicionalmente, pero no de forma exclusiva), las normas de seguridad pueden funcionar como reglas de precaución [739]. Creemos, en este punto, que a Bertolesi le asiste toda la razón cuando separa estas dos posibilidades y, a nuestro modo de ver, todos los problemas que en el capítulo anterior se predicaron respecto del principio de precaución en su vertiente dogmática se reproducen cuando se trata de trasladar el estándar de precaución (relacionado de forma muy patente con los postulados de la responsabilidad civil por producto defectuoso), a la imputación de daños sobre bienes jurídicos individuales.

Veremos, a continuación, por qué, a nuestro juicio, si se trata de seguir el camino del delito de resultado en estos escenarios, nos enfrentaremos una y otra vez a dificultades que, en algunos casos, son insalvables.

## 4. LOS PROBLEMAS QUE SUBSISTEN

### 4.1 **Causalidad**

La discusión que, al hilo de los clásicos problemas de responsabilidad penal por el producto, surgió tras la resolución de los grandes *leading cases,* se reproduce de una forma similar en el contexto tecnológico actual.

---

preliminary results of a literature survey», *TERECOM2018@ JURIX, Technologies for Regulatory Compliance, 2309*, 2018. Aires habla de *design test* or «Value Sensitive Design approach». Aires de Sousa, S., «The importance of AI for financial crimes», *op. cit. N*. 245, p. 174.

[739] Bertolesi, R., *op cit*., N. 423, p. 141.

Será de utilidad, pues, hacer un repaso por los argumentos de entonces (escépticos y favorables a la afirmación del vínculo causal), así como repasar los argumentos que hoy se exploran para comprobar si la necesidad de afirmar ese vínculo causal entre la acción humana y el evento dañoso supone un problema insalvable para la imputación penal en los *hard AI crimes.*

### a) El problema del vínculo causal entonces

Tras la afirmación de la causalidad en los polémicos casos arriba estudiados, toda una serie de críticas dogmáticas vinieron a poner de manifiesto que los tribunales estaban, de alguna forma, relajando los estándares de imputación, dando lugar a una flexibilización de los principios clásicos.

Para Hassemer, esto se traduce en una aceptación de una construcción de la causalidad basada en una lógica de caja oscura propia de un prestidigitador, donde están controladas la entrada y salida de la caja (el producto compuesto), pero su contenido permanece en la oscuridad. Es decir, se sabe que el producto en cuestión tiene una relación suficiente con determinados daños en su concreta configuración, y se sabe que además se puede excluir con cierta fiabilidad que haya otro factor nocivo distinto a ese producto, aunque no se sepa cuál es exactamente el factor nocivo dentro de los componentes [740]. Reconoce el autor alemán que esta concepción de la causalidad fue criticada por la mayoría de autores (así, Puppe [741]), y, de hecho, a él mismo le parece criticable, afirmando que cuando no es posible excluir de una forma completamente fiable que existan condiciones causales ajenas que contribuyen al resultado, la imputación así fundada es *contra reum* [742]. Critica, por tanto, que la lógica causal pase a estar basada en correlaciones estadísticas e hipótesis de simultaneidad no causal, que, más allá de considerarse constatadas, todo lo más pueden considerarse plausibles [743]. Ponen de manifiesto, además, que se produce, con esta aceptación, una subjetivización de la valoración de la prueba, pues esta se basa

---

[740] Hassemer, W./Muñoz Conde, F., *La responsabilidad por el producto…, op. cit* N. 409, tercera parte «La responsabilidad por el producto en el Derecho penal alemán», p. 133.

[741] En Alemania, además de Puppe, cuyo razonamiento se analizará a continuación, Kaufman. En España, ciertamente, la crítica fue, al menos en un primer momento, también mayoritaria (Paredes, Cuello Contreras, Torío, Romeo Casabona, Cerezo, o Muñoz Conde, entre otros). Las referencias completas a todas las obras críticas en, Gimbernat, E., «¿Las exigencias dogmáticas fundamentales hasta ahora vigentes de una parte general son idóneas para satisfacer la actual situación de la criminalidad, de la medición de la pena y del sistema de sanciones?», *ADPCP,* vol. LII, 1999, nota al pie 6.

[742] Hassemer, W./Muñoz Conde, F., *La responsabilidad por el producto…, op. cit* N. 409*,* p. 146.

[743] Así, de forma literal: «*Las correlaciones estadísticas no basadas en el caso concreto no son suficientes para la imputación individual que exige todavía nuestro Derecho penal*». *Ibid.,* p. 148.

en indicios débiles amparada bajo la libre valoración del juez, pero convirtiendo así a la jurisprudencia en menos segura y poco controlable.

La idea de Puppe, en un sentido similar, es que si no es posible individualizar la concreta sustancia que causó los daños, entonces no se estará legitimado a suponer que los daños causados no se pueden explicar por otras causas naturales [744]. La idea que late en la crítica de Puppe es que la cuestión causal debe reconducirse a una cuestión de probabilidad: es decir, que habida cuenta de que ya a finales del siglo XX no se sostiene la idea de que las ciencias de la naturaleza se basan en certezas absolutas, no queda más remedio que fundamentar la imputación del resultado sobe explicaciones probabilísticas, es decir, sobre incrementos de riesgo [745].

Por su parte, Paredes parte de una concepción crítica de la causalidad afirmada por la jurisprudencia, basada, según sus propios términos, en un «dudoso concepto normativo de causalidad», que choca con el principio de legalidad, pues fuerza sentencias condenatorias sobre la base de meros indicios, llevando a cabo una interpretación más allá de los términos de la ley [746]. Sin embargo, el autor acaba proponiendo una solución intermedia que, *de facto,* pasa por aceptar lo que en principio critica, como es determinar que bajo ciertas condiciones, cuando el proceso causal es complejo y opaco, puede darse por probado aunque no se puedan reconstruir todos sus detalles, exigiendo que existan, al menos, una ley causal suficientemente sólida desde el punto de vista científico que establezca *«no tan solo a título de mera posibilidad, ni siquiera con una probabilidad relevante, sino con una probabilidad rayana a la certeza, que a la acción le ha de seguir un resultado»,* y además que no haya otra explicación causal alternativa con base científica igual de sólida [747].

---

[744] Gimbernat Ordeig, E., «¿Las exigencias dogmáticas…», *op. cit*. N. 741, cita un primer escrito de Puppe, (nota al pie 7): Puppe, I., *Nomos Kommentar*, 1.ª ed., 1964, antes del parágrafo 13, n/m., 86, pero reconoce que más adelante (en el comentario de 1998 al StGB) ya no se encuentra esta afirmación.

[745] Así, Puppe, I., «Naturgesetze» vor Gericht. Die sogenannte generelle Kausalität und ihr Beweis, dargestellt an Fällen strafrechtlicher Produkthaftung», *JZ,* 1994, pp. 1147-1151. También, aunque no exactamente igual, pero en castellano: Puppe, I., «Problemas de imputación del resultado en el ámbito de la responsabilidad penal por el producto», en Mir Puig, S./ Luzón Peña, D. M., (Coords.), *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Bosch, Barcelona, 1996, p. 215 y ss.

[746] Paredes Castañón, J. M, «Responsabilidad penal por productos…», *op. cit*. N. 716, pp. 605 y ss. También, desarrollando concretamente la crítica a partir de la causalidad afirmada por el TS en el caso del aceite de colza, Paredes Castañón, J. M., «Caso del aceite de colza…», *op. cit*. N. 681. p. 432, afirmando directamente que el juez que realiza esta interpretación lleva a cabo una analogía *in malam partem*. En este trabajo, anterior en el tiempo, Paredes afirma con rotundidad que la constatación de la causalidad debe basarse en una conexión determinista, y no meramente probabilística, aunque como hemos podido comprobar no es esta la propuesta que acaba materializando. Crítica con la idea de la analogía *in malam partem* realizada por el TS en el caso de la colza: Corcoy Bidasolo, M. «La lección de la experiencia española …, o*p. cit*, N. 625, p. 455.

[747] Paredes Castañón, J. M., «Responsabilidad penal por productos…», *op. cit*. N. 716, p. 606. Esta idea es la que, a nuestro modo de ver, acaba por imponerse, aunque se parta de una concepción críti-

Por su parte, Kuhlen [748] sí se muestra favorable, de forma abierta, a que exista un criterio distinto por parte de la praxis judicial, de tal forma que el juez responda de forma distinta a como lo haría un científico, inclinándose este autor por aceptar la teoría de la causalidad general, pero con base en un fundamento normativo [749]. Así, afirma Kuhlen que están equivocados quienes objetan contra la confirmación de la causalidad que el juez no debe atribuirse una competencia mayor que los propios científicos en cuestiones empíricas, porque, en última instancia, esta objeción olvida que los requisitos mínimos para que una secuencia temporal de hechos repetidos sea considerada ley causal no deja de ser una cuestión normativa, y que esta cuestión puede responderse de forma diferente en el ámbito penal que en el científico, aunque reconoce que se necesitan criterios para concretar y precisar en qué se basa la decisión del juez [750].

Sin embargo, parece que según han avanzado los tiempos, esta lógica se ha ido revirtiendo, y es ya un lugar común encontrar autores que, partiendo de una concepción natural y no normativa del juicio de causalidad, arrojan algo de luz sobre los problemas mal entendidos en el pasado.

Ya Gimbernat puso de manifiesto en aquella temprana discusión que, contra las críticas esgrimidas, *«hay que decir que la relación de causalidad entre un comportamiento y un resultado consiste en un enlace cientificonatural, independientemente de si se conoce o no su mecanismo concreto de actuación»,* que igual que los casos difíciles como el de las enfermedades o muertes derivadas del caso de la colza, deben reconducirse no a una cuestión normativa (axiológica) o probable, *«sino material y necesariamente a algo cientificonatural»* [751]. Es decir, y por continuar con las clarificadoras palabras del autor, *«El tipo objetivo de los delitos de lesión exige únicamente que el autor haya causado el resultado típico, sin que sea necesario –porque la tipicidad no lo requiere– saber cómo lo ha causado».* [752] Se muestra Gimbernat crítico con todos los intentos de volver a una concepción superada de causali-

---

ca, y así la sostiene también Rodríguez Montañés cuando afirma que el juez debe poder dar por probada la relación causal aunque haya discrepancias entre la comunidad científica, siempre que acoja una hipótesis que, aunque controvertida, sea reconocida por un sector relevante de la ciencia empírica de que se trate. Vid.: Rodríguez Montañés, T., «Incidencia dogmática…», *op. cit*. N. 669, p. 126.

[748] Así, por ejemplo, Kuhlen, L., «Cuestiones fundamentales …, *op. cit*. N. 457, p. 240. Destacadamente en: Kuhlen, L., «*Grundfragen der strafrechtlichen... op. cit.* N. 457, pp. 1145-1146.

[749] También partidario de un concepto normativo de causalidad, pero mucho más ambicioso, capaz de superar todos los problemas que surgen en los casos de responsabilidad por el producto y superar las dificultades de la teoría de la imputación objetiva, Hilgendorf, E., «Relación de causalidad e imputación objetiva a través del ejemplo de la responsabilidad penal por el producto», *ADPCP*, Vol. LV. 2002, *passim.*

[750] Kuhlen, L., «*Grundfragen der strafrechtlichen... op. cit.* N. 457, p. 1145.

[751] Gimbernat Ordeig, E., «¿Las exigencias dogmáticas…», *op. cit*. N. 741, pp. 57-58.

[752] Gimbernat Ordeig, E., «La omisión impropia de la dogmática penal alemana: Una exposición», *ADPCP,* vol. L., 1997, p. 73.

dad (que él caracteriza, podríamos decir que de una manera mayoritaria hoy en España), como una cuestión que debía ser respondida solo con criterios ontológicos, para solo una vez afirmada la misma, proceder a completar el tipo objetivo con argumentos normativos [753].

Siguiendo la estela de Gimbernat, nos parece que Dopico [754] da con algunas de las claves de los problemas hasta entonces formulados. Sitúa el autor el origen de la discusión dogmática en la creación, por parte de Kaufmann, del concepto de «causalidad general», conforme al cual para poder afirmar la causación en un caso en concreto se requería hacer explícita la premisa mayor o, en sus palabras, ley causal general que explicaba que esa concreta sustancia producía esas consecuencias, de tal forma que sobre esa premisa mayor se aplicara el caso concreto como premisa menor [755]. Kaufmann se muestra, además, contrario a aceptar que las leyes estadísticas puedan dar lugar a leyes causales generales. Conviene Dopico que se produce aquí una confusión, pues una cosa es que se deba contar con una premisa mayor en el razonamiento causal, cosa que él no discute, y otra es que esta premisa mayor deba ser en todo caso una «ley causal general» [756]. Aunque señala Dopico que las discusiones que se reproducen en la doctrina española tienen diferentes características, se ha tratado aquí principalmente la crítica relativa a lo que el autor denomina «argumento metodológico» (el relativo a qué metodología sea válida para dar por buena la relación causal).

Resulta aquí fundamental la distinción que apunta Dopico entre la prueba del concreto suceso causal y la prueba de la ley causal general. Así, en opinión de una corriente ya mayoritaria, el tipo penal solo exige probar la relación causal concreta entre la conducta de un sujeto y el resultado, sin que sea necesario en todo caso que se pruebe, además, la ley causal general. El método científico para probar el primero de los extremos (es decir, el concreto curso causal), puede ser perfectamente uno basado en la estadística (en los casos de

---

[753] El desarrollo de este concepto que, como decimos, parece ser el aceptado mayoritariamente hoy en día, afirmando una causalidad mínima y basada en lo empírico, se caracteriza por completo en su fundamental obra: Gimbernat Ordeig, E., *Delitos cualificados por el resultado y causalidad*, Reus, Madrid, 1966, cuya tesis se puede encontrar, resumida, en p. 101: «*En lo que sigue vamos a tratar de fundamentar nuestra posición; posición que asigna a la causalidad un significado muy modesto – la causalidad (condicionalidad) del resultado es imprescindible para que pueda haber delito consumado, pero ninguna teoría causal puede decidir sobre la cuestión de si una acción es antijurídica o no–, y afirma que las acciones diligentes no son antijurídicas y, también, que aquí estamos ante una cuestión ajena por completo a la causalidad».*

[754] Dopico Gómez-Aller, J., «¿Dogmática o probatoria? La cuestión de la causalidad y su prueba en las intoxicaciones masivas con agentes tóxicos desconocidos», *La ley Penal*, 169, 2024, pp. 1-24.

[755] *Ibid.,* p. 6, con referencia al artículo de Kaufmann, A., *op. cit.* N. 404.

[756] *Ibid.,* p. 7.

intoxicaciones masivas analizadas aquí por Dopico es la epidemiología), que no deja de ser un método científico como otro cualquiera. En definitiva, y con la doctrina mayoritaria, señala Dopico que la cuestión de la causalidad es una *quaestio facti* [757], siendo lo importante que exista una premisa mayor científica, pero cuyo origen puede ser diverso (siempre y cuando sus razonamientos sean válidos desde el punto de vista de la lógica y de las ciencias naturales) [758].

Por lo demás, el reciente artículo de Dopico nos parece clarificador no tanto por contener una interpretación definitiva sobre las confusiones que se produjeron en sede de causalidad, sino porque pone en valor, en la última parte de su trabajo [759], una realidad que nos parece menos discutida a partir de los *leading cases* y muy valiosa a los fines que aquí nos mueven. Al centrarse la crítica en la causalidad, se descuidó otro aspecto esencial. No es que en todo caso solamente importe el «si» y no el «cómo» de la causalidad, sino que el «cómo» sirve para otro momento de la imputación. Si para afirmar el juicio de causalidad basta con realizar una afirmación genérica (este evento produjo ese resultado lesivo, sin importar de qué manera lo produjo, siempre que sepamos que así fue, con base en un método científico válido y suficientemente reconocido), no pasa igual para afirmar la imputación objetiva, es decir, el cómo se produjo ese daño (o más bien, en qué momento se produjo ese daño), importa, y mucho, para determinar quién responde de qué conducta. Sobre esta cuestión, clave para los casos de daños generados por herramientas opacas de IA, nos ocuparemos a continuación.

En definitiva, puede afirmarse que, si bien las dificultades de los casos explicados eran obvias (pues, en definitiva, pusieron sobre la mesa el debate sobre cómo debía actuar el derecho penal ante lo que se ha venido a llamar «cursos causales no verificables» o, al menos, difícilmente verificables a la vista de sus complejidades científicas), se puede concluir afirmando que la causalidad requiere, en estos casos, la subsunción de un caso particular bajo una ley general, pero con matices. Es cierta la crítica en ocasiones vertida de que sin la referencia a la ley causal general no es posible distinguir la conexión causal entre dos hechos o fenómenos *(propter hoc),* de su simple sucesión temporal *(post hoc),* ni por tanto vencer la falacia *«post hoc, ergo propter hoc»* («después de esto; por tanto, debido a esto») [760]. Pero lo que no se acepta ya es

---

[757] *Ibid.,* p. 9. Aunque acepta que hay ciertas voces autorizadas que aún son partidarias de normativizar el concepto, citando a Íñigo Corroza. Así, Íñigo Corroza, E. *La responsabilidad…*, *op. cit*. N. 697, pp. 101 y ss.

[758] Dopico Gómez-Aller, J., *op. cit*. N. 754, p. 10.

[759] *Ibid.,* pp. 17-19, criticando la condena de dos funcionarios de aduanas en la segunda sentencia del caso (STS 895/1997, de 26 de septiembre).

[760] Así, en Molina Fernández, F., «Acción y tipicidad…», *op. cit*. N. 253, nm 1350, con referencia al trabajo de Torío. *Vid*.: Torío López, A., «Cursos causales no verificables en Derecho penal», *ADP*-

que esta ley general deba identificarse con una ley científica de carácter universal, o conforme a leyes completas, capaces de explicar todas y cada una de las condiciones que dan lugar a un fenómeno. En definitiva, ni siquiera las leyes causales universales están sujetas a esta plena certeza sobre el cómo del resultado lesivo.

Lo que exigimos a la causalidad, llegados a este punto, es una mínima afirmación científico-natural, relegando todas las cuestiones normativas a un juicio ulterior.

### b) El problema del vínculo causal y la IA

El punto de partida del cuestionamiento de la causalidad cuando se trata de daños generados por tecnologías dotadas de amplios márgenes de autonomía operativa recuerda a los problemas arriba señalados. Así, ante un resultado dañoso, necesitamos una explicación causal lo suficientemente plausible (basada en leyes generales, si bien no universales sí reconocidas ampliamente por la comunidad científica), según las cuales se pueda excluir con probabilidad amplia que otras hipótesis contrarias son igualmente explicables.

Uno de los primeros planteamientos teóricos en cuanto a la causalidad en las herramientas de IA viene de la mano de los «*failures of causation*» que Bathaee localiza en las tecnologías de caja negra (*black box*). Este autor considera que los problemas derivados del *black box* implican que ciertos sistemas de IA llevan a cabo procesos de toma de decisiones completamente opacos para los seres humanos, sin que exista forma de determinar «(a) cómo llegó la IA a una decisión o predicción, (b) qué información fue determinante para el resultado, o (c) obtener una clasificación de las variables procesadas por la IA en orden de importancia» [761].

Para Giannini, que parte de que la lógica opaca de la caja negra (sumada al problema ya enunciado de las «muchas manos» y la posibilidad de que algunos sistemas den lugar a atajos [762] o predicciones inexactas que parecen correctas a primera vista), se debe afirmar que una vez que el sistema de IA ha

*CP,* 1983, pp. 221-238.

[761] Bathaee, Y., *op. cit.* N. 290, p. 906 (traducción propia). Este autor diferencia este tipo de cajas negras fuertes *(strong black box),* que asocia a determinadas técnicas de IA, de las cajas negras débiles *(weak black box).*

[762] Utiliza Giannini el término *«shortcut»* para referirse a estos errores difícilmente detectables. Giannini, A., *op. cit*. N. 143, p. 172 y ss.

sido desplegado, será difícil reconstruir *ex post* por qué actuó de una manera específica. [763]

También Pagallo pone de manifiesto este problema en el plano causal, afirmando que «los criterios cruciales para seleccionar, dentro de toda la cadena de eventos, la condición específica o el conjunto de condiciones que mejor expliquen un resultado determinado se verían desafiados por el comportamiento impredecible de estas máquinas y la complejidad de las aplicaciones centradas en redes» [764]. Partiendo del trabajo de Pagallo, concluye Giannini afirmando que este tipo de complicaciones o fallos de causalidad pueden dar lugar a dos tipos de consecuencias, que ella denomina el problema del *locus* de la responsabilidad y el problema del peso de la responsabilidad. En cuanto al primero *(locus),* se dificulta sobremanera la identificación de la causa o causas relevantes que condujeron a la realización del evento dañino o adverso, y en cuanto al segundo *(weight issue)* [765], aun suponiendo que todas las causas se identifiquen, se complica la determinación del grado de relevancia de cada factor individual dentro del evento dañino.

Otro de los argumentos innegables que nos permiten afirmar que la IA supone un reto a los postulados tradicionales sobre causalidad viene de la mano de la normativa europea en materia de responsabilidad civil extracontractual por daños causados por IA (Propuesta de Directiva sobre responsabilidad en materia de IA [766]). Esta propuesta, de mínimos, que no pretende armonizar el significado que respecto de cada uno de los conceptos se haga en los ordenamientos nacionales, (así, la culpa o la causalidad, como centrales), sí nos permite, en cambio, afirmar las dificultades que supondría no modificar el régimen tradicional que, para poder indemnizar a la víctima de un daño, requiere que la misma pruebe el nexo causal entre la conducta culposa del demandado y el daño. Lo que viene a proponer la nueva propuesta de Directiva es que, dado que la prueba de causalidad será especialmente difícil en estos casos, se podrá presumir bajo ciertos parámetros [767].

---

[763] *Idem*.

[764] Pagallo, U., *The laws of robots*, op. cit. N. 143, p. 74 (traducción propia).

[765] Giannini, A., *op. cit*. N. 143, p. 170. A nuestro modo de ver, mientras el problema del *locus* sí se debe reconducir al requisito de la causalidad (determinar si X conducta fue causa de Y resultado), el problema del peso *(weight)* puede enmarcarse ya en el siguiente estadio de valoración, que tiene que ver directamente con la idea de qué persona generó un riesgo lo suficientemente peligroso (o previsiblemente susceptible de dañar el bien jurídico en cuestión), que se analizará *infra.*

[766] Propuesta de Directiva del Parlamento y del Consejo, de 28 de septiembre de 2022, Bruselas, 28.9.2022 COM(2022) 496 final 2022/0303 (COD).

[767] Sin perjuicio de que esta norma se analizará en detalle en el capítulo séptimo, *vid*., por ahora, en detalle sobre el alcance de esta presunción, así como sobre la relación de esta norma con la nueva directiva de responsabilidad por el producto defectuoso y el Reglamento de IA: Muñoz García, C., *Regulación*

Afirma Muñoz García, en lo relativo a la génesis de este instrumento, que: *«hemos de admitir que, el PE sí avistó en muchos aspectos la relevancia de proporcionar una regulación diferenciada para los daños causados por sistemas de IA de los ocasionados por productos. La complejidad, el grado de conectividad, la opacidad o la falta de transparencia, la vulnerabilidad, la posibilidad de ser modificados mediante actualizaciones, o la capacidad de aprendizaje autónomo de los sistemas de IA, constituye un hándicap insoslayable. Tengamos en cuenta que, la opacidad y autonomía progresiva del sistema específico, per se, dificulta la trazabilidad de las variaciones o actuaciones precedentes en la causación del daño y deriva en una imposibilidad en la averiguación de la causalidad –fáctica o jurídica– y en la prueba de la conexión causal»* [768].

Así, en el considerando 28 de la citada Propuesta se explica que esta dificultad para probar la causalidad se dará incluso en sistemas de IA que no sean de alto riesgo, que podrían conllevar una dificultad de prueba: *«tales dificultades podrían evaluarse a la luz de las características de determinados sistemas de IA, como la autonomía y la opacidad, que hacen muy difícil en la práctica la explicación del funcionamiento interno del sistema de IA, lo que afecta negativamente a la capacidad del demandante para demostrar el nexo causal entre la culpa del demandado y la información de salida de IA. Los órganos jurisdiccionales nacionales deben aplicar la presunción cuando el demandante se encuentre en una situación excesivamente difícil para demostrar la causalidad por verse en la obligación de explicar la manera en que el acto u omisión humano determinante de la culpa llevó al sistema de IA a producir la información de salida que dio lugar al daño o a no producir la información de salida cuya ausencia dio lugar al daño»* [769].

También en España ha sido objeto de discusión esta cuestión en el campo penal. Así, Pérez Arias, afirmando de manera rotunda que en los escenarios de daños causados por IAs, no existiría técnicamente la relación causal entre el hecho primario del agente y el hecho final concreto, es decir, el resultado lesivo que se termina produciendo. Para el autor no es posible ni siquiera afirmar una eventual causalidad cumulativa, pues, aunque estemos ante varios actos que, en conjunto, conducen al resultado, el aprendizaje del algoritmo no es un

*de la inteligencia artificial en Europa. Incidencia en los regímenes jurídicos de protección de datos y de responsabilidad por productos*, Tirant lo Blanch, Valencia, 2023.

[768] *Ibid.,* p. 164.

[769] Considerando 28 de la Propuesta de Directiva del parlamento europeo y del consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA), de 28 de septiembre de 2022.

acto consciente de la persona natural, por lo que no se va a poder considerar ese acto algorítmico de aprendizaje como una conducta consciente y subsiguiente de la persona [770]. La consecuencia que deriva Pérez Arias de esta circunstancia es que en la mayoría de las «conductas» de las IAs, entraremos en el terreno de la predicción estadística, donde será necesario indagar sobre la combinación matemática de datos que sea capaz de explicar la causalidad y el margen de error.

Aunque el autor explicita en el mismo texto que la consecuencia a la que conduce esta afirmación es la negación de que estas conductas puedan ser objeto del juicio de imputación objetiva [771], no nos parece evidente que esto sea así. Quizá se puede entender que, tal y como otros autores sugieren, lo que Pérez Airas pretenda es establecer una suerte de ruptura del nexo de imputación llevada a cabo por una conducta autónoma asimilable a la humana, pero tampoco se llega a entender por qué, entonces, considera adecuado el método estadístico para alcanzar una explicación causal.

Algo más claro resulta Piergallini cuando alude al problema que él denomina de la red de causación *(web of causation)* [772], afirmando que la prueba del origen del defecto puede ser imposible, quedando ocultada tras la red de diversos agentes. En lo relativo a la causalidad, señala la existencia de un «gap cognoscitivo», es decir, hay dificultades para reconstruir *ex post* el fenómeno, planteando la posibilidad de que estemos una especie de «factor causal sobrevenido que rompe el nexo causal» [773], interpretándose que hay una interposición de la conducta de un tercero, dando lugar a un riesgo «cualitativamente diverso del que se creó en un origen» [774].

Para Piergallini hay dos posibles formas de interpretar la causalidad, ante este escenario: una interpretación restrictiva [775], según la cual esta causa sobre-

---

[770] Pérez Arias, J., «Algoritmos y *big data* en la responsabilidad penal», en Peris Riera, J./ Massaro, A. (Coords.), Derecho Penal, Inteligencia Artificial y Neurociencias, Roma Tre-Press, 2023, p. 180.

[771] *Idem,* de forma literal.

[772] Piergallini, C., «Intelligenza Artificiale: ¿De "mezzo" ad "autore"…, *op. cit.* N. 654, p. 1755.

[773] *Ibid.,* p. 1760 (*fattore causale sopravvenuto interrutivo del nesso causale*). En este sentido se remite al célebre trabajo de Gless y Weigend. Gless, S./ Weigend, T., *op.cit*. N. 64, pp. 561 y ss.

[774] Piergallini, C., «Intelligenza Artificiale: ¿De "mezzo" ad "autore"…, *op. cit.* N. 654, p. 1761.

[775] Restrictiva respecto de la interpretación del artículo pertinente (art. 41) del CP italiano, que regula de forma expresa la problemática de las causas sobrevenidas estableciendo que, con carácter general, las causas sobrevenidas no excluyen la relación de causalidad, salvo que hayan sido por sí solas suficientes para determinar el resultado. *«Il concorso di cause preesistenti o simultanee o sopravvenute, anche se indipendenti dall'azione od omissione del colpevole, non esclude il rapporto di causalita' fra l'azione od omissione e l'evento. Le cause sopravvenute escludono il rapporto di causalita' quando sono state da sole sufficienti a determinare l'evento. In tal caso, se l'azione od omissione precedentemente commessa costituisce per se' un reato, si applica la pena per questo stabilita. Le disposizioni precedenti si applicano anche quando la causa preesistente o simultanea o sopravvenuta consiste nel fatto illecito altrui».*

venida restablece el riesgo inicial, remplazándolo por uno nuevo que depende ya de la decisión auto-aprendida; y una interpretación amplia, según la cual la excepcionalidad del evento lesivo final estaría más allá de cualquier capacidad de control humano, y por tanto no sería atribuible a quien programó el sistema [776]. A Piergallini no le acaba por convencer esta afirmación, pues considera que esto conduciría a tener que afirmar la impunidad en todos los escenarios, cuestión que considera inviable. Sin embargo, sigue reconociendo que la prueba de la causalidad no va a ser fácil, pues en muchos casos será difícil rastrear, a partir de los paradigmas nomológicos más acreditados disponibles, la acción que es causa del daño. Esta podría estar determinada por defectos de programación, de construcción, falta de información, pero también depender de influencias ambientales externas, lo que nos sitúa ante una «masa de factores causales alternativos» que además pueden fusionarse entre sí, dando lugar a una red de causalidad *(web of causation)* con efectos interactivos (debido a la suma de algunas causas) y sinérgicos (capaces de dar lugar a un fenómeno cualitativamente diferente respecto de la mera suma) [777]. Le parece a Piergallini este un escenario que, si bien añade dificultades considerables, no son impeditivas para la imputación de responsabilidad penal [778].

También en el contexto italiano ha reflexionado Fragasso sobre el asunto, resaltando que hoy en día hay una dificultad para probar la cadena causal de los acontecimientos, y es difícil determinar qué hecho llevó a qué consecuencia. Aunque se reconoce que con técnicas más perfeccionadas en el futuro sí será posible, con lo que se llama *«reverse engineering»* [779], esto permitirá extraer la concreta causa en un caso concreto, pero no la ley de carácter general

---

[776] Si se analizan las referencias utilizadas por el autor, empero, se puede comprobar que los conceptos de causalidad manejados (entre los cuales destaca el de causalidad adecuada o causa penalmente relevante), contemplan también criterios de imputación que tienen que ver con la creación de riesgos previsibles, por lo que se entremezclan los dos estadios de análisis arriba diferenciados; el propio de la causalidad y el propio de la imputación objetiva del resultado.

[777] Piergallini, C., «Intelligenza Artificiale: ¿De "mezzo" ad "autore"…, *op. cit.* N. 654, p. 1762.

[778] Esto no significa que el autor acoja la posibilidad de imputar el delito de resultado sin más, pues afirma, de hecho, que el mayor problema para el derecho penal reside en la imputación subjetiva, que se analizará *infra*.

[779] La ingeniería inversa en inteligencia artificial consiste en analizar un sistema complejo de «caja negra» observando solo sus entradas y salidas para así reconstruir una explicación sobre cómo se generan sus decisiones, sin tener acceso al funcionamiento interno ni a los datos originales de entrenamiento. Para ello, se consulta el sistema para crear un conjunto de resultados (oráculo) que se usa para entrenar un modelo interpretable que explique el comportamiento del sistema. Algunos enfoques son generalizables y pueden aplicarse a cualquier sistema, mientras que otros están diseñados para modelos específicos. Este proceso es clave para lograr transparencia y comprensión en sistemas automatizados, y la literatura sobre IA explicable *(explainable AI)* lleva unos años trabajando sobre este tipo de explicaciones, denominadas en ocasiones como interpretabilidad *post-hoc*. Así, a título de ejemplo, sin ánimo de exhaustividad: Guidotti, R. *et al.,* «A Survey of Methods for Explaining Black Box Models», 51 *ACM Computing Surveys*, 51,5, Article 93, 2018, p. 18 y ss.

que se dará en todo caso, pues se tratará de un razonamiento más prospectivo que predictivo [780].

Lo cierto es que, tal y como apunta Bertolesi, cuando se trata de examinar los problemas de causalidad que conlleva la imputación de daños generados por IA, al menos dos concepciones se entremezclan.

En primer lugar, el problema más clásico, relativo a la dificultad de reconstruir el nexo causal debido a la complejidad tecnológica, ya que, como se ha venido afirmando, la capacidad de aprendizaje de los sistemas hace difícil individualizar la acción que provoca el daño, por lo que será complejo determinar si este proviene del momento de la programación, de la información sucesiva de la que se va alimentando, o de otras influencias ambientales [781]. Bertolesi es consciente de que esta no es la primera vez que estos problemas surgen, y de que en otras ocasiones se ha intentado superar estas dificultades epistémicas, o bien recurriendo a razonamientos de tipo indiciario, basados en la frecuencia de aparición de estos errores y la posible correlación con el sistema (esta es la solución propia de los *leading cases* en materia de responsabilidad por el producto [782]), o bien mediante otras propuestas que van más en la línea de la reconstrucción del evento *ex post* mediante técnicas de ingeniería inversa ya exploradas en otros campos, como la instalación de cajas negras (también denominadas aquí *Event Data Recorders),* que ayuden a determinar a qué agente interviniente se debe un determinado error [783].

---

[780] Fragasso, B., *op. cit*. N. 686, p. 34. Esto implica, para la autora, un problema para el paradigma nomológico-deductivo del razonamiento causal, pero, como se ha visto *supra*, lo cierto es que no es prescindible este paradigma de tipo deductivo; es decir, que de una premisa general se derive una concreta. Lo que es prescindible es que esa premisa general esté constituida por una ley de las llamadas «universales».

[781] Este aspecto es resaltado por Bertolesi, R., *op cit*., N. 423, p. 209, con cita a Beck, que ya se preocupaba por los problemas de la causalidad y las limitaciones de las explicaciones retrospectivas en un trabajo de 2012. Beck, S., «Roboter und Cyborgs – erobern sie unsere Welt» en Beck, S. (Coord.), *Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborg*s, Nomos, Baden-Baden, 2012, p. 15: «*Auch wird retrospektiv gerade nicht mehr ohne Schwierigkeiten feststellbar sein, ob sich ein bestimmtes „falsches» Verhalten direkt aus der Programmierung, aus der späteren Unterrichtung oder aus anderen Umwelteinflüssen ergibt. Bei neuronalen Netzen nachempfundenen Verknüpfungen ist die nachträgliche Veränderung dieser Verbindungen derart komplex, dass sie schon faktisch im Nachhinein nicht mehr nachvollzogen werden können».*

[782] Bertolesi, R., *op cit*., N. 423, p. 210. Apunta, de hecho, el autor italiano que existe alguna propuesta en el sentido de aplicar el concepto de causalidad derivado de las sentencias de los casos *Lederspray* y spray de la madera (*Holzschutzmittel)* en los supuestos de accidentes de vehículos autónomos. Así, *vid.:* Nehm, K., «Autonomes Fahren – Bremsen Ethik und Recht den Fortschritt aus», *JZ*, 8, 2018, p. 401.

[783] Bertolesi, R., *op cit*., N. 423, p. 211. Aunque el autor reconoce que es una propuesta común en el campo de la conducción de vehículos autónomos, también plantea los inconvenientes procesales que puede suponer, tanto por la cesión de privacidad que supone como, muy especialmente, la posibilidad de que la apertura de estas cajas negras (*scatole nere*) se convierta en una forma de autoincriminación contraria al *nemo tenetur.*

La segunda de las opciones exploradas tiene que ver con la novedosa posibilidad de reconocer que la acción del sistema de IA es un factor capaz de interrumpir la cadena causal [784], asimilando la actuación del sistema de IA al comportamiento de un tercero. Respecto a esto, el autor italiano presenta sus reservas, teniendo en cuenta al menos tres argumentos: primero, que esto implica dejar entrar en el derecho penal un argumento puramente analógico [785]; segundo, que no es pacífico para el derecho penal cuál es el alcance que se da a la conducta de un tercero en cuanto a su capacidad para interrumpir la imputación [786]; y, finalmente, un argumento de oportunidad, pues si así se afirmara se generaría una clara laguna de punición, puesto que sería imposible imputar el hecho al productor en ningún caso [787].

Finalmente, Aires de Sousa [788] expone que la intrínseca opacidad de los sistemas más complejos (pone el ejemplo paradigmático de las redes neuronales, por su estructura similar al cerebro humano en cuanto a su diseño y funcionamiento), puede implicar un obstáculo a la verificación del nexo de causalidad si este se pretende establecer según los criterios dominantes en la doctrina penal, señalando que puede resultar tan difícil como indagar dentro del funcionamiento de un cerebro humano.

No pone en duda la autora que cuanto más opaco, complejo y autónomo es el sistema, más difícil será acceder al mecanismo concreto que ha llevado al resultado. Pero también reconoce que las teorías que más problemas encontra-

---

[784] Reconoce Bertolesi que estas propuestas provienen más bien del campo civilista (Chopra, S./ White, L. F., «A legal theory for Autonomous Artificial Agents», *University of Michigan Press*, Michigan, 2011, p. 122), pero reconoce que también algunos penalistas las han explorado: Gless, S./ Weigend, T., *op.cit*. N. 64, p. 561 y ss., o Gless, S./ Silverman, E./ Weigend, T., *op. cit*. N. 280, pp. 431-432. También Seher, basándose principalmente en los postulados de las autoras anteriores, peguntándose por la posibilidad de cursos causales atípicos (*atypischer Kausalverlauf*) desde el punto de vista de los «agentes inteligentes», pero negando que lo sean desde el punto de vista del operador humano. Así, Seher, G., *Intelligente Agenten als «Personen» im Strafrecht*, en Gless, S./Seelmann, K. (Coords.), *Intellegente Agenten und das Recht,* Nomos*,* Baden-Baden, 2016, p. 53.

[785] Así, Seher, G., *ibid.,* reconociendo que las normas jurídicas tienen como requisito previo la capacidad de ser receptivo *(ansprechbar zu sein),* por lo que cualquier persona incapaz de comprender el contenido de las normas jurídicas queda excluida como destinataria. *Ibid.,* p. 50.

[786] En este aspecto, cabría señalar que, de hecho, parece latir aquí una especie de confusión, de nuevo, respecto de los planos de la causalidad y la imputación objetiva pues, en todo caso, y si se acepta un concepto de causalidad mínima del que aquí se parte, con la doctrina mayoritaria española, en todo caso la conducta de un tercero servirá para interrumpir el nexo de imputación, porque el riesgo iniciado por un curso causal anterior no es el que se concreta finalmente en el resultado, pero no porque sea posible afirmar que el nexo causal se rompe.

[787] Esto se pone de manifiesto por Gless, S./ Silverman, E./ Weigend, T., *op. cit*. N. 280, p. 432, que, además, en todo momento dejan claro que en el derecho penal alemán no hay un verdadero consenso sobre los efectos que tiene la llamada «prohibición de regreso» (*Regressverbot*), es decir, la eficacia excluyente que ejerce el comportamiento de un tercero. Así, en p. 431, notas al pie 64-65.

[788] Aires de Sousa, S., «Não fui eu...», *op. cit*. N. 236, p. 68.

rán para afirmar la causalidad serán aquellas de corte «normativo-causal», es decir, aquellas que se basan en la previsibilidad del resultado, como por ejemplo la teoría de la causalidad adecuada que es, de hecho, la predominante aún en el derecho penal portugués [789]. Reconoce, así, que si se acoge un criterio normativo que base la causalidad en la previsibilidad del daño, este enfrentará, sin duda, serias dificultades en relación a eventos sobre los que no existe una previa generalización [790].

### c) Toma de postura

A nuestro modo de ver, la discusión mantenida cuando aparecieron los primeros intentos de imputar delitos de resultado a eventos complejos encerraba algunas particularidades en lo relativo a la causalidad que, a día de hoy, parece que han sido superadas.

Podemos afirmar que, al menos en España, existe un consenso sobre la necesidad de separar el juicio puramente ontológico que requiere la constatación de la causalidad, del juicio normativo que requiere la imputación objetiva del resultado. Desde luego que si se adoptan conceptos de corte normativo (como Aires apunta que ocurre en el sistema portugués), donde el foco de la causalidad se ponga en lo previsible que era que una determinada conducta diera lugar a un determinado resultado, la afirmación de la causalidad, siempre que exista un sistema de IA con la suficiente complejidad y autonomía, será prácticamente imposible.

No consideramos que este sea, empero, un concepto adecuado. Por el contrario, si se parte de un concepto simplificado o mínimo de causalidad, podemos afirmar, con Dopico y Gimbernat, que lo que importa es el sí y no tanto el cómo del proceso [791]. En los supuestos en que haya una interacción de factores será difícil determinar qué concreto hecho dio lugar a qué concreta consecuencia, aislando el resto de factores que, sin duda, interactúan en la producción de ese resultado. Pero bastará, en esencia, con contar con un método científico válido (y esto presupone el reconocimiento amplio por parte de la

---

[789] Así referido en: Aires de Sousa, S., *Responsabilidade...*, *op cit*. N.628, p. 454 y ss.

[790] Aires de Sousa, S., «Não fui eu...», *op. cit*. N. 236, p. 68.

[791] Bastará, por usar la teoría más dominante en nuestro entorno, con aceptar un concepto de causalidad conforme a la teoría de la «condición ajustada a las leyes de la naturaleza», capaz de superar las razonables críticas a las que ha estado sujeta la también ampliamente defendida teoría de la *condinio sine qua non.* En este sentido, rotunda respecto de la viabilidad de la teoría de la *conditio sine que non,* ampliamente aceptada en el ámbito alemán, para la solución de daños generados por productos de IA defectuosos: Lohmann, A., *op. cit. N. 298*, p. 154 y ss.

comunidad científica) que sea capaz de determinar, incluso por métodos estadísticos, que, dada una determinada consecuencia dañosa, (en su concreta forma de aparición), es altamente probable que sea debida a un evento anterior, con exclusión de otras hipótesis alternativas. En tanto en cuanto esta afirmación se consiga (y esto no es una cuestión en absoluto sencilla de obtener en contextos donde la generación de daños aun es eminentemente opaca, condicionada por multiplicidad de factores y agentes interconectados, donde la comunidad científica apenas comienza a explorar el funcionamiento interno de los sistemas más complejos), la causalidad se podrá afirmar.

Cuestión distinta (y muchísimo más compleja, como enseguida se desarrollará), es la relativa al cómo de la causación de ese daño. Se hace necesario advertir que, si bien la verificación del nexo causal va a resultar complicada (y no será poco común que no sea posible verificar la causalidad en casos especialmente complejos [792], donde interactúen muchos factores interconectados), no consideramos que sea, siempre y en todo caso, un obstáculo insalvable a la imputación de resultados dañosos derivados de actuaciones donde se vean implicadas herramientas de IA.

## 4.2 **Imputación subjetiva y objetiva del resultado lesivo del bien jurídico individual**

Más allá del problema de la causalidad, si existe un lugar de crítica común en todos los trabajos que han explorado la posibilidad de imputar un delito tras la producción de un daño utilizando un sistema de IA es el problema que supone respecto de la previsibilidad.

Resulta, sin embargo, complejo, sistematizar todas las críticas en este sentido vertidas, por lo que se hacen necesarias aquí dos matizaciones previas.

[792] Por nombrar un ejemplo, el caso de la producción de daños en el mercado como consecuencia de la utilización de algoritmos de negociación de alta frecuencia, especialmente en los casos en que los daños finalmente causados tienen un resultado especialmente catastrófico (así, los denominados *flash crashes*). El *frash crash* más paradigmático de la historia reciente se produjo en mayo de 2010, cuando los mercados financieros de EE. UU sufrieron un desplome repentino debido, en parte, a la manipulación de un *trader* independiente, Navinder Sarao, quien utilizó tácticas de *spoofing*. Su actividad desencadenó una reacción en cadena amplificada por algoritmos de mercado, lo que reveló la vulnerabilidad sistémica de la negociación algorítmica. El caso marcó un punto de inflexión en la regulación de estos sistemas. *Vid.:* Scopino, G., *Algo bots and the Law. Technology, Automation, and the Regulation of Futures and Other Derivatives,* Cambridge University Press, 2020, p. 414. El autor se lamenta de que el final de este procedimiento, con una condena en conformidad en la que el propio Sarao aceptó el cargo criminal por *spoofing* y pagó una indemnización, hizo que se perdiera la oportunidad de llevar a cabo un procedimiento penal completo, donde se tenía que haber enfrentado la difícil cuestión de probar la causalidad de que su conducta fue la que efectivamente causó tales consecuencias.

Primero, los trabajos de los autores que exploran la responsabilidad derivada de los *AI crimes* beben de tradiciones jurídicas distintas, y esto implica que la previsibilidad se ubica como elemento esencial en muy diferentes estadios de análisis de la teoría del delito. Segundo, los trabajos que se han utilizado como apoyo metodológico para este apartado no pretenden, en absoluto, por lo prolijo de la discusión, agotar toda la literatura sobre imputación objetiva y subjetiva desde el punto de vista de la dogmática tradicional, sino más bien dar a conocer la literatura sobre IA que ha tratado estos problemas, poniendo en relación, la discusión tradicional con el contexto novedoso.

Así, por simplificar y traer la discusión al contexto español, hemos dejado claro que, según la doctrina mayoritaria, no es necesario analizar la previsibilidad en sede causal (pero esto no es así según las teorías normativas de causalidad, mayoritarias en otros países). Sin embargo, sí encontramos elementos relacionados con la previsibilidad en sede de imputación objetiva (así, el primer requisito relativo a la generación de un riesgo jurídicamente desaprobado no implica más que un juicio de pronóstico *ex ante,* de previsibilidad, sobre la peligrosidad de una conducta para un determinado bien jurídico); mientras que en sede de imprudencia [793], lo que se suele decir es que la previsibilidad forma parte esencial de la determinación del deber de cuidado objetivo [794].

Aunque somos conscientes de que no es esta una posición unánimemente aceptada, se parte aquí de la base de que, de hecho, el injusto objetivo del delito doloso y del imprudente comparte una misma estructura, de forma que se identifica la creación del riesgo jurídicamente desaprobado con la infracción

[793] Nos da igual en este aspecto si esta se considera una manifestación del elemento subjetivo del delito, como ocurre en la doctrina mayoritaria de corte español e influencia germana, o si se considera una forma de culpabilidad, como ocurre en los ordenamientos de influencia anglosajona y en las propuestas italianas, donde se consideran parte del *mens rea*; aunque nos hacemos cargo de la polémica que al efecto, aún en España, se mantiene, donde, aunque la doctrina mayoritaria es partidaria de estudiar estos conceptos dentro del tipo, voces muy autorizadas, como Molina Fernández, aún encuentran buenos argumentos para situarla como una forma de culpabilidad. Para un desarrollo de sus argumentos: Molina Fernández, F., «El razonable «regreso» del dolo a la culpabilidad», en AA. VV., *Estudios de derecho penal. Homenaje al profesor Santiago Mir Puig,* BdF, 2017, pp. 735-751.

[794] Los problemas de previsibilidad aplican, no obstante, también en las doctrinas de la imprudencia más apegadas al finalismo. En este sentido, resulta ilustrativo el reconocimiento de Gil Gil, que admite que, aun partiendo de una concepción según la cual rechaza que la determinación del cuidado debido deba comenzar exigiendo de entrada un determinado grado de probabilidad del resultado, la imputación del delito imprudente de resultado en los casos de daños generados por sistemas autónomos de IA (en el campo armamentístico, concretamente), sigue resultando imposible, pues «*establecer la ponderación entre la ventaja y el riesgo para decidir si se trata de una conducta imprudente, o, por el contrario, de un riesgo permitido, exige, de igual manera, la previsión del resultado y su mayor o menor probabilidad, lo que, como se ha señalado, resultará prácticamente imposible en los sistemas de IA entrenados con redes neuronales*.» Gil Gil., A., *op. cit*. N. 242, p 41.

del deber de cuidado [795]. Más allá de que consideremos que esta asimilación es válida con carácter general, nos resulta por lo demás mucho más útil para analizar los problemas que la IA plantea en este estadio objetivo.

Finalmente, hemos de afrontar la repercusión innegable que la previsibilidad tiene en la delimitación entre el dolo y la imprudencia [796], que ha llevado a afirmar, de una forma muy generalizada, que, en los casos de daños causados por herramientas de IA, el paradigma adecuado es solo el del delito imprudente.

Se tratará de poner a prueba cada uno de los argumentos aquí esgrimidos, adelantándose ya que la idea que aquí se defiende es que, habida cuenta de la innegable incidencia que tiene la autonomía de los sistemas de IA en la previsibilidad de su actuación, será extremadamente difícil (y requerirá, en el mejor de los casos, una adaptación forzosa de las instituciones), imputar delitos según el paradigma del delito de resultado (de una forma muy evidente en el caso de los delitos dolosos, pero aún patente cuando se trate de delitos imprudentes).

Podemos afirmar, con Íñigo, que existen ciertos principios asentados en nuestro ordenamiento penal que responden a la propia función asignada al mismo, y dentro de estos principios está precisamente la previsibilidad, de tal forma que, aunque sería posible sostener un derecho penal donde se castigase a los sujetos por hechos que no fueran previsibles, esto no tendría sentido si partimos de la base de que la función del derecho penal (o, al menos, una de ellas), es motivar a los ciudadanos para que no comentan delitos [797]. Es por ello por lo que la previsibilidad aparece, en muy diferentes momentos y lugares de la teoría del delito, como elemento esencial.

---

[795] Somos conscientes de que esto no es lo que defienden todos los autores. Nos remitimos a algunos de sus trabajos para ver las diferencias que se señalan, por no poder hacernos cargo de la discusión. Así, Molina Fernández, F., *La cuadratura del dolo: problemas irresolubles, sorites y Derecho penal*, Univ. Externado de Colombia, Bogotá, 2007, p. 11: *«la infracción del deber objetivo de cuidado en la imprudencia es intensionalmente equivalente a la primera parte del juicio de imputación objetiva: la creación de un riesgo no permitido»*. En mayor detalle, sobre la discusión: Molina Fernández, F., *Antijuridicidad penal y sistema del delito*, Bosch, Barcelona, 2001, pp. 376 y ss., y especialmente en nota 39, p. 384.

[796] Desde luego no es una discusión exclusiva de hoy, pues tan central como la discusión sobre la causalidad en las sentencias sobre responsabilidad penal por el producto fue la cuestión relativa a la delimitación entre el dolo eventual y la imprudencia consciente en la sentencia del TS que resolvió el caso de la colza (discusión, por lo demás, absolutamente complejizada y sobre la que han corrido ríos de tinta dentro y fuera de nuestro territorio, sobre la que no podemos más que señalar su existencia, sin ocasión para desarrollarla).

[797] Así, de una manera velada y con referencia a otros intereses (se refiere Íñigo en este pasaje a los criterios que fundamentan el riesgo permitido), aparece como innegable la conexión de ciertos principios con las bases del Derecho Penal. Íñigo Corroza, E., *op. cit.* N. 697, p. 124. No es necesario adscribirse a una concreta teoría de la pena en este punto para aceptar lo anterior, simplemente aceptar que todas las teorías que no sean puramente retributivas han de dar un lugar relevante a la motivación del ciudadano por la norma penal.

a) La preocupación por la previsibilidad, antes

Ya Stortoni advertía, a colación de la posibilidad de castigar penalmente los riesgos derivados del avance tecnológico, que esto podría realizarse a costa de las garantías penales. El autor parte de la idea de que aquí el problema es que se considera que una determinada actividad es beneficiosa para la sociedad en su conjunto, por lo que no es posible prohibirla y, ante esto, (partiendo de que hay un desconocimiento sobre los mecanismos que dan lugar al daño), se producen varias distorsiones en el derecho penal [798]. Considera el italiano que, en sede de culpabilidad, se produce una desvaloración absoluta de la previsibilidad del resultado específico. Es decir, sería algo así como afirmar que para la imprudencia no es necesario que el resultado específico se derive de la propia conducta, sino que bastará una previsibilidad genérica, sin necesidad de especificación de la actuación [799].

También Fornasari adopta una crítica similar, poniendo en duda que se deba aceptar la fundamentación de un deber cautelar de comportamiento sobre la existencia de una incertidumbre o duda científica acerca de la genérica posibilidad de resultados lesivos distintos de los previsibles [800]. Lo que viene a criticar el autor es la afirmación, que bebe de una incorrecta aplicación del principio de precaución, de que la previsibilidad respecto de un hecho peligroso genéricamente para la salud baste para afirmar la previsibilidad específica de una consecuencia lesiva de mayor gravedad, dando lugar a la imputación de un evento gravísimo, cuando *ex ante* era conocida solamente la posibilidad de resultados menos graves, denunciando que de un «cuadro cognitivo incompleto sobre la idoneidad lesiva de la conducta» se derive una responsabilidad por culpa (imprudente) excesiva [801].

---

[798] Stortoni, L., «El riesgo procedente …», *op. cit*. N. 453, pp. 87 y ss. Critica este autor también la idea de la «elefantiasis» del bien jurídico, así como la reconstrucción de la causalidad en términos de incremento de riesgo. Ídem.

[799] *Ibid.,* p. 91. Cabe matizar que para Stortoni, como es lugar común en la literatura italiana, es en la culpabilidad donde se sitúa la imprudencia, como una de sus formas. Una idea similar encontramos en Consulich cuando denuncia lo que ocurre con el fenómeno de la IA, donde se corre el riesgo de que el dolo se convierta en una especie de dolo genérico. Vid *infra.*

[800] Fornasari, G., «El principio de precaución en la experiencia legislativa, jurisprudencial y doctrinal italianas. Aspectos de parte general» en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estad*o, Tirant Lo Blanch, Valencia, 2014, p. 166.

[801] *Idem.* En este punto, acoge el autor los postulados de Castronuovo. En realidad, lo que vemos aquí es una de las concretas críticas respecto de la función del principio de precaución que aquí se ha calificado de dogmática; si se deja entrar plenamente la lógica precautoria en la determinación del deber de cuidado de la imprudencia, diríamos, se está imputando una conducta de gravedad excesiva con un conocimiento incompleto respecto del riesgo que comporta.

No parecen estas críticas, empero, demasiado alejadas de la realidad, pues algunos autores (en España, Romeo Casabona como principal exponente), han abogado por la posibilidad de que, en aplicación del principio de precaución, la lógica de la imprudencia se vea modificada. Parte de la base este autor de que la imprudencia se orienta al paradigma de la prevención a través de la previsibilidad objetiva [802], pero que, dado que la precaución parte de la incertidumbre, para poder aplicar este principio sería necesario remodelar el criterio de previsibilidad objetiva del delito imprudente, llegando a admitir que, si bien de manera limitada, cuando una disposición establezca la obligación de observar ciertas cautelas en torno a una actividad cuyas concretas consecuencias dañosas no son objetivamente previsibles, las pautas de cuidado debido se establecerían conforme a este principio de precaución [803].

Es esta una postura claramente polémica entre la doctrina española, que ha tratado en todo caso de especificar cómo de concreta debe ser la previsibilidad de un determinado resultado para poder hablar de esta vulneración del deber de cuidado, pero siempre partiendo de la importancia de la previsibilidad real, y no la derivada de la incertidumbre.

Así, Íñigo, que en su estudio sobre la responsabilidad penal por el producto defectuoso parte de la importancia central de la infracción del deber de conducta, desarrolla cuáles son los elementos que configuran dicho deber, dando por sentado que (a pesar de las variadas denominaciones), la previsibilidad o cognoscibilidad del riesgo es central [804]. Lo que importa es establecer a qué grado de peligrosidad ha de referirse la previsibilidad por el sujeto; es decir, si ha de ser una peligrosidad en abstracto o en concreto, señalando Íñigo que *«no basta con la mera previsibilidad genérica de producción de un resultado disvalioso, sino que debe hacer referencia al peligro concreto que la conducta lleva consigo ex ante de su producción»* [805]. Es decir, para determinar si la creación de riesgos era previsible se debe atender a pautas ideales que hagan referencia a todos los sujetos que ocupan una determinada posición jurídica. Lo difícil, reconoce la autora, es establecer cómo de concreta ha de ser

---

[802] Romeo Casabona, C. M., «Aportaciones del principio de precaución …», *op cit*. N. 622, p. 410.

[803] *Ibid.,* p. 411.

[804] Íñigo Corroza, E., *op. cit.* N. 697, nota 468. Dentro de los elementos típicos ella diferencia entre los internos y los externos; siendo los internos los que configuran el deber de conducta, que para ella son la previsibilidad o cognoscibilidad del riesgo, y la evitabilidad, o controlabilidad del riesgo. En todo caso, reconoce que hay otros autores que no distinguen y hablan solo de evitabilidad (así, Martínez Escamilla o Corcoy), mientras Paredes sí distingue. Para algunos autores, la previsibilidad se enmarca implícitamente en el concepto de peligrosidad. Así, por ejemplo, Corcoy, en lugar de hablar de previsibilidad habla de cognoscibilidad, que fundamenta el deber de conocer.

[805] *Ibid.,* p. 185.

esa norma de conducta, pero lo que sí afirma es que, si la norma de conducta se individualiza excesivamente, también deberá individualizarse la determinación de la previsibilidad [806]. Se exige, así, una correlación entre la norma que ha de guiar la conducta (ya sea esta dolosa o imprudente), y la previsibilidad, de forma que, si la norma es detallada y concreta, también la previsibilidad lo deberá ser.

También Paredes se pronuncia sobre los problemas de previsibilidad del régimen de la responsabilidad penal por productos defectuosos. Reconoce el autor que, una vez fijada la relación causal, será normalmente sencillo establecer la previsibilidad objetiva del resultado, pero esto se exceptúa al menos en dos supuestos, donde la imputación del resultado lesivo se complica sobremanera [807]. En primer lugar, cuando el proceso de fabricación esté muy fragmentado, y, en segundo lugar, cuando ha habido varios errores de varios intervinientes [808]. Sitúa Paredes el verdadero problema en una concreta dinámica comisiva, como es la posibilidad de que no todas las personas hayan conspirado al efecto, sino que se dé un cúmulo de imprudencias aisladas que interactúan entre sí. Para el autor, el principio de responsabilidad individual supone aquí un obstáculo a la imputación conjunta, impidiendo la imputación del resultado final en muchos casos, o bien porque no se podrá afirmar la conexión de previsibilidad bastante entre acción y resultado, o bien, aun existiendo tal conexión, porque el resultado lesivo finalmente producido se aleja con mucho de aquello que constituía el ámbito natural de protección del deber de cuidado infringido [809].

Aunque Paredes no es explícito en este sentido, parece que la lógica detrás de su argumento es negar la posibilidad de imputar responsabilidad por el resultado en casos de la denominada «concurrencia de culpas» o, si se quiere, concurso de participaciones imprudentes. Es esta una cuestión,

---

[806] *Ibid.,* nota 477.

[807] Paredes Castañón, J. M, «Responsabilidad penal por productos…», *op. cit*. N. 716, p. 608.

[808] Este problema, que tiene que ver con las dificultades de la individualización de la autoría dentro de estructuras complejas es, desde luego, otro de los comunes en la discusión de la responsabilidad por el producto que se ve extrapolado a los escenarios de IA, donde el «*many hands problem*» se repite. Sin embargo, de la individualización de la responsabilidad dentro de la estructura colectiva, como problema con entidad propia y que sería objeto, por sus muchas complicaciones, de una investigación en su conjunto, no nos haremos cargo aquí, por entender, fundamentalmente, que es posible solucionar lo que aquí nos interesa de una forma aislada, sin perjuicio de que, en un ulterior estadio de individualización de la responsabilidad, haya que acometer esta tarea. El desarrollado aparato instrumental que maneja la doctrina penalista en este punto puede servir de base de partida sólida. Así, sin ánimo de exhaustividad: Lascuraín Sánchez, J. A., «La responsabilidad penal individual en los delitos de empresa», en De la Mata Barranco, N./ Dopico Gómez-Aller, J./ Lascuraín Sánchez, J. A./ Nieto Martín, A., *Derecho penal económico y de la empresa*, Dykinson, Madrid, 2018, pp. 87-128.

[809] Paredes Castañón, J. M, «Responsabilidad penal por productos…», *op. cit*. N. 716, p. 608.

desde luego, con entidad propia dentro de la teoría de la participación delictiva, pero baste decir que es una solución que, por la propia dinámica de los *hard AI crimes* ha sido puesta de manifiesto por varios autores. Así, Manes, preguntándose si las actuales formas de atribución de responsabilidad penal pueden adaptarse a un delito en el que la acción es obra autónoma de una máquina guiada por programas informáticos y algoritmos, sólo «asistida» por la «inerte y única posible presencia humana» y «cómo pueden formularse hipótesis en las que la conducta sea obra compartida *(opera condivisa)* de la acción humana y la inteligencia artificial» [810]. Se plantea el autor la posibilidad de dar entrada a nociones nuevas de culpa (habla de la culpa de programación o de automatización), o bien si se debe dar entrada a particulares formas de cooperación imprudente *(«contributory negligence»),* capaz de adaptarse a las interacciones humano-máquina y a afrontar el llamado riesgo de interconectividad. Lo cierto es que no da el autor una única solución, más allá de poner sobre la mesa todas las opciones posibles. Al margen de que la cuestión de la tipificación de las cooperaciones imprudentes sea teóricamente viable, sí nos gustaría dejar claro que, antes de su viabilidad dogmática debe plantearse la cuestión de su conveniencia práctica pues, a nuestro juicio, los inconvenientes derivados de afirmar tal posibilidad tienen que ver mucho más con el exceso de punición que con su encaje teórico [811].

No pretendemos agotar, desde luego, los problemas que dentro de la doctrina moderna se han puesto de manifiesto, sino más bien dejar apuntados los inconvenientes, que, como enseguida se verá, se reproducen en el ámbito de los resultados lesivos causados por sistemas de IA. Algunas de las cuestiones arriba apuntadas surgen, aún con más fuerza, cuando se trata de este tipo de tecnologías.

---

[810] Manes, V., «L'oracolo algoritmico», *op. cit*. N. 284, p. 3 (traducción propia) El mismo trabajo puede encontrarse publicado en Ruffolo, U. (Ed.), *Intelligenza Artificiale – Il diritto, i diritti, l'etica*, Giuffré, Milano, 2020.

[811] Esta es la postura que ha sido defendida, en esencia, por autores como Robles, Peñaranda, o Feijoo, entre otros. La postura que aquí se sostiene coincide, en esencia, por lo expuesto en Robles Planas, R., «Participación en el delito e imprudencia», *Revista de Derecho penal y criminología*, 9, 2000, señalando que es doctrina mayoritaria y que, el artículo 12 del CP, al fijar el sistema de *numerus clausus* del delito imprudente, no conlleva la exclusión de la participación imprudente en aquellos delitos para los cuales está prevista esta modalidad de incriminación. Esta conclusión, empero, no constituye un dato definitivo «sobre la posibilidad dogmática y la admisibilidad político-criminal del castigo de la participación imprudente», pp. 236-237.

### b) Los problemas de previsibilidad y la IA: un punto de quiebra estructural

Fragasso da en la clave del problema cuando habla de los problemas que acarrean los algoritmos de aprendizaje automático en sede de imputación subjetiva. Parte la autora, como aquí se hace, de que el escenario más probable será aquel en el que el daño se produzca de manera involuntaria, por lo que explora la imputación subjetiva a título de imprudencia, aunque sin descartar la posibilidad teórica del dolo eventual. Afirma la autora que en los eventos lesivos derivados de sistemas de IA no habrá demasiada dificultad para afirmar el «si» (es decir, la previsibilidad respecto de que ocurran), pero muchos más a la hora de determinar el «cuánto» (magnitud del daño derivado) y el «cómo» [812].

Se puede dificultar enormemente la imputación de eventos lesivos a título imprudente si se tiene en cuenta que lo que normalmente se exige en estos casos es que el evento dañoso reprochable al agente sea precisamente la realización concreta del riesgo que la norma de precaución pretendía evitar [813].

La alternativa (más bien, a nuestro parecer, el dilema), que plantea Fragasso es evidente: si se adopta una concepción amplia respecto de la previsibilidad exigida (es decir, afirmando que basta con una previsibilidad genérica de eventos y no en su concreta configuración según el resultado lesivo), será fácil afirmar la imprudencia, pero, de hecho, esta imprudencia estará presente en cada uno de los casos en que se dé un resultado lesivo, cualquiera que sea su forma de aparición o magnitud. Es decir, siempre habrá imprudencia (Fragasso se refiere en este caso al productor, pero nos parece extrapolable al usuario). La segunda de las posibilidades pasa por adoptar una concepción más restrictiva respecto de la previsibilidad, en cuyo caso será muy difícil afirmar la concurrencia de culpa, dado el carácter intrínsecamente imprevisible del evento lesivo en su concreta configuración *(quantum y quomodo).* En definitiva, o todo es previsible o nada lo es. La primera de las opciones nos conduce de forma irremediable a la responsabilidad objetiva, mientras la segunda conduce

---

[812] Fragasso, B., *op. cit*. N. 686, p. 34. Nos recuerda esta afirmación, quizá por su redacción casi literal, a la realidad señalada por Dopico cuando analizaba los problemas de causalidad de los *toxic cases*, con especial atención al del aceite de colza. *Vid supra.* Una cosa es que sea fácil determinar que ciertos resultados fueron, con toda probabilidad (y según el mejor método científico) causados por un grupo de conductas, y otra será individualizar en qué momento se produce la conducta objetiva y subjetivamente imputable (y, por tanto, a qué personas competía encargarse de ese momento). Es este el problema que, en esencia, ocurre ante los algoritmos de aprendizaje.

[813] Alude la autora al criterio típico de cobertura del riesgo, *idem.*

a la impunidad. ¿Es esta una brecha insalvable? Parece que, si el esquema escogido es el del delito de resultado, estamos ante un callejón sin salida [814].

Un dilema similar plantea Consulich, utilizando como ejemplo un contexto muy concreto y frecuentemente discutido dentro de los *Hard-AI crimes:* la causación de daños en el mercado tras la utilización de algoritmos de alta frecuencia con IA. Diferencia el autor dos escenarios. En el primero de ellos el algoritmo de negociación, debido a su autonomía, reacciona ante una situación contextual del mercado (una inestabilidad ocasional de las cotizaciones debida, paradigmáticamente, a una noticia como puede ser un atentado, sea esta falsa o no), de tal forma que el mercado se distorsiona gravemente, pero sin que sea posible afirmar ninguna tipicidad en la conducta del inversor que operó con este algoritmo, pues la persona física que desarrolla o introduce este algoritmo en el mercado está demasiado alejada respecto de la perturbación, pues esta se produce por la interacción ocasional con una determinada condición del mercado. [815] Despacha, así, Consulich, de un plumazo, la posibilidad de imputar este resultado lesivo en el mercado, dado que no hay posibilidad de afirmar que exista creación de un riesgo típico (habida cuenta de que el delito de manipulación de mercado es un delito doloso), aunque lo cierto es que su argumentación nos parece, en este punto, algo pobre.

También le parece problemático al autor italiano el segundo de los escenarios posibles. En este caso el operador, en uso del algoritmo de negociación, le proporciona una serie de instrucciones ilícitas al sistema, pero este último decide cómo ejecutar en concreto estas operaciones. Así, si bien es cierto que el algoritmo ejecuta las instrucciones manipulativas, estas son demasiado genéricas para dar lugar a una imputación por dolo, pues, al entender de Consulich, solo con una interpretación del dolo entendido como mera intención de perturbar el mercado, que prescinda de la concreta representación y voluntad respecto de las modalidades concretas de manipulación y el tipo de títulos objeto de la conducta, se podrá afirmar la condena. Dicha interpretación, que

[814] Para Fragasso, una de las opciones para liberar de este escenario de culpa generalizada es explorar la posibilidad de que se reconozcan áreas de riesgo permitido que caracteriza como «impermeables» al juicio de culpa. Empero, aunque la autora indaga en la formalización de estas normas de cuidado (explorando brevemente la posible existencia de reglas cautelares escritas y no escritas), su solución coincide, en lo esencial, con nuestro pensamiento; la mejor opción es, por el momento, la lógica del «*reati culposi de mera condotta*». *Ibid.,* p. 39. A él dedicaremos el próximo capítulo, si bien en una formulación mucho más común en el entorno español.

[815] Consulich, F., «Il rastro di möbius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato», *Banca Borsa Titoli di credito*, 2, 2018, pp. 218-219.

el autor denomina *«dolus generalis»,* resulta, a su entender, incompatible con la aplicación de los delitos contra el mercado [816].

Hemos de realizar, en este punto, algunas matizaciones. Lo que tiene de valioso el argumento de Consulich es que pone sobre la mesa la insuficiencia de algunos tipos penales de nuestros códigos actuales en su concreta configuración para hacer frente al castigo de conductas con altos niveles de imprecisión en cuanto a su resultado final. No es cierto que la falta de previsibilidad sobre la concreta configuración de todos los extremos del hecho dañoso impida en todo caso una imputación a título de dolo (sería tanto como decir que solo un dolo de primer grado es aceptado), pero sí es cierto que este tipo concreto de dolo es el que está presente en determinados tipos de nuestro Código penal, incluso exigiendo, en algunas ocasiones, algo más que este elemento subjetivo, requiriendo otros elementos subjetivos adicionales. Es el caso paradigmático de los delitos de abuso de mercado, que, debido a su concreta configuración dolosa y manipulativa, harán extremadamente difícil su aplicación en los casos en que la actuación humana se vea complementada por herramientas dotadas de autonomía e imprevisibilidad estructural.

Una cuestión similar parece derivarse del trabajo de Aires de Sousa, cuando cuestiona los problemas relativos al elemento subjetivo, concretamente, la cuestión de si el resultado disvalioso se torna imprevisible para el humano (ya sea este fabricante o usuario). Acepta la autora que surgirán problemas en los casos en que el daño es extraño a cualquier representación humana, en cuyo caso la imputación solo podrá tener amparo en una conducta imprudente (y solo en la medida en que suponga una violación del deber objetivo de cuidado) [817], pero por ello esto supone un problema en todas aquellas figuras delictivas que tipifican conductas dolosas o intencionales [818].

Aunque no es explícita respecto de las consecuencias que tiene, para Aires es esencial distinguir entre la previsibilidad respecto del resultado final y la previsibilidad respecto del proceso que permite alcanzar este resultado [819]. Nos parece valiosa esta distinción pues, a nuestro modo de ver, precisamente es en

---

[816] *Ibid.,* p. 219.

[817] Aires de Sousa, S., «Não fui eu...», *op. cit*. N. 236, pp. 68-69.

[818] Se refiere la autora a delitos económicos en general, pero hace hincapié, una vez más, en los delitos de manipulación de mercado, en los que es cada vez más común que interaccionen algoritmos con capacidad de aprendizaje. *Ibid.,* p. 69.

[819] *Idem.* La conclusión que adopta Aires de Sousa en este trabajo implica que la responsabilidad penal debe reservarse solo para aquellas situaciones en las que se conoce, de alguna forma, con algún grado de probabilidad o previsibilidad, aunque sea aun de forma abstracta, que una determinada tecnología genera un riesgo para determinados bienes jurídicos. Toma el ejemplo de Beck relativo a la prohibición de las armas autónomas, o los supuestos que sí tienen un marco regulatorio concreto (así, Francia respecto de la prohibición de análisis de datos referentes a magistrados y decisiones judiciales para la previsión

ese proceso donde hay que poner el énfasis de la responsabilidad, como más adelante se argumentará.

También Cappellini pone el énfasis en los problemas que para la imputación a título de impudencia suponen las nuevas tecnologías, (a las que, como arriba se ha señalado, caracteriza de «objetos subjetivizados»). En lo relativo a la imputación culposa (por imprudencia), afirma que la imprevisibilidad propia de estas tecnologías genera un *responsibility gap* [820]. Para llenar este vacío se puede intentar atribuir responsabilidad a dos sujetos: o bien al usuario o bien al productor.

Empezando con el productor, lo que se señala es que los eventos adversos que se pueden dar en este campo no son similares a la tipología de culpa que se suele asociar a los fabricantes o productores, relacionada con defectos de fabricación, de mantenimiento, o falta de actualización. Todos estos sí pueden encuadrarse como formas ordinarias de responsabilidad del productor, pero la casuística más compleja es la que tiene que ver con hipótesis de daño por un comportamiento imprevisible del sujeto artificial, atribuido al margen de autonomía que tienen reservado [821]. En este caso, se dice, solo se puede superar la dificultad de individualización del sujeto responsable dentro de una red plurisubjetiva de productor-programación, con una complejidad considerable, con el problema de fondo por el cual el comportamiento de la máquina es imprevisible.

En cuanto a la responsabilidad del usuario, introduce una reflexión sobre el cambio del papel del usuario y su responsabilidad en los últimos tiempos. Hoy en día es imaginable un escenario donde los usuarios pierden cada vez más control sobre los productos, debido a la mayor asistencia y autonomía de todas las máquinas de las que se sirven, pero conservando, a su vez, una parte de ese control (lo ejemplifica con el caso de los vehículos autónomos). Por eso el vacío de responsabilidad puede quedar un poco oculto [822], pero ello no implica que no exista aquí un dilema. Dicho dilema está en la cuestión de que todavía hay un poder material y efectivo de control, pero la cuestión en el plano político (o más bien, diríamos, normativo) es si una posible o eventual ruptura de este control puede o debe ser castigado como ruptura de una obligación jurídica de control respecto de la actividad de la IA. Si así se decide puede haber responsabilidad por un acto omisivo culposo, pero esta imputación

---

de futuros fallos judiciales). *Ibid*. pp. 87-88, optando por una tipificación separada de lo que alude como «*AI-Crimes*», en áreas donde se haya alcanzado, entendemos, cierta seguridad jurídica.

[820] Cappellini, A., *op. cit*. N. 662., p. 25.

[821] *Ibid.,* p. 26.

[822] *Ibid.,* p. 28.

solo se va a poder mantener en los casos en que todavía permanezca, aunque solo en abstracto, la posibilidad de intervención sustitutiva del humano [823]. Por el contrario, en los casos en que la tecnología utilizada excluya del todo la posibilidad material de intervención del humano, la laguna se muestra de forma clara [824], sin que sea posible imputar un eventual resultado lesivo a un usuario que no participa en la acción peligrosa, y que está privado de la capacidad de gobernarla [825].

Para Piergallini, de hecho, el verdadero obstáculo para la imputación penal de delitos de resultado por daños causados por IA se sitúa en el nivel de la imputación subjetiva, debido a la imprevisibilidad del sistema. No le parecen válidos los argumentos que aluden a que cada uno de los eventos a que da lugar el sistema es reprochable porque, *ab initio,* era genéricamente predecible, porque esto salta el criterio de que ese evento esté cubierto por el riesgo típico, que está en el ADN de la culpa. Según Piergallini, esta interpretación de las exigencias de previsibilidad conduciría a una responsabilidad objetiva en estado puro [826].

Ahonda en estos mismos problemas Bertolesi, reconociendo que ya es un tema recurrente afirmar que, aunque sí sea previsible que la introducción de una determinada tecnología cause algún daño, no lo es saber el cómo y cuándo ocurrirá [827]. Reconoce el autor, empero, que otra parte de la doctrina ha mani-

---

[823] Parece adecuado resaltar en este punto que la idea puesta de manifiesto por Cappellini, y que más adelante se analizará en profundidad, cuando se aluda al «*control dilemma*», puede llevar a pensar que la distinción entre previsibilidad o cognoscibilidad del riesgo, por un lado, y la evitabilidad, por otro, pueda tener sentido. Una cosa es conocer la posibilidad de que se produzcan eventos lesivos y otra distinta tener la capacidad real de evitarlos.

[824] *Ibid.,* p. 29.

[825] Se trata, en definitiva, entendemos, finalmente, de un problema de evitabilidad, como una cuestión ulterior a la existencia de un deber. Para el autor, esto hace que la acción se convierta en un hecho propio (*fatto proprio*) atribuible al mismo sujeto artificial, o bien, simplemente, que se considere un caso fortuito. Es interesante, en todo caso, cómo Cappellini apunta que, en la necesaria tarea de concreción de los deberes de cuidado en este campo (cuestión aún pendiente), se deben considerar cuestiones de carácter político, y alude a la idea, que consideramos muy acertada, de que, llegados a un punto de complejidad y desarrollo, la actuación del humano no añadirá, necesariamente, una mayor seguridad. *Ibid.,* p. 30.

[826] Piergallini, C., «Intelligenza Artificiale: ¿De "mezzo" ad "autore"…, *op. cit.* N. 654, pp. 1762-1763. Para Piergallini, por ello, el derecho penal tiene muy poca fuerza aplicativa, a menos que se intente desfigurar. Su propuesta pasa por un papel muy prudente del derecho penal: «más allá del método reactivo, hay que atender a lo proactivo», p. 1722, en el sentido de que este fenómeno debe abordarse desde una pluralidad de reglamentaciones: normas éticas, técnicas, e instrumentos de *self-regulation.* Se puede definir este papel del estado y los organismos de regulación como de *cooperative compliance*. Es decir, castigar a empresas si no cumplen con sus deberes de información y quizá aplicar medidas cautelares; se requiere un «*gobierno condiviso*» (p. 1773) de la generación del riesgo, haciendo un uso prudente del derecho penal.

[827] Uno de los primeros autores en tratar el tema es Balkin, J. B., «The Path of Robotics Law*», California Law Review,* 6, 2015, p. 52*: «although the risk of some kind of injury at some point in the future is foreseeable whenever one introduces a new technology, how and when an injury occurs may not be particularly foreseeable to each of the potential defendants in the above list».* Debemos señalar que Ber-

festado que esto no puede constituir un impedimento para la afirmación de la responsabilidad del productor, pues esta imprevisibilidad es la que hace nacer precisamente los deberes de cuidado [828]. El problema en este caso, nos permitiríamos añadir, es que tales deberes de cuidado no se encuentran, todavía, determinados con una certeza suficiente. Precisamente por esto resulta de utilidad el repaso que realiza Bertolesi sobre lo que ha manifestado la doctrina civilista al respecto.

En primer lugar, se destaca que existe una relación evidente entre la técnica de programación del sistema y la previsibilidad de los peligros [829]. En segundo lugar**,** que la previsibilidad no debe interpretarse como un concepto

---

tolesi trata todos los problemas relativos a la previsibilidad dentro de la constatación de la culpabilidad (*colpevolezza*), pero esto no debe confundir, como se ha declarado arriba, su importancia, sea donde sea que se sitúe la imputación del elemento subjetivo. De hecho, ya Beck, en el contexto germano, reconoce que la robótica es un terreno útil para discutir cuánto de específica ha de ser la previsibilidad. Beck, S., *op. cit.* N. 259 p. 139: *«The foreseeability is only connected to the general possibility of harming; the specific conditions and situations become more and more unforeseeable. Robotics is therefore an opportunity to discuss how specific the foreseeability has to be: Does it have to be directed towards specific circumstances, causalities, harms, or is it sufficient to foresee the possibility of violating humans as such».*

[828] En este sentido, Gless, S./ Silverman, E./ Weigend, T., *op. cit*. N. 280, p. 427. De hecho, los autores germanos comparan la problemática de la imprevisibilidad de la IA con la eventual responsabilidad del dueño de un zoológico si se escapa un tigre del mismo, no pudiendo fundamentarse una exención de responsabilidad basada en la imprevisibilidad del animal pues, precisamente esta impone deberes especiales de cuidado. Estamos de acuerdo con la óptica, pero consideramos que el ejemplo no es, tal vez, el más adecuado. En primer lugar, porque respecto de las normas a las que ha de ajustarse el responsable de un zoológico existe bastante seguridad jurídica, y en segundo lugar, porque la causación de daños por un sistema de IA que se encuentra debidamente comercializado en el mercado se asemeja más a la situación en la que el dueño de un pequeño caniche que cumple la normativa que le permite sacarle a disfrutar de un paseo por un parque canino, deba responder ante el ataque (imprevisible) de su pequeño perro frente a quien pasea por ese mismo lugar, en ese espacio que, en principio, tenía reservado para su disfrute. En un sentido también crítico con la idea de Weigend y el ejemplo del tigre, se muestra Gil Gil, que aporta muy buenas razones para entender que no es posible equiparar la previsibilidad (el juicio de peligrosidad objetiva como elemento que restringe las conductas típicas en los delitos de resultado a las conductas consideradas *ex ante* peligrosas), con la mera posibilidad fáctica, porque esto sería como equiparar la causalidad con la previsibilidad. Así, *«No es lo mismo que la actuación de la maquina no sea predecible en el sentido de que su autonomía introduzca un alto grado de incerteza o indeterminabilidad, pero siendo científicamente posible (y por tanto ya previsible) la desviación de la máquina hacia un resultado dañoso no querido, que hablar de falta de previsibilidad en el sentido de la previsibilidad objetiva, que rechaza de entrada un determinado grado de probabilidad (lo absolutamente improbable). Si lo previsible equivaliera a lo científica o naturalmente posible, el criterio de la previsibilidad objetiva no añadiría nada al de la causalidad, y es sobradamente conocido que aquel fue creado para limitar el tipo objetivo frente a este último*.». Gil Gil., A., *op. cit*. N. 242, nota al pie 152. Nos parece acertada su conclusión, en el sentido de que aquí lo importante es determinar el grado de probabilidad que exijamos para entender que algo es previsible y no tanto que algo sea meramente posible según las leyes de la causalidad.

[829] Es decir, la decisión sobre un aspecto técnico en la fase inicial de creación ya implica una decisión relevante. En este sentido, se alude a las redes neuronales como técnica que va más allá de lo convencional, pues más que programarse se entrena. Bertolesi, R., *op cit*., N. 423, p. 217. Así, Kowert, W., «The Foreseeability of Human-Artificial Intelligence Interactions», *Texas Law Review*, 96, 2017, pp. 183: «*Machine learning departs from software coding in the conventional sense and begins to look more like coaching than it does programming».*

estático, sino que es capaz de adaptarse a las nuevas exigencias del cuadro social y tecnológico [830]. Uno de los factores que, según Bertolesi, condiciona sobremanera este nuevo concepto de culpa es la hipercomplejidad relacional [831]. Dentro de esta complejidad relacional, como hemos podido ver, se ha prestado especial atención a ese paradigma de interacción humano-máquina, haciendo hincapié en los deberes de vigilancia debida del usuario. El problema está aquí, según señala acertadamente Bertolesi, en establecer los confines de los deberes de vigilancia, es decir, cuándo el conductor debe retomar el control o cuándo el médico debe dejar de poder confiar en el robot [832]. Incluso se ha llegado a plantear la admisibilidad de un deber inverso; esto es, que, en un escenario de conducción autónoma, un «conductor modelo», al darse cuenta de que sus capacidades están mermadas, tenga la obligación de ceder el control al coche autónomo [833].

En definitiva, y como acertadamente señala el autor italiano, para determinar el alcance de la previsibilidad en la determinación de la atribución de la imprudencia, se debe tener en cuenta el aspecto relacional de las interacciones, sin las cuales no se entiende el contexto de uso de las tecnologías. [834]. De esta interacción múltiple en contextos donde las imprudencias

---

[830] En este punto cita a Bertolini, A., *op. cit*. N. 94, pp. 214-247.

[831] Bertolesi, R., *op cit*., N. 423, p. 217, «(*iper) complessità relazionale».* Aunque el autor utiliza este término no cabe duda de que se refiere al fenómeno que otros autores identifican con la interconectividad (así, Teubner y Beckers), o incluso «*web of causation*», (así, Piergallini), o, si se quiere utilizar la terminología más clásica, «*many hands problem*», que, en palabras de Hilgendorf, en este contexto (concretamente, en el de los vehículos autónomos) significa una explosión de sujetos potencialmente responsables (*eine Explosion er Zahl potenzieller Haftungssubjekte*). Hilgendorf, E., Bayerischer Landtag, Ausschuss für Wirtschaft und Medien, Infrastruktur, Bau und Verkehr, Energie und Technologie. (2015, 29. Oktober). *Anhörung zum Thema: «Autonomes Fahren»* (Audiencia parlamentaria en el parlamento de Baviera), p. 48, disponible en: <https://www.bayern.landtag.de/fileadmin/Internet_Dokumente/Sonstiges_P/Protokoll_WI_Autonomes_Fahren.pdf >[Accedido por última vez el 27/01/2026].

[832] Bertolesi, R., *op cit*., N. 423, nota al pie 51.

[833] *Idem*., nota 51. Se dice, usando la expresión de Hilgendorf, que estamos aquí ante un *«control dilemma»*, pues se plantea si es posible que el conductor delegue una parte de sus deberes/responsabilidad, pero a la vez mantenga un deber de vigilancia o monitoreo en todo caso. Un desarrollo más completo sobre el papel del usuario supervisor en el capítulo 8, *infra.* Hilgendorf, E., «Automated Driving and the Law» en Hilgendorf, E./Seidel, U., (Eds.), *Robotics, Autonomics and the Law*, Nomos, Baden-Baden, 2017, pp. 181-182.

[834] Resulta curioso cómo el autor resalta el papel que debe ejercer en este campo el principio de confianza *(principio di affidamento).* Nuestra sensación es, empero, la contraria, pues parece que el principio del que se parte cuando se trata de una cooperación grupal innegable (presupuesto, sin duda, de aplicación de este principio de confianza), la regla a seguir no es la propia de este principio, sino más bien todo lo contrario, es decir, nos parece que en la interacción humano-máquina ocurre la situación inversa que en la interacción humano-humano, siendo el protagonista absoluto el principio de desconfianza. Consideramos, de hecho, que el principio de supervisión humana, sobre el que se hablará más adelante, no es más que una concreción de este principio de desconfianza en esta nueva clase de acción grupal o colectiva. Sobre el principio de confianza en nuestro ordenamiento, imprescindible: Maraver Gómez, M., *El prin-*

serán habituales extrae el autor italiano la posibilidad de aplicar la institución de la *cooperazione colposa* [835].

Finalmente, y poniendo el énfasis en el último de los requisitos de la imputación objetiva, a saber, el requisito de realización de riesgo en el resultado o relación de riesgo, (es decir, en lo relativo a la imputación de resultado en sentido estricto), Ibold emplea un argumento que es, a nuestro juicio, definitivo, para negar utilidad al modelo de delito de resultado y que pone en relación la falta de previsibilidad o incertidumbre derivada de la naturaleza de los procesos técnicos de los sistemas de IA, que traen causa en lo que ella denomina opacidad sistémica [836]. La idea es que si un sistema es epistémicamente opaco, las conclusiones necesarias para afirmar la relación de riesgo (que el riesgo jurídicamente desaprobado sea el mismo que se ha concretado en el resultado), no se puede deducir qué reglas o factores han intervenido en la causación del daño concreto, pues no es posible conocer el proceso de transformación técnica del sistema. La autora reflexiona críticamente sobre las fuentes actuales para conocer esta relación causal técnica, donde la técnica de la comparación estadística con otros sistemas no es válida en tanto cada sistema de IA, precisamente por sus capacidades de autoaprendizaje y modificación constantes, impide las reconstrucciones genéricamente aplicables [837].

Ibold explica que las dos teorías enfrentadas para resolver los clásicos problemas de riesgos concurrentes, cuando no es posible saber con certeza cuál de ellos se concreta en el resultado (el jurídicamente desaprobado o uno interpuesto) son, esencialmente, la teoría del incremento del riesgo y la teoría de la evitabilidad. Aunque la primera (la teoría del incremento del riesgo), que ha encontrado firmes defensores entre la doctrina [838], es capaz de resolver alguna laguna (por facilitar enormemente la imputación, bastando con la afirmación de que el comportamiento del autor, contrario al deber y superando el riesgo permitido, ha incrementado la probabilidad de producción del resultado), es completamente ilegítima, a pesar de ser más fácilmente aplicable en el contexto de la IA, donde las decisiones a pesar de ser opacas epistémicamente,

---

*cipio de confianza en derecho penal*, *Un estudio sobre la aplicación del principio de autorresponsabilidad en la teoría de la imputación objetiva*, Civitas, Cizur Menor (Navarra), 2009.

[835] No es esta, sin embargo, la propuesta principal de Bertolesi que se decanta por un modelo más parecido a nuestro delito de peligro. Sin embargo, plantea esta posibilidad como otra hipótesis, que no desarrolla demasiado. Bertolesi, R., *op cit*., N. 423, pp. 220-221. Nos remitimos, respecto de este modelo de la cooperación culposa, a lo afirmado *supra*, en nota al pie 811.

[836] Ibold V., *op. cit*. N. 104, pp. 384 y ss (habla de „*Unsicherheit über KI-technische Prozesse» afectando a la „Pflichtwidrigkeitszusammenhang»).*

[837] *Ibid.,* p. 386.

[838] *Ibid.,* p. 390, entre los cuales se encuentran, destacadamente, Roxin y Greco.

sí son transparentes estadísticamente [839]. Se comparten en este sentido los argumentos esgrimidos en cuanto a la ilegitimidad de esta teoría, que Ibold califica de responsabilidad por un *versari in re illicita* [840], pues pasa por delante de los postulados del *in dubio pro reo,* convirtiendo los delitos de resultado en delitos de peligro condicionados al resultado. Si una vez que se ha superado el riesgo permitido, siguiendo la teoría de Roxin, se crea un riesgo, en su conjunto desaprobado, con independencia del resto de factores o cadenas causales interpuestas, ningún sentido tiene acudir al juicio de imputación de resultado, pues todo riesgo desaprobado se tendrá por materializado en el resultado, de forma que el criterio de relación del riesgo perdería toda virtualidad restrictiva, convirtiéndose esta teoría en una presunción en contra del reo.

De todo ello deriva Ibold una consecuencia, que se comparte plenamente, y es que el requisito de imputación del resultado propio de la teoría de la imputación objetiva puede, de hecho, impedir que se impongan sanciones por un delito imprudente, pues dentro del contexto de tecnificación y de IA, es previsible que aumenten estas constelaciones de casos donde la relación técnica incierta no permite la imputación legítima. De este resultado «insatisfactorio» para el derecho penal [841], que aquí se ha denominado laguna de responsabilidad, no puede oponerse, como se ha tratado de argumentar, una solución innovadora que conlleve una aplicación forzada o ilegítima de las categorías dogmáticas, sino tan solo pedir que sea el legislador el que ponga solución al problema mediante la creación de tipos especiales de peligro abstracto [842].

---

[839] Pone el ejemplo la autora, en p. 389, de que sería suficiente para imputar el resultado constatar que un sistema de IA sin deficiencias presenta un porcentaje menor de decisiones erróneas (por ejemplo, el 9.5% de los casos) que uno con deficiencias (por ejemplo, en el 5% de los casos), para afirmar la imputación del resultado, bastando con conocer las estadísticas generales de funcionamiento de una tecnología.

[840] *Ibid.,* p. 393.

[841] *Ibid.,* p. 393 *(unbefriedigende Ergebnis).* Algo más matizado en sus conclusiones, pero también señalando las dificultades que para el régimen de la imprudencia supone la autonomía de los sistemas, reconociendo que la determinación necesaria para ello del desarrollo hipotético del «comportamiento» de un sistema de IA con una programación conforme a sus obligaciones resulta especialmente difícil en el caso del aprendizaje automático: Fateh-Moghadam, B., «Innovationsverantwortung im Strafrecht: Zwischen strict liability, Fahrlässigkeit und erlaubtem Risiko – Zugleich ein Beitrag zur Digitalisierung des Strafrechts», *ZStW*, 131, 2019, p. 885.

[842] Ibold, V., *ibid,* p. 393. Conviene precisar en este punto que somos conscientes de la existencia de otros criterios delimitadores a tener en cuenta en la determinación de la imputación objetiva del resultado, tales como el criterio del fin de protección de la norma, ampliamente reconocido por la doctrina como elemento de la imputación objetiva del resultado. Consideramos, sin embargo, que cumple una función esencialmente delimitadora o negativa, en cuanto permite excluir la imputación de resultados ajenos al sentido de la norma infringida, pero no constituye un fundamento autónomo para imputar resultados cuya evitabilidad no haya quedado acreditada con el grado de certeza exigido por el principio *in dubio pro reo,* especialmente en contextos de incertidumbre causal estructural como los derivados del uso de sistemas de inteligencia artificial opacos. Sin ánimo de exhaustividad, uno de los defensores del fin de protección de

c) Propuestas de reconstrucción del concepto de previsibilidad

En la idea de añadir a la discusión penal los postulados de los civilistas que, sin duda, han reflexionado largamente sobre las implicaciones que la IA presenta para el régimen de responsabilidad por imprudencia, resulta especialmente útil traer a colación el trabajo de Selbst [843], que caracteriza con maestría las necesidades para un nuevo concepto de previsibilidad. El punto de partida, que consideramos acertado, es que la previsibilidad, cuya función esencial es limitar la responsabilidad, pasa a ser en estos contextos la excepción, y no la regla, y esto distorsiona enormemente todo el esquema de la imprudencia.

La idea señalada por Selbst es que la responsabilidad por negligencia se verá esencialmente modificada por el hecho de que la IA es una herramienta donde la posibilidad de usarla sin errores no depende del nivel de cuidado desplegado por el usuario [844]. Y ello es así por la propia concepción de los sistemas, que están diseñados no solo para replicar las habilidades humanas, sino para superarlas, lo que determina una evidente dificultad para que un humano sea capaz de corregir el desempeño de una IA en tiempo real, sin que sea posible saber, por tanto, si está ajustándose a su deber de cuidado en su tarea de supervisor. No solo aplica este problema al usuario, pues Selbst destaca que también la labor de los productores, como supervisores del producto, se complica sobremanera, cuando dicho producto tiene un funcionamiento impredecible por su propia naturaleza. Por ello, para este autor lo que se debe hacer es no abandonar el concepto de previsibilidad que permite afirmar la imprudencia, sino reconfigurarlo.

El primero de los factores que hay que tomar en cuenta, según Selbst, para reformular este concepto es el relativo al error. En lo relativo al error, nos remitimos a lo apuntado *supra,* cuando tratábamos de caracterizar la naturaleza «defectuosa» de los sistemas, pero merece la pena reiterar en este punto la idea de base. Como una vez que el sistema lleva a cabo una decisión (en forma de predicción basada en un análisis de datos), que el usuario (o el productor, da lo mismo) decida confiar en esta decisión, tomándola por válida o correcta, es una cuestión de fe, pues lo cierto es que la razonabilidad *(reasonableness)* de una acción reside más en la decisión acerca del uso de la IA que acerca de

---

la norma en España: Gimbernat Ordeig, E., «Fin de protección de la norma e imputación objetiva», *ADPCP*, 2008.

[843] Selbst, A., «Negligence …», *op. cit*. N. 718.

[844] *Ibid.,* p. 1331.

la corrección de la decisión de la IA [845]. Será el nivel de desarrollo científico el que influya en si es o no razonable utilizar inteligencia artificial. Pero cuando, en un determinado sector, se determina que la IA logra reducir los errores por debajo de cierto umbral, podría considerarse una elección adecuada. Para cuando se evalúa si el profesional actuó de manera razonable o no, la decisión de implementar la IA ya ha sido tomada, por lo que no puede incluirse en ese juicio. Es decir, si en un inicio se determina que el uso de la IA fue apropiado (y esto, nos tememos, tiene que ver con que la propia normativa permita, con carácter general, el uso de sistemas en ciertos sectores), no se puede considerar posteriormente como negligente el no haber identificado errores que son imposibles de detectar por seres humanos. De lo contrario, estaríamos equiparando la imprudencia con una responsabilidad automática, de carácter objetivo, sin importar las posibilidades reales de control del profesional.

Reformula directamente Selbst el dilema o paradoja arriba apuntado: no es cierto que ningún tipo de error sea previsible, de hecho, y habida cuenta de la tecnología que se usa actualmente, que está circunscrita a una (o unas pocas) funciones, es posible hacerse una composición sobre los posibles tipos de daños que puede causar, pero esto, para Selbst, al igual que lo señalan otros autores, no puede significar que sea posible afirmar que, como algunos errores ocurren, todos los errores son previsibles en cada uno de los casos [846]. Se pone de manifiesto que, una vez tomada la decisión sobre que el uso de la IA es razonable en un determinado sector, la dicotomía será la siguiente en caso de que se produzca un daño: o bien siempre es imputable (pues todo daño era remotamente previsible), lo cual implica un régimen de responsabilidad objetiva, o bien nunca dará lugar a la responsabilidad.

En caso de que todas estas reflexiones, que beben esencialmente de la interpretación del régimen de responsabilidad por negligencia civil, no fueran lo suficientemente convincentes, se ha de destacar cómo algunos penalistas han puesto de manifiesto problemas similares. En este sentido, el trabajo de

---

[845] *Ibid.,* p. 1339: «*This implies that the reasonableness of an action in individual cases must be tied to the decision to use AI in the first place*».

[846] *Ibid.,* p. 1346. En un sentido similar, afirmando que esta paradoja no puede conducir a la redundancia del requisito de previsibilidad objetiva, y exigiendo que la cuestión de la previsibilidad objetiva se resuelva, en el caso de la IA, caso por caso, Lohmann, A., *op. cit. N. 298*, p. 161. Esto implica, necesariamente, una inseguridad jurídica evidente para el receptor de la norma de cuidado. La misma autora concede, finalmente, que, aunque ni la doctrina ni la jurisprudencia alemanas exigen una previsibilidad detallada sobre la configuración del hecho, el hecho de que los sistemas autónomos de IA sean más imprevisibles que otros productos tradicionales puede llevar a la afirmación automática de la previsibilidad objetiva. Más bien, concluye, deben exigirse indicios de que se va a producir un daño en el caso concreto para que se pueda afirmar. Ibid, p. 163.

Zhao [847] resulta esclarecedor. Parte este autor de la existencia de lagunas de responsabilidad en el derecho penal, considerando que uno de los principales problemas se presenta, precisamente, en el régimen de imputación por imprudencia, partiendo de la base de que el autor de los hechos no tenga la opción de prever las consecuencias de sus acciones [848]. Realiza un repaso por la evolución teórico-dogmática del delito imprudente, para tomar partido por una concreta concepción, que nos parece en lo esencial correcta, según la cual más que concebir la imprudencia con base en la voluntad del actor [849], se debe centralizar el régimen de la imprudencia dentro del análisis de la imputación objetiva (él llama a este modelo imputación normativa), lo que nos permite caracterizar su postura con la de aquellos que, como aquí se ha optado, consideramos que en su parte objetiva, el injusto del delito imprudente y del delito doloso responden al mismo esquema de imputación (basado, en esencia, en la creación de un riesgo jurídicamente desaprobado). La idea defendida aquí por Zhao es que las lagunas de responsabilidad en los crímenes de IA tienen que ver con que se pone demasiado énfasis en los elementos subjetivos, cuando es en el plano objetivo donde está la mayor batalla pendiente. El mayor problema está, para el autor, en el estado aun incompleto de determinación del deber de cuidado.

Afirma Zhao que riesgo permitido y deber de cuidado, están, por tanto, directamente relacionados, siendo el gran obstáculo para la imputación del delito imprudente que en materia de IA las leyes, regulaciones y normas industriales estén aún en proceso de perfeccionarse, por lo que esta ausencia parcial de deber de cuidado convencional genera claras lagunas de imputación, y solo haciendo una evaluación (hemos de entender que en un sentido más laxo, o menos apegado a estándares formales) del riesgo permitido, se podrá avanzar en la imputación [850].

Por el momento, a pesar de esta incertidumbre sobre los estándares de cuidado debido, el desarrollo tecnológico continúa, por lo que el desconocimiento sobre el deber de cuidado supera la capacidad cognitiva de la mayoría de personas, que, aunque puedan ser generalmente capaces de vislumbrar los peligros abstractos de las tecnologías, pueden ser generalmente incapaces de prever los peligros concretos. Introduce aquí Zhao un matiz que consideramos importante, afirmando que la posibilidad de prever un resultado no equivale necesariamente a la posibilidad de dominarlo [851]. La consecuencia es que el

---

[847] Zhao, S., *op cit*. N. 160.

[848] *Ibid.,* p. 47, afirmando que este es el principal impacto de las tecnologías de IA en el delito imprudente, llegando a calificarlo de «*attribution dilemma*» (dilema de atribución).

[849] *Ibid.,* pp. 71-73, hablando el autor de los modelos de «*will imputation*».

[850] *Ibid.,* p. 81

[851] *Ibid.,* p. 82. «*The main reason here is that the foreseeable possibility of an outcome does not necessarily equate to the possibility of dominating such outcome*».

elemento de la dominabilidad debe tener un papel, de forma que sustituya de alguna forma un concepto de negligencia muy apegado a lo subjetivo, para afirmarla en un sentido de dominio efectivo de las esferas de riesgo [852].

Lo que propone Zhao es, en definitiva, aplicar en todo su esplendor la teoría de la imputación objetiva como forma de determinar el injusto del delito imprudente, donde la creación de un riesgo jurídicamente desaprobado se verá influida (ante la ausencia de normas de cuidado claras), por el rol o posición de cada sujeto, y donde la normativa existente actúe como filtro, estableciendo una orientación sobre el fin perseguido por ellas.

Aunque la propuesta de Zhao resulta coherente y dogmáticamente fundada, nos parece que relegando todo el problema eminentemente subjetivo al plano objetivo no soluciona la verdadera dificultad, que no es capaz de negar en su trabajo. Si la imprevisibilidad es un hecho, optar por una imputación normativa en lo relativo al injusto objetivo del delito imprudente, tomando como base el primer juicio de análisis de la teoría de la imputación objetiva, lo único que hace es desplazar el problema a la esfera subjetiva que, inevitablemente, va a tener que examinarse. Podemos compartir el esquema que identifica la creación de riesgo con el dominio de una esfera de responsabilidad asignada, (siempre y cuando, desde luego, esta asignación se base en estándares que ofrezcan la suficiente certeza jurídica, aunque el propio Zhao se lamenta de que aún son insuficientes), pero ello no impide que, en un ulterior momento, haya que examinar que, más allá de esa infracción del cuidado debido objetivamente, se deba comprobar la infracción del deber personalmente exigible, esfera que el autor deja sin resolver, y donde, una vez más, parece ineludible que, en la determinación de los estándares de lo exigible (o, si se quiere, en la determinación de la frontera entre la imprudencia y el dolo), haya que atender a aquello que se espera de un sujeto ideal.

Y en este punto, nos tememos, no nos queda más que afirmar, con Selbst, que estamos muy lejos de saber cuál es ese sujeto ideal «supervisor» de la IA [853].

---

[852] Habla, en sentido literal, del modelo «*risk responsibility*» (*Ibid.,* p. 83 y ss.). En este contexto, la responsabilidad por riesgo implica que la persona que controla un riesgo es responsable tanto de dicho riesgo como de sus consecuencias. En otras palabras, quien tiene dominio sobre un hecho debe ser responsable de su ocurrencia. Según este principio, cada individuo debe gestionar su entorno de manera que no transfiera peligros a otras personas, ya que, de hacerlo, asumirá la responsabilidad de las consecuencias. En el caso de los delitos de negligencia, esta responsabilidad por riesgo se convierte en la base para la imputación. Así, la atribución de responsabilidad al autor debe basarse en si gestionó razonablemente el riesgo dentro de su ámbito de control, siendo especialmente importante identificar a quién corresponde gestionar cada riesgo. En definitiva, la responsabilidad por riesgo sustituye el concepto de previsibilidad en la imputación de la negligencia, favoreciendo la imputación basada en el dominio del riesgo, en detrimento del examen del elemento subjetivo. El autor se remite a textos de Duff o Lacey sobre la responsabilidad en sentido amplio, pero lo cierto es que sus propuestas se acercan más, entendemos, a una forma de entender la responsabilidad organizacional a partir del rol asumido en la línea de Jakobs.

[853] Selbst, A., «Negligence …», *op. cit*. N. 718, p. 1346 y ss.

La alternativa es, en este caso, la siguiente: o bien exigimos al humano (en calidad de supervisor de los riesgos del producto inherentemente peligroso, ya sea este supervisor-productor o supervisor-usuario), un estándar de atención y conocimiento muy alto (que, en determinados sectores, supera, de hecho, el estándar del hombre prudente que espera el derecho penal), o bien nos apoyamos en una idea de razonabilidad y afirmamos que, si los estudios psicológicos afirman que es difícil superar ciertos sesgos o condicionamientos naturalmente presentes [854] en nosotros, el derecho no puede legítimamente exigir a una persona superar estos condicionamientos.

### d) Intentos de aplicación práctica del delito imprudente: el ejemplo del contexto médico

Desde luego, somos conscientes de que otras voces autorizadas de la doctrina penal actual han defendido el modelo de imputación del delito de resultado, y lo han hecho de forma especialmente convincente en contextos altamente sensibles, como es, paradigmáticamente, el de causación de lesiones o muertes en el ámbito de la salud. Sin embargo, creemos que sus razonamientos, que enseguida analizaremos, no superan las objeciones aquí manifestadas.

Así, muy recientemente, Valerius [855] y Rueda Martín [856], respectivamente, han analizado la posible responsabilidad por los daños causados por IA en el contexto médico, llegando a conclusiones similares.

En el caso de la autora española, uno de sus puntos de partida es que estas herramientas de IA (partiendo ya de la definición que otorga el Reglamento), *«toman decisiones o proponen actuaciones que quedan fuera de la capacidad de*

[854] Nos referimos, en este punto, a los estudios, ya amplios, sobre las pobres habilidades de control o reacción que implica una labor de supervisión de herramientas que toman decisiones en espacios de tiempo muy cortos (paradigmáticamente, en casos de vehículos autónomos, pero es aplicable a otros contextos de supervisión), así como a todos los estudios que afirman la difícil superación del sesgo de automatización, naturalmente presente en nosotros. Lo mismo respecto de los estudios más recientes que muestran cómo los humanos, en nuestra interacción con máquinas, somos absolutamente sensibles a absorber, de forma inconsciente, sus sesgos. En un sentido genérico y clásico, sobre los sesgos: Kahneman. D., *Pensar rápido, pensar despacio,* Debate, Barcelona, 2012. Un estudio sobre nuestra tendencia a aprehender de forma inconsciente los sesgos de los sistemas: Glickman, M./Sharot, T., «How human-AI feedback loops alter human perceptual, emotional and social judgements», *Nature Human Behavior*, 9, 2024, pp. 345-359.

[855] Valerius, B., «Strafrechtliche Grenzen lernender Künstlicher Intelligenz», *GA,,* 3/2022. En algún trabajo anterior, Valerius se ha encargado de la presente cuestión, si bien desde otros ámbitos, como el relativo a la conducción autónoma, con una argumentación paralela a la aquí presentada. Valerius, B., «*Sorgfaltspflichten beim autonomen Fahren*», en Hilgendorf E., (Coord.), *Autonome Systeme und neue Mobilität,* Nomos, Baden-Baden, 2017, pp. 9-22.

[856] Rueda Martín, M. Á., «La concreción …», *op. cit*. N. 687, pp. 178-231.

*previsión del ser humano»* [857] y que, aunque el Reglamento establece una firme obligación de supervisión humana [858], lo cierto es que aún queda por delante toda una tarea de concreción del deber objetivo de cuidado. En todo caso, reconoce que la doctrina mayoritaria sí considera que para determinar este deber de cuidado es necesario partir como «presupuesto lógico» de la previsibilidad objetiva [859].

Compartimos en lo esencial lo que sigue, a saber, que en la averiguación de esta previsibilidad objetiva del resultado hay que tener en cuenta (dentro de la actividad médico-quirúrgica), los conocimientos propios del autor (en este caso, el médico) y toda la experiencia o saber acumulado en una determinada época (el saber nomológico). Sin embargo, a partir de aquí, la idea que defiende Rueda, y que a aquí se discute, es que, de los sistemas de inteligencia artificial usados en el contexto médico es posible inferir un conocimiento nomológico más específico, que conlleva un saber experimental más preciso, lo que conduce a una concreción más adecuada sobre la producción de resultados lesivos objetivamente imputables, dando lugar a la elevación del nivel de diligencia o cuidado necesario exigido por el derecho [860]. No compartimos, en absoluto, esta aseveración. Que el uso de una determinada técnica (en este caso, de ayuda al diagnóstico), resulte en un mayor conocimiento del usuario profesional que se vale de ella, es, a nuestro parecer, una afirmación incorrecta. De hecho, como se ha expuesto más arriba, la razón por la cual se decide usar una herramienta en el contexto médico es, precisamente, porque se presupone que existen ciertas deficiencias (cognitivas) en los profesionales que pueden ser suplidas por los sistemas, por su mayor capacidad de procesamiento. Que una herramienta más compleja, creada para superar las habilidades de quien es asistido por ella (recordemos que, de hecho, es normal la referencia a herramientas de asistencia al diagnóstico, implicando esta idea de asistencia la de

---

[857] *Ibid.,* p. 185.

[858] El mismo se recoge, de una forma bastante exhaustiva, en el artículo 14. Su contenido y alcance será objeto de análisis infra, en capítulo séptimo. Por ahora, baste apuntar que algunos autores han tratado sobre la importancia de este principio. Así, y concretamente explorando el papel en el ámbito médico: Lazcoz Moratinos, G., «Sistemas de inteligencia artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos», Premio de Investigación en Protección de Datos Personales «Emilio Aced», 2022, Agencia Española de Protección de Datos, parte de la tesis doctoral: «Gobernanza y supervisión humana de la toma de decisiones automatizada basada en la elaboración de perfiles», 2023. También, extensamente, sobre la nueva configuración de los deberes jurídico penales en el ámbito del diagnóstico médico con IA, desde una perspectiva sociológica y penal: Amore, N./ Rossero, E., Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità. Una ricerca sociologica e giuridico-penale, Il Mulino, Bolonia, 2024.

[859] En este punto, las referencias que usa Rueda son a varios autores, todos ellos partidarios de la corriente del finalismo, que incluyen a Welzel, Cerezo, Hirsch, Romeo, Luzón, Gracia o Gil. Rueda Martín, M. Á., «La concreción …», *op. cit*. N. 687, nota al pie 30.

[860] *Ibid.,* p. 195.

ayuda a quien no tiene los medios suficientes), conduzca necesariamente a unas mayores expectativas normativas sobre quien necesita del asistente, nos parece un punto de partida inadecuado.

No negamos, desde luego, el papel esencial de la supervisión humana (no sería posible, por otra parte, habida cuenta del contenido de la reciente normativa europea), pero sí nos parece discutible la extrapolación que se realiza de dicho principio. Rueda, que coloca a la supervisión humana como elemento consustancial a la definición del deber objetivo de cuidado, parece que considera aplicable, de forma directa, los postulados del artículo 14 del Reglamento (en esencia, los deberes de entender las capacidades y limitaciones del sistema de IA, ser conscientes de la posible tendencia a confiar automáticamente en los sistemas, interpretar correctamente la información de salida y decidir no utilizar el sistema ante cualquier situación concreta, invalidando la información de salida).

Nos parece, empero, atrevido poner una carga de estas características en un usuario como el médico que, si bien está dotado de amplios conocimientos respecto de su campo de experticia, no los tiene (ni, nos atrevemos a decir, se le presuponen), conocimientos lo suficientemente amplios en el funcionamiento de los sistemas de IA como para cumplir con las obligaciones arriba indicadas.

Los planteamientos del artículo 14 suponen un *desideratum* que es difícil de alcanzar. En cuanto a la competencia de entender las capacidades y limitaciones del sistema (por quien es simple usuario, con formación médica, pero sin capacidad de superar la capacidad predictiva del sistema, que ha sido precisamente diseñado para actuar allí donde el error humano no llega), y porque va en contra de los estudios empíricos más básicos. Una cosa es que podamos ser conscientes de nuestro sesgo de automatización, pero otra muy distinta (y así lo ponen de manifiesto todos quienes han estudiado el sesgo humano en profundidad [861]), es que podamos realmente superarlo. ¿Con qué herramientas decidir si una máquina ha fallado, si no sabemos en qué parámetros se ha basado? No se pone en duda que al personal médico se le presuponga normativamente una serie de conocimientos especializados y elevados (así, conocimientos jurídicos, éticos y técnicos) [862], pero no consideramos que los conocimientos que se presumen

[861] De nuevo, por todos, Kahneman, D., *op. cit*. N. 854. Entre la literatura española: Muñoz Aranguren, A., «La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación», *InDret*, 2/2011.

[862] En este sentido, Obregón Fernández, A./Lazcoz Moratinos, G., «La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional Humanitario y el Derecho de la Unión Europea», *Revista Electrónica de Estudios Internacionales*, 42, 2021, p. 17. Resulta curioso que la propia Rueda reconoce (p. 85), que los sistemas de IA ofrecen técnicas más eficaces en el diagnóstico, el tratamiento o la medicación, «*difíciles de hallarse por la inteligencia humana*», afir-

exigibles en una situación como la presente, donde este tipo de técnicas están en sus primeras fases de implementación, y ni siquiera los expertos en el campo de la inteligencia artificial son capaces de desentrañar su funcionamiento. Es evidente que este *desideratum* se debe tomar en consideración, pero los parámetros que la normativa administrativa impone, en este punto, no parece que sean extrapolables al campo penal, o al menos, no en los términos aquí propuestos, pues exigen lo imposible a quien no está en posición de hacerlo, si partimos de la base del hombre medio ideal [863].

No podemos estar de acuerdo, tampoco, con la imagen de «hombre sensato y prudente» que propugnan Romeo y Lazcoz Moratinos [864], dando un papel primordial a las máximas de experiencia en la determinación del cuidado debido, apelando a la lógica de la persona inteligente y sensata que será fundamental para ayudar a determinar el margen de automatización asumible en cada situación. Solo será predicable, en la idea arriba apuntada, un alejamiento evidente de ese estándar de persona sensata en los casos en que el sistema devuelva errores patentes (errores que cualquier médico ideal, diríamos, sería capaz de detectar), pero allí donde el error es sutil, velado, o difícilmente comprensible, al estar basado, por ejemplo, en correlaciones cuya comprobación escapa al entendimiento humano (recordemos que estas herramientas no van a explicar los factores por los cuales dan sus resultados), será atrevido pensar que apelando a la sensatez bastará para detectar estos errores. Ante escenarios de tanta complejidad, el mejor detector de errores no es, desde luego, el hombre más sensato, sino el que posee conocimientos más especializados.

---

mando que constituyen un complemento adecuado del personal sanitario. En definitiva, hay un reconocimiento de que los sistemas llegan donde no puede llegar el humano en cuanto a sus capacidades. Rueda Martín, M. Á., «La concreción …», *op. cit*. N. 687, p. 85.

[863] Cuestión distinta es que se considerase, por ejemplo, que, habida cuenta de las enormes dificultades que plantea la implementación de estas herramientas en un campo tan sensible como el médico se plantease la obligación de obtener una capacitación teórica específica (algo así como un permiso de manejo de herramientas de IA), donde se hiciera una formación que garantizara tener estos conocimientos. De hecho, la propia Rueda reconoce la conveniencia de algo parecido, hablando de la posibilidad de establecer un examen previo a la implantación como una suerte de acreditación oficial, Rueda Martín, M. Á., «La concreción …», *op. cit*. N. 687, p. 218, hablando de un «*examen general y minucioso del sistema por técnicos expertos acreditados e independientes*».

[864] Véanse Romeo Casabona, C. M./Lazcoz Moratinos, G., «Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?», *Revista de Derecho y Genoma Humano*, 52, 2020, p. 161. *«La dificultad de ofrecer criterios practicables y más precisos [*para determinar el deber de cuidado, se refieren*] empieza precisamente aquí, al presentarse en la realidad variables de difícil subsunción normativa. Ello nos devuelve a los criterios normativos propios de la imprudencia, como el prototipo del hombre sensato y responsable y el principio de confianza, principalmente»,* y en p. 162 que «*al enfrentarnos a tecnologías en constante desarrollo y perfeccionamiento, las máximas de experiencia desempeñarán una relevante función a la hora de señalar el cuidado debido, que apelará fundamentalmente a la figura ideal del hombre sensato y responsable y ayudará a determinar el margen de automatización asumible en cada situación».* Rueda comparte esta idea, y así lo expresa en p. 198.

Desde luego, reconoce la autora un margen para la exclusión de la intervención penal (que para ella deriva de un criterio valorativo extra sistémico de ponderación de intereses), pero diferencia entre dos supuestos: una cosa es que el comportamiento del médico se ajuste al estado actual del conocimiento nomológico, donde no se podrá afirmar infracción del deber objetivo de cuidado [865], pero otra cosa es que el comportamiento del médico no responda a este actual conocimiento, en cuyo caso la atribución de responsabilidad por imprudencia dependería de dos factores estrechamente relacionados: de la corrección o incorreción de la información proporcionada por el sistema de IA, y de la supervisión efectiva del facultativo [866].

Una vez más, no podemos estar del todo de acuerdo, pues, ante algunos supuestos (nos atreveríamos a decir que la mayoría de ellos), el resultado del diagnóstico no será evidentemente incorrecto, y por lo tanto la supervisión del facultativo se produce, sí, pero solo con los medios que tiene a su alcance, con unas herramientas de supervisor, si se quiere, intrínsecamente imperfectas [867].

En todo caso, no en todo estamos en desacuerdo con las conclusiones de Rueda, pues destaca que se hace precisa una regulación específica que asegure que se cumplan requisitos de calidad y eficiencia antes de su implementación [868], y que solo con el establecimiento de pautas estrictas de actuación en caso de discrepancias entre el juicio del profesional y del sistema se podrá avanzar hacia un necesario nuevo perfil de profesional de la salud [869]. Conside-

[865] Rueda Martín, M. Á., «La concreción …», *op. cit*. N. 687, p. 214, afirma que en este caso se puede reconocer que, aunque haya un error o fallo técnico estaríamos ante un error invencible.

[866] *Idem.*

[867] Plantéese el lector, para poner a prueba la validez de este juicio, si estaría dispuesto a imputar por un homicidio imprudente al funcionario (policía o juez) que, ante la denuncia de una mujer presuntamente maltratada por su pareja, después de usar un sistema de IA de caja negra, tiene que decidir si da por bueno ese diagnóstico o no. El único de los pronunciamientos judiciales que, en un caso muy similar (aunque la herramienta era de IA «predeterminada» y no incorporaba IA en sentido estricto), se pronuncia al respecto acaba condenando por mal funcionamiento de la Administración en la vía administrativa. Vid. Sentencia de la sección 5.ª de la Sala de lo contencioso administrativo de la Audiencia Nacional de 30 de septiembre de 2020. Si siguiéramos al pie de la letra la propuesta de Rueda (y dando por hecho que no sería descabellado considerar a un funcionario como un policía garante), deberíamos condenarlo por el correspondiente resultado en su modalidad imprudente.

[868] Rueda Martín, M. Á., «La concreción …», *op. cit*. N. 687, p. 217, aceptando, con Romeo y Lazcoz, la implementación de un marco sobre una estructura próxima a la idea de *compliance.*

[869] Estas ideas, apuntadas en p. 219, se vinculan, acertadamente, con la idea señalada por Tretter, M./ Samhammer, D./ Dabrock, P, «Künstliche Intelligenz in der Medizin: Von Entlastungen und neuen Anforderungen im ärztlichen Handeln», *Ethik in der Medizin,* 36, 2024, p. 20 y ss., en el sentido de que al personal facultativo se le habrán de exigir, en el futuro, otro tipo de habilidades, principalmente comunicativas. La interesante tesis que defienden los autores es que la IA médica permitirá diferenciar en mayor medida los deberes de tipo técnico de los deberes de tipo social de los médicos, siendo el médico el único agente que aún estará sujeto a la relación de confianza con el paciente, pero partiendo de la base de que la decisión final será, en todo caso, compartida, siguiendo el esquema del *shared-decision-making.*

ramos, en tanto en cuanto no haya seguridad jurídica sobre las pautas que aquí se perfilan, que una imputación por el resultado lesivo resulta, en un estadio como el actual, excesiva, habida cuenta de que no debe realizarse, aún, una asimilación concreta, al menos en este punto, entre los deberes de supervisión tal y como los presenta el Reglamento y el cumplimiento del deber de cuidado.

En un sentido similar, apunta Valerius que, en estos mismos contextos médicos donde el profesional se ve asistido por herramientas de IA dotadas de autonomía, la imputación penal no puede rechazarse aludiendo a la imprevisibilidad del resultado. Analiza, sin embargo, el autor alemán la posible responsabilidad penal del fabricante (aunque es consciente de que otros sujetos, incluidos los usuarios médicos, pueden ser responsables también) [870]. Parte en este caso Valerius de forma expresa del paradigma de la responsabilidad por el producto del fabricante, puesto que en estos contextos (paradigmáticamente imprudentes), el deber de cuidado se debe remitir a los principios de responsabilidad civil y penal por el producto, pues la IA es, a juicio del autor, a fin de cuentas, un producto colocado en el mercado [871].

En este sentido, más allá de destacar los posibles defectos en el momento de diseño o fabricación, se refiere a todos aquellos que surgen tras su comercialización, pues con ella no acaba la responsabilidad del fabricante, sino que, a partir de ese momento, afirma Valerius, el producto se convierte en una «fuente independiente de riesgo» *(verselbstständigte Risikoquelle).*

La idea de partida es que el hecho de que la IA presente ciertas particularidades no debe excluir el régimen general; de hecho, reconoce su falta de previsibilidad *(Mangel an Vorhersehbarkeit),* pero no considera que eso (que no es más que una característica propia de su complejidad), deba impedir la imputación. Afirma, además, que según la interpretación que se viene realizando sobre el grado de previsibilidad exigido para afirmar imprudencia por la jurisprudencia, solo los casos más extraños que queden fuera de la experiencia vital podrán servir para afirmar que no hay previsibilidad. Es decir, si se usa un sistema en un contexto médico (él usa el ejemplo de un sistema para detectar tumores), será previsible que los posibles errores tengan consecuencias potencialmente morta-

---

[870] Así, VALERIUS, B., «Strafrechtliche Grenzen …», *op. cit*. N. 855, p. 124, donde plantea la posibilidad teórica de que sea el director del hospital, responsable de la atención médica, responsable por no actualizar el software o bien el propio médico que rechaza un diagnóstico ofrecido por el sistema o, todo lo contrario, aceptarlo sin verificar su corrección.

[871] *Ibid.,* p. 125: *„Um die vom Hersteller zu beachtende Sorgfaltspflicht zu ermitteln, liegt es sodann nahe, an die Grundsätze der zivil- wie strafrechtlichen Produkthaftung anzuknüpfe».*

les, siendo esta previsión genérica suficiente [872]. Habrá que estar, entonces, a todas las posibles fuentes para determinar el deber de cuidado exigible.

A continuación reconoce el autor un papel al riesgo permitido en este sector donde, tal y como ocurre en muchos otros, habrá que hacer una necesaria ponderación previa entre el coste y el beneficio de una actividad, aceptando que en los casos de productos sobre los que siempre pueda predicarse un determinado riesgo (lo que aquí hemos calificado como productos de riesgo inherente), ese «riesgo residual» que necesariamente se materializará no puede dar lugar a imputación, porque estará dentro del margen del riesgo permitido. Solo en este caso se podría negar la responsabilidad penal del médico/usuario del sistema cuando este emite un diagnóstico que finalmente es errado, pues en la decisión previa de permitir el sistema ya se habrá calibrado y tenido en cuenta que, a pesar de que algunos pronósticos sean equivocados, en general el beneficio global será mayor [873].

Es consciente Valerius de que los estándares de diligencia son, aquí, inusuales, afirmando que el estándar de calidad que debe mantenerse en la IA no puede simplemente basarse en las habilidades de una persona prudente y diligente en una situación comparable [874], y que, además, el comportamiento hipotético de un homólogo humano no puede servir como referencia, porque la acusación de comportamiento negligente contra el fabricante no se refiere a la decisión del software que ha puesto en circulación, sino que se relaciona con su lanzamiento con una programación mejorable [875].

---

[872] *Ibid.,* p. 127, con extensa referencia en notas 29 y 30 a sentencias y trabajos académicos que defienden esta postura. Destacadamente, HILGENDORF, E., «Autonome Systeme, künstliche Intelligenz und Roboter», en AAVV., *Festschrift für Thomas Fischer*, Beck, Múnich, 2018, p. 105, sobre la previsibilidad de la muerte de una persona causada por el sistema autónomo, afirmando que, respecto a la probabilidad de un suceso como este (se refiere al fallecimiento de dos ocupantes de un vehículo autónomo, en este caso por un sistema automático de mantenimiento de carril), cabe señalar que un accidente como el presentado es, en detalle, extremadamente improbable. Sin embargo, cuanto más general se describa el escenario, mayor será la probabilidad de que ocurra, sobre todo considerando que no solo un vehículo, sino una gran cantidad de ellos están equipados con estos sistemas autónomos. Con cifras tan grandes, la muerte o lesión de personas por estos sistemas se vuelve estadísticamente casi segura. Todo esto, para Hilgendorf, respalda la idea de que se puede afirmar que existe un incumplimiento de un deber de cuidado (*einer Sorgfaltspflichtzverletzung).* En definitiva, refleja este autor la idea ya comentada *supra* de que nos enfrentamos al dilema de que o bien todo es previsible o nada lo es.

[873] Así, afirma Valerius de forma literal que *«Si se cumple con este riesgo permitido, no habrá violación del deber de diligencia y, por lo tanto, no habrá responsabilidad por imprudencia**, incluso si debido al uso del producto en el caso concreto se ponen en peligro o se lesionan bienes jurídicos, como en el ejemplo en el que el diagnóstico erróneo de un sistema informático médico podría causar la muerte de una persona»*** (traducción propia, la negrita es nuestra). Así, VALERIUS, B., «Strafrechtliche Grenzen …», *op. cit*. N. 855, p. 129.

[874] *Ibid.,* p. 130: *«gewissenhaften Menschen in einer vergleichbaren Situation».*

[875] *Idem,* con referencias en nota al pie 43 a un «estándar de atención antropocéntrico» *(eines solchen anthropozentrischen Sorgfaltsmaßstabes).*

La idea que se resalta es, por tanto, la necesidad de que en su labor de supervisión continua, el productor no deje de atender a los avances posibles, implementados según los mejores conocimientos disponibles (se propugna la comparación con sistemas diagnósticos similares, como otros sistemas médicos, con la obligación de modificar la programación continuamente según los últimos avances científicos y las experiencias prácticas), obligando al fabricante a una vigilancia cuidadosa, continua y regular [876]. Esta regulación debe extenderse no solo a la supervisión del «resultado», es decir, del posible diagnóstico, sino a todo el proceso de entrenamiento y toma de decisiones. Para Valerius la mejor forma de lograr esta supervisión es transformar las cajas negras en «cajas blancas» *(whitebox),* atendiendo al paradigma del *«Explainable AI».*

La imputación penal, para el autor, podría pasar o bien por el lanzamiento al mercado de un producto con un error de diseño, o bien por la omisión del deber de monitoreo, pero ambos supuestos (entendemos, en su versión activa u omisiva), podrían dar lugar a la imputación de un homicidio imprudente [877].

Valerius reconoce la dificultad que, en términos de causalidad, introduce la complejidad de los sistemas, pero insiste en que no debe ser impeditivo para la atribución de responsabilidad, que debe basarse fundamentalmente en las obligaciones de transparencia del fabricante. Reconoce, empero, que será difícil aplicar esta responsabilidad, pero que no es teóricamente imposible.

Llegados a este punto, es importante resaltar que no nos parece criticable en todo la conclusión a la que llega Valerius sobre la importancia fundamental del proceso de la supervisión (elemento esencial también en el trabajo de Rueda), lo que sí nos parece aquí cuestionable es que la concreta forma de protección (acudiendo a los delitos de resultado, en este caso solo planteándose el castigo por lesiones u homicidio imprudente, respectivamente), sea la más adecuada.

Si la importancia primordial está en asegurar el proceso de producción y el de supervisión (ambos sujetos están implicados; productor y usuario, respectivamente), la forma óptima de hacerlo es incidir en obligaciones sobre las que exista una mayor seguridad jurídica. En definitiva, el derecho penal debe poder asegurar que los sistemas sean lo más perfeccionados posibles (de acuerdo con los conocimientos existentes en el momento), pero quizá la forma más eficaz de lograr esta supervisión sea, precisamente, clarificar cuál debe ser el papel de cada actor en cada momento. Consideramos que esta seguridad jurídica (y esta protección efectiva), se consigue de una forma mucho más completa (amén de sin complicaciones respecto de interpretaciones retorcidas so-

[876] *Idem.*
[877] § 222 del Código Penal alemán (StGB).

bre la causalidad y la previsibilidad), por medio de otro modelo de protección, cual es el del del delito de peligro.

La cuestión de la previsibilidad que está inevitablemente unida al delito imprudente parece ser insalvable atendamos a la teoría que atendamos, al menos en un último escalón, como es la imputación directamente personal, por la infracción del deber de cuidado subjetivamente exigible al humano (productor o usuario) en su labor de control de un riesgo jurídico penalmente relevante. Que sea posible determinar la infracción del cuidado debido dependerá de cómo de concreta sea la tipificación de dichos deberes (esto es una cuestión que, directamente unida con la del riesgo permitido, se tratará en el capítulo siguiente). Pero lo que resulta especialmente difícil en esta tarea de control efectivo es que, no en vano, el control humano significativo es el principal escollo cuando se trata de IA con autonomía. La verdadera dificultad del régimen de la imprudencia reside, a nuestro juicio, en la complejidad inherente a la infracción personal en el caso concreto. Los criterios normativos que aluden a un hombre ideal, medio, cuidadoso o razonable, quedan muy lejos de ser satisfactorios cuando se trata de supervisar sistemas que actúan con una racionalidad que se nos escapa. Salvo errores muy patentes en esta labor de supervisión personalmente exigible ante deberes de cuidado perfectamente cognoscibles y determinados *ex ante,* la imputación imprudente encontrará obstáculos en tanto resulta muy complejo saber cuál era la actuación debida en el caso concreto y, más allá de eso, resultará de hecho muy complejo desempeñar esa labor. Por decirlo claramente, si el deber de cuidado exigible al supervisor humano es de resultado (evitar que se produzca un determinado resultado), será ciertamente difícil afirmar que se trata de un deber realmente realizable, mientras que, si se trata de un deber de medios, podremos estar de acuerdo en que su infracción podrá dar lugar a la generación de un injusto penal. Sin una delimitación capaz de otorgar seguridad jurídica suficiente sobre el deber objetivo y subjetivo exigible al humano que interactúa con una herramienta, la labor de imputación del resultado será, nos tememos, extremadamente arbitraria, y se corre el riesgo de hacerla depender, como ha ocurrido en muchas ocasiones, de lo catastrófico de los resultados y no tanto de lo exigible que era un comportamiento cognoscible de antemano.

### e) Toma de postura

La dificultad central que plantea la imputación penal de delitos de resultado en contextos mediados por sistemas de inteligencia artificial no se limita únicamente al problema de la previsibilidad objetiva del daño. Más bien, tiene que ver

con la estructura misma del delito imprudente, que exige (para su afirmación legítima) una infracción doble: del deber objetivo y del deber subjetivo de cuidado.

En lo relativo al deber objetivo, si bien doctrinalmente ha existido un debate constante sobre cuán concreta debe ser la previsibilidad del resultado (si basta con una previsibilidad genérica o se requiere una más específica), lo cierto es que este debate se ha mostrado estéril frente a los retos que plantea la IA. En lugar de seguir discutiendo sobre el grado de probabilidad o el nivel de especificidad del resultado, parece más útil (y sobre todo más seguro jurídicamente) anclar el juicio de ilicitud objetiva en deberes normativamente definidos, puestos en relación con el marco de riesgo permitido del sector en cuestión. Esta reorientación proporciona una guía más clara tanto para los operadores jurídicos como para los potenciales sujetos activos.

Sin embargo, incluso si se lograra identificar una infracción del deber objetivo de cuidado, el obstáculo verdaderamente estructural aparece en el plano subjetivo. Y es que, por más que doctrinalmente se hayan propuesto múltiples estándares (el del hombre medio, el profesional razonable, el cuidadoso en condiciones normales, etc.), todos ellos naufragan cuando el contexto exige supervisar comportamientos que exceden la capacidad humana de comprensión o control. En otras palabras, el derecho penal no puede, sin perder legitimidad, exigir al sujeto lo que no puede razonablemente evitar. No se trata solo de una cuestión de injusto penal, sino de la base misma de la culpabilidad: no hay deber subjetivo de cuidado donde no hay posibilidad real de evitar el daño.

Aplicado al contexto de los sistemas de IA, esto implica que el modelo clásico del delito imprudente se revela incapaz de ofrecer una solución legítima y operativa. La exigencia de previsión y control al supervisor humano (ya sea profesional o usuario), carece de sentido si lo que se le pide es que detecte o evite errores que, por diseño, están fuera de su alcance cognitivo. Se le impone, de hecho, un deber de resultado sin capacidad real de actuación, lo que vacía de contenido el juicio de reproche y hace depender la imputación de la mera gravedad del resultado, no de una infracción personal efectivamente evitable.

En consecuencia, el intento de mantener la imputación por imprudencia en estos contextos no solo es dogmáticamente frágil, sino que arrastra el riesgo de producir responsabilidad penal objetiva encubierta, en la medida en que desplaza el foco desde la conducta al desenlace. Frente a esta disfunción, la solución pasa por reconfigurar los marcos de imputación y centrar la atención penal no en los resultados, sino en el incumplimiento *ex ante* de deberes claramente delimitados, adecuados a las capacidades humanas y contextualizados en el marco del riesgo permitido. Solo un modelo así puede sostenerse desde un derecho penal de acto, motivador y garantista.

### 4.3 **Dificultades del esquema de la comisión por omisión (las polémicas en torno al deber de retirada)**

Habida cuenta de lo apuntado hasta el momento, podemos afirmar que la imputación del delito de resultado, especialmente en su modalidad dolosa, atendiendo a la necesidad de previsibilidad de un riesgo concreto [878], pero también en su modalidad imprudente (allá donde la ley lo permita), con todos los problemas que sobre la determinación del deber objetivo y subjetivo de cuidado se han puesto de manifiesto, así como de afirmación del riesgo en el resultado (segundo nivel del juicio de imputación objetiva), resulta extremadamente compleja y, en muchas ocasiones, implica realizar presunciones de conocimiento que en contextos de complejidad como el aquí descrito, corren el riesgo de acabar por convertirse, *de facto,* en espacios de responsabilidad objetiva [879].

No queríamos, empero, terminar el presente capítulo sin hacer referencia a otra de las posibles respuestas que la doctrina ha desarrollado (una vez más, en atención a lo propuesto por la moderna responsabilidad penal por el producto), y que tiene que ver con la posibilidad de imputar el resultado lesivo (en los casos primigenios, las muertes y lesiones), a ciertas conductas de carácter omisivo, fundamentalmente las que tienen que ver con la omisión del deber de supervisión y retirada de los productos, basándose en una supuesta posición de garante del fabricante.

No será objeto de este apartado realizar un estudio profundo sobre la dogmática del delito en comisión por omisión en casos de responsabilidad por el producto, no solo por exceder con mucho las pretensiones de la investigación, sino porque, en lo esencial, en nada modifica lo afirmado hasta el momento (es decir, el principal problema, que tiene que ver con las distorsiones que la complejidad y autonomía de los sistemas impregnan en la previsibilidad, tanto de cara a la imputación por imprudencia como de cara a la imputa-

---

[878] Tal y como se desprende, desde la primigenia sentencia del TS en el caso de la colza, hasta nuestros días, por la jurisprudencia y doctrina como criterio mayoritario.

[879] Selbs ha calificado a estos espacios de «focos de responsabilidad objetiva» *(«pockets of strict liability»),* que se dan en situaciones en las que se impone responsabilidad dentro de un régimen de negligencia, incluso cuando no existe ningún nivel de cuidado que el autor del daño pudiera haber alcanzado para prevenir la lesión. Un ejemplo de ello es que, al aplicar un estándar objetivo del «hombre razonable», la ley impone responsabilidad objetiva a personas (como los niños, las personas con discapacidades o los ineptos) que tal vez no sean capaces de cumplir con dicho estándar. Selbst, A., «Negligence …», *op. cit*. N. 718., p. 1349. Un ejemplo claro de uno de estos «focos», lo podemos encontrar en el derecho penal del mercado de valores, donde precisamente creemos que se reproduce el problema en todo su esplendor; se presupone al usuario (al negociador físico que interactúa en un ambiente plagado de sistemas de IA, propias y ajenas), unos estándares de racionalidad o razonabilidad que no son en absoluto ciertos. Sobre este problema, en mayor profundidad, Bravo Bolado, A., «Negociando en la sombra…, *op. cit*. N. 245, pp. 243-288.

ción objetiva del resultado), se mantienen intactas tanto si la conducta precedente es activa, como si es omisiva. Sin embargo, sí quisiéramos poner de manifiesto algunas cuestiones.

Como es bien sabido, cuando se empezó a discutir sobre la posibilidad de imputar resultados lesivos a los productores en comisión por omisión, una de las mayores polémicas dentro de la doctrina tuvo que ver con la discusión sobre cuál era la concreta fuente de la posición de garante del productor, pues era difícil asumir que, ante una conducta inicial que era perfectamente ajustada a derecho (la puesta a disposición de un producto que, en ese momento, era acorde con todas las normas y los deberes de cuidado), pasa a transformarse en una conducta generadora de una posición de garantía tras un cierto tiempo [880].

Las respuestas, en este sentido, han sido variadas. Hay quienes han entendido que la injerencia podía justificarse a pesar del actuar previo correcto, o quienes entendieron que la posición derivaba del deber de control de una fuente de peligro [881], y quienes, finalmente, consideraron la creación de una nueva fuente de esta peculiar posición [882]. Como bien es sabido, además de la existencia de la posición de garante, en virtud del artículo 11 del Código penal,

---

[880] Esta era, precisamente, la principal dificultad en el caso del *spray* de cuero pues, durante muchos años, se había comercializado de una forma completamente acorde con la ley. No así en los casos de la talidomida (Contergan) o el aceite de colza donde, desde un inicio, la conducta puede considerarse contraria a derecho, por lo que permite fundamentar, sin demasiados problemas, una imputación en su modalidad activa. Así, Aires de Sousa, S., *Responsabilidade...*, *op cit*. N.628, p. 222. En profundidad sobre el caso del Lederspray: Gimbernat Ordeig, E., Ensayos penales…, *op. cit.* N. 679, p. 304. También: Juanatey Dorado, C., «Responsabilidad penal omisiva del fabricante o productor por los daños a la salud derivados de productos introducidos correctamente en el mercado», *ADPCP*, vol. LVII, 2004, p. 54, nombrando también el caso Degussa S. A. (un material usado para empastes dentales en Alemania).

[881] Así, destacadamente en Alemania, Hilgendorf, que afirmó que los productores asumen un deber de garante por control de una fuente de peligro que persiste aunque en el momento de la comercialización dichos peligros no sean cognoscibles, cuyo fundamento encuentra en el deber de seguridad del tráfico que se afirma respecto de todos quienes poseen industrias peligrosas, lamentándose de que el BGH desperdiciara la oportunidad de declarar abiertamente esta fuente de garantía desde la perspectiva de la responsabilidad por una fuente mayor de peligro (*eine Quelle gesteigerter Gefahr*) por parte del operador de instalaciones peligrosas. Hilgendorf, E., *Strafrechtliche Produzentenhaftung..., op. cit*. N. 671, p. 142.

[882] Las ideas más «innovadoras» o propias de este topos de problemas son las que tienen que ver con la idea de que el productor está en una posición privilegiada para poner fin a las consecuencias dañinas derivadas de sus productos, por ser quien está en mejor posición, pues su dominio sobre la cosa genera, además, confianza en terceros. En este sentido, exponente de estas ideas, Juanatey Dorado, afirmando que el productor es quien mejor conocimiento tiene para poder prevenir el riesgo, con mejor información y con la mejor posición. Dicho dominio que el productor aún mantiene tras la comercialización es el que fundamenta su deber de garante. Juanatey Dorado, C. *op. cit*. N. 880, p. 72. Algo más restrictivo es Demetrio Crespo, que rechaza convertir esa posición privilegiada del empresario en un criterio autónomo y suficiente, porque conduciría a una responsabilidad casi vicaria del empresario. Para este autor, el deber de garante del directivo solo se justifica cuando existe dominio efectivo sobre el fundamento del resultado, esto es, sobre una fuente de peligro típicamente empresarial que se integra en su ámbito organizativo y supone la creación o el control de un riesgo no permitido. Así, en: Demetrio Crespo, E., «Fundamento de la responsabilidad en comisión por omisión de los directivos de las empresas», *Revista Penal México*, núm., 3, 2012.

se exige, más allá del estatus de garante, una equivalencia entre la no evitación del resultado y la causación del mismo [883].

Muchos de los autores que se han ocupado de esta concreta problemática se han remitido expresamente a la normativa civil, aceptando, abiertamente, (como nos parece, adelantamos, más sensato), que el surgimiento de estos deberes encuentra un fundamento extrapenal [884]. Los deberes de vigilancia y retirada del producto constituyen fuentes de obligación emanadas del ordenamiento civil. Ordenamiento que contempla, asimismo, que ciertos riesgos, (los que llama riesgos de desarrollo), deben quedar excluidos del régimen ordinario de responsabilidad. Dichos riesgos, que tienen un importante papel cuando hablamos de la comisión por omisión del fabricante, son directamente tomados de la regulación civil para darles un contenido propiamente penal, afirmando los autores (ya sea con base en una u otra posición de garante) que no pueden eximir en ningún caso de responsabilidad [885]. Se produce, así, una curiosa situación: allá donde el orde-

[883] Dentro de las posturas que buscan otra fuente de legitimación para la imputación (al margen de la posición de garantía e incluso de la estructura del delito omisivo), se debe destacar la de Jakobs, que fundamenta la imputación en la equiparación entre deberes positivos y negativos derivados de ciertas posiciones especiales, convirtiendo a las conductas (donde ya da igual si hay acción u omisión, pues se equiparan en su significado), en delitos de infracción de deber. Dentro de los deberes negativos se encuentran los deberes de salvamento, donde podríamos encontrar incluidas las tradicionales situaciones de injerencia. Para determinar este deber de salvamento, más que una conducta previa contraria ilícita, Jakobs considera que cuando se genera un riesgo especial, aunque sea tras una conducta lícita, nacen estas obligaciones. Así, en Aires de Sousa, S., *Responsabilidade...*, *op cit*. N.628, p. 226, y, en versión española, en: Jakobs, G., *La imputación penal de la acción y de la omisión en derecho*, Universidad Externado de Colombia, 1996, *passim.*

[884] El propio tribunal de instancia en el caso Lederspray así lo reconoció, fundando la condena por una omisión impropia en el deber jurídico privado del productor, apoyándose, de forma directa, en los principios desarrollados por la jurisprudencia civil del Tribunal Supremo alemán. Aires de Sousa, S., *Responsabilidade...*, *op cit*. N.628, p. 223 y Hilgendorf, E., *Strafrechtliche Produzentenhaftung..., op. cit*. N. 671, p. 135. La sentencia del BGH, que resuelve el caso de forma definitiva, da un paso más, reconstruyendo la ampliación del concepto de injerencia hasta entonces dominante, permitiendo fundamentar la acción previa como contraria a deber, pero cuya constatación se produce *ex post*. Dicha interpretación fue objeto de críticas dentro de la doctrina (así, por ejemplo, Kuhlen, L., «Strafhaftung beim unterlassenem Rückruf gesundheitsgefährdender Produkte», *NStZ,* 1990, pp. 566-570), considerando que se desnaturaliza el concepto de injerencia, pero sin llegar a mostrarse en desacuerdo con la decisión final.

[885] Esta traslación de los deberes civiles directamente al derecho penal para fundamentar la posición de garantía no ha estado exenta de críticas. Así, el propio Schünemann expresó reservas sobre la argumentación del BGH en el caso *Lederspray*, criticando el uso abusivo de la figura de la injerencia para fundamentar la responsabilidad jurídico-penal por omisión. En su opinión, la posición de garante en sentido jurídico-penal, distinta de las reglas civiles, solo determina la advertencia a los consumidores. Schünemann, B., «Cuestiones básicas de dogmática jurídico-penal y de política criminal acerca de la criminalidad de empresa», *ADPCP,* vol. XLI, 1988, p. 538 y ss. Sin embargo, para este autor la decisión final sí es materialmente correcta, pero solo entendida la posición de garante dentro de la estructura empresarial y del ámbito de dominio del dirigente o superior jerárquico, separando la posición de garante de los respectivos ámbitos de poder asumidos. *Idem.* Por supuesto, los trabajos que se han dedicado a la problemática propia de la omisión en la estructura empresarial son inabarcables. A título de ejemplo: Peñaranda Ramos, E., «Sobre la responsabilidad en comisión por omisión respecto de hechos cometidos en la empresa (y en otras organizaciones)», en AA. VV. *Liber amicorum en homenaje al profesor Antonio González-*

namiento civil reconoce la dificultad de imputar, aceptando (implícitamente) que implicaría un exceso hacer cargar con estos sorpresivos daños al productor, (y solo reconociendo que debe afirmarse la responsabilidad puramente objetiva en casos muy concretos), el derecho penal se empeña en castigar [886], pero haciendo cambiar de naturaleza la imputación sin que, paralelamente, al menos a nuestro entender, se exija ningún requisito adicional.

En esencia, en aquellas áreas vitales donde el derecho civil, por los bienes en juego, considera que debe existir responsabilidad sin importar cuál sea la conducta del productor (son, precisamente, los daños generados por medicamentos y productos alimentarios los excluidos del régimen de exoneración general de los riesgos de desarrollo), el derecho penal afirma exactamente lo mismo, pero lo que era abiertamente responsabilidad objetiva en el derecho privado, se torna curiosamente en responsabilidad subjetiva para el derecho público.

Ambos sistemas parten de la obligación del productor de estar atento a los posibles riesgos de desarrollo (aquellos no conocidos en el momento en que el producto se lanza al mercado), pero no queda muy claro qué se debe hacer cuando, a pesar de esa vigilancia, el defecto se da [887]. ¿Hemos de presumir que, de haberse generado el daño, la labor de vigilancia no fue lo suficien-

---

*Cuéllar García*, Colex, 2006, pp. 411-430; Lascuraín Sánchez, J. A., «Fundamento y límites del deber de garantía del empresario» en *Hacia un derecho penal económico europeo. Jornadas en honor al profesor Klaus Tiedemann,* BOE, Madrid, 1995, pp. 209-227; Demetrio Crespo, E., *Responsabilidad penal por omisión del empresario,* Centro Mexicano de Estudios en lo Penal Tributario, 2017.

[886] En este sentido, reconoce Contreras Chaimovic que, cuando se trata de los riesgos de desarrollo, la legislación civil chilena (lo mismo puede predicarse de la española), excluye la responsabilidad civil, pues «*no es razonable que el fabricante esté obligado a impedir aquellos cursos peligrosos para bienes jurídicos de los consumidores que no puedan reconocerse desde una perspectiva ex ante»,* por lo que, seguidamente afirma: «*Los riesgos de desarrollo están tolerados por el orden jurídico pre-penal y, en consecuencia, no pueden fundamentar una responsabilidad criminal, en atención al carácter siempre accesorio del Derecho Penal»* y que *«el Derecho Penal sólo puede castigar conductas desaprobadas por el orden normativo primario. Por consiguiente, las conductas del fabricante que están toleradas por el Derecho Civil o el Derecho Administrativo no pueden estimarse nunca como penalmente relevantes».* Contreras Chaimovic, L., *op. cit*. N. 683, p. 279 y nota 50. Sin embargo, aun reconociendo esto, Contreras no tiene problema en reconocer que ello no impide que en un momento posterior no surjan otros deberes especiales (deberes de vigilancia), que permitan fundamentar la responsabilidad penal. No se entiende muy bien, en todo caso, la compatibilidad de ambas afirmaciones; desde luego que el legislador civil está pensando en que la exoneración por riesgos de desarrollo aplique en un momento posterior a la comercialización, y no durante la misma, porque queda claro que ningún riesgo de desarrollo se producirá en fase temprana, por su propia naturaleza. En un sentido similar, Silva Sánchez, J. M., *El riesgo permitido en el derecho penal económico*, Atelier, Barcelona, 2022, p. 128.

[887] En este sentido, muy crítico con la afirmación de la posición de garante en el caso *Lederspray*, Hassemer, poniendo de manifiesto que se produce una vulneración de las exigencias de certeza del principio de legalidad, en la medida en que es indeterminado el momento en el cual recae sobre el productor el deber de retirada, llegando a afirmar que mantener en todo caso la posición de garante es lo mismo que decir que aquellos que ejercen determinadas profesiones naturalmente peligrosas pueden estar permanentemente sometidos a responsabilidad penal. Hassemer, W./Muñoz Conde, F. *La responsabilidad por el producto…, op. cit* N. 409*,* p. 168.

temente óptima? La respuesta debe ser, evidentemente, negativa para el derecho penal, que no puede basarse en presunciones en contra del reo [888].

Aires de Sousa presenta un recorrido por los diferentes argumentos y posiciones enfrentadas tras el caso *Lederspray,* pero acaba señalando que, a pesar de que muchos autores buscaron soluciones para legitimar la condena por un delito de omisión impropio, algunos otros sugirieron la necesidad de una intervención legislativa en este ámbito, ya fuera por medio de un delito de peligro (frente a la solución del delito de resultado) [889], ya fuera por la introducción de un ilícito de riesgo estructurado como delito de omisión [890]. En este último sentido, destacadamente, Piergallini [891]. La propuesta del italiano implica abandonar directamente el paradigma del derecho penal del evento, para dar entrada al paradigma del ilícito de riesgo. De hecho, no es de extrañar que, partiendo de esta línea de pensamiento, en sus trabajos más recientes, donde se ha encargado de examinar la responsabilidad por los daños generados por IA, sus postulados sean similares [892].

Para Aires de Sousa, esta confrontación evidenció las dificultades que el aparato dogmático de la omisión enfrenta en la responsabilidad penal por el producto. No podemos más que estar de acuerdo [893].

En todo caso, en los supuestos en que la doctrina ha reconocido eficacia penal a los deberes de vigilancia y retirada, se ha destacado que su afirmación no debe realizarse de forma genérica, sino que hay que atender a cómo de gravosa resulta cada conducta para el productor frente a cuánta protección se puede ofrecer al consumidor, dando entrada aquí a un juicio de proporcionalidad [894].

---

[888] Considerando estos deberes de aviso y retirada del mercado como una excepción al principio de lesividad, y aplicación directa del principio de precaución, que tilda de «peculiar solución» para el derecho penal: Silva Sánchez, J. M., *Derecho Penal. Parte General, op cit*. N. 521, 48/10.

[889] Sobre esta propuesta de incriminación, sin ir más lejos, se hace eco Gimbernat, siendo consciente de la polémica sobre la fundamentación de la posición de garante en el caso *Lederspray*, aceptando la opción de que el legislador tipifique una omisión impropia, que no es más que un particular delito de peligro donde se impongan deberes de acción que pusieran directamente fin a la comercialización, pues, en palabras de Gimbernat, «*una imposición de un deber jurídicopenal de estas características puede ser considerado, en nuestra sociedad química de riesgos, como plenamente adecuado y positivo*», dando así cabida a la propuesta de autores como Freund, que propugnaba un delito de peligro para la vida o salud como consecuencia de productos, siempre con la salvedad de que existiese una fuerte sospecha de que es apropiado para lesionar la vida o salud de otros. Gimbernat Ordeig, E., *Ensayos penales..., op. cit.* N. 679, p. 324-325.

[890] Destacadamente, Piergallini.

[891] Piergallini, C., *Danno da Prodotto..., op. cit.* N. 732*,* p. 467 y ss. y en particular p. 534.

[892] Así, en su reciente trabajo: Piergallini, C., «Intelligenza Artificiale: ¿De "mezzo" ad "autore"…, *op. cit.* N. 654, p. 1773, mostrándose en este caso partidario de un uso aún más prudente del derecho penal, habida cuenta de los problemas de imputación (principalmente de imputación subjetiva), siendo partidario de un derecho basado en un modelo de «*cooperative compliance*», con importancia de la coregulación.

[893] Aires de Sousa, S., *Responsabilidade...*, *op cit*. N.628, p. 232.

[894] Así, Contreras Chaimovic, L., *op. cit*. N. 683.

Bertolesi afirma, por su parte, que se debe diferenciar este primer tipo de obligaciones de las surgidas después de la introducción del producto en el mercado, donde surgen deberes de vigilancia, de información, de advertencia y de retirada [895], afirmando que estos últimos deberes tienen un contenido elástico, es decir, no está predeterminado cuál es su contenido ni presupuestos.

Llegados a este punto, más allá de señalar la polémica sobre el alcance de los deberes de retirada, y la negativa que aquí se sostiene a aceptar que sean suficientes para justificar una imputación del resultado final, es oportuno señalar que aceptarlo sería aceptar una vertiente fuerte del principio de precaución (en el sentido dogmático que ha sido rechazado), en cuanto directamente permitiría imputar un daño con base en un conocimiento incierto sobre sus efectos. No estamos ante una fuente que aporte suficiente seguridad jurídica (no se ha producido la previa codificación o cognoscibilidad del principio de actuación), y, si ya ha habido polémica en casos anteriores, en este puede afirmarse que debe rechazarse *prima facie* esta forma de imputación como vía preferente.

Otro de los problemas puestos de manifiesto precisamente en los casos de responsabilidad derivada de los riesgos de desarrollo (defectos surgidos con posterioridad a la comercialización permitida), tiene que ver con la inseguridad que se deriva respecto de la actitud exigida por el derecho (penal) al productor. Esta inseguridad, unida a la «elasticidad» de los deberes de vigilancia, advertencia y retirada, es puesta de manifiesto por Bertolesi [896]. El italiano, tomando como referencia la obra de Piergallini [897], señala que en lo relativo a la situación objetiva que actúe como presupuesto para el deber de actuación del productor hay que tener en cuenta elementos como: indicios como la frecuencia y difusión de los efectos lesivos, el perfil espacial y temporal de los mismos, la gravedad y homogeneidad del daño, y la posible existencia de hipótesis alternativas [898]. Pero también señala que, una vez establecido el presupuesto objetivo, aún queda pendiente la segunda cuestión: es decir, determinar la actuación debida. Aunque reconoce que ya la doctrina ha propuesto criterios que pasan por, en la medida de lo posible, recuperar el nivel de seguridad al cual aspiraba la regla de cuidado que luego resultó ser fallida [899], tales como el principio de proporcionalidad (Chaimovic, como acaba de verse), que exigirá el deber de retirada como última de las soluciones en caso de que no haya otra

---

[895] Bertolesi, R., *op cit*., N. 423, p. 192. Son los supuestos que la doctrina civilista califica de «riesgos de desarrollo *(rischio di sviluppo).*

[896] *Ibid.,* p. 193: *«Questi obblighi si caratterizzano per avere un contenuto elastico, non determinando precisamente quali siano i presupposti e il contenuto del dovere di (re)agire».*

[897] Piergallini, C., *Danno da Prodotto..., op. cit.* N. 732.

[898] Bertolesi, R., *op cit*., N. 423, p. 196.

[899] *Ibid.* p. 197

menos gravosa e igual de eficaz, o un análisis coste-beneficio (en este caso, Piergallini [900]), al autor le parece preferible una tercera opción. Cree Bertolesi que lo más óptimo en las situaciones de daños generados por sistemas de IA, son lo que califica de «mecanismos de cogestión del riesgo», es decir, el establecimiento de mecanismos cautelares muy precisos, a decidir entre la parte privada y los decisores administrativos, que incluyan autorizaciones o licencias, con el objetivo de descargar al sujeto privado de que cargue con todas las consecuencias derivadas de la incerteza [901].

Es necesario destacar que, en lo relativo a la forma de aparición de los defectos, una vez más debemos mostrar un punto de discordancia, pues, aunque Bertolesi señala con acierto que el deber de actuación del productor (ya sea en forma de deber de información o retirada) se habrá de colegir de ciertos indicios, tampoco creemos que los mismos indicios que normalmente eran indicativos de un defecto entendido en un sentido clásico funcionen de igual forma con los sistemas de inteligencia artificial. Precisamente por la lógica de producto estático frente a la del producto dinámico arriba explicadas, los defectos propios de los productos estáticos serán, en principio, defectos en serie, y darán lugar, ante las mismas condiciones, a similares defectos (así ha ocurrido, como puede comprobarse, en los casos clave de la discusión: una misma sustancia tóxica da lugar a varias intoxicaciones similares, o un defecto en la fase de producción da lugar a fallos en serie, que se repiten una y otra vez en los diferentes modelos). La lógica de la intoxicación masiva o el fallo masivo no es tan fácilmente trasladable al ámbito de la IA pues, por su dinamismo e interacción con el entorno, los errores que produzca en unos casos no tienen por qué reproducirse en otros. Por poner solo algunos ejemplos: el error en el diagnóstico incorrecto puede solucionarse tras un *feedback* negativo (el sistema aprende tras su intento y consiguiente corrección), o bien puede deberse a la introducción de los propios sesgos o deseos del propio usuario (así, si un bot de IA interactúa con un usuario racista se acabará comportando como si fuera racista, mientras que si interactúa con un usuario ecologista se acabará comportando como ecologista [902]), por lo que los errores serán más bien «indivi-

[900] Piergallini, C., *Danno da Prodotto..., op. cit.* N. 732, pp. 537-538, y 544-546. Adviértase que, como arriba se ha indicado la propuesta de Piergallini no lo es con intención de aplicar el delito de resultado, sino un eventual delito de peligro.

[901] Bertolesi, R., *op cit*., N. 423, p. 200. Nos recuerda en este punto la propuesta de Bertolesi a la que desarrolla Piergallini, cuando habla de que el mecanismo más idóneo para los daños generados por IA sea el propio de una co-gestion, un *compliance cooperativo*. *Op. cit*. nota 892.

[902] Nos referimos con este ejemplo al conocido caso del bot Tay, chat conversacional que fue lanzada por Microsoft al entorno de la red social Twitter (hoy X), y que acabó siendo bloqueada y retirada tras unas horas de interacción por la publicación de *twitts* de contenido racista y misógino. Para ver el conte-

dualizados», y no tanto en serie, dificultando encontrar patrones o, en su caso, indicios fuertes que aconsejen su retirada (¿habrá que retirar todos los *bots* conversacionales del ejemplo, o solo aquellos que hayan interactuado con un entorno racista?). La inseguridad jurídica que se genera en estos entornos cambiantes es evidente.

No se trata, en este caso, de que según el estado de la ciencia estos riesgos no fueran previsibles, pues la autonomía de los sistemas (derivada de sus capacidades de adaptación tras el análisis del entorno), lleva implícita esta característica. La respuesta, claro está, puede pasar por obligar a las empresas productoras a actualizar de manera constante (tras cada uno de los posibles sesgos, errores, alucinaciones o *shortcuts* [903]), la programación, empleando los mejores datos y las mejores técnicas [904]. Algunas técnicas, sin embargo, no estarán exentas de estos riesgos, y nunca lo estarán por sus propias características.

La pregunta que nos debemos hacer es la relativa a qué peso queremos poner a los productores. En este sentido, y de forma similar a lo propuesto por Bertolesi, Íñigo realiza una reflexión valiosa respecto del papel de la Administración en la determinación de la conducta exigida al fabricante. En los sectores ampliamente regulados la posición de estos sujetos está muy bien delimitada en la normativa (ya sea mediante órdenes, decretos, o leyes), que, en definitiva, sí orientan de forma clara sobre cuáles son las actuaciones que han de llevar a cabo los productores [905]. Habida cuenta de que estos sectores indus-

---

nido de las publicaciones: HUNT, E., «Tay, Microsoft's Al Chatbot, Gets a Crash Course in Racism from Twitter», *The Guardian,* 24 de marzo de 2016, disponible en: <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter> [Accedido por última vez el 27/01/2026].

[903] Una alucinación, en el contexto de un modelo fundacional, se refiere a una situación en la que el modelo genera contenido que no está basado en información factual o precisa. La alucinación puede ocurrir cuando el modelo produce texto que incluye detalles, hechos o afirmaciones que son ficticios, engañosos o completamente inventados, en lugar de proporcionar información confiable y veraz. RAWTE, V./ SHETH, A., / DAS, A., «A survey of hallucination in large foundation models», *arXiv preprint arXiv:2309.05922*, 2023, p. 1 (traducción propia). Aunque las causas principales de alucinación incluyen problemas de recolección de datos y distorsiones en los datos que pueden surgir en el proceso de entrenamiento y en la elección de los modelos (JI, Z. ET AL., «Survey of Hallucination in Natural Language Generation», 55 *ACM Computing Surveys*, 2022, pp 1-59) existen estudios recientes que están estudiando formas de detectar cuando un LLM (*Large Language Model*) alucina por confabulación, siendo los sistemas de cálculo de entropía semántica los más representativos para avanzar en la fiabilidad de los modelos lingüísticos grandes. Así, FARQUHAR *et al.,* «Detecting hallucinations in large language models using semantic entropy», *Nature*, 630 (8017), 2024, pp. 625-630.

[904] En este sentido, por ejemplo, VALERIUS, B., «Sorgfaltspflichten beim autonomen Fahren*», op. cit.* N. 855*,* p. 13, estableciendo, en el campo de la conducción autónoma, que solo se puede hablar de infracciones del deber de diligencia si la programación correspondiente o las medidas de protección ya no correspondían al estado actual de la ciencia y la técnica en el momento de su implementación.

[905] ÍÑIGO CORROZA, E., *op. cit.* N 697, p. 178, señalando que, en materia de responsabilidad por productos defectuosos *«El perfil que del fabricante o distribuidor muestra la normativa administrativa es un claro reflejo de lo que se espera socialmente del productor o del vendedor».*

triales están fuertemente regulados (por la Administración y por otras instituciones, en esta lógica de la gestión compartida del riesgo), surgirá un problema cuando el fabricante sepa más que la Administración, pero la idea defendida por Íñigo es que si es la Administración quien asume la responsabilidad de regular las normas de comportamiento de un determinado sector, no se debería exigir más a las empresas, pues *«Si la Administración posee la facultad de determinar cuál es el nivel científico y tecnológico que debe exigirse a los distintos sectores de elaboración de productos […] entre otras cosas, para controlar el mercado, debe, asimismo, cargar con los posibles defectos que de una regulación deficiente de estas actividades puedan derivar»* [906], actuando en este caso la Administración como árbitro o moderador de mercado, de forma que el coste de un sistema intervencionista, será que no sea el fabricante quien corra con los riesgos [907].

De cualquier modo, y siendo consciente Íñigo de que la imputación de los resultados por omisión no se debe excluir en todo caso (cosa que tampoco se niega aquí), pone de manifiesto que lo verdaderamente importante es, más allá de afirmar posiciones de garantía de alto contenido formal, buscar el fundamento material que permita afirmar la equivalencia requerida por nuestro Código [908]. En este sentido, la mejor doctrina [909] señala que lo difícil no será afirmar que exista una posición de garante en muchos de los casos de responsabilidad penal por el producto, sino que habrá que comprobar si, en cada uno de los casos con-

---

[906] *Ibid.,* p.179.

[907] La autora propone esta idea para justificar la exclusión de los conocimientos especiales de los productores, pero nos parece que tiene plena aplicación si se pone en relación con la idea del sistema compartido de gestión del riesgo señalado por Bertolesi y Piergallini. Sin embargo, Íñigo no reconoce la solución que Piergallini propone en uno de los casos de responsabilidad por el producto célebres en Italia (el caso del paracaidista), y, aunque en lo relativo a los riesgos de desarrollo el derecho penal no puede intervenir allí donde se ha cumplido la norma de conducta que regía en el momento de fabricación y/o comercialización, sí afirma el deber de retirada. *(ibid.,* p. 237). Yendo un paso más allá, Rodríguez Montañés considera que, de hecho, este deber de retirada constituye, en todo caso, un deber civil, y así lo expone y justifica en su artículo dedicado al análisis de los casos sobre responsabilidad por el producto. Rodríguez Montañés, T., «Incidencia dogmática…», *op. cit*. N. 669, p. 130: *«Tal afirmación (*se refiere a la posibilidad de hacer responder al productor por eventuales eventos lesivos en comisión por omisión*) de carácter genérico lleva a una enorme ampliación del ámbito de la posición de garantía y me parece insostenible desde los postulados del Derecho penal (otra cosa es la responsabilidad por el riesgo que obliga a reparar el daño en el Derecho civil), por violar principios tan elementales como el de certeza y seguridad jurídica, pues, de este modo, podemos ir creando nuevas fuentes de garantía en todos aquellos ámbitos donde detectemos dificultades de punición con arreglo a los criterios de imputación tradicionales.»*

[908] *Ibid.,* p. 131.

[909] Cita Rodríguez Montañés en este punto a Luzón, Silva o Gimbernat. Así, en: Luzón Peña, D. M., *Estudios penales,* PPU, Barcelona, 1991*;* Silva Sánchez, J. M.*, El delito de omisión: concepto y sistema,* Bosch, Barcelona, 1986*;* Gimbernat Ordeig, E., «Omisión propia e incremento del riesgo en el Derecho penal de empresa», *ADPCP,* LIV, 2001*,* pp. 5-112.

cretos es posible dominar ese peligro, *«teniendo en cuenta las competencias del sujeto, sus posibilidades de actuación y el significado material que tiene la omisión de su deber de actuación»,* sin que esto pueda hacerse de una forma automática, basándonos simplemente en la posición de garantía [910].

Precisamente sobre las dificultades que en los contextos de IA se generan para afirmar, como señala Íñigo, la efectiva posibilidad de actuación efectiva (foco, como hemos visto, de los problemas que en última instancia surgen para la imputación del delito imprudente), que pudieran colocar al productor de sistemas inteligentes en la necesidad de revisar constantemente cada uno de sus sistemas, hacen difícil afirmar que la imputación a título de comisión por omisión de sus eventuales daños pueda ser la forma más óptima de atajarlos y, sobre todo, de prevenirlos. No negamos, en la línea seguida por Íñigo, que la imputación del resultado lesivo se produzca también en supuestos de clara omisión respecto de los sistemas de IA, pero solo se dará en los casos en que quepa apreciar un cumplimiento escrupuloso de los requisitos genéricos para la imputación [911].

---

[910] Rodríguez Montañés, T., «Incidencia dogmática…», *op. cit*. N. 669, p. 131. En un sentido similar, Lohmann, que, tras examinar los deberes civilmente impuestos al fabricante de productos (esencialmente, los derivados de defectos de diseño, defectos de fabricación, errores en las instrucciones y falta de vigilancia sobre el producto), afirma que aunque pueden transferirse al derecho penal en cuanto a su contenido, no deben hacerlo sin reservas, sino adaptándose a los fines perseguidos en cada ámbito, lo que obliga, en derecho penal, a examinar detenidamente si hay, caso por caso, un comportamiento contrario a los deberes del fabricante, contrario también a los requisitos generales de diligencia del derecho penal. Lohmann, A., *op. cit. N. 298*, p. 176.

[911] Así, por plantear un caso donde sería perfectamente posible la imputación del resultado lesivo en comisión por omisión, imagínese un contexto en el que el proveedor de un sistema de armamento autónomo (arma autónoma letal) conserve un control en tiempo real sobre la actuación de dicha arma, y, ante una situación inminente de error en su actuación, omita accionar el botón de apagado. Lo mismo cabe predicarse si similares errores de «diagnóstico» o «identificación» se producen una y otra vez y, tras ser advertidos, no causan ningún tipo de actuación por parte del fabricante. Siendo consciente de la problemática que introducen este tipo de tecnologías en el campo armamentístico, la incapacidad de control en los casos de técnicas como el *deep learning*, y la imposibilidad de aplicar los delitos clásicos según la aplicación estricta de los principios dogmáticos del derecho penal, Gil se muestra partidaria de su prohibición, o (en caso de que no prospere) de la creación de delitos de peligro abstracto. No podemos estar más de acuerdo con las conclusiones de este trabajo. Gil Gil., A., *op. cit*. N. 242.

# CAPÍTULO SEXTO. VIEJAS POLÉMICAS, NUEVOS RIESGOS: LOS BIENES JURÍDICOS AMENAZADOS Y LA TÉCNICA DEL PELIGRO ABSTRACTO

> *Sentimos una alegría especial al observar que nuestros productos que contenían radio concentrado se volvían espontáneamente luminosos. Mi marido, que esperaba ver hermosas coloraciones, tuvo que estar de acuerdo en que esta otra característica inesperada le dio aún más satisfacción… [estos productos] fueron dispuestos en mesas y tableros [en el laboratorio]: por todas partes podíamos ver siluetas ligeramente luminosas y ese brillo, que parecía suspendido en la penumbra, despertó en nosotros nuevas emociones y encantamiento*[912].

Corría el final del siglo XIX cuando la aun joven investigadora polaca Maria Skłodowska-Curie, encandilada por el brillo de aquel fascinante producto desconocido, su querido radio, se pasaba los días encerrada en un laboratorio con su marido Pierre, en los albores de un descubrimiento que lo cambiaría todo. Narra en una de sus biografías Rosa Montero la fascinación del matrimonio, al que tilda de inocente e irresponsable, ante esa novedosa sustancia, cuya dañosidad parecía evidente, pero a la que se exponían una y otra vez, con tal de seguir explorando su potencialidad.

Siendo conscientes de los peligros que entrañaba su manipulación, y de los riesgos a los que estaban expuestos los trabajadores del laboratorio, ellos mismos y sus colegas, los experimentos no cesaban. Anomalías en la sangre, quemaduras en la piel, un debilitamiento patente de la salud propia y de su marido, que condujeron a la muerte prematura de ambos y a la de su hija Iréne, también apasionada (y laureada) por la misma sustancia fascinante y peligrosa.

---

[912] Fragmento del diario de Marie Curie, extraído de la novela de Rosa Montero, *La ridícula idea de no volver a verte,* Seix Barral, Barcelona, 2013, p. 127.

Cuenta Montero cómo las medidas de seguridad, insuficientes para los requerimientos actuales, incluían análisis de sangre periódicos para aquellos expuestos a la radiación, pero que la joven y tozuda Marie, absurdamente empeñada en defender los beneficios de su amado elemento, escribió en respuesta a un informe que resaltaba los peligros del mismo, que, aunque era necesario advertir del riesgo, ella no era consciente *«de ningún accidente grave debido al radio o al mesotorio entre el personal de otras factorías […] ni entre el personal de mi Instituto»* [913]. Seis años más tarde, un tercio de los trabajadores del Instituto mostraba anomalías en la sangre. Narra la biógrafa las consecuencias devastadoras, pero negadas, de este nuevo descubrimiento, al que tilda de criminal seductor silencioso.

Hubieron de pasar unos años para que la sociedad en su conjunto fuera consciente de que aquella nueva y fascinante sustancia, a la que innegablemente habría que asociar beneficios sociales incuestionables, requería ser tratada como lo que era: un arma peligrosa y susceptible de causar perjuicios que bien valían su inclusión en las legislaciones penales de la época. Los comienzos de la técnica son, inevitablemente, confusos. La fascinación y los riesgos se mezclan, y los argumentos de los detractores y defensores se contraponen en el siguiente paso inevitable: el de su regulación legal. Legislar sobre lo nuevo no es, pues, una opción, sino una necesidad inherente al desarrollo.

## I. LA PROBLEMÁTICA RELECTURA DE LOS BIENES JURÍDICOS COLECTIVOS

### 1. ¿SIGUE SIRVIENDO EL BIEN JURÍDICO COMO APARATO CONCEPTUAL?

A pesar de las críticas que la teoría del bien jurídico ha recibido en los últimos tiempos, parece inevitable que en un intento de estudiar cuál es la forma de tipificar un delito de peligro adecuado a las necesidades de la sociedad del riesgo tecnológico, debamos enfrentarnos, siquiera brevemente, a la polémica sobre qué bienes jurídicos se ven afectados con la irrupción de esta nueva tecnología, que, si bien continúa con los postulados del riesgo, tal y como se ha tratado de poner de manifiesto *supra* en el capítulo cuarto, los acaba, de algún modo, superando, para dar lugar a una forma de ataque particular, y que ha reproducido, una vez más, las preguntas que se venían haciendo entonces:

[913] Montero, R., *ibid.,* p. 122.

¿son suficientes los bienes jurídicos tradicionales para afrontar los desafíos penales de la IA?

A contestar esta pregunta se dedicará la presente sección, para, seguidamente, proceder a determinar cómo los protegemos.

Como se adelantaba, somos conscientes de que la doctrina del bien jurídico ha encontrado críticas autorizadas, y, aunque las esbozaremos brevemente, adelantamos ya que nuestro punto de partida será asumir la utilidad dogmática de esta categoría, si no como exclusiva fundamentación de la punición penal, sí como aparato conceptual que sirve a la interpretación teleológica de los tipos.

Ello no obsta, empero, a conceder un punto de razón a quienes han explorado otros modelos, que se apuntaban ya al inicio de la investigación, tales como el anglosajón principio del daño *(harm principle),* como forma de reconstruir una teoría del bien jurídico que se percibe como superada en una actualidad, donde la lesividad (y el daño) se perciben en un contexto donde apariencia y realidad se entremezclan [914]. La crítica ha sido acogida dentro y fuera de nuestras fronteras [915], pero, en lo que nos interesa, y como Miró ha determinado recientemente entre nosotros, quizá la crisis del principio del bien jurídico como único y suficiente principio legitimador del castigo penal no sea tan difícil de compartir [916]. Así lo reconocen también los autores que siguen utilizando este aparato conceptual, por considerar que sigue siendo el punto de partida aun mayoritario. Coincidimos, pues, plenamente, con la precisión que realiza Maldonado al respecto, afirmando que cuando el legislador opta por proteger un bien jurídico lo hace partiendo de una doble consideración: primero, la del merecimiento, es decir, el interés o trascendencia de dicho interés para posibilitar la convivencia social; y, segundo, la necesidad, esto es, la cons-

---

[914] Agustina, J. R., *op. cit*. N.62.

[915] Así, por nombrar una de las voces más relevantes que, desde Alemania, ha explorado y puesto en cuestión, comparándolos, la teoría del bien jurídico y el principio del daño anglosajón, Hörnle, T., «La protección de sentimientos en el StGB», en Hefendehl, R./Von Hirsch, A./ Wohlers, W. (Eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?,* ed. en español a cargo de Rafael Alcácer Guirao, María Martín Lorenzo e Íñigo Ortiz de Urbina, Marcial Pons, 2016, pp. 375-392. Una referencia más completa a su bibliografía reciente en inglés y alemán, en: Miró-Llinares, F., «La criminalización de conductas «ofensivas». A propósito del debate anglosajón sobre los «límites morales» del Derecho penal», *RECPC*, 17-23, 2015, pp. 1-65, nota al pie 15.

[916] Véase, a estos efectos, el completo análisis que realiza Miró sobre las formas de entender las versiones más y menos exigentes de las teorías anglosajonas basadas en el *harm principle* y, especialmente, en el *offence principle*, donde se dibujan con acierto las limitaciones de la teoría del bien jurídico en cuanto fundamentación suficiente y necesaria de la criminalización penal (en especial cuando se trata de conductas que son socialmente cuestionables). Miró-Llinares, F., *ibid*. También, el mismo, en: «Lo que queda (y debe quedar) del bien jurídico en Derecho penal», en González Cussac, J. L./ León Alapont, J. (eds.), *Estudios jurídicos en memoria de la Profesora Doctora Elena Górriz Royo*, Tirant lo Blanch, Valencia, 2020, pp. 599-620.

tatación de la insuficiencia de otros medios de control alternativos [917]. En definitiva, el bien jurídico solo opera como condición necesaria de la protección, pero no suficiente, pues el derecho penal requiere de la presencia de principios como el de subsidiariedad [918].

Una aproximación parecida realiza Silva, que, si bien reconoce los límites de esta teoría (admitiendo que es incluso difícil afirmar que exista una única teoría del bien jurídico, habida cuenta de las variantes sobre la misma) [919], hace hincapié en su falta de capacidad crítica, y de su incapacidad para contener la expansión de la criminalización del derecho penal [920], señalando que incluso hasta alguno de los padres de esta teoría han acabado por aceptar sus limitaciones [921].

Nos parece más interesante, empero, centrarnos entonces en la vertiente de este aparato conceptual que aún resulta irrenunciable, que no es otra que su orientación teleológica. En palabras de Silva, constituye *«un criterio rector de la categoría sistemática de la tipicidad como ámbito de significado y valoración»* [922], además de un elemento importante para la interpretación te-

---

[917] Maldonado Fuentes, F., «Reflexiones sobre las técnicas de tipificación de los llamados «delitos de peligro» en el moderno derecho penal», *Revista de Estudios de la Justicia*, 7, 2006, p. 27. También aquí muy destacadamente Álamo Alonso, frente a la crítica de Frisch, aceptando que el hecho de que haya otros criterios importantes no ha de negar la importancia al bien jurídico. Alonso Álamo, M., «Fundamentación pre-positiva de los bienes jurídico-penales y derecho penal mínimo de los derechos humanos», *RGDP*, 15, 2021, p. 13.

[918] Una visión, ligeramente diferente, por su conexión directa con los valores derivados de la Constitución, pero igual en cuanto a la insuficiencia de la teoría del bien jurídico para explicar lo legítimamente protegible en: Lascuraín Sánchez, J. A., «Bien jurídico y objeto protegible», *ADPCP*, vol. LX, 2007, pp. 119-163. Similar, pero defendiendo la autonomía conceptual e importancia del bien jurídico, Álamo Alonso, M., *ibid*., p. 13: «*bien jurídico es el primer filtro, necesario aunque no suficiente, del que depende el sí de la intervención. Nada más, pero también nada menos»*. Mendoza ponía de manifiesto ya en su monografía sobre sociedad del riesgo que el concepto de bien jurídico, que tradicionalmente había operado en un sentido restrictivo, estaba perdiendo, con la creación incontrolada de bienes nuevos, su capacidad limitadora, usándose, ahora, como criterio de agravación en lugar de, como corresponde, criterio de limitación. Mendoza Buergo, B., *El derecho penal en la sociedad del riesgo*, *op. cit.* N. 384, p. 77.

[919] Silva Sánchez, J. M., *Derecho Penal. Parte General, op cit*. N. 521, 45/2.

[920] *Ibid.,* 53/2, poniendo de manifiesto cómo una de estas manifestaciones es, precisamente, la creación de bienes jurídicos intermedios, cuya lesión, para el autor, no es más que una vulneración de un orden de policía administrativa, legitimando esta expansión que confirma al derecho penal como un mecanismo de «función promocional».

[921] Así, el propio Roxin, que viene reconociendo, a pesar de ser unos de sus padres intelectuales, que la teoría del bien jurídico necesita como complemento imprescindible del principio de subsidiariedad, teniendo ambos un rango equivalente. Roxin, C. «El concepto de bien jurídico como instrumento de crítica legislativa sometido a examen», *RECPC,* 15-01, 2013, p. 15. Ya en Roxin, C., *Derecho Penal, Parte General, Tomo I, Fundamentos. La estructura de la teoría del delito*, Civitas, Madrid, 1997, nm 28/ § 2. También Hefendehl, uno de los más modernos defensores y reconstructores de la teoría, Hefendehl, R., «¿Debe ocuparse el Derecho penal de riesgos futuros? Bienes jurídicos colectivos y delitos de peligro abstracto», *RECPC,* 04-14, 2002, *passim,* aunque en p. 7 y ss., destacadamente, partiendo de que las definiciones anteriores del concepto de bien jurídico si no son defectuosas son, al menos, insuficientes.

[922] Silva Sánchez, J. M., *Derecho Penal. Parte General, op cit*. N. 521, 56/2.

leológica de los tipos, en cuanto sirve de referencia para realizar el juicio de peligro *ex ante* y valorar la lesividad *ex post* de las conductas, para cuantificar la lesión concreta y, por tanto, para determinar la pena en función de la gravedad del desvalor de resultado [923]. Y, en esencia, lo que aquí más importa, es una herramienta válida para excluir la tipicidad de aquellas conductas que, aunque en una primera lectura puedan ser formalmente subsumibles en un tipo penal, deban quedar extramuros de la intervención, al no lesionar el bien jurídico que, en teoría, el tipo protegía [924].

Será esta función de exclusión la que tendrá más capacidad de rendimiento en nuestra propuesta, y merece la pena resaltar, por tanto, que partir de la delimitación del bien jurídico, resultará, aun a sabiendas de las críticas vertidas, útil a los efectos que aquí se persiguen.

### 1.1 **La tradicional polémica sobre el bien jurídico colectivo**

No pretende constituir este apartado nada parecido a una exposición histórica detallada sobre la evolución de las teorías del bien jurídico, no por su falta de interés, sino porque nos distraería demasiado de nuestro objetivo principal, que no es otro que examinar la posible existencia de nuevos objetos de protección.

Resulta adecuado, no obstante, situar la discusión en su contexto, para poner de manifiesto que, desde luego, la tradición de poner en duda el surgimiento de bienes objeto de protección no nace con las nuevas tecnologías, sino que se ha instituido, más bien, en un punto común. Ya en la década de los 70 comienza a plantearse una funcionalización del concepto de bien jurídico, donde empieza a admitirse que se protejan penalmente funciones del derecho, desde una lógica más alejada de la puramente valorativa, para aceptar que la dañosidad social puede fundamentar el injusto penal, si así lo decide el legislador en uso de sus prerrogativas [925]. También surgen las posturas de aquellos que consideran que el

---

[923] *Idem.*

[924] Usando, una vez más, las palabras de Silva, *«permite manejar los casos de sobreinclusión de supuestos en el tenor literal de un tipo legal». Idem.*

[925] El padre de la teoría de la dañosidad social es Amelung, en su famosa obra *Rechtsgüterschutz und Schutz der Gesellschaft. Untersuchungen zum Inhalt und zum Anwendungsbereich eines Strafrechtsprinzips auf dogmengeschichtlicher Grundlage. Zugleich ein Beitrag zur Lehre von der «Sozialschädlichkeit» des Verbrechens,* Athenäum, Frankfut am Main, 1972, pp. 330 y ss. Una actualización de la postura de este autor en: Amelung, K., «El concepto de «bien jurídico» en la teoría de la protección penal de bienes jurídicos», en Hefendehl, R./Von Hirsch, A./ Wohlers, W. (Eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?,* ed. en español a cargo de Rafael Alcácer Guirao, María Martín Lorenzo e Iñigo Ortiz de Urbina, Marcial Pons, 2016, p. 256, apun-

bien jurídico debe beber de los intereses constitucionalmente protegidos [926]. Entre las posturas más positivistas y las más iusnaturalistas [927] se ha ido moviendo el debate, que, en lo que nos interesa, ha sido especialmente encarnizado cuando se ha tratado de determinar si, frente a una interpretación extrema o formal del bien jurídico personal, cabe la protección de bienes jurídicos de carácter colectivo, pregunta que está en el ADN del derecho penal moderno.

Podemos diferenciar la existencia de dos posturas enfrentadas. La de los monistas, defensores de una concepción estrictamente individual del bien jurídico [928], y la de los dualistas, que son aquellos que aceptan la coexistencia de

---

tando a que, junto a una concepción del bien jurídico como objeto de protección que ha de inferirse del Derecho positivo, cabe una teoría sociológica, sobre la lesividad social, que entiende el daño social no como el deterioro de objetos, «*sino que retoma la vieja idea de que el delito ataca las condiciones de organización de la convivencia humana*», y que por ello se apoya en la idea de que la sanción penal es el instrumento para la protección de la vigencia fáctica de las normas. Este proceso de funcionalización acaba por convertirse en toda una postura que niega relevancia total al concepto de bien jurídico, teniendo como exponente principal a Jakobs. Un esbozo de esta teoría, en clara contestación a los postulados de Amelung, en: Jakobs, G., «¿Daño social? Anotaciones sobre un problema teórico fundamental en el Derecho Penal», *Cuadernos de Política Criminal*, 100, 2010, pp. 29-45.

[926] Paradigmáticamente: González Rus, J. J., *Bien jurídico y Constitución (Bases para una teoría)*, Fundación Juan March, Madrid, 1983. Yendo un paso más allá, separándose del constitucionalismo, pero afirmando como marco referencial los derechos humanos, fundamentalmente (aunque no solo) en: Alonso Álamo, M., *Bien jurídico penal y derecho penal mínimo de los derechos humanos*, Ediciones Universidad de Valladolid, 2014. No obstante, también existen sólidos críticos a las posturas de corte constitucionalista, considerando que la Constitución no basta como base para determinar los bienes que ha de proteger el derecho penal. En este sentido, vid.: Hefendehl, R., *Kollektive Rechtsgüter im Strafrecht*, Heymans, 2002, pp. 44 y ss., o Silva Sánchez, J. M., *Aproximación al Derecho penal contemporáneo*, Bosch, Barcelona, 1992, p. 274, reconociendo que, si bien exigir una consagración constitucional directa o indirecta de los bienes penalmente protegidos es justificada, la Constitución es, aún, *«un marco flexible y ambiguo, demasiado amplio»*. Otras posturas esencialmente constitucionalistas: Muñoz Conde, F., «Protección de bienes jurídicos como límite constitucional del Derecho penal», en AA. VV, *El nuevo Derecho penal español. Estudios penales en memoria del Profesor José Manuel Valle Muniz*, Aranzadi, Cizur Menor, 2001, pp. 562 y ss., pero insistiendo en que más importante aún que el derecho fundamental necesitado de protección es el proceso por el cual son reelaborados normativamente; Schünemann, B., «El principio de protección de bienes jurídicos como punto de fuga de los límites constitucionales de los tipos penales y de su interpretación», en Hefendehl, R./Von Hirsch, A./ Wohlers, W. (Eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?,* ed. en español a cargo de Rafael Alcácer Guirao, María Martín Lorenzo e Íñigo Ortiz de Urbina, Marcial Pons, 2016, pp. 191-220, explicando que no tiene importancia que la propia Constitución no recoja expresamente el concepto de bien jurídico. En profundidad, sobre las diferentes concepciones existentes, Alonso Álamo, M., «Bien jurídico penal: más allá del constitucionalismo de los derechos», *Estudios penales y criminológicos*, 29, 2009, pp. 71 y ss., aclarando que, en realidad, la ampliamente reconocida concepción constitucionalista del bien jurídico no es una, sino varias.

[927] Así, Valls Prieto, J., «La inteligencia artificial frente a la colectividad. Una protección de los bienes jurídicos colectivos frente al auge de los sistemas inteligentes», *Estudios penales y criminológicos*, 44, 2023, p. 3.

[928] Dejamos aquí de lado, conscientemente, otra rama del monismo, que ha sido bautizada de colectivista (así, Pérez-Sauquillo), y que está más relacionada con las primeras manifestaciones de la Escuela de Kiel, donde lo importante era la protección del pueblo por encima incluso de la protección del individuo, desarrolladas al albur de ideas nacionalsocialistas. En todo caso, otros autores posteriores desarrollan estas ideas de protección de la función antes que del individuo, pero a través de la idea de proteger la norma. Así,

bienes jurídicos individuales y colectivos, partiendo de la idea de que, con la evolución de las sociedades, las relaciones y las interacciones, no queda más remedio que reconocer importancia a ciertos elementos comunes a la comunidad o la colectividad [929].

Lo cierto es que la discusión sobre los bienes jurídicos colectivos, tras las críticas que Hassemer y sus discípulos realizan, en global, al derecho penal moderno, aparecen más bien respondidas desde un punto de vista sectorial, sin que la defensa de los bienes jurídicos colectivos llegue a encontrar una teoría completa, sino que se van contestando a las críticas más bien desde la interpretación de grupos de delitos contenidos en la parte especial [930].

Podemos afirmar, empero, que, a pesar de la discusión, lo que prevalece en la actualidad es un concepto personal de bien jurídico moderado, aceptando esencialmente la legitimidad de los bienes jurídicos supraindividuales o colectivos, pero siempre y cuando estos permitan una referencia al individuo [931]. Lo que determinará las distintas posiciones de los autores será la mayor o menor

---

el propio Binding, valiéndose de la teoría de la norma. Para Pérez-Sauquillo constituyen una forma de perder de vista el necesario individualismo ético que debe regir la toma de decisiones en una sociedad democrática, consideración que compartimos. PÉREZ-SAUQUILLO MUÑOZ, C., *Legitimidad y técnicas de protección penal de bienes jurídicos supraindividuales*, Tirant Lo Blanch, Valencia, 2019, p. 79.

[929] Uno de los mayores exponentes es Tiedemann, pero, a diferencia de este, que adopta una visión que podríamos tildar de esencialmente comunitaria (concibiendo la comunidad como un fin en sí mismo), otros autores, aun partiendo de la coexistencia de bienes individuales y colectivos, asumen que para la legitimidad de los mismos es necesario este criterio de, en palabras de Pérez-Sauquillo, «individualismo ético». Así, para esta autora, son capaces estas posturas de superar la crítica de Hassemer de que entre los bienes jurídicos individuales y colectivos se produce una separación tal que acaban convirtiéndose en «dos columnas paralelas carentes de una clave de bóveda». PÉREZ-SAUQUILLO, C., *ibid*., p. 74. La postura del autor alemán en: TIEDEMANN, K., *Tatbestandsfunktionen im Nebenstrafrecht: untersuchungen zu einem rechtsstaatlichen Tatbestandsbegriff entwickelt am Problem des Wirtschaftsstrafrechts,* Mohr, Tübingen, 1969. Otros de los autores defensores del dualismo son Schünemann, Kuhlen, Hefendhel o Greco, y, en España, Gracia Martín, Corcoy Bidasolo, Soto Navarro o Mayo Calderón. Todas las citas a sus obras concretas donde desarrollan este pensamiento en la obra de Pérez Sauquillo, notas al pie 12 a 19.

[930] PÉREZ -SAUQUILLO, C., *ibid*., pp. 75-76, aunque matizando Pérez-Sauquillo estos reproches. La discusión se puede encontrar, de hecho, en cada uno de los sectores del derecho penal «moderno» tal y como aquí se ha caracterizado, que abarca, en esencia, una parte amplísima del Código penal de 1995, que va desde la discusión sobre los delitos contra la salud pública, los delitos económicos, los delitos medioambientales, los delitos contra la seguridad del tráfico, y, finalmente, los delitos contra la seguridad colectiva. No trataremos de realizar, como se ha expuesto, una sistemática sobre el bien jurídico colectivo en cada uno de estos sectores, pues cada uno presenta sus particularidades, pero sí se traerán a colación aquellos aspectos que consideramos comunes a todas las modalidades.

[931] Así, incluso el propio Hassemer, muy crítico en un inicio con el derecho penal moderno, pero que acaba por aceptar una concepción según la cual se admitirán los bienes supraindividuales siempre que estén «funcionalizados» en atención a los bienes jurídicos individuales. HASSEMER, W., «Rasgos y crisis …, *op. cit*. N. 409, p. 248. El mismo, en alemán: «Grundlinien einer personalen Rechtsgutslehre», en SCHOLLER, H./PHILIPPS, L. (Coords.), *Jenseits des Funktionalismus: Arthur Kaufmann zum 65. Geburtstag*, Decker & Müller, Heidelberg, 1989, p. 85 y ss. También es el caso, aunque moderadamente, de Silva, pues en principio prefiere referirlos a un derecho penal periférico o de segunda velocidad. SILVA SÁNCHEZ, J. M., *La expansión …, op. cit.* N. 383, *passim*.

autonomía de unos respecto de otros [932], que examinaremos más adelante, por sus consecuencias respecto de la configuración de los tipos (cuestión que tiene una importancia no desdeñable, pues es necesario saber si la técnica de protección debe exigir una lesión sobre uno u otro).

Aunque la discusión se ha extendido hasta el extremo, cuestionando la legitimidad de tipos largamente asentados en los códigos penales, tales como los delitos contra la administración de justicia, debemos dejar de lado todo el debate relativo a los bienes jurídicos puramente institucionales para centrarnos en la que tiene que ver con los bienes jurídicos calificados como intermedios (aunque la terminología ha sido diversa, utilizaremos esta denominación, por considerarla la más gráfica) por ser los que más relación han tenido con los tipos de peligro abstracto que más características comparten con los tipos que exploraremos [933].

Nos centraremos, pues, en la discusión sobre los denominados bienes intermedios o instrumentales, cuyo nacimiento debemos a Tiedemann en el campo económico y que Schünemann desarrolla más adelante con la denominación de bienes intermedios espiritualizados [934].

---

[932] Así, como autor que ha defendido la autonomía de los bienes colectivos de forma abierta y sin condicionarlos a su vinculación con bienes jurídicos individuales: Gracia Martín, L., «Contribución al esclarecimiento de los fundamentos de legitimidad de la protección penal de bienes jurídicos por el Estado social y democrático de Derecho», en Demetrio Crespo, E. (Dir.), *Crisis financiera y derecho penal económico*, Edisofer, Madrid, 2014, pp. 339-380. También, destacadamente, Corcoy que, pese a afirmar su autonomía de forma rotunda y explícita, en cierta forma acaba renunciando a la misma, pues no deja de reconocer que dichos bienes han de encarnar un interés esencial para el desarrollo del hombre como ser social. Corcoy Bidasolo, M., *Delitos de peligro y protección de bienes jurídico-penales supraindividuales: nuevas formas de delincuencia y reinterpretación de tipos penales clásicos*, Tirant lo Blanch, Valencia, 1999, p. 372. La segunda y más actual edición de esta obra, en la que se mantiene el contenido sustancial de la misma y en la que la autora se reafirma en sus postulados, fue publicada por BdeF en 2024.

[933] No debe dejar de señalarse, no obstante, nuestra posición al respecto, que parte de la posibilidad (y necesidad) de reconocer autonomía propia a los bienes jurídicos institucionales, acogiendo en lo esencial el argumento manifestado por Greco, relativo a que, de aceptar una concepción extremadamente personal del bien jurídico, se deberían destipificar algunos de los delitos más asentados de nuestros Códigos, como el cohecho. Greco, L., «Gibt es Kriterien zur Postulierung eines kollektiven Rechtsguts?», en Heinrich, M. (Coord.), *Strafrecht als Scientia Universalis: Festschrift für Claus Roxin zum 80. Geburtstag*, Walter Mouton de Gruyter, 2011, vol. 1, pp. 205-206. Tampoco puede dejar de señalarse, empero, que algunos autores, pese a reconocer la legitimidad de dichos bienes jurídicos, reconocen en ellos la misma lógica del bien jurídico intermedio que a continuación se explicará.

[934] Así, inicialmente en Schünemann, B., «Moderne Tendenzen in der Dogmatik der Fahrlässigkeits und Gefährdungsdelikte», *Juristische Ausbildung (JA),* 1975, pp. 793 y ss. Destaca Pérez-Sauquillo, C., *Legitimidad y técnicas…, op. cit.* N. 928, p. 156, que, para otros autores, como Rodríguez Montañés, se trata de bienes con función representativa, pero que se ha usado mayoritariamente la denominación de bienes jurídicos intermedios o instrumentales. Así, Rodríguez Montañés, T., *Delitos de peligro, dolo e imprudencia*, Centro de Estudios judiciales, Madrid, 1994, p. 301. El propio Tiedemann ha acabado reconociendo que los bienes jurídicos colectivos supraindividuales, en el marco de los delitos de peligro abstracto, se han impuesto en el extranjero, afirmando que se ha acabado imponiendo la lógica del contenido de deber o los deberes de tráfico, pero que estos no son más que un correlato de los bienes jurídicos intermedios. Originalmente, ya en Tiedemann, K., *Tatbestandsfunktionen*, *op. cit*. N. 929, pp. 119 y ss. Y

La interpretación sobre estos bienes ha sido diversa, relatando Pérez-Sauquillo que una parte de la doctrina los entiende como una forma de proteger, a su vez, otros bienes de naturaleza supraindividual y no individualizables, con gran nivel de abstracción, de manera que se protegería de forma mediata el bien jurídico supraindividual y de forma inmediata el instrumental, de tal forma que los tipos que protegen estos bienes pueden entenderse como delitos de lesión respecto del bien jurídico instrumental [935].

No obstante, nos parece mucho más útil atender a la concepción restrictiva de Mata y Martín [936], que se aleja de esta idea de proteger mediatamente

---

más recientemente en: Tiedemann, K., *Manual de Derecho penal económico: parte general y especial*, Tirant Lo Blanch, Valencia, 2010, p. 67.

[935] Sería el caso de los delitos que protegen instituciones económicas, así, los delitos tributarios, donde mediante la protección de la recaudación tributaria se protege, mediatamente, el orden económico-tributario. El ejemplo, puesto de manifiesto por Martínez-Buján, da lugar, empero, a diversas manifestaciones, pues la naturaleza indeterminada o colectiva de ambos bienes hace dudar sobre si el interés de protección principal es uno u otro. Así, Pérez-Sauquillo (*op. cit*. N. 928, p. 161), señalando que la interpretación que hace Buján es contraria a la mayoritaria, evidenciando, a nuestro juicio, un déficit de claridad si de lo que se trata es de proteger bienes con perfiles ya difusos mediante la creación de otros bienes con contenidos más difusos, en lugar de afirmar, llanamente, que en ese caso se está protegiendo una institución como bien jurídico independiente y con entidad propia. Merece la pena, en este punto, resumir la postura de Martínez-Buján Pérez, C., *Derecho penal económico y de la empresa. Parte general*, Tirant Lo Blanch, Valencia, 6.ª ed., 2022, pp. 203-210, explicando la diferenciación entre el bien jurídico inmediato y el bien jurídico mediato. Para este autor, la distinción entre bien jurídico inmediato y bien jurídico mediato resulta fundamental para una adecuada comprensión dogmática del tipo penal. El bien jurídico inmediato es aquel que la norma penal protege de forma directa y específica, integrándose en el tipo de injusto como un elemento normativo cuya lesión o puesta en peligro constituye presupuesto indispensable de la conducta típica, y que debe ser abarcado por el dolo o la imprudencia del autor. Por su parte, el bien jurídico mediato remite a la *ratio legis*, esto es, a los fines político-criminales más amplios que justifican la incriminación de la conducta, sin formar parte estructural del tipo penal. No solo en los delitos económicos esta cuestión encuentra amplio consenso, sino también en otras figuras delictivas, como es el caso de los delitos de tráfico. Así, sin ánimo de exhaustividad: Molina Fernández, F., «Delitos contra la seguridad del tráfico», en Bajo Fernández, M., (Dir.), *Compendio de Derecho Penal (Parte Especial)*, vol. II, Ceura, Madrid, 1998, p. 709.

[936] Mata y Martín, R. M., *Bienes jurídicos intermedios y delitos de peligro,* Comares, Granada, 1997. Las características que de estos bienes destaca el autor son: se corresponden con intereses supraper-sonales (p. 23), se vinculan con un bien personal, su configuración depende de una decisión legislativa (p. 26), que los bienes colectivos y personales sean intereses homogéneos (p. 30), la complementariedad entre ambos bienes (p. 31) y el riesgo potencial para una pluralidad de víctimas (p. 33). Se muestra algo crítica sobre las conclusiones de este autor Santana Vega, que afirma que, aunque la obra tiene la bondad de restringir los tipos de peligro que pueden dirigirse contra bienes jurídicos colectivos, acaba por no respetar la autonomía entre bienes individuales y colectivos, reduciendo los colectivos a su mera función instrumental. Esto, dice la autora, puede correr el riesgo de disfrazar verdaderos tipos de peligro abstracto contra bienes colectivos bajo la apariencia de tipos de lesión. Así, Santana Vega, D. M., «Reseña bibliográfica relativa a la obra: «Bienes jurídicos intermedios y delitos de peligro. Aproximación a los presupuestos de la técnica …», *Rev. cienc. Juríd*, 04, 1999, pp. 530-531. Creemos, no obstante, que una de las fortalezas de la obra reside en no mostrar una distinción tajante (ya superada) entre bienes individuales y colectivos y, en la búsqueda de soluciones de protección complementarias entre estas dos distinciones, acaba por limitar los problemas propios del delito de peligro abstracto (si bien, a nuestro juicio, no es necesaria la afirmación de que todos los delitos de lesión-peligro son de peligro concreto, sino afirmando la pertinencia de los delitos abstractos de aptitud).

otros bienes supraindividuales y propone entender que solo habrá bienes intermedios cuando se vinculan expresamente por el legislador con intereses que sean individuales, de tal forma que se lesiona el bien jurídico colectivo (bien instrumental o intermedio), pero solo se pone en peligro el bien jurídico individual. Se justifica, así, que realmente son conductas con un grado de lesividad suficiente, que las aleja de la pura dicotomía entre lesión y peligro, y permite afirmar una verdadera posición intermedia.

Nos parece interesante, en este punto, la propuesta de Paredes respecto de los requisitos exigibles a los bienes jurídicos intermedios. Habrán de cumplirse tres condiciones para que estos bienes alcancen una lesividad propia: la relevancia del ámbito social (en el que se hayan generado expectativas de seguridad importantes); que en este campo haya una estandarización importante (es decir, un amplio espacio para el riesgo permitido); y que se produzcan frecuentemente cuestionamientos a estas expectativas de seguridad, sin que las soluciones extrapenales sean eficaces y preferibles [937].

Esta concepción de los bienes jurídicos intermedios, a pesar de haber tenido una acogida generalizada, ha sido objeto, asimismo, de críticas, tanto por su aparente fuerza legitimadora de los delitos de peligro abstracto [938], como por su formalidad [939]. No nos parece, empero, que la mera afirmación de que existen ciertos bienes jurídicos como objeto de protección haya de preconizar la legitimidad de los tipos, pues lo que habrá que examinar será, en su caso, la concreta legitimidad de cada una de esas concretas formas de protección. Compartimos en este sentido la concepción de otros autores [940], que afirman que en el caso de los bienes jurídicos instrumentales se habrán de aplicar todas las restricciones

---

[937] Paredes Castañón, J. M., «Los delitos de peligro como técnica de incriminación en el derecho penal económico: bases político-criminales», *RDPC*,11, 2003, pp. 138-189. En todo caso, señala el autor que estas tres características no prejuzgan, aún, cómo va a ser la protección de dicho bien intermedio.

[938] En este sentido, Corcoy Bidasolo, M., *op. cit.* N. 932, p. 262, afirmando que lo que buscan es que los delitos de peligro abstracto estén dotados de una suficiente lesividad u ofensividad, de la que en realidad carecen.

[939] Así, Hefendehl, R., *Kollektive Rechtsgüter..., op. cit*. N. 926, p. 175. También Pérez Saquillo, C., *op. cit.* N. 928, p. 172, que critica que su aplicación será entonces casi automática, dando lugar a verdaderos delitos de infracción de deber.

[940] Destacadamente, Rodríguez Montañés, T., *Delitos de peligro…*, *op. cit*. N. 934, p. 307. También Martínez-Buján Pérez, C., *Derecho penal económico*, *op. cit*. N. 935, p. 202-203, explicando que la delimitación del bien jurídico es un requisito necesario, pero no suficiente, para legitimar la intervención penal. El bien jurídico cumple una función técnico-jurídica que permite estructurar los tipos penales, pero no fundamenta por sí solo la selección normativa de conductas punibles. La verdadera legitimidad del Derecho Penal exige identificar qué aspectos de la realidad social resultan dañinos, quién debe definir las conductas intolerables y, además, justificar moral y políticamente la intervención punitiva conforme a los valores de una sociedad democrática. Asimismo, es indispensable una justificación instrumental que demuestre que la respuesta penal es el medio más adecuado y eficiente frente a otras formas posibles de intervención estatal.

propias de los delitos de peligro abstracto, pues ciertamente su forma de configuración apunta directamente hacia esta técnica de protección, pero eso no significa preconizar, como se pretende, su falta de legitimidad. Que un determinado tipo respete o no los postulados del principio de lesividad (u ofensividad), y que no peque de ser un tipo estrictamente formal dependerá más bien de la concreta técnica de tipificación, de la elección de los bienes jurídicos finales y de las matizaciones que se realicen en cuanto a su interpretación, pero no debemos excluir, de entrada, su legitimidad, de la que aquí se parte.

### 1.2 La discusión sobre la salud pública y la seguridad colectiva en el desarrollo del bien jurídico intermedio o de referente individual

Si hay un terreno donde la discusión sobre la legitimidad del bien jurídico alcanzó cotas insospechadas fue el campo de la salud pública. No en vano los primeros *leading cases* del moderno derecho penal, examinados *supra,* afectaron directamente a bienes jurídicos relacionados con la salud (integridad física y vida, esencialmente). Tal y como señala Pérez-Sauquillo, la discusión sobre la posible existencia de un bien demasiado abstracto, que fue tildado en ocasiones de bien jurídico aparente, no fue exclusiva en este campo, sino que se extendió también a todos los delitos contra la seguridad colectiva [941].

La doctrina recuperó, así, un debate de larga tradición, que, tal y como expone Aires de Sousa, venía produciéndose en Alemania desde antiguo, de la mano de los delitos de peligro común, cuyo exponente más clásico lo representa, seguramente, el delito de incendio [942]. En efecto, con este delito se castigaba la amenaza de un daño difuso para un número indiferenciado e indeterminado de objetos o bienes jurídicos. Explica esta misma autora cómo el paralelismo es obvio en cuanto a la fabricación y comercialización de productos defectuosos, pues estas actividades conllevan un potencial peligro difuso. Así es como, de hecho, se configura el antecedente más inmediato a nuestros delitos de peligro contra la seguridad colectiva: el entonces delito de envenenamiento peligroso de la comunidad *(Gemeingefährliche Vergiftung)* [943]*,* don-

[941] Pérez-Saquillo, C., *op. cit.* N. 928, pp. 191-192.

[942] Aires de Sousa, S., *op. cit.* N. 628, pp. 173 y ss.

[943] Que configura el actual § 314 del STGB, cuyo contenido actual es: § 314 Envenenamiento peligroso para la comunidad: (1*) Será castigado con pena privativa de libertad de uno a diez años quien: 1. Envenene agua en fuentes canalizadas, pozos, conducciones o depósitos de agua potable, o 2. envenene objetos destinados a la venta o consumo públicos, o mezcle en ellos sustancias nocivas para la salud, o quien venda, ofrezca para la venta o de cualquier otra manera introduzca en el comercio objetos conforme al número 2 que estén envenenados o mezclados con sustancias nocivas para la salud.* (traducción propia).

de, en lo que aquí nos interesa, se incriminaban aquellas conductas que suponían inseguridad para un elevado número de personas o bienes cuando los efectos de las conductas eran muy amplios [944], integrándose dentro de los llamados delitos de peligro común. La discusión que se siguió en Alemania fue, por tanto, si debía aplicarse un delito controvertido como el mencionado a los nuevos casos de responsabilidad por el producto [945]. Las propuestas enfrentadas reproducen, en lo esencial, la discusión aquí mantenida. Mientras Kuhlen criticó su aplicación por la difícil verificación de la destrucción del bien jurídico [946] (entendemos que en una crítica similar a la ya identificada relativa a su falta de lesividad), autores como Hilgendorf expresaron su posible aplicación para la imputación a los directivos en el caso *Lederspray* [947], sin cuestionar la legitimidad del tipo, argumentando que, dado que los directivos conocían su carácter de producto destinado al consumo público y, a partir de un determinado momento, conocían su carácter nocivo, el tipo era perfectamente aplicable.

El debate fue similar en lo relativo a la salud pública en la experiencia española, donde, si bien no existía una tradición en cuanto a la protección de delitos de peligro en general, sí estaba asentada la protección penal del consumidor [948]. A pesar de su tradición, explica Aires de Sousa que este tipo de delitos fue también objeto de discusión en cuanto al bien jurídico protegido, dividiéndose en este caso la doctrina española, esencialmente entre quienes consideraban esta forma de tipificación como una mera anticipación de la tutela de bienes individuales, aquellos que consideraban que lo novedoso aquí no era tanto el objeto de protección como la forma especial de ataque contra esos bienes, por su potencialidad para afectar a un número indeterminado de perso-

---

[944] Aires de Sousa, S., *op. cit.* N. 628, p. 179. Explica Aires cómo, de hecho, la interpretación de este delito (antiguamente ubicado en los §§ 319 y 320, referentes a los delitos de peligro común de envenenamiento doloso y envenenamiento negligente, respectivamente) supuso una división entre quienes interpretaban que se trataba de un delito de peligro abstracto puro o quienes consideraban que debía interpretarse como un delito de aptitud (*Eignungsdelikte*), discusión que ha sido determinante también en la discusión española, y que se desarrollará *infra.*

[945] La doctrina estaba dividida, pues, aunque estos delitos existían, eran escasamente aplicados, y además contenían unas penas que podían llegar a ser especialmente elevadas.

[946] Kuhlen, L., *Fragen einer strafrechtlichen Produkthaftung*, Müller C. F., Heidelberg, 1989, p. 166.

[947] Hilgendorf, E., *Strafrechtliche Produzentenhaftung..., op. cit*. N. 671, p. 168.

[948] Es en este sentido revelador el estudio realizado por Díaz-Maroto y Villarejo, señalando que podemos encontrar en el CP de 1870 el antecedente inmediato de la adulteración nociva de bienes alimentarios. Díaz-Maroto y Villarejo, J., *El Derecho Penal ante los Fraudes Alimentarios. Responsabilidad por el Producto en la moderna sociedad del riesgo*, Civitas/Thomson Reuters, Cizur Menor, 2010, p. 64. Una reflexión actualizada sobre estos delitos, por el mismo autor, enmarcándolos dentro del paradigma de la sociedad del riesgo en: Díaz-Maroto y Villarejo, J., «Sociedad del riesgo y fraudes alimentarios» *op. cit*. N. 632, pp. 1099-1109.

nas, y finalmente, aquellos que consideraban que la salud pública tenía una autonomía conceptual propia [949].

Destacan dentro de este último grupo aquellos que, más allá de reconocer una conexión con ciertos bienes individuales, ponen el énfasis en que la autonomía del bien jurídico de la salud pública viene configurada por la «protección de la confianza de los consumidores en la inocuidad de los productos que se encuentran en el mercado» [950], no exentos de críticas por su excesiva formalidad o la importancia concedida a la vertiente subjetiva de la confianza [951].

La discusión doctrinal ha pivotado aquí, una vez más, en torno a los problemas de ofensividad (por la lejanía respecto de la materialización del daño y el amplio abanico de conductas penalizadas), unidos inevitablemente a las críticas sobre la formalidad, es decir, que la relevancia típica de las conductas penalizadas no provenga de su peligrosidad, sino de la mera inobservancia de deberes administrativos que la ley fija [952]. Vemos cómo los términos son, en esencia, iguales que los desarrollados entre la doctrina alemana.

---

[949] Nos referimos, fundamentalmente, a los delitos hoy tipificados bajo la rúbrica *«De los delitos contra la salud pública»*, entre los artículos 359 y 367 CP. De esta opinión, entre otros: Díaz-Maroto y Villarejo, J., *El Derecho Penal ante los Fraude …, ibid*., p. 27 y ss., hablando de la salud pública como bien jurídico «colectivo, autónomo, independiente y diferenciado». El autor cita también diversa jurisprudencia que considera la salud pública, distinta de la salud individual, como bien jurídico protegido. Así, la temprana STC 62/1983, de 11 de julio, FJ2, donde se afirma que «*el bien jurídico protegido en los delitos de riesgo en general contra la salud pública es el bien común en que la misma consiste, que se refleja en definitiva en la salud personal de los ciudadanos*», o la STS 1210/2001, de 11 de junio, donde se afirma que la salud pública se refiere al *«conjunto de condiciones positivas y negativas que garantizan la salud de los integrantes de una comunidad»* (ibid., p. 28). Precisamente señala Aires de Sousa cómo detrás de las tesis de estos autores no deja de haber una concepción monista-personalista del bien jurídico colectivo, donde la tutela penal de la salud pública encuentra su fundamentación en la protección de los intereses del individuo, «asegurando las posibilidades de supervivencia del ser humano». Aires de Sousa, S., *op. cit*. N.628, nota al pie 457.

[950] Corcoy Bidasolo, M., «Delitos contra la salud de los consumidores», en Corcoy Bidasolo, M./ Gómez Martín, V. (Coords.), *Manual de derecho penal, económico y de empresa, parte general y parte especial: (adaptado a las LLOO 1/2015 y 2/2015 de Reforma del Código Penal): doctrina y jurisprudencia con casos solucionados*, tomo 1, Tirant lo Blanch, Valencia, 2016, pp. 182 y ss. Igual en Corcoy Bidasolo, M., *op. cit*. N. 932, pp. 228 y ss.

[951] Crítico con la postura de Corcoy, señalando su «excesiva psicologización de los bienes jurídicos», Feijoo. Así, Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto. Sobre la normativización del peligro», en AAVV., *Homenaje al profesor Dr. Gonzalo Rodríguez Mourullo*, Civitas, Madrid, 2005, nota al pie 44. Nos parece, que, en este punto, la postura de Corcoy, está mucho más vinculada al bien jurídico individual de lo que ella misma reconoce. Nos hacemos cargo de la distinción que la autora propone entre el objeto de protección y el bien jurídico, pero, a pesar de ello, nuestra percepción es que Corcoy no sostiene una postura del bien jurídico tan autónoma como la que afirma, y que, por tanto, no se le pueden hacer las mismas críticas que a aquellos que, de una forma más evidente, sí la han sostenido. En este sentido, la propuesta de Kindhäuser, que se comentará *infra.*

[952] Especialmente crítico aquí Muñoz Conde. Hassemer, W./Muñoz Conde, F. *La responsabilidad por el producto…, op. cit* N. 409, p. 65 y ss., estableciendo que la vulneración de normas legales o reglamentarias que regulan la distribución y venta de productos destinados al consumo humano, cuando pueda poner en riesgo la salud de las personas o cuando se realice sin la debida autorización administrativa, constituye un indicio de una posible conducta sancionable penalmente. Sin embargo, para que dicha in-

La dificultad interpretativa que ponen de manifiesto todos estos tipos penales tiene que ver con que el tipo suele exigir un peligro que afecte a una pluralidad de consumidores, dando lugar a la discusión sobre si la salud pública es verdaderamente algo más que la suma de las saludes individuales. La opinión de Aires de Sousa resulta, en este punto, valiosa, pues expresa que aquello que diferencia al bien jurídico colectivo de los individuales tiene que ver con que, más allá de que exista un daño individual, se sumará necesariamente un daño comunitario en cuanto la colectividad se ve amenazada, de forma que se evidencia una dimensión social del bien jurídico colectivo protegido [953], entendiendo que la amenaza a la colectividad deriva del peligro para el «conjunto de condiciones positivas y negativas que garantizan la salud de los miembros de la comunidad» [954]. Advierte la misma autora, empero, que afirmar la importancia de la dimensión colectiva no basta, pues debe adoptarse en todo caso una concepción restrictiva de la misma, señalando que más allá de tutelar estas condiciones, deben observarse principios como el de *ultima ratio* y el de proporcionalidad entre la gravedad de hecho y la sanción [955], principios que deben guiar en todo caso al legislador en la tipificación de las conductas, so pena de acabar por criminalizar conductas que no suponen más que la mera infracción de deberes administrativos [956]. Estamos, en este sentido, de acuerdo con la conclusión alcanzada por la autora portuguesa: que un bien jurídico sea legítimo no prejuzga la legitimidad de la incriminación siempre y en todo caso.

No es esta, en cualquier caso, la única forma de concebir la salud pública, y para completar el cuadro de interpretaciones nos parece especialmente expresiva la que ofrece Doval, tanto en su prominente investigación doctoral como en trabajos posteriores [957].

---

fracción adquiera la categoría de delito y sea merecedora de una sanción penal, es necesario que cumpla un requisito adicional: debe ser nociva, es decir, debe implicar un perjuicio efectivo para la salud pública y, en consecuencia, para la vida y salud de las personas que forman parte de la sociedad. De no establecerse esta distinción, se corre el riesgo de confundir una simple infracción administrativa con una infracción penal, lo que supondría una distorsión de los principios fundamentales del derecho penal en un Estado de Derecho moderno: su carácter de *ultima ratio,* el principio de mínima intervención y la exigencia de proporcionalidad entre la gravedad de la conducta y la severidad de la sanción.

953 Aires de Sousa, S., *op. cit*. N.628, p. 205.

954 *Ibid.,* p. 206.

955 *Ibid.,* p. 207.

956 *Idem*: *«[…] um entendimento demasiado amplo do bem jurídico saúde pública diluiu ao nível daquelas normas a distinção entre crime e infracção administrativa, em clara violação daqueles princípios basilares, elevando à categoria de delito a mera infracção de deberes*».

957 Nos referiremos, en lo esencial, al contenido de la obra fruto de su tesis doctoral, sin perjuicio de que el autor haya tenido ocasión de desarrollar en trabajos posteriores su postura, que venía esencialmente esbozada en su primigenio trabajo de 1995. Las referencias serán hechas al contenido de su tesis doctoral: Doval Pais, A., *Fraudes alimentarios. Evolución histórica y elementos esenciales del sistema de*

Realiza Doval un repaso por las concepciones de los autores que consideran la salud pública como el conjunto de condiciones que posibilitan la conservación de la salud de los ciudadanos [958], en el sentido que acabamos de ver exponía Aires, así como de aquellos que, en un concepto parecido, hablan de la protección de los «presupuestos objetivos que garantizan la salud de todos» [959], pero discrepando de esta concepción, asegurando que erigir estas condiciones en bien jurídico es como *«pretender garantizar sólo el medio por el cual se asegura la conservación de la salud»* [960], formalizando, en su opinión, el bien jurídico en exceso.

En cuanto a la concepción que afirma que es posible hablar de dos bienes jurídicos, uno mediato, de carácter básico, (la salud individual) y otro de carácter complementario (la salud pública), destaca la opinión de Bustos, cuando afirma que realmente solo en el plano material cabe hacer esta distinción, sin que en el plano formal nada impida considerar un bien jurídico, entendemos que propio, a partir de esta formulación [961].

Tras realizar un repaso por las otras opciones que doctrina y jurisprudencia han ido manejando, (expectativas del consumidor, sanidad, o suma de saludes individuales, respectivamente), aclara su propia apreciación. Doval considera que la confusión generada en torno al bien jurídico «salud pública» tiene que ver con la cuestión de que se han acabado introduciendo problemas ajenos al bien jurídico en la discusión sobre el mismo, que pertenecen más bien a la

*protección penal*, Valencia, 1995, tesis doctoral en abierto en: < https://roderic.uv.es/items/09bce9b9-1728-4b4e-a7ee-5e3736e547e2 > [Accedido por última vez el 27/01/2026]. Más adelante publicada, en una versión reducida, en: Doval Pais, A., *Delitos de Fraude Alimentario. Análisis de sus Elementos Esenciales*, Aranzadi, Pamplona, 1996.

[958] Doval Pais, A., *Fraudes alimentarios. Evolución histórica…*, ibid. p. 271.

[959] Explica Doval que el concepto de «condiciones» como parte del bien jurídico «seguridad colectiva» ya fue abordado por autores como Jiménez de Asúa y Antón Oneca, quienes lo definieron como un conjunto de condiciones protegidas por el orden jurídico que garantizan la vida, la integridad, el bienestar y la propiedad, considerados como bienes comunes. Esta idea ha sido desarrollada por otros autores (como Rodríguez Ramos, Corcoy Bidasolo, Pérez Álvarez, Octavio de Toledo y García-Pablos), quienes analizan los fraudes alimentarios desde esta perspectiva. Para los detalles y referencias sobre estos autores, Doval Pais, A., ibid, nota 78.

[960] Doval Pais, A., ibid., p. 275.

[961] *Ibid.,* p. 285. Se refiere el autor a Bustos Ramírez, J.: «Los bienes jurídicos colectivos (Repercusiones de la labor legislativa de Jiménez de Asúa en el Código penal de 1932)», *Revista de la Facultad de Derecho de la Universidad Complutense*, 11, 1986, p. 159. Más adelante, mismo trabajo de Bustos Ramírez, J., «Los bienes jurídicos colectivos» *Revista de Derecho penal,* 27, 2019, p. 472-473, afirmando que en muchas ocasiones las críticas a los bienes jurídicos (colectivos) excesivamente vagos tienen razón, pues se comete el error de formalizar el bien jurídico, «*olvidando que el bien jurídico solo tiene sentido como concepto material; de otro modo se convierte en un mero instrumento interpretativo teleológico […]».*

estructura de los delitos [962]. Y así, aclara, que uno de los conceptos más controvertidos, el relativo al hecho de que se apele a la salud «pública» indica no una particularidad en la titularidad del bien (pues es evidente que el tipo protege la salud de todas las personas), sino que indica algo sobre la forma de ataque. En definitiva, se trata de obtener una protección adecuada ante conductas que no tienen intención de lesionar y que, por el medio en el que se producen (paradigmáticamente el mercado), comportan un peligro para un número indeterminado de personas [963]. Estas peculiares características hacen que no sirva ya la estructura de la tentativa, porque no se dará un dolo de lesionar, y que tampoco quepa el castigo de la imprudencia, (aun cuando se admitiera el castigo de la conducta imprudente sin resultado lesivo), porque, según el autor, sería necesario entonces tener un marco típico con el que poner en relación la genérica previsión de la imprudencia [964].

Lo que se da aquí, por tanto, a juicio de Doval, es una indiscriminada puesta en peligro y una ausencia de resolución lesiva en el autor, y esta es la verdadera característica fundamental de las situaciones que el legislador pretende proteger cuando alude a la salud pública. Es la clase de ataques (diríamos, las características específicas de la forma de atacar esos bienes), lo que hace necesario acudir a unas nuevas formas de protección, cuales son los delitos de peligro, pero lo cierto es que los bienes jurídicos no cambian en lo esencial [965].

La naturaleza colectiva del bien jurídico, por tanto, se pone de manifiesto no en la fase de protección, sino esencialmente en la fase de agresión [966]. Es importante resaltar que el propio Doval es consciente de que esta postura (que se basa en una concepción personal del bien jurídico) no se aleja demasiado de la fórmula del conjunto de condiciones que posibilitan la salud individual, y que, en esencia, se trata de fórmulas equivalentes en un plano conceptual [967].

---

[962] Doval Pais, A., *op. cit.* N. 957, p. 332: «*Considero que buena parte de la confusión que se ha generado en torno al concepto de «salud pública» procede de que se han introducido en el terreno del bien jurídico protegido aspectos que no son propios de él, sino que se refieren, más bien, a la inevitable estructura de los correspondientes delitos. Así sucede con las notas de la publicidad y del peligro, con las que se trata de compatibilizar a toda costa la configuración del bien jurídico».*

[963] *Ibid.,* p. 337, hablando de «absoluta indeterminación *ex ante* del sujeto pasivo».

[964] *Idem.*

[965] *Ibid.,* pp. 340 y ss. Aclara Doval que esta indeterminación de sujetos pasivos en los delitos contra la salud pública se produce *ex ante* y también e*x post*, es decir, sigue siendo indeterminado el número de sujetos que han sido puestos en peligro. Alude (en nota 172) a la nomenclatura, utilizada ya en su momento por Rodríguez Mourullo, de delitos «vagos» o «vagantes». Rodríguez Mourullo, G., *Derecho penal. Parte general*, Civitas, Madrid, 1978, p. 283.

[966] Doval Pais, A., *op. cit.* N. 957, p. 344, aclarando que la titularidad colectiva del bien jurídico se afirma por lo general partiendo de la indeterminación del bien jurídico afectado por el delito.

[967] *Ibid.,* p. 347.

Estos bienes jurídicos se han concebido de distintas formas por la doctrina en cuanto a su relación con los bienes de referente individual, a veces considerándose una forma anticipada de protección de estos últimos, otrora como bienes de carácter instrumental o medial, que actúan como complemento en cuanto al sistema integral de protección, aludiéndose en este sentido a la idea de las «normas obstáculo» [968]. Aclara el autor que, por ello, la única forma válida de hablar de un adelantamiento en la protección es entendiendo que la protección de los medios es necesaria y suficiente para asegurar el verdadero objeto protegido. En el caso de los delitos de fraude alimentario, que él analiza, implicaría que la protección de las condiciones de seguridad de los alimentos y los productos es la única vía mediante la cual puede garantizarse la salud individual [969].

Recoge también Feijoo [970] la reflexión relativa al bien jurídico cuando, en el homenaje a su maestro, el profesor Rodríguez Mourullo, trae a colación una célebre cita del homenajeado, donde se deja entrever ya la vieja discusión sobre la problemática del bien jurídico en los delitos de peligro, que, como se pone de manifiesto con estas páginas, sigue siendo de actualidad [971].

---

[968] *Ibid.,* p. 350. Hablando también de esta concepción de los delitos obstáculo, Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 311, nota al pie 10. En el campo de la IA, apunta brevemente la aplicación del esquema del delito obstáculo Pérez-Arias, J., «Algoritmos y big data en la responsabilidad penal: el reto de la cibercriminalidad en el Derecho Penal», en Peris Riera, J. M./ Massaro, A., (Coords.), *Derecho Penal, Inteligencia Artificial y Neurociencias*, Roma Tre-Press, Roma, 2023, p. 183, definiéndolos como «fenómeno de criminalización anticipada mediante el que se castigan conductas en un momento anterior a la lesión del bien jurídico e incluso con anterioridad a que se genere un peligro concreto o abstracto para el bien jurídico. Son por lo tanto delitos que se configuran como auténticos obstáculos que tienen como función impedir que lleguen a producirse los actos delictivos futuros que se tipifican en otros preceptos. En ellos el principio de ofensividad cede ante la necesidad de prevención general y se presentan como tipos penales formales o de mera desobediencia mediante los que se anticipa, si bien en algunos casos de forma muy cuestionable, la barrera defensiva que supone la aplicación de toda norma penal (STS 2315/2017)». La obra de obligada referencia sobre este modelo de delitos: Cuesta Pastor, P., *Delitos obstáculo. Tensión entre política criminal y teoría del bien jurídico*, Comares, Granada, 2002.

[969] Doval Pais, A., *op. cit.* N. 957, p. 352. Reconoce el autor, a lo largo de su obra, por tanto, que asegurar el conjunto de condiciones que posibilitan la salud es un medio necesario para asegurar la salud individual, pero que adoptar la protección del medio como bien jurídico resultaría en un concepto demasiado formal de bien jurídico. Discrepamos, en este punto, con él, pues formal sería una invocación a un concepto vacío, por lo difuso, ambiguo o general, sin que quepa tachar de formal, a nuestro juicio, la protección de un medio (medio que el propio autor califica de necesario y suficiente) directamente encaminado con la finalidad pretendida. De hecho, el propio Doval concede a Tiedemann que la salvaguarda de los bienes jurídicos colectivos pueda referirse, en última instancia, a la protección del individuo, sin que esto implique negar la autonomía (si bien parcial) de los bienes colectivos. Así, en nota 187. Aclara, en todo caso, y en esto hemos de concederle la razón, que una cosa es lo que se diga respecto de los bienes colectivos con referente individual, de los que no cabe predicarse una autonomía (o al menos no una autonomía absoluta), y otra cosa lo que se predique de los bienes propiamente colectivos o difusos, aquellos en los que no es posible encontrar referente individual directo.

[970] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951.

[971] Así, haciendo nuestra la cita de Feijoo, traemos a colación la postura de Rodríguez Mourullo, que ya en *La omisión de socorro en el código penal*, Tecnos, Madrid, 1996, señalaba cómo en los delitos de

Feijoo examina en el mencionado trabajo conmemorativo la problemática de los delitos contra la seguridad colectiva, a los que califica de herederos directos de los anteriormente llamados delitos de riesgo en general, caracterizando, con la doctrina dominante, dicho bien jurídico (la seguridad colectiva), como bien jurídico intermedio de referente individual [972].

Resalta a continuación Feijoo un rasgo también manifestado por Doval, relativo no tanto al bien protegido como a la técnica de protección, reconociendo que la estructura del delito imprudente de resultado no da mucho más de sí [973], y que, por lo tanto, cuando se trata de examinar los delitos que protegen normalmente estos bienes (entre los que se encuentran, como ejemplo paradigmático, los siempre polémicos y discutidos delitos de peligro abstracto), merece la pena situar el problema en términos de injusto (formulando la pregunta de si existe un injusto penal válido) [974], buscando un fundamento normativo para la tipificación más que un objeto de protección. Compartimos en este sentido lo que sostiene Feijoo cuando afirma que en algunos de estos delitos (paradigmáticamente aquellos que protegen funciones de control y ordenación de materiales o sustancias peligrosas), es evidente que el derecho penal protege estas funciones porque las considera condiciones de seguridad para los bienes jurídicos individuales (al igual que, recordemos, para Doval era evidente que la protección de las condiciones de salud era la forma idónea de obtener una protección integral de la salud individual), concluyendo que «*la abstracción que presentan estas tipificaciones tiene que ver más con la técnicas de protección que con el objeto de protección*» [975]. No es este un rasgo que sea predicable exclusivamente de aquellos delitos en que hay una protección de funciones de ordenación (Feijoo pone el ejemplo de los delitos de tráfico de sustancias, o los relacionados con los materiales radioactivos o nucleares), sino también con lo que llama delitos de peligro común para

peligro, la ley convierte a la seguridad en bien jurídico respecto de otro bien, fijando así una valiosa reflexión, que nos sitúa en la lógica de la protección dual (mediata e inmediata) de bienes jurídicos de distinto nivel, p. 148.

[972] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 307, con cita a Doval País, Mata y Martín o Soto Navarro. En contra de la concepción de Feijoo, aunque analizando los delitos contra la seguridad vial, caracterizándolos de tentativas imprudentes sin resultado: Molina Fernández, F., «Delitos contra la seguridad vial», en Molina Fernández, F. (Dir.), *Memento práctico Penal,* Francis Lefebvre, 2025 nm. 15560 y ss., negando que los delitos contra la seguridad colectiva estén en todo caso circunscritos a un riesgo colectivo, aunque aceptando que, si bien no en todo caso las personas afectadas deben ser un colectivo, sí deben ser, al menos, indeterminadas, pues, de lo contrario, debería aplicarse el delito de resultado (nm. 15565). En un sentido similar, oponiéndose directamente a las tesis de Feijoo: Molina Fernández, F., «Delitos contra la seguridad del tráfico», *op. cit*. N. 935, p. 712.

[973] *Ibid.,* p. 308. En este sentido, citando a Hefendehl, Jakobs, o Silva. Todos ellos *op. cit.*

[974] *Ibid.,* p. 309.

[975] *Ibid.,* p. 310.

bienes jurídicos amplios (tales como la vida y la salud), donde, una vez más, el autor afirma que *«lo que hay es genuinamente una técnica de protección diferente más que un bien jurídico diferente»*.

Concreta su postura respecto de los bienes jurídicos en los delitos contra la seguridad colectiva Feijoo en otro trabajo, donde afirma rotundamente que en estos delitos (Título XVII del CP), aunque un sector importante de la doctrina ha considerado que se protegen delitos de titularidad colectiva, supraindividuales o difusos, lo que se protege son los mismos bienes que se protegen ya en otros lugares del Código (vida, salud, patrimonio, incluso medio ambiente), siendo la única particularidad la forma de protección [976]. Señala el autor que estos delitos pretenden combatir una forma de ataque, cual es un riesgo de carácter expansivo que afecta a un número indeterminado o difuso de personas, siendo precisamente la indeterminación la que preocupa al legislador [977].

Estos supuestos nuevos bienes jurídicos (Feijoo se muestra reacio a reconocerles una autonomía, a riesgo de confundir el bien jurídico con la interpretación teleológica, es decir, el fin perseguido con el objeto protegido, entendemos), tratan de solucionar un problema que se produce todavía en el campo de la política criminal, como es la expansión del potencial lesivo de los avances técnicos en el marco de la sociedad postindustrial [978]. Explica cuáles son los contornos de estos hechos, afirmando que se trata, en general, de hechos contra los que los delitos tradicionales de resultado no pueden luchar de forma satisfactoria, constituyendo la mayoría supuestos en los que una única persona puede causar una auténtica catástrofe. Por ello, explica, el derecho penal intenta evitar dichos resultados catastróficos cambiando su estrategia de protección, adelantándose al momento en el que el riesgo deviene incontrolable [979].

Se explicará, *infra*, cómo desarrolla Feijoo una teoría completa y coherente para interpretar los delitos de peligro abstracto, precisamente al hilo de los deli-

---

[976] Feijoo Sánchez, B., «Delitos contra la seguridad colectiva: ¿tentativas imprudentes?», *La Ley*, 22432/2001,1997, p. 1.

[977] Pone Feijoo varios ejemplos de delitos que buscan atacar esta dinámica: conductas susceptibles de producir infinidad de resultados de muerte, lesiones, daños al patrimonio y/o, en algunos casos (por ejemplo, arts. 348 y siguientes del Código Penal), afectar al medio ambiente, sin que tenga el autor un control sobre el alcance de su hecho. Afirma Feijoo que, en este sentido, se trata de delitos de los denominados de peligro común o general. Feijoo Sánchez, B., *ibid.,* nota al pie 2.

[978] *Ibid.,* p. 1 y nota 3.

[979] *Ibid.,* p. 2. Explica el autor que la estandarización y alta reglamentación de las conductas tiene pleno sentido en este contexto, pues el riesgo que se pretende evitar es abierto, incontrolado y con potencial indefinido de víctimas. De ahí que para Feijoo no quepa interpretar los delitos de estas características como meras tentativas imprudentes. Así, p. 3.

tos contra la seguridad colectiva, pero hasta aquí queríamos dejar sentadas las ideas relativas al bien jurídico objeto de protección en este tipo de delitos.

Terminaremos, pues, este primer apartado con una reflexión y una recapitulación. Feijoo destaca cómo, a diferencia de lo que ocurre con los delitos hasta ahora existentes contra la seguridad colectiva, no existe ningún delito contra la seguridad colectiva relacionado, por ejemplo, con la actividad médico-quirúrgica, donde, afirma textualmente el autor, «la actuación imprudente suele afectar a un único paciente». Cabe preguntarse si, precisamente, con la aparición de algunas formas de ataque nuevas (precisamente relacionadas con las tecnologías de IA en sectores como el médico, donde es ya absoluta protagonista), no hará cambiar esta consideración. Siguiendo la lógica expuesta por Feijoo, que en lo esencial compartimos, si en el sector médico se generalizan riesgos indeterminados por el uso de nuevas técnicas que ponen en peligro de forma difusa diversos bienes jurídicos individuales de los pacientes, en esta lógica del riesgo de carácter expansivo que afecta a un número indeterminado de personas, nos veríamos obligados a reproducir esta dinámica también en sectores como el médico. ¿No es, acaso, una herramienta de diagnóstico que utiliza IA, en manos de un médico, una fuente potencial de errores de diagnóstico y discriminación que pueden dar lugar a situaciones incontroladas de riesgo para la vida y la integridad física de los pacientes a ellas expuestos?

### 1.3 **Recapitulación**

Cuando nos enfrentamos a la pregunta acerca de qué bienes pueden proteger (legítimamente) los códigos penales, la discusión viene de lejos. Desde concepciones más monistas o apegadas al bien jurídico en un sentido estrictamente personal, parece común afirmar que la doctrina ha ido paulatinamente reconociendo la existencia (y conveniencia) de otros bienes de carácter supraindividual.

Cuando se trata de hablar de estos bienes las críticas suelen ir más bien dirigidas a los tipos a los que dan lugar, es decir, a los delitos de peligro abstracto, señalándose esencialmente su falta de lesividad y su excesiva formalidad.

Empero, frente a estas críticas pueden considerarse suficientemente sólidos los argumentos que defienden la existencia de un tercer género de bienes, entre el individual y el colectivo, que ha venido a denominarse de forma variada (intermedio, espiritualizado, o instrumental), pero que, en definitiva, permite entender la protección de los bienes jurídicos finales (vida e integridad física, en esencia), cuando la forma de atacar los mismos responda a unas características que justifiquen la alusión a técnicas de peligro.

No nos parece tan importante ese matiz de si lo que se protege es el bien jurídico colectivo en forma de condiciones necesarias para el disfrute efectivo o el bien jurídico final mediato, pues creemos que ambas afirmaciones son ciertas y no incompatibles. Proteger las condiciones del sistema de salud es, de hecho, un medio efectivo para proteger la salud, de igual forma que proteger las condiciones de seguridad de la industria de la IA es una forma de proteger los bienes jurídicos que daña la IA. En otras palabras: protegiendo los medios se consigue proteger el fin. Que esto se llame de una forma u otra, es decir, colectivo en cuanto suma de las condiciones o individual protegido mediatamente, es algo que no nos interesa demasiado, pues creemos que no es la discusión esencial aquí.

Es menester, por tanto, dejar sentado que, hasta ahora, o al menos antes de que la IA irrumpiera, es posible sostener que los bienes objeto de ataque en la moderna sociedad del riesgo son, esencialmente, los mismos, y que una forma efectiva y adecuada de protegerlos es protegiendo este paraguas de condiciones objetivas que permiten su disfrute. Sin embargo, afirmar la legitimidad del bien no basta para afirmar la legitimidad de la técnica, pues para salvar las acusaciones de formalidad habremos de indagar en el fundamento de los delitos de peligro abstracto.

Sin embargo, antes de iniciar dicha tarea queda pendiente resolver una cuestión previa, relacionada con los objetos de protección de la nueva sociedad del riesgo tecnológico. ¿Podemos decir que tras la irrupción de una tecnología tan potente y lesiva como la IA, los bienes jurídicos actuales son suficientes? A contestar a esta pregunta dedicaremos las siguientes páginas.

## 2. ¿NECESITAMOS NUEVOS BIENES JURÍDICOS? LA «SEGURIDAD COLECTIVA ALGORÍTMICA O DIGITAL»

Si ante el auge de las nuevas técnicas y formas de producción propias de la sociedad del riesgo el debate sobre la necesidad de proteger nuevos bienes jurídicos fue inevitable, algo similar lleva pasando desde que la IA irrumpe como una de las técnicas más prometedoras y omnipresentes en todos los sectores de la vida. Las promesas de la técnica, igual que sucedía con el entonces prometedor radio, en los albores de su descubrimiento, hacen surgir no solo aplicaciones prometedoras (promesas de reducción de accidentes de tráfico, mejores diagnósticos médicos e incluso reducción de la criminalidad). Vienen, asimismo, acompañadas de los temores de la doctrina, que se plantea, de nue-

vo, con esta nueva oleada de descubrimientos técnicos, si no estaremos necesitados de nuevos objetos de protección [980].

Se explorará, en esta sección, cuál es la discusión en torno a los bienes amenazados por las tecnologías de IA [981]. Se hará en un sentido no constreñido al derecho penal propiamente dicho, sino abriendo algo más el horizonte y fijándonos esencialmente en la discusión generada desde el constitucionalismo, pues, pese a que no acogemos una equivalencia entre el objeto de protección del derecho constitucional y el derecho penal (como sí realizan, por otra parte, aquellos abiertamente partidarios de reconocer una conexión entre la teoría del bien jurídico y los derechos reconocidos en la Constitución), no podemos dejar de ser conscientes de que una cierta influencia será inevitable y que aquellos derechos que por su relevancia ocupan un papel en las cartas magnas, de alguna forma acabarán influyendo, de manera inevitable, la protección penal, aunque no sea de forma directa, partiendo de la idea de que el texto constitucional y, esencialmente, los principios constitucionales, pueden servir tanto para aclarar el alcance de los bienes jurídicos protegidos, como la interpretación sobre el alcance de los preceptos penales [982].

---

[980] Una clara exposición sobre la evolución de la discusión sobre la aparición de nuevos bienes jurídicos a lo largo de las diferentes etapas de la evolución tecnocientífica, desde los delitos contra la salud pública hasta las manipulaciones genéticas, puede encontrarse en Solari Merlo, M. N., «Desarrollo tecnocientífico y bienes jurídicos. Cuando la tecnología revela lo importante», *ADPCP*, VOL. LXXVII, 2024, pp. 255-303.

[981] Cabe matizar que no se va a explorar la problemática desde el punto de vista, ya largamente explorado, de la ciberdelincuencia, pues se considera que no son exactamente los mismos fenómenos los que se están tratando de caracterizar en uno y otro caso. Sobre el cibercrimen, largamente desarrollado y discutido desde la promulgación del Convenio de Budapest, muchos son los autores españoles que han planteado la opción de crear nuevos bienes, tales como la seguridad en la red. La discusión ha tenido sus propios contornos, que, si bien pueden presentar alguna similitud, no deben mezclarse en la discusión aquí mantenida. Así, por todos, se menciona la reciente obra exhaustiva y recopilatoria de la doctrina de López Gorostidi, J., *Ciberdelincuencia: proporcionalidad y bienes jurídicos protegidos*, Comares, Granada, 2023. Una síntesis de su pensamiento que, en lo esencial, compartimos, lo encontramos en su reciente trabajo de 2025. Aunque el ciberespacio como espacio delictivo implica una redefinición en las formas de afectación, esto no significa que la protección penal en el mundo virtual deba implicar la creación de nuevos bienes jurídicos. Algunos apoyan esta idea basándose en normativas internacionales como el Convenio de Budapest (que apunta hacia bienes tales como la confidencialidad, integridad y disponibilidad de datos y sistemas). Sin embargo, como no todas las categorías de ciberdelitos se agrupan en torno a un mismo bien jurídico, se sugiere que esta protección puede ser intermedia, orientada a defender bienes jurídicos tradicionales. Además, los principios fundamentales del derecho penal desaconsejan ampliar la protección penal con cada nueva modalidad delictiva. López Gorostidi, J., «Sobre la posibilidad de comisión de ciberdelitos contra la vida a la luz de la teoría de la imputación objetiva» *RGDP*, 43, 2025, pp. 13-14. Para mayor profundización criminológica sobre la ciberdelincuencia en relación con las nuevas tecnologías: Miró-Llinares, F., *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*, Marcial Pons, Madrid, 2012. Insistiendo en la relación de la ciberdelincuencia con el paradigma de la sociedad del riesgo, recientemente: Anarte Borrallo, E., «Incidencia de las nuevas tecnologías en el sistema penal: aproximación al Derecho penal en la sociedad de la información», *Derecho y conocimiento: anuario jurídico sobre la sociedad de la información y del conocimiento*, 1, 2001, pp. 191-257.

[982] En este sentido, nos remitimos a lo ya explicado respecto de la fundamentación constitucional del bien jurídico en nota 926. También, recientemente, en profundidad: De la Fuente Cardona, F., *La teoría del bien jurídico y su función de crítica legislativa. Una contribución a la discusión sobre la evaluación constitucional de las normas penales,* Marcial Pons, 2025.

La velocidad propia del desarrollo de las nuevas tecnologías obliga, asimismo, a atender a la discusión de la forma más actual posible, pues lo cierto es que cada nuevo desarrollo presenta riesgos de contornos novedosos, y es esta novedad de las más recientes formas de aparición de los daños la que justifica que nuestro estudio se vaya a centrar en algunas de las obras más recientes. Así, se explorará la monografía que Valls dedica en exclusiva a la protección de bienes jurídico-penales ante los ataques propios de la IA (en un sentido estrictamente penal), para ponerla en relación con los trabajos de hasta siete autores del campo del derecho constitucional que en un reciente trabajo publicado en 2025 responden a un cuestionario dirigido específicamente a explorar la interconexión entre la inteligencia artificial y los derechos fundamentales. El resultado del trabajo, publicado en forma de encuesta donde se exploran hasta seis cuestiones, tiene la ventaja de, por su actualidad, tomar en cuenta la regulación europea ya en vigor. Se pondrán, de esta forma, en diálogo, las dos monografías de un penalista y un constitucionalista (los profesores Valls Prieto [983] y Presno Linera [984]), y el resultado de la encuesta respondida por Ana Aba Catoira, Francisco Balaguer Callejón, Lorenzo Cotino Hueso, Mario Hernández Ramos, Miguel Presno Linera, Lucrecio Rebollo Delgado y José Tudela Aranda [985].

## 2.1 **¿Nuevos bienes jurídico-penales?**

El planteamiento general que ofrece Valls parte de que la mayoría de los estudios realizados hasta el momento en torno a la IA se enfocan demasiado

---

[983] Valls Prieto, J., *Inteligencia artificial…, op cit*. N.46. Una versión algo reducida, esencialmente igual, en: Valls Prieto, J., «La inteligencia artificial frente a la colectividad..., *op. cit.* N. 927. Solo se van a comentar, esencialmente, las secciones II y la V de la monografía, que tratan los nuevos supuestos de riesgo y la protección de los derechos humanos mediante el derecho penal, respectivamente, cuyo desarrollo se actualiza en el artículo de 2023. Cuando no se diga otra cosa, las referencias serán al segundo de los trabajos aquí citado, por entender que recoge la visión más actualizada del autor, en un momento en que la normativa europea estaba ya a punto de ver la luz.

[984] Presno Linera, M. A., *Derechos fundamentales e inteligencia artificial*, Marcial Pons, Madrid, 2022.

[985] Todos ellos renombrados profesores de derecho constitucional de distintas universidades del entorno geográfico español, incluyendo, finalmente, al politólogo y doctor en Derecho José Tudela Aranda. El trabajo se publica en la revista *Teoría y Realidad Constitucional*, núm. 55, 2025, pp. 11-86. Debemos remitirnos, en este punto, a la reciente monografía de uno de los autores de la encuesta (Mario Hernández Ramos), pues en ella profundiza en todos los aspectos relevantes desde el punto de vista del derecho constitucional y la democracia. Especialmente interesante el pasaje relativo a cómo la IA impacta en la democracia representativa, tanto en lo relativo al condicionamiento de la formación de la opinión pública, como en cuanto a la posibilidad de que la IA dé lugar a lo que se ha bautizado como formas de «democracia aumentada». Hernández Ramos, M., *El impacto de la inteligencia artificial en el derecho constitucional*, Centro de Estudios Políticos y constitucionales, Madrid, 2025, pp. 159 y ss.

en su impacto en los derechos o bienes individuales, descuidando el análisis de la dimensión colectiva. Parte de un reconocimiento del concepto de bien jurídico, que, aunque nace imbricado en el iusnaturalismo y el idealismo alemán, pasa a verse influenciado por el positivismo constitucional y el neoconstitucionalismo. Reconoce, empero, un punto de consenso, pues ambos enfoques anclan de alguna forma los bienes jurídicos en los principios constitucionales [986].

Tras esta aclaración inicial, y sin llegar a tomar partido acerca de la teoría del bien jurídico acogida, parece que parte de un consenso apenas esbozado cuando afirma que *«independientemente de si tenemos un acercamiento iuspositivista, iusnaturalista o iussociológico, podemos afirmar que la doctrina comparte la existencia de bienes jurídicos penales que protegen intereses colectivos, con ayuda o no de un bien jurídico individual»* [987]. El verdadero interés de Valls reside, pues, en hacer un acercamiento a los bienes colectivos amenazados con la IA.

Comienza entonces una exposición donde hace un repaso por todos los posibles bienes jurídicos colectivos afectados por la IA, destacando varios ámbitos, usando ejemplos de supuestos ya acaecidos, que enmarca dentro de las categorías propias de la parte especial (aunque él agrupa en tres grandes secciones: bienes jurídicos que protegen a la sociedad, que protegen al Estado y que protegen la Constitución).

Dentro del primero de estos grupos, habla de una clara afectación a los derechos de los trabajadores en conductas cada vez más generalizadas de vigilancia extrema en el puesto de trabajo, discriminación en la contratación, e incluso trae a colación supuestos de sistemas de IA que, en su empeño por evaluar a trabajadores (en este caso del sector docente), acaban por favorecer una evaluación sesgada que conlleva a la pérdida de empleo basada en motivos discriminatorios [988]. Concluye Valls que se ve afectado, así, el bien jurídico protegido por los delitos contra los derechos de los trabajadores [989].

---

[986] Alonso Álamo, M., «Bien jurídico penal…», *op. cit*. N. 926, p. 66. Se hace cargo Valls de la tensión doctrinal existente entre los que consideran que solo los derechos expresamente reconocidos por la Constitución pueden ser bienes jurídicos (González Rus) y quienes creen que puede haber bienes jurídicos más allá de lo constitucional (Hefendehl, Silva).

[987] Valls Prieto, J., «La inteligencia artificial frente a la colectividad..., *op. cit.* N. 927, p. 6. Igual en Valls Prieto, J., *Inteligencia artificial…, op cit*. N.46, pp. 90-91.

[988] El ejemplo se toma de la obra de O'Neill, C., *Weapons of math destruction, op. cit. N. 543*, pp. 135-137.

[989] Concretamente afirma Valls que el uso de sistemas para la valoración, vigilancia, gestión, clasificación y, en definitiva, control extremo de los trabajadores, afectará a sus condiciones de trabajo de una forma ya protegida por el artículo 311 del CP, es decir, suponiendo un abuso de las condiciones laborales mediante la supresión o restricción de derechos que implican un trato digno. Valls Prieto, J., «La inteligencia artificial frente a la colectividad..., *op. cit.* N. 927, p. 8, con referencias a: Terradi-

El segundo de los ámbitos afectados es el orden socioeconómico, donde el autor se encarga de mencionar conductas con diferente impacto y gravedad, que van desde la publicidad personalizada que sesga las elecciones de los consumidores (el ejemplo del algoritmo de Amazon que beneficia sus propios productos en perjuicio de los de la competencia [990]), hasta la utilización de algoritmos de negociación en el mercado financiero, donde la información se ve debilitada por la comunicación autorreferencial entre algoritmos, dando a entender que se producen daños, si bien de variable gravedad, tanto para los consumidores que ven mermadas sus capacidades de decisión racional en los distintos mercados, como para los propios competidores, debilitándose paulatinamente la libre competencia [991]. En definitiva, bienes protegidos tradicionalmente por los delitos que protegen el buen funcionamiento del mercado y la libre competencia, sin que los tipos penales previstos sean capaces de abordar estas nuevas formas de ataque. Aunque Valls no lo pone de manifiesto, sí nombra algunos de los tipos penales cuya aplicación podría plantearse en estos supuestos, desde los tipos de abuso de información privilegiada, hasta los delitos de falsedad en la publicidad. Tipos todos ellos, como ocurre en su mayoría en los delitos que protegen el orden socioeconómico, dolosos, cuya modalidad de ataque no parece responder bien a los supuestos narrados por el autor.

Dentro de los bienes que protegen al Estado (entendemos, en un sentido institucional) Valls habla de la afectación a la Administración pública y la Ad-

---

llos Basoco, J. M.: «Delitos contra los derechos de los trabajadores: veinticinco años de política legislativa errática» *Estudios Penales y Criminológicos*, 41, 2021. Otros autores han destacado, asimismo, la posible vulneración de otros derechos fundamentales en el ámbito laboral con el paradigma de la IA, siendo la discriminación uno de los más invocados. Fundamental, en este sentido: Abadías Selma, A., «La protección penal de los trabajadores frente a la inteligencia artificial en el ámbito del delito de discriminación laboral»; *Revista General de Derecho Penal*, 39, 2023, señalando especialmente en pp. 27 y ss. las dificultades de aplicación del artículo 314 CP en este ámbito. En todo caso, cabe señalar que no es en absoluto pacífica la regulación de los denominados delitos contra los trabajadores, habiéndose señalado por una parte de la doctrina que pecan, en muchas ocasiones, de excesivos e indeterminados. En este sentido, Lascuraín Sánchez, J. A., «Los delitos contra los derechos de los trabajadores: lo que sobra y lo que falta», *ADPCP,* VOL. LVII*,* 2004, pp. 19-52.

[990] Valls Prieto, J., *ibid.,* p. 13: «Amazon determina el precio de sus productos mediante un algoritmo que en un 80% de los casos beneficia a sus productos en contra de los de la competencia, así como las compañías de vuelos también discriminan sus precios para un mismo vuelo y usuario dependiendo de la información de compra que tiene del usuario». Más referencias en: Angwin, J./ Mattu, S., «Amazon Says It Puts Customers First. But Its Pricing Algorithm Doesn't», *ProPublic*a, 2016, disponible en: < https://www.propublica.org/article/amazon-says-it-puts-customers-first-but-its-pricing-algorithm-doesnt > [Accedido por última vez el 28/01/2026].

[991] Valls Prieto, J., *ibid.,* p. 10. Relata el autor cómo ya se han producido situaciones en que una empresa ha etiquetado erróneamente en sus bases de datos a otra como autora de un delito, con la consecuencia de que se rechazaran sus solicitudes de préstamo, denegándose al acceso a créditos porque la información había sido vendida a otros corredores de datos sin posibilidad de control. En detalle, sobre este caso (ChoicePoint): Pasquale, F., *op cit.* N. 539, p. 33.

ministración de justicia por la introducción de sistemas de IA para la asistencia en las decisiones públicas, señalando en este caso que delitos como la prevaricación y el cohecho protegen precisamente que no se produzcan ese tipo de resoluciones injustas. Se pone como ejemplo el célebre caso del algoritmo Syri [992], sistema de apoyo a la decisión pública de concesión de ayudas sociales en Holanda, en el que se descubrió que el sistema discriminaba a los inmigrantes, impidiéndoles recibir estas prestaciones esenciales. Si se trata de proteger la integridad de la función pública, la imparcialidad en la toma de decisiones y, en esencia, la legalidad en el uso de la función pública, la IA supone una nueva forma de ataque de estos bienes jurídicos tradicionales.

No menciona Valls, empero, pues parece implícito en su razonamiento, que los delitos a los que él directamente apunta (prevaricación, malversación y cohecho), no serán aplicables en la mayoría de estas decisiones [993]. Una vez más, a nuestro juicio, Valls ofrece una caracterización de las formas en que estos bienes se ven atacados, pero muestra la insuficiencia de los tipos penales existentes.

Finalmente, en el tercer bloque de bienes menciona aquellos que protegen la Constitución, entre los que destaca la dinámica de discriminación sistémica que introducen algunas prácticas de IA, en todos aquellos sectores donde se han empezado a implementar. Habla de la posibilidad de discriminaciones en el entorno de la salud, la justicia, los seguros y los préstamos [994], tanto en el sector público como en el privado. Se pone en cuestión que, habida cuenta de la protección penal de la discriminación en el Código Penal, sea posible obtener una protección suficiente ante casos de denegación de servicios esenciales o de resultados discriminatorias patentes [995].

---

[992] Sobre este famoso caso y sus implicaciones, Oubiña Barbolla, S., «Límites a la utilización de algoritmos en el sector público. Reflexiones a propósito del caso SyRI», en Barona Vilar, S., (Dir.), *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, Tirant lo Blanch, Valencia, 2021, pp. 655-683.

[993] Reconoce, tan solo, que se abre la puerta a la prevaricación judicial, pero «*dependiendo del conocimiento que tenga el juez de la injusticia de su decisión*». Valls Prieto, J., «La inteligencia artificial frente a la colectividad..., *op. cit.* N. 927, p. 16.

[994] Aunque en este punto hemos de reconocer que las referencias a ejemplos empleados por el autor se refieren, en algunos casos, a técnicas que, pese a estar fundamentadas en el análisis masivo de datos para inferir correlaciones, no se puede afirmar que lleguen a constituir una IA en el sentido acogido por las definiciones más recientes, y tampoco el que se adopta en este trabajo. Es el caso del ya célebre ejemplo del programa COMPAS, usado para determinar la peligrosidad de los delincuentes en EEUU, ejemplo usado ya como paradigmático en la doctrina para mostrar los efectos perniciosos del sesgo algorítmico. Angwin, J./ Larson, J./Mattu, S/ Kirchner, L., *op. cit*. N. 724.

[995] Así, explica Valls (*Inteligencia artificial, derechos* humanos…*op. cit*. N. 983, pp. 94-95) que, aunque la discriminación en genérico se protege penalmente mediante los delitos tipificados en los arts. 511 y 512, son delitos escasamente aplicados e interpretados (por todos: Garrocho Salcedo, A., «Delitos de denegación de una prestación por motivos discriminatorios de los artículos 511 y 512 CP», *Eunomía. Revista en Cultura de la Legalidad*, 5, 2013, pp. 240-247), destacando este autor como principal inconveniente de los delitos existentes que, además de no incluir todos los tipos de discriminación posibles (dejando fuera, por ejemplo, la discriminación económica), *«no abarca toda discriminación que pueda incluir la protección de*

Tras esta exposición, en la que el autor deja claro que los ataques a bienes que vienen protegiéndose por el derecho penal son una realidad en multitud de sectores, Valls sugiere que quizá con la IA algunos de los bienes jurídicos tradicionalmente individuales (como es la privacidad), adquieran una dimensión que escape de ese individualismo tradicional, para pasar a adquirir una dimensión colectiva [996]. El autor afirma que los bienes jurídicos colectivos se ven afectados de forma clara, pero de esta afirmación deriva una premisa que nos parece difícilmente compartible, como es que todos estos ataques generan un impacto lo suficientemente grande como para justificar la intervención penal [997]. En este punto hemos de precisar ya que, a nuestro modo de ver, una cuestión es que haya ataques más o menos intensos a los bienes arriba expuestos, de cuya realidad cabe difícilmente dudar habida cuenta de los numerosos ejemplos, pero de ello no deriva, necesariamente, siempre y en todo caso, la necesidad de responder penalmente, pues no nos parece que todos los ejemplos mencionados tengan una gravedad comparable. Por poner un ejemplo: no se entiende que pueda justificar la intervención penal el hecho de que Amazon anuncie sus productos por encima de los de los competidores (por mucho que, en su caso, esto pudiera constituir una infracción de la normativa de competencia), pero sí nos parece mucho más grave el hecho de que un ciudadano pueda quedar excluido de un seguro de vida o salud porque los datos usados para entrenar al algoritmo se basaban en información sesgada, dando lugar a una discriminación con consecuencias fatales.

Ejemplifica el autor el paso del bien jurídico individual al colectivo hablando de la privacidad, que está considerada como un bien jurídico individual que es disponible por la propia persona. Sin embargo, apunta, en la actualidad, con el uso de grandes cantidades de datos y de sistemas inteligentes, ya no sólo se afecta únicamente a las personas individuales, pues al compartirse la información, que en principio se considera individual, *«se afecta a colectivos, por*

---

*una convivencia plural y multicultural ni la protección de grupos determinados, puesto que se protege sólo al individuo»*. Propone el autor una interpretación más amplia, de forma que la discriminación penalmente punible abarcase también los casos en que se produce una discriminación de un grupo de personas por la creación de perfiles algorítmicos. Finalmente, y para aquellos contextos en que no esté expresamente prevista, sugiere Valls la aplicación de la agravante genérica del artículo 22.4. En todo caso, cabe decir que, por mucho que fuera deseable una interpretación más amplia, lo cierto es que, a nuestro modo de ver, ni los tipos actuales ni la agravante del artículo 22 permitirían castigar los casos aquí descritos, donde la dinámica discriminatoria no solo se da en ámbitos muy variados y por motivos no expresamente previstos en la ley, sino con una lógica subjetiva muy alejada de la aplicación tradicional.

[996] Valls Prieto, J., «La inteligencia artificial frente a la colectividad..., *op. cit.* N. 927, p. 23.

[997] Implícitamente, afirmando que es necesaria una nueva redacción de los delitos que protegen la intimidad y la privacidad mediante la inclusión de la modalidad imprudente en aquellos tipos en que no esté prevista. *Ibid.,* p. 24.

*el análisis que se puede hacer de ellos y los modelos con que trabajan la inteligencia artificial»* [998].

No podemos estar del todo de acuerdo con esta conclusión, que nos parece apunta hacia un lugar adecuado, pero yerra en su conclusión. La dinámica de creación, entrenamiento y uso de una gran mayoría de sistemas de IA (esencialmente de todos aquellos que vienen a tomar decisiones que pueden afectar a derechos de los ciudadanos en forma de asistentes de decisión humana), hace que no sea posible esperar que el uso de datos que contengan información valiosa (sobre la privacidad o intimidad de una persona), vaya a quedarse circunscrita al ámbito personal, sino que servirá de base para la toma de decisiones que afectarán a personas ajenas. Estas decisiones tendrán repercusiones globales, llegando a afectar, de hecho, a colectivos indeterminados de personas (en efecto, quizá los afectados son los acusados racializados como en el caso COMPAS, o quizá lo es el colectivo de personas inmigrantes, como ocurrió en el caso Syri), por lo que la intuición sobre el impacto colectivo no es en absoluto desdeñable. Esto no implica, empero, que la privacidad cambie en su naturaleza, pues en esencia lo que se trata de proteger es el derecho de cada una de las personas que (pertenezcan a un colectivo discriminado o no) son expuestas a decisiones injustas. No es que la privacidad deje de ser individual, es que la forma de ataque a la misma presenta otros contornos, esencialmente colectivos. No hay un nuevo bien que proteger, sino el mismo bien, solo que la forma de aparición del ataque al mismo es diferente a la que se venía produciendo.

La idea de la dimensión colectiva no es exclusiva de Valls [999]. Pero merece la pena finalizar el repaso por esta influyente monografía realizando un comentario

---

[998] *Ibid.,* p. 23. Se matiza que el hecho de compartir un dato en principio individual afecta no solo al individuo, sino al grupo, hablando de descendientes y ascendientes, o a varias generaciones; afirmando que por ello la privacidad debe cambiar y ser concebida como un bien global (quizá en relación con la interconexión). Trae a colación el caso de *Cambridge Analytica* como un ejemplo de ataque más importante a la esfera íntima de los últimos años y que quedó, sin embargo, impune. Afirma Valls que «*Hay que replantearse seriamente una reformulación del bien jurídico privacidad para protegernos de estas nuevas amenazas por el uso de la tecnología que afectan directamente a nuestros derechos más personales como ser humano.*» Cabe resaltar que el caso citado tuvo, de hecho, una importante repercusión en España, dando lugar a la introducción del artículo 58 bis de la Ley Orgánica de Régimen Electoral General, tras la aprobación de la LOPD, que permitía recopilar datos personales de opiniones políticas libremente expresadas en redes sociales para usarlos en campañas electorales, aunque finalmente fue suprimido en parte por STC 76/2019, de 22 de mayo, exigiendo el consentimiento expreso del ciudadano. Así, Castillo Parrilla, J. A., «Algunos retos éticos, y sobre todo jurídicos, de la inteligencia artificial», *Anales de la Cátedra Francisco Suárez*, 57, p. 318.

[999] Así, Véliz, C., *Privacy is power: Why and how you Should take back control of your data*, Bantam Press, Londres, 2020. También, en la recensión a la obra de Valls, Castillo, ibid: «*las reflexiones del Profesor Valls en este punto son un buen punto de partida de cara a la necesidad de repensar el funcionamiento de un derecho fundamental (intimidad, privacidad, protección de datos; entendámoslos conjuntamente a estos efectos) que ha pasado de ser fundamentalmente un derecho de no injerencia a ser algo parecido a un «derecho ambiental»*. Queda mucho por desarrollar en relación con la creación de «*zonas de bajas emisiones contaminantes de consentimientos al tratamiento de datos, pero es un buen comienzo*».

sobre las propuestas normativas. El autor concluye dejando apenas apuntadas las soluciones ante los ataques arriba caracterizados, que se pueden resumir en las siguientes ideas: algunos bienes están correctamente protegidos en el CP, así, la discriminación y el derecho a la igualdad (art. 511, 314, 153.1 y agravante 22.4). No, así, la «autonomía de las personas», la dignidad personal (sin protección directa) [1000], el derecho a la libertad de expresión en las redes y el derecho de reunión, y que respecto de la vida y la integridad física se hace necesaria una regulación algo más sofisticada. Las carencias más importantes se sitúan, empero, respecto de los bienes de carácter económico, (publicidad, protección del mercado, libre competencia, derechos de los trabajadores). En los delitos contra la administración de justicia y contra los derechos fundamentales se propugna directamente el castigo de la modalidad imprudente, y respecto de la administración pública también se detectan carencias de protección. Finalmente, en cuanto a la administración pública, analiza la eventual incidencia que la inteligencia artificial puede tener en relación con los delitos de prevaricación, cohecho, tráfico de influencias, malversaciones, y delitos cometidos por los funcionarios, respecto de los que aprecia también carencias de protección que deberían ser corregidas [1001].

---

Castillo Parrilla, J. A., ibid., pp. 319-320. En este sentido, destacadamente, Isabel Turégano, que reflexiona sobre la privacidad como un bien colectivo estructural para las democracias digitales, en la medida en que el procesamiento masivo de datos reconfigura el reparto de poder, condiciona la igualdad en el acceso a oportunidades y moldea la propia esfera pública, de modo que su protección ya no puede articularse solo como autogestión individual, sino mediante normas y controles institucionales de los sistemas sociotécnicos. Vid.: Turégano, I., «Los valores detrás de la privacidad», *Doxa,* 43, 2020. Sin embargo, puede objetarse que esta ampliación conceptual tiende a absorber bajo la rúbrica de «privacidad» afectaciones que corresponden primariamente a otros derechos fundamentales –autonomía en su dimensión de libre desarrollo de la personalidad, no discriminación, derechos de participación política, debido proceso–, cuya creciente exposición a riesgos informacionales justificaría, más que un cambio de naturaleza individualcolectiva de la privacidad, un reforzamiento general de la tutela pública y de la eficacia frente a particulares del catálogo clásico de derechos fundamentales. No obstante, hemos de reconocer que esta posición va ganando adeptos y que el propio Miró ha reconocido, recientemente, que la concepción de privacidad como derecho puramente individual debe, quizá, modificarse, por las *«implicaciones que su control y uso por parte de instituciones o individuos puede tener para el conjunto de la sociedad»*. Es este el único de los puntos que, por los motivos arriba señalados, nos hace diferir de la visión de conjunto del autor expresada en uno de sus más recientes trabajos, que en lo demás, compartimos: Miró-Llinares, F., «Derecho penal y desafíos de la inteligencia artificial (en el contexto del nuevo marco regulatorio europeo)» en Casas Baamonde, M. E., (Dir.), *Derecho y tecnologías*, Fundación Ramón Areces, 2025, p. 210.

[1000] Veremos cómo este aspecto es uno de los más señalados en las propuestas de los constitucionalistas, inmediatamente *infra.*

[1001] La obra, recensionada por diversos autores, entre los que se cuenta el autorizado González Rus, convence en cuanto al análisis, pero no en cuanto al diagnóstico, mostrándose reticente este último en lo relativo a la creación de tipos penales que califica de «*ad hoc*» para superar las dificultades de la dinámica comisiva. No estamos, sin embargo, de acuerdo con la equiparación que hace González Rus entre la propuesta de Valls y la propuesta de un «derecho de las máquinas», pues Valls parte en todo caso de que los bienes a proteger no son, en lo esencial, novedosos, por mucho que afirme la existencia de una pretendida dimensión (colectiva) novedosa y discutible. González Rus, J. J., «Recensión del libro de Javier Valls Prieto, Inteligencia artificial, Derechos Humanos y bienes jurídicos», *Revista electrónica de ciencia penal y criminología*, 24, 2022, p. 10.

Considera, por tanto, que será necesario crear nuevos tipos penales en algunos sectores, incorporando nuevos tipos de peligro, pero haciendo especial hincapié en la necesidad de llenar los vacíos señalados acudiendo al castigo de la modalidad imprudente en los delitos arriba mencionados, y redefiniendo el deber de cuidado en relación con el uso ético de la IA [1002].

## 2.2 **IA y derechos fundamentales**

Otro reciente trabajo, esta vez de la mano de un constitucionalista con largo recorrido en el estudio de la IA, nos permite acercarnos, también mediante problemas ya acaecidos, a la afectación que la IA presenta, esta vez desde la perspectiva de los derechos fundamentales. Presno Linera presenta, así, la afectación en hasta siete bloques de derechos [1003].

En cuanto a la afectación de la IA sobre los derechos a la libertad personal y a un juicio justo, Presno se refiere a las implicaciones que para estos derechos suponen los usos de IA en el ámbito policial y judicial, explicando la problemática del sesgo implícito en estas decisiones. Resalta el autor también la dimensión relacionada con la explicabilidad de las actuaciones, destacando la necesidad de favorecer los modelos explicativos de sus decisiones, pues es la única forma de huir de correlaciones espurias [1004].

En lo relativo a los derechos relacionados con la intimidad (propia imagen, protección de datos o secreto de las comunicaciones), se acude a casos ya sucedidos en que grandes empresas tecnológicas, a partir de la información masiva extraída de los datos agregados de los consumidores, se explotaban con fines comerciales, para personalizar anuncios [1005]. La idea que parece poner de manifiesto Presno en esta sección es que los tribunales han comenzado ya a

---

[1002] Valls propone, en este punto, (Valls Prieto, J., «La inteligencia artificial frente a la colectividad..., *op. cit.* N. 927, p. 24), aprovechar los principios de las regulaciones éticas para determinar claramente cuáles son los elementos que se deben tener en cuenta para considerar un buen uso de la inteligencia artificial y, por ende, determinar qué debemos entender por el deber de cuidado, que va a permitirnos establecer cuándo se ha realizado una utilización peligrosa de esta tecnología. Un desarrollo más completo de la metodología a emplear para unir los principios éticos con los problemas jurídicos derivados del uso de sistemas de IA, en: Valls Prieto, J., «Sobre la responsabilidad penal …», *op. cit*. N. 48, p. 23 y ss.

[1003] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984.

[1004] Ibid. p. 36, con cita a Miró Llinares, F./ Castro Toledo, F. M., «¿Correlación no implica causalidad? El valor de las predicciones algorítmicas en el sistema penal a propósito del debate epistemológico sobre «el fin de la teoría», en Demetrio Crespo, E./ de la Cuerda Martín, M./ García de la Torre, F. (Coords.), *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant Lo Blanch, Valencia, 2022, pp. 507-530.

[1005] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 40. El ejemplo se extrae de un caso real protagonizado por Google, y descrito en la obra de Zuboff, S., *op. cit*. N. 553.

reconocer el potencial lesivo del paradigma de la vigilancia y de las redes sociales, por lo que puede entenderse que la intimidad (ya sea protegida mediante el paradigma más amplio del artículo 8 del CEDH [1006] o de las legislaciones nacionales [1007]), servirá como freno esencial y derecho protagonista en la era de la supervigilancia.

Continúa con los retos planteados para la libertad de expresión e información, señalando en este caso la situación de conflicto evidente, pues la IA está generando contenido capaz de determinar el comportamiento de los usuarios de redes, pero produciéndose, al mismo tiempo, un tenso equilibrio con derechos como la libertad de expresión e información, sin que sean pacíficas las opciones que abogan por restringir acceso a contenidos en la red [1008]. La necesidad de moderar contenidos en la red no es algo novedoso de la IA (pues el debate se planteó, esencialmente, con el Reglamento de Servicios Digitales [1009]), pero las cuestiones entonces planteadas vuelven con fuerza al debate, pues los riesgos se ven exacerbados, en tanto en cuanto no se trata ya de la

---

[1006] Se cita el caso Big Brother Watch y otros contra Reino Unido (caso Snowden), resuelto por STEDH de 25 de mayo de 2021, donde el TEDH condenó por violación del derecho a la vida privada y familiar y al secreto de las comunicaciones. *Ibid.,* p. 48.

[1007] Así, nuestro TC reconociendo el cambio de paradigma que suponen las redes sociales, en reciente sentencia STC (27/2020, de 24 de febrero), que en su fundamento jurídico 3 hace una exposición gráfica del alcance de las redes sociales, hablando del «*incremento de niveles de intercambio de contenidos*», donde los usuarios han pasado de ser meros consumidores a ser «*sujetos colaborativos, ciudadanos que ponen en común en redes de confianza lo que tienen»; […] «de modo que en pocas décadas ha pasado de ser un sujeto pasivo receptor de información a un sujeto activo que elabora, modifica, almacena y comparte información*». Y muy destacadamente, la reciente sentencia del BVerfG alemán, donde se condena a dos *länder* por el uso de vigilancia masiva basada en IA sin cumplir el requisito de proporcionalidad. Un comentario más extenso de este caso en Bravo Bolado, A., «El guardián de Gotham. …», *op. cit*. N. 57. También en: Cotino Hueso, L., «Los requisitos del Tribunal Constitucional alemán para el análisis automatizado de datos o con inteligencia artificial, que no se cumplen en España», *blog de la Revista Catalana de dret públic*o,12 de abril de 2023, disponible en: <https://eapc-rcdp.blog.gencat.cat/2023/04/12/los-requisitos-del-tribunal-constitucional-aleman-para-el-analisis-automatizado-de-datos-o-con-inteligencia-artificial-que-no-se-cumplen-en-espana-lorenzo-cotino-hueso/>[Accedido por última vez el 28/01/2026].

[1008] Se señala, en este caso, el debatido derecho de las personas a acceder a la red para comunicarse, discutido ya en sentencia Packingham vs. Carolina del Norte, resuelto por sentencia del Tribunal Supremo de Estados Unidos de 19 de junio de 2017, donde se trató la constitucionalidad de las limitaciones legales de acceso a redes a personas condenadas por abusos sexuales. Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 54. En nuestro territorio, de la dificultad que plantean este tipo de nuevas penas se ha encargado recientemente Gallego Arribas, D., «La prohibición de acceso a las redes sociales como consecuencia jurídica del delito. A propósito de la STS 547/2022, de 2 junio«, *Diario La ley*, 10332, 2023, pp. 1-11.

[1009] Nos referimos al: Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales). Analizando esta problemática: Santisteban Galarza, M., «Garantías frente a la moderación de contenidos en la propuesta de Reglamento único de servicios digitales», *CESCO*, 41, 2022, p. 162 y ss., poniendo de manifiesto el delicado equilibrio entre censura y control necesario de la información. También, más actual, en: Santisteban Galarza, M., «La criminalización de las ultrafalsificaciones», *op. cit.* N. 219, *passim*. También, monográficamente, Galán Muñoz, A., *Libertad de expresión y responsabilidad penal por contenidos ajenos en internet, Un estudio sobre la*

posibilidad de verter expresiones injuriosas en la red, sino que los algoritmos, al determinar qué ven los usuarios, suponen una fuente directa de manipulación del pensamiento y la libertad de información, e incluso de formación de la propia ideología, condicionando la creación de las llamadas «burbujas de filtros» [1010], que ponen en peligro el propio pluralismo político [1011].

Continúa Presno tratando la problemática de derechos como los de reunión, manifestación y huelga, poniendo de manifiesto cómo la IA se utiliza, en muchas ocasiones, como un medio de control de multitudes, aprovechando las ingentes cantidades de información de la que se alimenta, de tal forma que pone en marcha discriminaciones difícilmente controlables [1012]. El campo laboral ha sido ya objeto de este tipo de controles silenciosos, pues la empresa aprovecha los medios que ofrece la IA para ofrecer la imagen de una vigilancia aséptica, mientras que los trabajadores corren el riesgo de exponerse a decisiones adoptadas por algoritmos disfrazadas de neutralidad en su gestión efectiva de los recursos humanos [1013].

En cuanto a la prohibición de discriminación, la problemática del sesgo vuelve a traerse a colación, aunque lo cierto es que es este un derecho que se ve afectado en cada una de las decisiones donde la IA, de alguna forma, interfiere en la decisión del humano. Lo que destaca Presno en este sentido es la atención que el Reglamento de Inteligencia Artificial presta al sesgo, señalando que, precisamente con el objeto de evitar discriminaciones injustas, se han prohibido prácticas como la evaluación o clasificación de las personas aten-

*incidencia penal de la ley 34/2002 de Servicios de la Sociedad de la Información y el Comercio Electrónico*, Tirant lo Blanch, Valencia, 2010.

[1010] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 57, con cita a Stasi, M., L., «La exposición a la diversidad de contenidos en las redes sociales: entre la regulación o la desagregación en la curación de contenidos», *Teoría y Derecho*, 32, 2022, pp. 130-165.

[1011] Así ha sido señalado por el propio Consejo de Europa. Recomendación CM(Rec (2018) 1 [1], del Comité de Ministros del Consejo de Europea a los EEMM, abordando los problemas del pluralismo de los medios de comunicación, así como la transparencia sobre la propiedad de los mismos.

[1012] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, pp. 62-63.

[1013] Trae a colación Presno la STC 17/2017 de 2 de febrero, donde el Tribunal es consciente de los riesgos de sustituir el trabajo humano por la tecnología en un caso de «esquirolaje tecnológico», en cuyo voto particular (formulado por el magistrado Fernando Valdés Dal-Ré y firmado por Adela Asúa Batarrita y Juan Antonio Xiol Ríos) se resalta que «*No puede soslayarse el empleo y los efectos del uso de las nuevas tecnologías por parte del empresario para potenciar sus cauces de vigilancia y control*». Presno Linera, M. A., ibid., p. 63. Más ilustrativo nos parece, aún, el caso resuelto por el Tribunal de Bolonia, condenando a la empresa Glovo porque su algoritmo de reparto del trabajo perjudicaba a aquellos trabajadores que se habían ausentado de su puesto sin distinguir si se debía a motivos reconocidos por sus derechos laborales y constitucionales, considerando que el algoritmo estaba dando lugar a una discriminación indirecta. Sentencia dictada por el Tribunal de Bolonia, Sección de Trabajo, en fecha 31 de diciembre de 2020, (sentencia n. 2949/2020). Más información en: Fava, G., «L'ordinanza del Tribunale di Bologna in merito alla possibile discriminatorietà dell'algoritmo utilizzato da Deliveroo», *Lavoro Diritti Europa*, 1/2021, pp. 1-4.

diendo a su conducta social, así como la explotación de las vulnerabilidades de grupo [1014].

Finalmente trata la afectación de la IA al derecho de participación en procesos políticos, señalando una vez más la preocupación del Consejo de Europa sobre el riesgo que presentan las cámaras de eco para los procesos democráticos, señalando que, *«aunque el impacto real de las burbujas de filtros y de la desinformación selectiva en la formación de la opinión política es difícil de determinar con precisión, las cámaras de eco totalmente automatizadas plantean el peligro de crear «burbujas ideológicas»* [1015], y de los riesgos para el acceso a prestaciones sociales, apuntando de nuevo cómo unos datos de base sesgados han dado lugar a situaciones injustas para colectivos desfavorecidos en el acceso a ayudas a la infancia [1016].

Una vez realizado el repaso por todos los derechos constitucionales afectados, procede al autor a analizar algunas de las consecuencias que se pueden derivar de estas nuevas dinámicas de ataque a los derechos fundamentales. Podemos destacar tres ideas esenciales.

En primer lugar, Presno es consciente del debate acerca de la necesidad de incluir nuevos derechos en las constituciones al hilo de las nuevas tecnologías, especialmente el debate relativo a los neuroderechos [1017], apoyados no solo por parte de la doctrina [1018], sino también en algunas resoluciones nacio-

---

[1014] Sobre los principios que tratan de proteger los diferentes preceptos del Reglamento se profundizará *infra.* Baste dejar apuntado que, en su redacción, los preceptos que proscriben estas prácticas utilizan una terminología propia de las legislaciones antidiscriminación, hablando de que determinados usos provocan un trato perjudicial o desfavorable hacia determinadas personas o colectivos, pudiendo dar lugar a tratos injustificados o desproporcionados. La relación, por tanto, es más que evidente. Vid., especialmente, artículo 5 del Reglamento, *infra.*

[1015] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 69.

[1016] Se trata del conocido caso en que el algoritmo de Países Bajos da lugar a la caída del primer ministro ante el escándalo de sesgo. En España, el caso del algoritmo de asignación del bono social eléctrico (BOSCO) y la denegación de las autoridades de una respuesta sobre los factores que analizaba ha sido fuertemente criticado por la doctrina. Presno Linera, M. A., ibid., p. 77. La sentencia que resuelve la petición de información es Sentencia 143/2021, del Juzgado Central de lo contencioso administrativo n.º 8, de 30 de diciembre de 2021. Un estudio crítico sobre el impacto del algoritmo de Países Bajos en: Aministía Internacional, «El escándalo de los subsidios para el cuidado infantil en Países Bajos, una alerta urgente para prohibir los algoritmos racistas», 26 de octubre de 2021, disponible en: <https://www.amnesty.org/es/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/>[Accedido por última vez el 28/01/2026]. Una explicación crítica de ambos casos en Rivas Vallejo, P., *op. cit*. N. 57, pp. 45 y ss.

[1017] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 83. Se cita a Rafael Yuste como precursor de los mismos. Yuste, R./Genser, J./Hermann, S., «It's time for Neuro-rights», *Horizons*, Center for International Relations and Sustainable Development, 18, 2021.

[1018] Véase, por todos, la brillante exposición y defensa realizada por Pérez Manzano, M., «¿Un estatuto constitucional singular para el cerebro y las neurotecnologías? Sobre los neuroderechos», *RGDC*, 41, 2024, pp. 136-190. La autora se muestra partidaria de proteger el cerebro más allá de los marcos tradicionales de la intimidad y la integridad física, reconociendo la singularidad del cerebro como objeto de protección, pero diferenciando en función del modo de acceso a los datos, así como el tipo de datos objetivos, entre otros

nales y supranacionales [1019]. No se muestra, empero, el autor, favorable a la inclusión de nuevos derechos, sino más bien a la promoción o aplicación efectiva de los existentes, superando de una vez el enfoque basado en las meras declaraciones de derechos sin contenido normativo vinculante.

Esto nos lleva al siguiente punto, cual es la importancia de la dignidad como referente ético central de estas tecnologías. La dignidad aparece también como una de las principales afectadas cuando se trata de tecnologías prohibidas como la vigilancia masiva, y su importancia se ha puesto de manifiesto por Cotino [1020]. Se parte de la dignidad, en un sentido constitucional, como un mínimo invulnerable, concibiendo el libre desarrollo de la personalidad como la forma de maximización de los derechos [1021].

Finalmente, este autor pone de manifiesto la importancia que cobra la alusión a la doble dimensión de los derechos fundamentales, plenamente consolidada en la sentencia Lüth [1022], donde afirma que, junto a los derechos fundamenta-

---

motivos por los peligros que puede suponer hoy en día el tratamiento de datos mediante inteligencia artificial. Propone Pérez Manzano que en el marco de protección gradual del cerebro la última de las cautelas sea el reconocimiento de un derecho a la identidad neuronal que proteja el núcleo más íntimo de actividad mental.

[1019] En el ámbito europeo, la Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital, (2020/2266(INI)), párr. 247, pidiendo a la Comisión «*que estudie la posibilidad de presentar una iniciativa relativa a los neuroderechos, con el objetivo de proteger el cerebro humano contra la injerencia, la manipulación y el control por parte de la neurotecnología impulsada por la IA […]».* De las propuestas a nivel nacional destaca especialmente la impulsada por Chile, que en 2021 modificó su texto constitucional para incluir, en su nuevo artículo 19.1 (que tradicionalmente protegía el derecho a la vida e integridad física y psíquica) una referencia directa al resguardo de la actividad cerebral. Un desarrollo del proceso chileno en: Reche Tello, N., «Nuevos derechos frente a la neurotecnología: la experiencia Chilena», *Revista de Derecho Político*, 112, 2021, pp. 415-446. Finalmente, aunque sin impacto constitucional, cabe mencionar la aprobación de la Carta de Derechos digitales en España, el 14 de julio de 2021, que, aunque reconoce una amplia gama de derechos, no goza de fuerza normativa, como el propio texto reconoce.

[1020] Cotino Hueso, L., «Nuevo paradigma en la garantía de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial», en Cotino Hueso, L., (Dir.), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, pp. 69-105, reconociendo a la dignidad papel de fundamento jurídico y dogmático para abordar los retos que presentan las nuevas tecnologías, defendiendo la necesidad de una dimensión colectiva de los derechos, considerando el impacto en grupos y colectivos.

[1021] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 103, con base en STC 120/1990, de 27 de junio (Sentencia del caso GRAPO), que en su FJ 4 aclara que: *«Proyectada sobre los derechos individuales, la regla del artículo 10.1 C. E. implica que, en cuanto «valor espiritual y moral inherente a la persona» (STC* 53/1985*, fundamento jurídico 8.º), la dignidad ha de permanecer inalterada cualquiera que sea la situación en que la persona se encuentre […], constituyendo, en consecuencia, un minimun invulnerable que todo estatuto jurídico debe asegurar, de modo que, sean unas u otras las limitaciones que se impongan en el disfrute de derechos individuales, no conlleven menosprecio para la estima que, en cuanto ser humano, merece la persona. Pero sólo en la medida en que tales derechos sean tutelares en amparo y únicamente con el fin de comprobar si se han respetado las exigencias que, no en abstracto, sino en el concreto ámbito de cada uno de aquéllos, deriven de la dignidad de la persona, habrá de ser ésta tomada en consideración por este Tribunal como referente.»*

[1022] BVerfGE 7, 198 (*Lüth), de* 15 de enero de 1958. La emblemática sentencia del Tribunal Constitucional Federal alemán estableció que los derechos fundamentales no actúan solo como límites frente al

les como derechos subjetivos tradicionales frente a los poderes públicos, deben concebirse también como normas objetivas con un contenido que se irradia en todos los ámbitos del ordenamiento (incluyendo, entendemos, desde luego, el penal). Este efecto de irradiación se extiende a la función conformadora de Derecho por parte del legislador, a la actuación del ejecutivo en el ámbito de sus funciones y a la interpretación y aplicación de normas por los órganos jurisdiccionales. No presenta dudas, para Presno, que, habida cuenta de la importancia de los bienes en juego con estas tecnologías (tales como la salud y los derechos fundamentales), esta dimensión objetiva deba verse especialmente potenciada, viéndose proyectada también en las relaciones entre particulares, citando especialmente los ámbitos de la asistencia sanitaria, derechos del trabajador e incluso el consumo, por los riesgos que se introducen en la suscripción de seguros [1023].

La conclusión que alcanza Presno es que la IA puede cambiar el objeto de los derechos fundamentales, tanto en lo relativo al ejercicio del poder público (la IA permite nuevas formas de afectación a la vida privada), pero también las facultades que forman parte del objeto del derecho a no sufrir discriminación, destacando que estos dos sectores son algunos de los que corren mayor riesgo, pues la aplicación de sistemas de IA puede generar mayores desafíos, destacando que los mismos pueden darse en relación o conexión con otros derechos fundamentales, como el derecho a la libertad personal y prohibición de detenciones arbitrarias [1024].

### 2.3 **Los nuevos retos constitucionales de la IA**

Finalmente, exploraremos los resultados de la encuesta arriba citada [1025], que pretende abordar la relación entre la IA y los derechos constitucionales a

---

poder estatal (eficacia vertical), sino que también proyectan eficacia indirecta en el ámbito de las relaciones entre particulares, expresando un orden objetivo de valores que debe informar a todo el ordenamiento. Se concluye que una sentencia civil que prohibía al recurrente (el senador Lüth) promover el boicot de una película dirigida por un cineasta vinculado al nazismo vulneraba su libertad de expresión, aunque ambas partes implicadas fuesen actores privados. Los efectos de dicha sentencia han acabado irradiando otros ordenamientos, habiéndose reconocido por nuestro TC su doctrina. Así, en STC 25/1981, de 14 de julio que, en FJ 5, afirma que «*Los derechos fundamentales son así un patrimonio común de los ciudadanos individual y colectivamente, constitutivos del ordenamiento jurídico cuya vigencia a todos atañe por igual. Establecen por así decirlo una vinculación directa entre los individuos y el Estado y actúan como fundamento de la unidad política sin mediación alguna*».

[1023] Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, p. 112.

[1024] *Ibid.,* pp. 123-126. Entendemos que con esta última alusión hace Presno referencia a la posible vulneración que, sobre otros derechos sustantivos, puede producir una vulneración de la prohibición de discriminación. Piénsese, sin ir más lejos, en la posibilidad de que un sistema tome una decisión discriminatoria que, como consecuencia, lleve a la detención de una persona inocente.

[1025] Aba Catoira, A./Balaguer Callejón, F./ Cotino Hueso, L./Hernández Ramos, M./ Presno Linera, M./ Rebollo Delgado, L., / Tudela Aranda, J., «Encuesta sobre Inteligencia artificial y

partir de la formulación de seis preguntas [1026] cuyo contenido pretende indagar sobre: la incidencia general de la IA en los derechos fundamentales; la implicación de la IA en el proceso democrático; los posibles cambios en la dogmática de los derechos fundamentales; la posibilidad de incorporar nuevos derechos o nuevos instrumentos de garantía; y la opinión sobre el Reglamento en lo relativo a la protección de estos derechos. Trataremos brevemente cada uno de los bloques.

En cuanto al *impacto general de la IA en los derechos fundamentales,* todos los autores coinciden esencialmente en que la IA afecta de manera transversal a casi todos los derechos fundamentales, aunque con distinta intensidad. Los derechos especialmente afectados son la privacidad y protección de datos personales (por el uso masivo e indiscriminado de datos); la igualdad y no discriminación (por los sesgos algorítmicos), las libertades públicas (libertad de expresión, información, reunión y participación política [1027]); la tutela judicial efectiva y el debido proceso (por las decisiones automatizadas en entornos judiciales o administrativos) y, finalmente, la dignidad humana (en tanto la persona corre el riesgo de ser reducida a perfiles o datos).

En lo relativo a la *afectación del proceso democrático,* los riesgos identificados por los autores tienen que ver con la existencia de burbujas de filtro, cámaras de eco y generación de *deepfakes,* el *microtargeting* político (capaz de poner en entredicho el principio de igualdad en procesos electorales), la libertad de sufragio y confianza en los procesos electorales, el riesgo de manipulación subliminal, así como los peligros para la libertad de expresión cuando la información es filtrada por los sistemas sin que el usuario sea consciente.

---

Derechos fundamentales», *Teoría y Realidad Constitucional*, (55), 2025, pp. 11-86.

[1026] Las preguntas en cuestión son: 1. *¿Qué consideraciones generales puede hacernos sobre la incidencia de la inteligencia artificial en los derechos fundamentales reconocidos en la Constitución? ¿Hay algún derecho en particular o grupo de derechos en los que, a su juicio, esa incidencia tenga especial impacto?* 2. *De forma más concreta ¿qué reflexiones puede hacernos sobre las consecuencias de la inteligencia artificial en los derechos fundamentales implicados en los procesos democráticos y, por tanto, en esos mismos procesos?* 3. *Desde el punto de vista de la dogmática de los derechos fundamentales ¿Cree que la inteligencia artificial puede llegar a producir algún cambio en la forma en la que concebimos hoy los derechos fundamentales (su objeto, contenido, estructura)?* 4. *¿Considera preciso incorporar a la Constitución algún nuevo derecho fundamental para poder hacer frente a los riesgos que comporta la inteligencia artificial?* 5. *En un contexto de creciente presencia de los sistemas de inteligencia artificial ¿sería conveniente o necesario prever nuevos instrumentos de garantía de los derechos fundamentales? 6. Desde la perspectiva de los derechos fundamentales ¿qué opinión le merecen el Reglamento europeo de inteligencia artificial y el Convenio Marco del Consejo de Europa sobre inteligencia artificial y derechos humanos, democracia y Estado de derecho?*

[1027] Resulta en este sentido original el enfoque que realiza Callejón, con su interpretación de la nueva construcción o paradigma del espacio público, dando lugar a transformaciones culturales «de alcance imprevisible». *Ibid.,* p. 15. Para un mayor desarrollo de la original propuesta de este profesor, Callejón Balaguer, F., La constitución del algoritmo, Fundación Manuel Giménez Abad, Zaragoza, 2022.

Cotino pone de manifiesto la importancia no solo del Reglamento, sino de los avances generados ya por medio del Reglamento de Servicios digitales [1028]. Por su parte, Hernández da un paso más al proponer que sea la propia IA la que se utilice para combatir los efectos negativos de la propia IA (es decir, detección de *deepfakes* o bots) [1029].

En la pregunta relativa al *cambio en la concepción de los derechos* (objeto, contenido y estructura), la respuesta general es afirmativa, al menos destacando tres dimensiones: se detecta una dimensión colectiva o supraindividual de los derechos fundamentales, destacando que hay un daño social o estructural que excede al individuo [1030]. En segundo lugar, se requiere una reinterpretación de derechos como la privacidad, que deja de ser solo negativa (ausencia de intromisión), para tener una dimensión proactiva (de control sobre los datos [1031]); y finalmente se destaca la erosión de la distinción clásica entre derechos negativos y positivos, destacando que la IA impone nuevas obligaciones al Estado y también a los particulares. Balaguer observa, en este punto, que la Constitución, tradicionalmente interpretada como motor de progreso, comienza a verse con recelo o como obstáculo para los operadores del mercado global, mientras Hernández, en un sentido similar, señala la necesidad de regulación como freno al poder de las grandes empresas, y Aranda resalta la complejidad de la titularidad de los derechos frente a la relevancia de actores privados multinacionales [1032].

En un sentido similar al ya apuntado por Presno, es común entre los autores abogar por la adopción de la dignidad como un concepto operativo para articular nuevos límites, garantías y respuestas regulatorias.

---

[1028] *Ibid.,* p. 31, destacando que la *Digital Services Act* introduce obligaciones específicas de transparencia y responsabilidad para grandes plataformas, exigiendo implementar mecanismos de gestión de riesgos sistémicos, evitando tanto el modelo de control o dirección de contenidos público como el modelo de evaluación de su constitucionalidad por los gobiernos.

[1029] *Ibid.,* p. 32.

[1030] *Ibid.,* p. 42, señalando que: «*es necesario adoptar un enfoque que reconozca los efectos colectivos y que permita evaluar jurídicamente el daño social, una categoría análoga al daño medioambiental, difícil de aprehender desde una visión tradicional subjetiva. El uso masivo de datos y la capacidad de los algoritmos para crear perfiles y predecir comportamientos colectivos requieren mecanismos de protección que vayan más allá del caso concreto y aborden los efectos sociales del uso de IA*». En este sentido, respecto de la privacidad, MANTELERO, A., «Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection», *Computer Law & Security Review*, 32, 2016, pp. 238-255.

[1031] Así, destaca Aba Catoira la importancia, ya nada discutida, de la autoderminación informativa. ABA CATOIRA, A. ET AL., *op. cit*. N. 1025, p. 40.

[1032] *Ibid.,* pp. 50-51, señalando que uno de los desafíos más relevantes para la efectividad de los derechos fundamentales es que muchos de los nuevos riesgos y oportunidades están anudados a sujetos privados de carácter multinacional, cuestionando la relación jurídica sobre la que tradicionalmente se ha construido la teoría de los derechos y libertades entre ciudadano y poder público y manifestando que la importancia que adquieren los actores privados en el campo de la IA acaba afectando a la elaboración teórica de la titularidad de derechos.

A la pregunta relativa a la *necesidad de nuevos derechos,* la respuesta absolutamente predominante es la negativa. Se opta, en cambio, por la ampliación de contenidos de los derechos existentes, optando por una interpretación evolutiva, o, en todo caso, por la creación de derechos «instrumentales», como es la explicabilidad algorítmica (los términos son diversos, refiriéndose algunos autores al derecho a la revisión humana, o bien a la transparencia o explicabilidad), pero que, en todo caso, están orientados a fortalecer derechos clásicos ya reconocidos. Aunque algunos autores mencionan la discusión sobre los neuroderechos, consideran que ya están cubiertos por el marco actual [1033].

Sobre si se deben crear *nuevos instrumentos de garantía,* la respuesta es afirmativa en general, proponiéndose esencialmente la supervisión de entidades especializadas (en la línea de las agencias nacionales o europeas de supervisión), las evaluaciones de impacto, los enfoques que priman el diseño responsable (destacando las propuestas que hablan de la integración de valores constitucionales «*by design*» [1034]), y, finalmente, un esfuerzo de alfabetización digital que sitúe al ciudadano en posición de controlar y tener conocimiento efectivo sobre la operatividad de los sistemas [1035].

Finalmente, en cuanto a la pregunta sobre la *valoración de los dos instrumentos vinculantes a nivel internacional* (Reglamento de IA y Convenio del Consejo de Europa, respectivamente [1036]), la opinión es, en general, positiva, mencionando que el Reglamento establece mecanismos preventivos adecuados que proporcionan seguridad jurídica, mientras el Convenio del Consejo de Europa ofrece un marco mucho más flexible, con un enfoque más colectivo y centrado en los derechos humanos. Sí se señala, empero, que el Reglamento tiene un carácter marcadamente mercantilista, lo que limita su

---

[1033] Entrando a formar parte de la lista de personas que, si bien reconocen que las amenazas que plantean las nuevas tecnologías son reales, bastará con una interpretación adaptativa o evolutiva, capaz de integrar los derechos ya existentes con la nueva realidad. En esta misma línea, en el plano internacional: Ligthart, S. et al, «Forensic Brain-Readind and Mental Privacy in European Human Rights Law: Foundations and Challenges», *Neuroethics,* 14, 2021, pp. 191-203, en lo relativo a la adquisición no consentida de datos cerebrales.

[1034] Así, Hernández Ramos, en Aba Catoira, A. et al., *op. cit*. N. 1025, p. 71.

[1035] Llama aquí la atención la propuesta de Rebollo, en Aba Catoira, A. et al. *op. cit*. N. 1025, p. 73, que propone que los instrumentos de garantía de los derechos sean tecnológicos, siendo otras IAs las que, una vez más, garanticen la esencia de los derechos y libertades, desde la óptica del algoritmo jurídico encargado de controlar vigilar y revertir las posibles vulneraciones.

[1036] Consejo de Europa, *Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho*, 17 de mayo de 2024.

ambición garantista [1037]. Señala Cotino que *«el Convenio pone la lírica allí donde el RIA pone la prosa»* [1038].

En conclusión, el estudio analizado refleja un consenso doctrinal amplio: se parte del reconocimiento de que la IA reformula riesgos y patrones de afectación, ensancha el círculo de sujetos afectados (de individuos a colectivos), y requiere un nuevo marco garantista, no necesariamente mediante nuevos derechos sustantivos, sino a través de una interpretación y fortalecimiento de los existentes y de la creación de instrumentos de garantía efectivos. Según señala Balaguer, las garantías deben adaptarse a las transformaciones sociales y a las nuevas actividades potencialmente lesivas, y que la protección de la democracia y los derechos exige medidas de muy diverso tipo, que van desde la educación digital y la regulación de la competencia, pasando por medidas fiscales, sancionadoras e incluso, en casos extremos, de carácter penal [1039].

### 2.4 **Recapitulación**

Tras el análisis de estos trabajos, estamos en posición de extraer algunas conclusiones relevantes, con el objetivo de contestar, finalmente, a la pregunta que guiaba este apartado, a saber: ¿existen nuevos bienes jurídicos atacados por la inteligencia artificial?

En primer lugar, cabe destacar la existencia evidente de puntos en común:

a) La IA genera formas inéditas de afectación a derechos y bienes jurídicos preexistentes que son, además, absolutamente transversales.

Mientras Valls muestra la vertiente de los bienes jurídico penales afectados (privacidad, mercado, derechos de los consumidores, función púbica, no discriminación), Presno y los encuestados (entre los que se en-

[1037] Aba Catoira, A. et al., *op. cit*. N. 1025, p. 79. Se señala, de hecho, como una desventaja que no haya entrado a regular los sistemas militares o de seguridad nacional, y que se permita exportar fuera de la UE sistemas prohibidos. *Ibid.,* p. 82.

[1038] *Ibid.,* p. 79.

[1039] *Ibid.,* p. 68. Se refiere el autor, como ejemplo de caso extremo que podría justificar la entrada del derecho penal, el uso de propaganda subliminal en procesos electorales. De hecho, en la Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial (122/000011) presentada por el Grupo Parlamentario Sumar publicada en el BOE el 13 de octubre de 2023 se planteaba, entre otras cuestiones, la modificación de la LOREG con el objetivo de introducir un delito de distribución maliciosa de imágenes y voces generadas mediante IA durante el proceso electoral. Sin embargo, esta cuestión no ha sido introducida finalmente en ningún proyecto de ley oficial, estando en la actualidad legalmente paralizada, pese a las voces académicas que continúan poniendo el foco en la problemática.

cuentra, de nuevo, este último), ponen el acento en los cambios estructurales que introduce la IA en todos los derechos constitucionales ya reconocidos. Se trata, en efecto, de señalar los problemas desde dos ópticas estrechamente relacionadas, si se quiere, dos caras de una moneda. En definitiva, podríamos concluir afirmando que no se trata tanto de que la IA genere nuevos derechos o bienes jurídicos, sino que cambia la dinámica de afectación de los ya existentes.

b) Emergencia de una dimensión colectiva en derechos tradicionalmente individuales.

Mientras Valls acentúa la dimensión colectiva de la privacidad, Presno hace algo similar desde una perspectiva constitucional, identificando una reconfiguración del objeto de protección, especialmente en relación con la dignidad, la privacidad y la no discriminación, poniendo el foco en la interrelación entre unos y otros derechos, que ya no pueden interpretarse de una forma aislada. En el bloque de los entrevistados, una vez más encontramos una especial incidencia de la idea de la colectividad del daño.

c) Insuficiencia de los medios de protección (ya sean tipos penales o garantías constitucionales).

Mientras Valls expresa una necesidad de incluir nuevos tipos que sean capaces de reflejar las dinámicas de ataque, el bloque constitucionalista insiste en la importancia no de crear derechos, sino de reforzar garantías. Evidentemente, las diferencias de uno y otro campo de estudio afloran cuando se trata de proponer medidas, pues mientras la reflexión de Valls se limita al Código penal, el resto de autores (sin descartar, como hemos visto *supra*, la necesidad de reacción penal), propugnan mecanismos de gobernanza y supervisión, pero eso sí, destacando en todo caso la necesidad de que la regulación tenga una fuerza realmente vinculante, sin quedar en la mera declaración de principios. Cabe destacar, aquí, cómo parece tomar forma la idea de crear derechos «instrumentales», orientados directamente a proteger los clásicos, pudiéndose entrever un paralelismo con la lógica del bien jurídico intermedio aquí acogida.

d) Privacidad y no discriminación, en el centro.

De entre todos los derechos afectados, cabe resaltar la preocupación especial que generan dos de ellos, como punto común de todos los trabajos: la privacidad y

la no discriminación, precisamente como consecuencia, a nuestro modo de ver, de la dinámica propia de la IA, pues se trata de controlar estructuras complejas de decisión automatizada que pueden condicionar radicalmente nuestras vidas.

En lo relativo a la privacidad, parece innegable que, ante el paradigma de la vigilancia y el tratamiento masivo de datos en todos los sectores, la afectación a grupos indeterminados es clara. Las técnicas más avanzadas pueden llegar a poner en entredicho, incluso, una nueva dimensión de la intimidad (llegando a mostrar Cotino una preocupación por la privacidad mental, la identidad personal y la libertad cognitiva [1040]), negando el papel de la privacidad en un sentido clásico (podríamos entender que vinculado exclusivamente al derecho de abstención del poder público, ya largamente superado, para pasar a integrar otros derechos).

En segundo lugar, y de forma muy destacada esta vez por su novedad, pues, a diferencia de lo que ocurre con la privacidad, regulada de una forma reforzada por la prolija legislación de protección de datos, la discriminación clásica comienza a verse, con el paradigma del *big data* y la IA, superada. En este sentido son transversales también las preocupaciones de generación de resultados injustos para colectivos en todos los campos [1041] (empleo, justicia, publicidad, salud, seguros…), con una preocupación patente sobre cómo el sesgo puede acabar afectando a derechos sustantivos, comprometiendo el derecho a la igualdad y el acceso a servicios o, incluso, a la indemnidad física. Una vez más, una consideración de la discriminación en este sentido estructural o colectivo parece inevitable, sin que el marco actual ofrezca una protección suficiente.

e) La necesidad de horizontalidad.

Finalmente, una idea destacada cuando se trata de hablar de derechos fundamentales desde hace unos años es la de reconocer la eficacia horizontal que despliegan frente a particulares. Más que nunca hoy resulta esencial tener en cuenta esta dimensión, y, nos permitimos añadir, sin perder de vista que estas afirmaciones han de tener un reflejo necesario en la protección penal. Si la lógica de los derechos fundamentales es que estos deben configurar esferas protegidas de forma activa también frente a los poderes

[1040] Aba Catoira, A. et al., *op. cit*. N. 1025, p. 58, pero reconociendo que se pueden integrar dentro de la libertad ideológica, la integridad física y moral o la protección de datos personales y el derecho a la privacidad.

[1041] Son paradigmáticos, en este sentido, los estudios sobre discriminación (algorítmica) en el ámbito laboral, rama en que este derecho ha sido especialmente explorado. Vid., en este sentido: Abadías Selma, A., *op. cit*. N. 989, o Rivas Vallejo, P., *op. cit*. N. 57.

privados, la lógica de las grandes multinacionales tecnológicas que ostentan un poder mayúsculo en una situación quizá sin parangón donde el poder político y el tecnológico se entremezclan, no habrá de parecer descabellado que también el aparato de poder más contundente con el que cuenta el Estado (que no es otro que la amenaza penal) se despliegue allá donde sea necesario para hacer frente a este mandato de protección frente a todo poder que amenace con destruir bienes, por mucho que estas técnicas sean novedosas y, en ocasiones, queden enmascaradas bajo la apariencia de neutralidad tecnológica.

En definitiva, tras el recorrido realizado por la discusión dogmática tradicional, que nos ha permitido acercarnos al rechazo inicial a la generación de bienes colectivos, así como a las posiciones de aquellos autores que propugnaban un modelo intermedio capaz de conectar al individuo con los intereses sociales colectivos, se advierte una línea evolutiva clara. En estas propuestas se ponía el énfasis en que, ante determinadas formas de ataque novedosas, vinculadas al progreso industrial y técnico, resultaba necesario concebir nuevos mecanismos de protección. Desde esta perspectiva, parece que dicho discurso es trasladable, sin mayores inconvenientes, a este nuevo episodio de la evolución tecnológica, por mucho que apreciemos sus contornos novedosos.

No se trata tanto de que estemos ante nuevos bienes jurídicos, sino más bien que las formas de ataque son nuevas. La idea no es tanto que el bien sea colectivo como que la «nueva seguridad colectiva» (si se quiere, la seguridad colectiva digital) es una seguridad que incluye la protección de la dimensión colectiva de la privacidad y la discriminación, pues los ataques siguen la lógica de la difusión ya prevista en otros bienes tradicionales (salud pública o seguridad colectiva, sin ir más lejos, estando estos nuevos ataques, probablemente, a medio camino entre ambas lógicas). Se trata de protegernos frente a ataques indiscriminados en un doble sentido: las víctimas de estos ataques son indeterminadas, y los intereses objeto de ataque también lo son (sin que esto signifique que estén absolutamente indeterminados, pues hemos identificado que, en esencia, son los tradicionales, añadiendo como particularmente relevantes la privacidad y la no discriminación).

Resultará interesante explorar cómo el instrumento de regulación más importante hasta el momento (el Reglamento Europeo de Inteligencia Artificial), ha canalizado estas nuevas formas de ataque, para comprobar si en algún punto estos bienes se ven lo suficientemente protegidos con las garantías preventivas que impone la regulación administrativa. Sin perjuicio de que el contenido del Reglamento se analizará en el capítulo siguiente, baste dejar apuntada una reflexión.

El Reglamento parece reducir sus objetos de protección a tres dimensiones: salud, seguridad (entendemos que se refiere a la seguridad como integridad física, aunque no lo explicita), y aunque nombra derechos fundamentales varios, sin duda presta especial atención a la discriminación y el sesgo, así como la privacidad, que de alguna forma acaba vinculándose con la libertad. Sin embargo, también resulta llamativo que el Reglamento no entra a regular de forma evidente, probablemente por las tensiones que genera respecto de otros derechos fundamentales como la libertad de expresión o de prensa, otros derechos más delicados, como los de participación política. Resulta destacable cómo las normas que regulan la IA generativa (técnica idónea para ser usada como arma de manipulación electoral o de pensamiento) son escasas, y que se excepcionan las muy pocas que existen cuando, por ejemplo, es posible localizar a una persona que tiene responsabilidad editorial, en lo que podría interpretarse como una deferencia a la libertad de prensa. Es bastante evidente que estos sectores delicados quedan deliberadamente fuera del Reglamento, por lo que aquí no contaremos con la ventaja que ofrece éste en cuanto al avance en seguridad jurídica, al delimitar el riesgo permitido, pero esto no significa que no se deba pensar con la misma lógica. En lo relativo a lo que queda fuera del Reglamento, el legislador nacional es libre para actuar, y, de hecho, es aconsejable que lo haga donde el silencio es total pero las necesidades de protección son idénticas [1042].

[1042] Nos referimos a casos como el de la negociación algorítmica de alta frecuencia. En este caso, nuestra propuesta sería esencialmente la misma, solo que si las normas que determinan el riesgo permitido y el cuidado debido no están en el Reglamento de IA, habrá que buscarlas en la normativa administrativa aplicable. Más desarrollado, respecto de la negociación algorítmica de alta frecuencia, en: Bravo Bolado, A., «Negociando en la sombra…, *op. cit*. N. 245. Lo mismo se puede decir respecto de la delicada discusión y calibración de la libertad de expresión frente a la moderación de contenidos, que tiene una complejidad propia que la hace diferir de la dinámica aquí tratada, y que impregna toda la discusión sobre la regulación de la desinformación y los delitos electorales de elementos que van, necesariamente, no solo más allá de la normativa administrativa, sino que tienen imbricaciones directamente constitucionales. Se trata, empero, de cuestiones que han sido reguladas por otros ordenamientos. Más detalles en: Guerini, T., «Disinformation, Fake news and Deep fakes», special report para la AIDP publicado en Miró- Llinares, F./ Duvac, C./ Toader, T./ Santisteban Galarza, M. (Eds.), «Criminalisation of AI-related offences», *RIDP,* 95, 2024, pp. 111-138, especialmente a partir de p. 118, comparando modelos de regulación, donde la tónica general en Europa, al amparo del artículo 10 del CEDH es la de proteger la libertad de expresión con límites estrictos y sanciones penales limitadas a casos excepcionales, con el ejemplo de Francia y Alemania adoptando leyes específicas (Francia sancionando manipulaciones informativas en periodos electorales y Alemania, por su parte, imponiendo obligaciones de borrado de contenido a las plataformas digitales). En contraste, el modelo de países asiáticos como China, Singapur o Vietnam, que han introducido leyes severas que criminalizan la difusión de noticias falsas otorgando poderes discrecionales al Estado para controlar la información, o Rusia combinando multas y penas graves con un sistema intenso de censura y control estatal.

## II. LA TÉCNICA DE LOS DELITOS DE PELIGRO ABSTRACTO: NUEVAS FORMAS DE ATAQUE, NUEVAS FORMAS DE PROTECCIÓN

### 1. EL DELITO DE PELIGRO COMO PARADIGMA DE LA MODERNIDAD

El delito de peligro, por oposición al delito de lesión, es aquel en el que la intervención penal no se basa en la destrucción o menoscabo del bien jurídico, sino que la intervención penal se produce antes, en el momento en que el bien jurídico se haya puesto en riesgo de lesión, entendiéndose, así, como delitos de consumación anticipada, pues la intervención se produce antes de que se dé el resultado lesivo [1043].

Como se ha ido apuntado a lo largo de la investigación, uno de los emblemas de la moderna expansión del derecho penal es, junto con la proliferación de los bienes jurídicos de corte colectivo o supraindividual [1044], el uso común del delito de peligro frente al delito de lesión [1045]. Según varios autores, es innegable que existe una crisis del paradigma del resultado [1046]. En palabras de Feijoo, la doctrina viene señalando desde hace un tiempo cómo,

---

[1043] Existe consenso en cuanto a esta definición genérica, que se considera pacífica. Ver, por todos, MOLINA FERNÁNDEZ, F., «Acción y tipicidad…», *op. cit*. N. 253, nm 1316.

[1044] No nos es ajena la discusión sobre la diferenciación entre bien jurídico colectivo o bien jurídico supraindividual, así como la polémica para caracterizar unos y otros y su legitimidad. Nos remitimos en este punto a lo sentado en el apartado anterior en cuanto a los bienes jurídicos supraindividuales protegidos en los delitos contra la seguridad, pues solo a estos nos interesa referirnos. Su denominación, en este punto, como colectivos o supraindividuales no tiene una importancia esencial para lo aquí establecido. Nos remitimos a la monografía de Pérez-Sauquillo para una revisión exhaustiva sobre las tipologías y terminologías empleadas por los autores (*op. cit*. N. 928).

[1045] No se desconoce, empero, que el nacimiento y la teorización sobre el delito de peligro puede situarse mucho más atrás en el tiempo y que, de hecho, ya Binding, en su monumental *Die Normen und ihre Übertretung,* preconiza una teoría del peligro que ha servido de base a la reflexión teórica posterior, con postulados aún vigentes. Sin embargo, remontarnos a la discusión sobre el delito de peligro en su construcción dogmática inicial excedería, con mucho, los propósitos de esta investigación. Baste dejar apuntada la importancia de este trabajo, al que nos remitimos. Una explicación profunda, con abundantes referencias, al pensamiento de Binding en este aspecto en: MENDOZA BUERGO, B., *Límites dogmáticos y político-criminales de los delitos de peligro abstracto*, Comares, Granada, 2001, pp. 95-112.

[1046] En este contexto, merece la pena aclarar que puede crearse cierta confusión terminológica cuando se habla del delito de resultado frente al delito de peligro, y, por ello, en este capítulo trataremos de referirnos siempre al delito de lesión como oposición al delito de peligro. Sin embargo, no todos los autores así lo denominan, pues el término «resultado» resulta polivalente, sin que deba confundirse el resultado en un sentido material, como objeto separado de la acción, al que ésta da lugar, y el resultado en sentido jurídico. Si hablamos de delito de resultado como oposición al delito de peligro, habrá de ser siempre refiriéndonos al resultado en primera acepción (es decir, resultado material), pues resultado jurídico en cuanto afectación, siquiera remota, al bien jurídico objeto de protección, podrá producirse también en los delitos de peligro. Así, paradigmáticamente, en los delitos de peligro concreto, se habla de resultado de peligro.

desde el punto de vista político-criminal, los delitos imprudentes de resultado no dan mucho de sí [1047].

Es entendible, por tanto, que el legislador haya recurrido a la lógica de la intervención basada en la generación de peligros para bienes jurídicos en tanto en cuanto, como se ha visto en el capítulo quinto, las dificultades para afirmar el juicio de causalidad respecto del resultado de lesión eran evidentes en los denominados riesgos tecnológicos, y las formas de ataque a los bienes jurídicos, incluso aquellos más clásicos y conocidos, como la vida y la integridad física, adquieren unos caracteres difusos y generalizados, siendo esta la forma más habitual para reaccionar al riesgo tecnológico en sentido amplio [1048].

Es necesario, pues, analizar los tipos de delitos de peligro que usa el legislador, así como sus ventajas e inconvenientes, revisar las críticas tradicionales y determinar cuál de esos tipos es el más adecuado para abordar el riesgo tecnológico de la IA, creando figuras penales que reflejen el reproche de los ataques a los bienes jurídicos más vulnerables en la sociedad algorítmica.

## 2. LA TRIADA DE LOS DELITOS DE PELIGRO

Es absolutamente mayoritario en la doctrina reconocer la existencia de hasta tres categorías de delitos de peligro, pacíficamente aceptadas, si bien, como suele ocurrir con muchos conceptos establecidos como supuestamente categóricos, un examen algo más detallado permite afirmar que la realidad es algo más permeable, moviéndonos en un espectro de carácter continuo [1049].

---

[1047] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 308, con citas en nota 3, a Hefendehl (*op. cit*. N. 926, p. 168), Silva (op. cit. N. 931, p. 35), o Jakobs, G., «La Ciencia del Derecho Penal ante las exigencias del presente», en *Estudios de Derecho judicial,* 20,1999, pp. 131 y s.

[1048] Así, Solari, haciendo un repaso por la lógica y los elementos comunes y diferenciales en los delitos de peligro fruto del desarrollo tecnológico, *op. cit*. N. 980, pp. 269 y ss.

[1049] Esta es, desde luego, la opinión reiteradamente defendida por Molina Fernández que afirma que, en el caso de los delitos de peligro, de la misma forma que con la distinción entre el dolo y la imprudencia, existe una gradualidad reconducible a la paradoja sorites. En este sentido, si el peligro no es más que la probabilidad de que acontezca un mal, en función de los conocimientos que se tomen en cuenta para realizar ese juicio o pronóstico, es difícil afirmar que exista ontológicamente algo así como un peligro concreto o abstracto, sino que la mayor concreción o abstracción del peligro dependerá, en última instancia, de cuáles sean los conocimientos que se introducen en el juicio de valoración, así como en el momento en que se realice este juicio, siendo que en los delitos de peligro concreto lo que verdaderamente se da es un *«indicador cualificado de peligrosidad ex ante de la acción»*. Así, Molina Fernández, F., «Acción y tipicidad…», *op. cit*. N. 253, nm 1320. De forma desarrollada en cuanto a su pensamiento global sobre la paradoja sorites en algunas categorías dogmáticas del derecho penal, el mismo: *La cuadratura del dolo, op. cit.* N. 795*, passim.* Una visión similar parece adoptar Teijón Alcalá, que propone concebir las categorías de delitos de peligro de una forma no lineal. Teijón Alcala, M., «Los delitos de peligro en el derecho contemporáneo», *RECPC*, 25, 2023, p. 21, afirmando que la ventaja de su modelo, que no dista tanto de

La clásica distinción es la que se refiere a la dicotomía entre delitos de peligro concreto y delitos de peligro abstracto [1050].

Los delitos de peligro concreto son aquellos en que el legislador requiere, además de la peligrosidad de la acción, un resultado de peligro, es decir, y según una formulación pacífica, debe valorarse si alguno de los objetos de la acción ha entrado en el radio de acción de la conducta peligrosa. Como apunta Terradillos Basoco, ya Binding los caracterizó como aquellos en que se produce una verdadera «conmoción» en el bien jurídico. La verdadera particularidad de estos delitos, apunta Terradillos, es la constatación de que se produjo efectivamente una situación de riesgo para un objeto material o bien jurídico determinado y, para ello, la constatación que permite afirmar esta situación próxima a la lesión, se debe realizar un análisis *ex post*, donde no solo se valore la entidad del peligro creado, sino su proximidad en cuanto al bien jurídico [1051].

El ejemplo más clásico de este tipo de delitos lo encontramos en el artículo 380 del CP español, relativo a la conducción manifiestamente temeraria poniendo en concreto peligro la vida o integridad física de las personas [1052]. La estructura del tipo delictivo del delito de peligro concreto se corresponde, pues, con el delito de resultado, siendo necesario imputar, como en cualquier delito de resultado, objetiva y subjetivamente, dicho resultado de peligro concreto al autor.

Por oposición, en el delito de peligro abstracto, el legislador no exige que un concreto objeto haya entrado en el radio de la acción típica, sino que el peli-

---

las anteriores clasificaciones, es que: «*en vez de descomponer los delitos de peligro abstracto en varias modalidades intermedias, como suele ser habitual en la doctrina, partimos del riesgo como un continuo que va desde los supuestos en los que el riesgo de producción de un resultado es cierto (delitos en los que el riesgo se materializa en un resultado de lesión) a los supuestos en los que es imposible (delitos en los que se castiga la mera desobediencia de la norma). Todo ello, como decíamos, en función del nivel de afectación material (probable) que supone la conducta típica para el bien jurídico tutelado*».

[1050] No es esta, desde luego, una categoría exclusiva del derecho español, siendo que en el derecho alemán es posible distinguir entre *Konkretes Gerfärdungsdelikt* y *Abstraktes Gefährdungsdelikt*. Así, recientemente, Lohman ha acudido a esta doble categoría para explorar la posibilidad de aplicar este modelo en los casos de daños causados por sistemas de IA, concluyendo que, si bien la posibilidad de aprobar delitos de peligro abstracto está presente, su conveniencia político-criminal ha de discutirse, por el freno que supondría para determinadas técnicas. Lohmann, A., *op. cit. N. 298*, p. 257 y ss. En la doctrina anglosajona, si bien la distinción no responde a los mismos límites, también es posible afirmar la existencia de dos grupos: *implicit* y *abstract endangerment offences*. Duff, R. A., «Criminalizing endangerment», *op. cit*. N. 344, p. 959, donde los delitos de riesgo «explícito» requieren, para su comisión, la creación de un riesgo relevante, mientras que en los «implícitos» la definición del delito no especifica el riesgo.

[1051] Terradillos Basoco, J., «Peligro abstracto y garantías penales», *Nuevo Foro Penal*, 62, 1999, p. 73, con cita a Rodriguez Montañés, que lo formula exigiendo que se coloque al bien jurídico en el *«ámbito de eficacia de la acción típica peligrosa»,* de forma que ya no sea segura la evitación de la lesión. Rodríguez Montañés, T., *Delitos de peligro*…, *op. cit*. N. 934, pp. 37-38.

[1052] El ejemplo y la definición, aunque ampliamente pacíficos, se extraen de Molina Fernández, F., *op. cit.* N. 1043.

gro existe solo en cuanto cualidad inherente a la acción. Es decir, el legislador prohíbe una acción porque es generalmente peligrosa para un bien jurídico, sin que sea necesario constatar que, además, el objeto de acción se ha puesto en evidente peligro en el caso concreto. Es, de hecho, el delito de peligro abstracto el verdadero protagonista de la tipificación del moderno derecho penal, y no tanto el delito de peligro concreto, pues este segundo grupo de delitos permiten intervenir penalmente en un momento anterior, funcionando como verdadero resorte preventivo o muro de contención a las conductas sobre las que no se sabe qué efectos pueden causar, sin necesidad de esperar a la producción de un resultado de peligro, que puede ser, en ocasiones, de contornos catastróficos.

Como tercera categoría, y a medio camino entre las anteriores, surge (en Alemania de la mano de Schröder [1053], y, en España, paradigmáticamente, de la mano de Torío López [1054]), un *tertium genus,* que ha recibido denominaciones de lo más variadas (delito de aptitud [1055], delito de peligro hipotético, delito de peligro abstracto-concreto, etc. [1056]). Este tipo de delitos, aunque no exigen un resultado de peligro concreto, sí exigen la comprobación de que la acción enjuiciada es, *ex ante*, peligrosa. Los delitos claramente hipotéticos o de aptitud fijados por el legislador en el código incluyen, así, un elemento indicativo de esa peligrosidad cuya concurrencia ha de evaluar el juzgador en el caso concreto.

En palabras de Cerezo: *«En los delitos de aptitud para la producción de un daño, o delitos de peligro abstracto-concreto, es mucho mayor la gravedad del desvalor de la acción. Para que la acción sea típica es preciso que, desde un punto de vista ex ante, aparezca como no absolutamente improbable que de ella se derive la lesión del bien jurídico. En los delitos de aptitud para la pro-*

---

[1053] Schröder, H., «Abstrakt-konkrete Gefährungsdelikte?», en *JZ*, 1, 1967, pp. 523 a 525.

[1054] Torío López, A., «Los delitos del peligro hipotético (Contribución al estudio diferencial de los delitos de peligro abstracto)», *ADPCP,* vol. XXXIII*,* 1991. Bebe la construcción de Torío de la propuesta anterior de Gallas (Gallas, W., «Abstrakte und konkrete Gefährdung»*,* en AAVV*., Festschrift für Ernst Heinitz zum 70. Geburtstag,* Walter de Gruyter, Berlín, Nueva York, 1972, pp. 171-184). Torío pone el énfasis en que el juez debe advertir no solo la peligrosidad genérica de la conducta (desvalor real de acción), sino, además, la posibilidad del resultado peligroso (lo que se puede entender como desvalor potencial de resultado). Torío López, Á., ibid., p. 846. Lo cierto es que, como acertadamente apunta Maldonado, parece que Torío, con su construcción, exige algo más, pues esa acreditación potencial del resultado de peligro no puede añadir más que lo que implica ya el juicio de peligro sobre la conducta, so pena de convertir el delito de peligro abstracto en uno de peligro concreto. Maldonado Fuentes, F., *op. cit*. N. 917, p. 49.

[1055] Concretamente, y usando denominación de Escrivá, «delitos de aptitud para la producción de un daño», Escrivá Gregori, J. M., *La puesta en peligro de bienes jurídicos en Derecho penal*, Bosch, Barcelona, 1976, pp. 31-32.

[1056] En Alemania la terminología típicamente utilizada es la de *Eignungsdelikte*. Por todos, Hoyer, A., Die *Eignungsdelikte,* Duncker & Humboldt*,* Berlín, 1987, *passim*. Más referencias a las obras de autores germanos en Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 312, nota 14. Todos ellos, empero, hacen referencia a una misma realidad, estando la polémica más bien centrada en si deben considerarse como un tipo de delito de peligro concreto o como una modalidad de peligro abstracto, siendo mayoritaria la última postura y la que, ya se adelanta, se defenderá aquí.

*ducción de un daño aparece como un elemento del tipo la peligrosidad de la acción desde un punto de vista ex ante* [1057]». Las formulaciones que pueda adoptar esta necesidad de aptitud son de lo más variadas. Véanse, a título de ejemplo, el elemento de «nocividad» en los mal llamados delitos de fraude alimentario, o la referencia a que los delitos medioambientales han de poder causar daños «sustanciales» en el aire, suelo o aguas [1058].

La doctrina, empero, en su empeño por dotar de legitimidad a los siempre polémicos delitos de peligro abstracto puros, ha acabado, en ocasiones, por tratar de reconducir estos últimos a la tercera categoría, en busca de un elemento legitimador capaz de dotar de antijuridicidad material a los altamente formales delitos de peligro abstracto. Ello, no obstante, no ha encontrado una aceptación unánime por parte de la doctrina, como enseguida veremos, pues, si bien este tipo de delitos pueden resultar criticables por su excesiva formalidad y contrariedad a principios tan básicos como la lesividad o antijuridicidad material, la arrogación de competencias que corresponden al legislador ha sido interpretada, a nuestro juicio con buenas razones, como contrarias al principio de legalidad [1059]. Así, Cerezo admite que solo de *lege ferenda* será admisible la interpretación como peligro hipotético, pero de *lege lata*, en los casos en que no está así previsto, supondría una nivelación o asimilación de todas las figuras delictivas de peligro, apuntado que el problema de legitimidad que en su caso presenten los delitos de peligro abstracto solo puede ser resuelto por el legislador [1060]. Por su parte, Feijoo asegura que aunque desde el punto de vista dogmático no solo deben considerase delitos de peligro abstracto-concreto los delitos de peligro en los que el tipo contempla expresamente una cláusula de idoneidad o aptitud, esto se deberá resolver como un problema interpretativo de la parte especial, pero nunca convirtiendo todos los delitos de peligro abs-

[1057] Cerezo Mir, J., «Los delitos de peligro abstracto en el ámbito del derecho penal del riesgo», *Revista de Derecho Penal y Criminología*, 10, 2002, p. 63.

[1058] Artículos 363.2 CP (poner en peligro la salud de los consumidores «Fabricando o vendiendo bebidas o comestibles destinados al consumo público y nocivos para la salud»), y el 325.1 CP, respectivamente.

[1059] Aunque se desarrollarán a continuación ambas posturas, déjese ya apuntado que, entre los defensores más férreos de la aplicación del paradigma del delito de aptitud a todo delito de peligro se encuentran Mendoza o Terradillos, que han sido abiertamente criticados por autores como Cerezo o Feijoo.

[1060] Cerezo Mir., J., *op. cit*., N. 1057, p. 71, pero insistiendo el autor en que «*Esta transformación de los delitos de peligro abstracto puros en delitos de aptitud para la producción de un daño debe realizarse únicamente, sin embargo, en los delitos contra bienes jurídicos colectivos, que tienen, como vimos, un carácter intermedio, carecen frecuentemente de contornos precisos y suponen una anticipación de la protección penal de bienes jurídicos individuales (delitos contra la seguridad del tráfico, delitos contra la salud pública). Idem*., pp. 71-72.

tracto, automáticamente, en delitos de peligro abstracto-concreto, pues el juez se encuentra, en todo caso, vinculado por el tenor literal de la ley [1061].

## 3. LOS PROBLEMAS DEL DELITO DE PELIGRO ABSTRACTO

El gran argumento esgrimido en contra de los delitos de peligro abstracto tiene que ver, en esencia, con su insuficiente lesividad. Las formulaciones que encuentra este principio son, también, variadas, y aunque es común encontrar una referencia genérica a la lesividad, hay quienes han puesto el énfasis en la antijuridicidad material o en el principio de ofensividad. Sin perjuicio de comprender sus matices, dejaremos sentado que, en lo que aquí nos concierne, la diferencia entre todas estas formulaciones puede ser sutil y que, aunque en ocasiones los argumentos aparecen entremezclados, consideramos oportuno hacer una distinción entre dos tipos de argumentos o problemas: al primero de ellos aludiremos con la denominación «injusto insuficiente», y lo circunscribiremos a la antijuricidad material, mientras el segundo de ellos lo denominaremos «injusto formal», y lo vincularemos más al principio de lesividad en sentido estricto [1062].

### 3.1 **Injusto material insuficiente: ¿solo desvalor de acción?**

En este punto, muchas son las advertencias que sobre el carácter de simbólico de este tipo de delitos se han vertido. Serán ilegítimos, por simbólicos, aquellos delitos donde el derecho penal cumple exclusivamente funciones simbólicas, sin tratar de proteger verdaderos intereses o bienes jurídicos [1063].

Especialmente insistente con esta vertiente es Terradillos Basoco, que destaca que el punto de partida o, más bien, de tensión inicial que provoca el recurso a estos delitos es la expansión de la intervención punitiva en un siste-

---

[1061] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 315.

[1062] No debe olvidarse que, en cuanto principio limitador del *ius puniendi* del Estado, el principio de lesividad ha de ponerse en relación con otros; siendo aquí especialmente importantes el principio de intervención mínima y el de subsidiariedad. En este sentido, cuando hagamos referencia a la supuesta insuficiencia de lesividad (injusto demasiado formal), lo haremos teniendo en mente la conjugación de estos tres parámetros.

[1063] Por todos, Silva Sánchez, J. M., *Aproximación…, op cit*. N. 926, p. 305: «*esta función simbólica o retórica de las normas penales se caracteriza por dar lugar, más que a la resolución directa del problema jurídico-penal (a la protección de bienes jurídicos), a la producción en la opinión pública de la impresión tranquilizadora de un legislador atento y decidido. Así, sería un proceder básicamente simbólico, y que en principio no garantizaría una mayor protección de los bienes jurídicos afectados, el trasladar al Código penal una norma ya existente en la legislación especial, sólo por el mayor efecto retórico que ello tiene*.»

ma penal presidido por el principio de intervención mínima [1064]. Sitúa el autor como correlato de la antijuridicidad (como concepto perteneciente a la teoría del delito), el principio político-criminal de lesividad, que no puede ser escindida, a su vez, de la necesidad de intervención. Sitúa, así, en contexto, este trinomio en ocasiones aludido indistintamente [1065].

Para poder responder a la cuestión acerca de la legitimidad real o el mero simbolismo de estos delitos, habrá que tomar, primero, postura, según advierte acertadamente Maldonado [1066], en lo relativo a la concepción del injusto. Ello es así porque uno de los argumentos esgrimidos para criticar la falta de legitimidad de los delitos de peligro abstracto se deriva de la postura dualista que entiende que todo delito debe tener un desvalor de acción y un desvalor de resultado, sin que sea posible en los delitos de peligro abstracto (como sí ocurre, por el contrario, con los delitos de peligro concreto) identificar un desvalor de resultado [1067].

Es posible, empero, superar la postura de los dualistas, para encontrar dos opciones más. Aunque ambas son monistas, sus matices son significativos, pudiendo distinguir una primera postura, según la cual es suficiente con afirmar la existencia de un desvalor de acción para fundamentar el injusto penal [1068], y una segunda, algo más matizada y con mayor apoyo general, según la cual el desvalor de acción puede fundamentar por sí solo el injusto, pero sin negarle un papel determinante al resultado, en tanto factor relevante para determinar el merecimiento, así como graduar la pena a imponer, teniendo en cuenta las necesidades de prevención [1069].

---

[1064] Terradillos Basoco, J., «Peligro abstracto…», *op. cit*. N. 1051, p. 68.

[1065] *Idem,* p. 69.

[1066] Maldonado Fuentes, F., *op. cit*. N. 917, p. 53. También insiste en la importancia de la discusión sobre la construcción del injusto material Mendoza Buergo, B.*, Límites dogmáticos, op. cit.* N. 1045*,* pp. 86-87.

[1067] En este sentido, para los defensores más convencidos del dualismo, incluso en los delitos puros de mera actividad, es posible (e imprescindible) identificar un desvalor de resultado que va más allá de la mera afectación al bien jurídico, que es necesario para marcar la barrera de lo penalmente relevante. Para profundizar en la cuestión relativa al rol del resultado y la discusión en cuanto a los delitos de mera actividad y peligro: Acale Sánchez, M., «Los delitos de mera actividad», *Revista de Derecho penal y Criminología,* 10, 2002, pp. 11-45. En p. 19 presenta la postura de quienes consideran, como Laurenzo, que delitos de mera actividad serán solo aquellos en que exista una mera desobediencia formal, sin identificar bien jurídico alguno, pero que, en el resto de los casos, cuando existe bien jurídico, habrá siempre desvalor de resultado. En extenso, la postura de Laurenzo desarrollada en Laurenzo Copello, P., *El resultado en Derecho Penal*, Tirant Lo Blanch, Valencia, 1992.

[1068] Algunas de estas posturas, según Maldonado las más extremas, «*concentran la relevancia del desvalor de la acción en la subjetividad de la conducta, radican el fundamento del injusto en el hecho de corresponder éste a una expresión de voluntad contraria a la norma o al derecho*», y pueden identificarse con las minoritarias tesis del finalismo extremo de Zielinski. Maldonado Fuentes, F., *op. cit*. N. 917, p. 54. Zielinsky, D., *Handlungs-und Erfolgsunwert im Unrechtsbegriff,* Duncker-Humblot, Berlín 1973.

[1069] Es decir, cumpliendo un papel en la evaluación de las necesidades de prevención, pero sin formar parte esencial del injusto, que solo debe estar integrado por conductas o comportamientos. Así, Silva

En este sentido, consideramos perfectamente posible y coherente con el propio sistema penal, rechazar el argumento en ocasiones esgrimido por los dualistas relativo a la insuficiente antijuridicidad material del injusto del delito de peligro abstracto, basándonos en la idea de que, ya se entienda que existe desvalor de resultado (resultado en sentido jurídico [1070]) en la mera afectación del bien jurídico protegido en el delito (bien jurídico que, como se ha tratado de establecer en el apartado inmediatamente anterior, forma parte ineludible de cualquier delito, sin duda también de los de peligro abstracto), sea por entender el resultado como un elemento condicionante de la pena pero no necesario para fundamentar el injusto, es perfectamente posible decir que el injusto del delito de peligro abstracto es, desde este punto de vista, suficiente.

Cabe destacar que las posturas mayoritarias en este sentido, tal y como Maldonado pone de manifiesto, son de corte dualista, partiendo de que la existencia de ambos desvalores es necesaria para afirmar la antijuridicidad material, pero uniendo el desvalor del resultado con la función de protección de bienes jurídicos, esencialmente limitadora del *ius puniendi* [1071], quedando colmadas por tanto las exigencias de una antijuridicidad material con la referencia a un bien jurídico merecedor de protección penal (en los términos exigidos por los principios de intervención mínima y subsidiariedad antes mencionados).

Mayores argumentos requerirá, sin embargo, aludir a la formalidad respecto de su lesividad, de la que a continuación nos ocupamos.

---

SÁNCHEZ, J. M., *Aproximación*, *op. cit*. N.926, p. 415 y ss. Concretamente, en p. 417: «*La ubicación del resultado fuera del ámbito sistemático de la antijuricidad penal, en el marco de los presupuestos de la sancionabilidad penal de los comportamientos penalmente prohibidos, no supone, como MIR PUIG ha señalado entre nosotros reiteradamente, que deje de tener importancia en el sistema del delito. Al contrario, el resultado, la constatación ex post de la presencia de una auténtica realización típica, desempeña una función político criminal y, por consiguiente, también dogmática, fundamental.»*. Vid.: MIR PUIG, S., «Sobre lo objetivo y lo subjetivo en el injusto», *ADPCP,* vol. 41, 1988, pp. 672-673, señalando que el resultado no ha de perder «un solo ápice de su *importancia* político-criminal». También: CANTARD, N., «Acerca de la función del disvalor de resultado en derecho penal», *Revista del centro de investigaciones de la Facultad de ciencias jurídicas y sociales de la Universidad Nacional del Litoral*, 16, 2015, pp. 9-30.

[1070] A estos efectos, y para tener una visión más clara sobre el concepto de resultado y su relación con los delitos de lesión y de peligro, nos remitimos a ACALE SÁNCHEZ, M., *op. cit*. N. 1067, especialmente en pp. 17 y ss., donde explica que el concepto de resultado puede usarse en dos planos (como efecto natural y separable de la acción o como desvalor jurídico), y que, en el plano de los delitos de mera actividad, aunque no exista resultado material, puede haber afectación al bien jurídico, por lo que este tipo de delitos pueden configurarse como delitos de mera actividad y lesión, o bien mera actividad y peligro, siendo posibles varias combinaciones, siendo el resultado natural un elemento típico que el legislador puede elegir si incluye o no en el tipo penal sin que ello determine la existencia de un desvalor de resultado, operando ambos conceptos en dos planos distintos.

[1071] MALDONADO FUENTES, F., *op. cit*. N. 917, p. 55.

### 3.2 **Injusto demasiado formal: ¿no hay forma de diferenciarlo de la infracción administrativa?**

La crítica ubicua cuando se trata el tópico del delito de peligro abstracto tiene que ver con lo que consideramos el verdadero escollo en la justificación de estos delitos, que no es otro que su presunta insuficiente lesividad en sentido material en lo relativo al desvalor de acción. Si damos por buena la aceptación de las posturas anteriores, que matizan la relevancia del desvalor de resultado en un sentido fuerte (o al menos material), será coherente, entonces, exigir, para alcanzar el umbral de la mínima intervención que requiere el derecho penal, a cambio, un desvalor de acción lo suficientemente elevado.

Traíamos a colación, ya en el capítulo tercero, a modo introductorio, las críticas que sobre este punto vertían voces autorizadas como Mendoza [1072] o Corcoy [1073]. Advertían que la lesividad no puede basarse en meras presunciones de peligrosidad, bajo la pena de mantener un injusto penal excesivamente formalizado en caso de entenderse que los tipos no han de exigir la concurrencia de una peligrosidad en el caso concreto para el bien jurídico protegido, bastando con aceptar que, en los casos en que el legislador así lo tipifique, la mera peligrosidad estadística de una serie de conductas generalmente peligrosas no es fundamento en absoluto adecuado para la tipicidad penal [1074]. Lo señala, asimismo, Pérez-Sauquillo, al poner de manifiesto que «*en los delitos de peligro abstracto se penalizan conductas que pueden haber sido inocuas en el caso concreto, y ello no sólo porque no haya habido ningún tipo de consecuencias –al no haber concurrido ninguna víctima potencial en el radio de acción de la conducta– sino porque el agente haya previamente adoptado medidas especiales de precaución. Como se ve, esta posibilidad pone de manifiesto que en los delitos de peligro abstracto podría llegar a ponerse en duda hasta la propia existencia de un desvalor de acción de la conducta individual, al no concurrir ni dolo –no ya de lesionar, sino tampoco de poner en peligro– ni imprudencia, en lo que al desvalor subjetivo de acción se refiere, ni tampoco una conducta idónea para lesionar o siquiera poner en peligro, en lo tocante al desvalor objetivo de acción*» [1075].

---

[1072] Mendoza Buergo, B., «Exigencias de la moderna política …, *op. cit.* N. 360.

[1073] Corcoy Bidasolo, M., «Principio de culpabilidad …, *op. cit.* N. 361, p. 576.

[1074] Damos por reproducidas aquí las críticas vertidas ya en el capítulo tercero, *supra,* apartado 3.1 «La pérdida de importancia del principio de lesividad o la crisis del principio del daño».

[1075] Pérez-Sauquillo, C., «Delitos de peligro abstracto y bienes jurídicos colectivos», *Revista Foro FICP,* 3, 2015, p. 118. Para referencias detalladas a los autores alemanes que tradicionalmente pusieron de manifiesto estas características (Schröder, Rudolphi, o Schünemann.), *vid*. nota al pie 13.

Nos tememos que esta es una crítica fundada si nos limitamos a aplicar de forma literal determinados delitos de peligro abstracto puros, donde la mera contravención de la norma no aporta indicio suficiente de lesividad en el caso concreto, haciendo imposible distinguir la protección de la norma con la protección del bien jurídico. A continuación, se verá cuáles han sido los intentos por superar esta interpretación al hilo de los delitos de peligro abstracto puro.

### 3.3 **Vulneración del principio de culpabilidad**

Finalmente, y retomando otra de las críticas que en el inicio de la investigación se exploraba desde el ámbito anglosajón, cabe traer a colación aquellos argumentos que tienen que ver con la posible vulneración del principio de culpabilidad. En este sentido, se ha apuntado que, en una aplicación estricta del delito de peligro abstracto, al autor se le suponen conocimientos de elementos del delito que no se dan, o bien que falla la premisa básica según la cual la pena debe poder ser atribuible personalmente, en un sentido de merecimiento individual al concreto sujeto al que se impone, sin que quepa aludir a argumentos de peligrosidad general para imponer una concreta pena a un concreto individuo [1076]. Ya Torío identificaba, de hecho, en su división de los delitos de peligro, una categoría asimilable a las *regulatory offences* angloamericanas, identificándolas con aquellos delitos de peligro que castigan meras desobediencias formales, afirmando, con rotundidad [1077], que este grupo debía desaparecer de las legislaciones penales.

Sin embargo, cabe destacarse que esta crítica no es compartida de forma unánime, pues hay quien afirma que no es posible ver infracción del principio de culpabilidad en tanto en cuanto la culpabilidad represen-

[1076] En este sentido, Feijoo. Diferencia, acertadamente, este autor, entre la posibilidad de prohibir conductas basadas en su peligrosidad genérica, de la posibilidad de imponer una pena «frente a un ciudadano concreto». Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 320. Es posible que en otros órdenes como el administrativo la prohibición sea legítima, pero las exigencias del principio de culpabilidad en sentido amplio no pueden justificar una pena en estos casos. Esta idea conecta, en lo esencial, con lo afirmado en el capítulo tercero relativo al principio de precaución y sus límites en su vertiente dogmática.

[1077] Torío López, A., «Los delitos de peligro…», *op. cit*. N. 1054, p. 837. Los tres grupos de los que se vale Torío para defender su concepción son: los que castigan la vulneración de principios ético-sociales o religiosos, los que castigan desobediencias formales, y los que él denomina «delitos de peligro hipotético». Solo el último de los grupos es considerado legítimo, en caso de interpretarse como delitos de peligro posible, es decir, siempre y cuando, a partir de una interpretación teleológica y normativa, se pudiera interpretar que la conducta era susceptible de dañar el objeto de protección fijado por el legislador.

ta la reprochabilidad personal de la conducta antijurídica, por lo que solo se podría ver infringida si falta el objeto de reproche, en caso de que la conducta carezca, por completo, de un contenido material de injusto, estando, por ello, exclusivamente en el campo del injusto el problema de estos delitos [1078].

Exploraremos, a continuación, los argumentos que se han ofrecido por la doctrina para justificar la existencia de estos siempre discutidos delitos.

En este punto es preciso aclarar ya una cuestión relativa a nuestra toma de postura. En lo que sigue se defenderá (sin que sea algo, de hecho, excesivamente atrevido, pues así lo sostiene la práctica totalidad de la doctrina), que, en caso de querer castigar conductas que responden al paradigma de la sociedad del riesgo tecnológico, el delito de peligro abstracto es el más adecuado, pero lo es en su modalidad de delito de aptitud, por ser el más respetuoso con los principios de lesividad y culpabilidad. Al tratarse nuestra propuesta de una iniciativa *de lege ferenda,* partimos, por tanto, de una posición inicial mucho más sencilla que la de quienes deben enfrentarse a la interpretación de tipos ya contemplados en el Código Penal. Sin perjuicio de que más adelante se profundizará en esta variante del delito de peligro, resulta necesario establecer, también, que la mayor parte de la doctrina considera a los delitos de aptitud como una modalidad inserta en los delitos de peligro abstracto puros [1079], y por ello no resulta baladí realizar un repaso más o menos completo por los argumentos de legitimación propuestos por la doctrina, pues nos parece que pueden ayudarnos a entender los riesgos e inconvenientes de una (eventual) tipificación de delitos de riesgo, como la que aquí se propondrá.

---

[1078] Así, Cerezo Mir., J., *op. cit*., N. 1057, p. 65. Sin embargo, ha de resaltarse que en Alemania esta cuestión sí ha sido traída a colación, entendiéndose el principio de culpabilidad en sentido amplio. Así, Roxin, C., *Derecho penal. Parte general*, *op. cit*. N. 921, p. 407, nm 119.

[1079] Así lo ha corroborado, además de la doctrina, la jurisprudencia en contadas ocasiones. Así, STS de 12 de noviembre de 1985 (ponente: Rodríguez López), definiendo los delitos de peligro como aquellos en que «*es necesaria, aunque también suficiente la ejecución de una acción peligrosa, idónea para producir un peligro para el bien jurídico protegido; la imposibilidad de la producción del peligro excluye la tipicidad*». Recientemente, respecto del delito contra el medio ambiente del artículo 325 CP, la STS 481/2020, de 29 de septiembre, FJ 3, equiparando a estos efectos, recogiendo las referencias de la jurisprudencia anterior, los delitos de peligro hipotético, los de peligro abstracto-concreto, de peligro potencial o de aptitud, donde se tipifica «*un comportamiento idóneo para producir peligro para el bien jurídico protegido»,* donde la situación de peligro no es elemento del tipo *«pero sí lo es la idoneidad del comportamiento efectivamente realizado para producir dicho peligro»*. Un listado exhaustivo de sentencias que así lo reconocen, analizando la figura del delito de peligro abstracto concreto respecto de los delitos contra el medio ambiente, contra la salud pública, o de incendio, en Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, nota el pie 17.

## 4. LOS ARGUMENTOS PARA REDUCIR EL IMPACTO NEGATIVO DE LOS DELITOS DE PELIGRO ABSTRACTO. TOMA DE POSTURA.

Se presentan a continuación los grandes posicionamientos acerca del delito de peligro abstracto puro [1080], con algunas variantes, que a continuación se explicarán. Se finalizará la exposición con una última postura, que, si bien menos apoyada, hasta lo que se nos alcanza, resulta convincente en cuanto pone el foco en lo verdaderamente esencial de este tipo de delitos, que, adelantamos ya, es la necesidad de referir el elemento subjetivo exclusivamente a los elementos contenidos en el tipo, huyendo de las propuestas que, por su consideración de la necesidad de previsión de un resultado sobre los bienes jurídicos finales, vuelven a reproducir los problemas inherentes al delito imprudente de resultado que, según lo aquí defendido, no presenta ninguna ventaja cuando se trata de castigar las conductas generadas por sistemas autónomos dotados de un alto grado de imprevisibilidad sobre el resultado final.

### 4.1 **Los delitos de peligro abstracto como delitos de peligro presunto (una posición superada)**

Es menester comenzar la exposición con la alusión a la primera de las teorías, aunque es de justicia conceder que la misma ha sido prácticamente abandonada [1081], pues sus postulados no encajan de ninguna manera con lo que cabe esperar en un ordenamiento respetuoso con los principios de lesividad y antijuridicidad material, referidos *supra.*

Se trata de la idea de que el delito de peligro se identifica con la presunción de peligrosidad llevaba a cabo por el legislador en el enunciado del delito. El peligro se configura, así, meramente como la *ratio legis* del normador, de manera que, si el legislador así lo establece, los tribunales habrán de presumir

[1080] Debe apuntarse, también, que la justificación de los delitos de peligro no ha sido, en todos los casos, uniforme, pues varios autores han hecho depender su defensa de los mismos al tipo de bien jurídico protegido. Así, Schünemann, en sus primeros escritos (no así en los más recientes: SCHÜNEMANN, B., «El principio de protección de bienes jurídicos…», *op. cit*. N. 926) diferenciando el ámbito de los delitos con bienes jurídicos intermedios de los bienes puramente institucionales. En un sentido similar, RODRÍGUEZ MONTAÑÉS, T., *Delitos de peligro*…, *op. cit*. N. 934, en su propuesta final, pp. 299-304. Como en este trabajo solo estamos valorando lo relativo al bien jurídico «seguridad colectiva tecnológica», en los términos arriba explicados, que lo convierten en uno del tipo bien jurídico intermedio, dejaremos de lado la problemática que tiene que ver con los bienes jurídicos institucionales, sin perjuicio de dejar aquí apuntadas las soluciones que los autores más señalados proponen.

[1081] Se comentará, enseguida, la defensa, matizada, de algunos autores en lo relativo a la posibilidad de que la presunción se convierta en refutable.

(*iuris et de iure*), que la conducta, tal y como se describe en el tipo penal, es, siempre y en todo caso, constitutiva de una infracción penal, con independencia de las concretas circunstancias que rodeen al delito.

Dicha postura, que bebe de su formulación italiana pero que tuvo su acogida en su momento en el derecho penal español [1082], lejos de ser abandonada, podríamos decir que ha sufrido alguna deformación. Así lo expresa Feijoo cuando alude a que, dentro de la literatura española más reciente, el término se usa, más que como concepto dogmático, como concepto político-criminal para criticar tipificaciones ilegítimas [1083].

La segunda de las variantes o deformaciones implicaría afirmar que habrá presunción, pero que será una presunción *iuris tantum* [1084]. Es decir, el juez sentenciador deberá valorar si, en el caso concreto, el acusado es capaz de desmontar la presunción del legislador, de tal forma que, en caso de que se logre demostrar por el acusado que, según el caso concreto, la conducta no era susceptible de poner en peligro el bien jurídico, habrá que absolver. Acertadamente, a nuestro juicio, señalan los críticos que no es admisible una argumentación de estas características, pues una presunción, por muy *iuris tantum* que sea, irá en contra de la presunción de inocencia del reo en el proceso penal [1085].

Una defensa, empero, valiosa (que no compartimos por las ideas que enseguida se revelarán) ha llevado a cabo recientemente Rusca [1086], poniendo en comparación los distintos argumentos que se han manejado, principalmente examinando los que se han hecho más comunes en el campo anglosajón. Así,

---

[1082] Beristain Piña, A., «Resultado y delitos de peligro», *Revista de la Facultad de Derecho de la Universidad de Madrid,* Vol. XIII, 1969, pp. 457-465; Rodríguez Mourullo, G., *La omisión de socorro en el Código Penal,* Tecnos, Madrid 1966, pp. 169-170; y Córdoba Roda, J., «Rapport sobre los delitos de peligro», *Revue Internationale de Droit Pénal,* 1969, p. 359.

[1083] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 318.

[1084] Así, narra Cerezo cómo en el X Congreso internacional de Derecho Penal, celebrado en Roma en 1969 se adoptó la conclusión de «*que el sistema de < peligro presunto> debe ser cuidadosamente dosificado y debe comportar la posibilidad legal de aportar prueba en contrario para rebatir la presunción, al menos en los casos especialmente previstos por el legislador*». Cerezo Mir., J., *op. cit*., N. 1057, p. 66, mostrándose completamente contrario a cualquier forma de presunción.

[1085] Entre otros, Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, con cita a muchos otros autores (Cerezo, Corcoy, García Rivas, Mendoza Buergo, Doval Pais), p. 318. Cita Feijoo en este punto las SSTC 111/1999, de 14 de junio, FJ 3: «*comoquiera que se califique a este tipo delictivo, bien de peligro simplemente bien de peligro abstracto o remoto, en ningún caso el derecho a la presunción de inocencia tolera que alguno de los elementos constitutivos del delito se presuma en contra del acusado, sea con una presunción iuris tantum sea con una presunción iuris et de iure*.», o la STC 87/2001, de 2 de abril, FJ 9: «*La primera modalidad de presunción iuris tantum no es admisible constitucionalmente ya que, como declaró la STC* 105/1988*, produce una traslación o inversión de la carga de la prueba, de suerte que la destrucción o desvirtuación de tal presunción corresponde al acusado a través del descargo, lo que no resulta conciliable con el artículo 24.2 CE».*

[1086] Rusca, B., «Los delitos de peligro abstracto como presunciones refutables. Nuevos argumentos en defensa de una teoría clásica», *Revista Chilena de Derecho*, 49, 1, 2022, pp. 101-126.

la teoría de la arrogancia cívica de Duff [1087], el argumento del atentado contra el estado de derecho defendido por Alexander y Ferzan [1088], la lesión de la seguridad de Kindhaüser y la teoría del delito de peligro como imprudencia sin resultado [1089], para llegar a la conclusión de que en los delitos de peligro abstracto se genera una tensión evidente e inevitable entre la seguridad jurídica y los derechos del acusado, resultando a juicio del autor más adecuado resolver la cuestión aceptando un papel de la presunción, habida cuenta de que no se trata del tipo de presunciones prohibidas por el ordenamiento penal.

Rusca propone [1090] entender los delitos de peligro abstracto como presunciones refutables sobre la peligrosidad de la conducta prohibida. Según esta perspectiva, cuando una persona realiza una acción tipificada (por ejemplo, conducir con una determinada tasa de alcohol en sangre), se presume que dicha conducta genera un riesgo no permitido. Sin embargo, esta presunción puede ser desvirtuada si el acusado demuestra que, dadas las circunstancias particulares, su comportamiento no produjo efectivamente ese riesgo. Esta concepción permite mantener los beneficios propios de las reglas penales, tales como la claridad normativa, la eficiencia y la uniformidad en su aplicación, sin por ello caer en el castigo de personas que no han realizado una conducta realmente reprochable. Rusca sostiene que esta solución no vulnera el principio de presunción de inocencia, siempre que se entienda como una exigencia

---

[1087] Duff, R. A., «Criminalizing endangerment», *op. cit.* N. 344, pp. 43-64. Según Duff, incluso cuando una persona realiza la conducta prohibida sin generar un riesgo real (por ejemplo, conducir apenas por encima del límite de alcohol en una situación controlada), merece ser castigada por su «arrogancia cívica». Esta consistiría en creerse con derecho a desobedecer la norma, actuando como si supiera más que el legislador y se colocara por encima de los demás ciudadanos. No se le reprocha haber causado un peligro, sino haber actuado con una actitud que socava el valor de la igualdad ante la ley. Rusca, sin embargo, rechaza este enfoque, argumentando que si lo reprochable es la arrogancia, entonces debería tipificarse y castigarse esa actitud directamente, no a través del delito de peligro abstracto. Además, cuestiona que tal comportamiento pueda considerarse verdaderamente arrogante, ya que una persona podría estar dispuesta a extender la excepción a todos los que se encuentren en su misma situación. Por último, señala que el derecho penal no debe intervenir sobre defectos de carácter como la arrogancia, ya que estos pertenecen a la esfera moral o privada, no al ámbito de injerencia del Estado. Rusca, B., ibid., pp. 110-111. Estamos esencialmente de acuerdo con estos argumentos.

[1088] Alexander, L. / Kimberly Kessler, F., «Beyond the Special Part», en Duff, R./Green, S. (Eds.) *Philosophical Foundations of Criminal Law,* Oxford University Press, Oxford, 2001, pp. 253-278. Estos autores defienden que quien comete un delito de peligro abstracto sin generar un riesgo real igualmente merece castigo por poner en peligro el valor del Estado de derecho, al no respetar la necesidad de normas generales que rijan para todos. Para ellos, la culpabilidad se basa en no valorar adecuadamente la posibilidad de estar equivocado y en desobedecer una regla común. Rusca critica esta posición por ser vaga y por llevar a consecuencias problemáticas, como justificar el castigo de quienes creen equivocadamente estar infringiendo la ley. Además, objeta que basar el reproche penal solo en la desobediencia formal a la norma no es compatible con un derecho penal centrado en la protección de bienes jurídicos. Rusca, B., *op. cit.* N. 1086, pp. 112 y 113.

[1089] Se explican inmediatamente *infra.*

[1090] *Ibid.,* pp. 117 y ss.

de que el Estado pruebe la culpabilidad más allá de toda duda razonable [1091]. En este esquema, el acusado no debe probar su inocencia de forma absoluta, sino solo introducir una duda razonable sobre la existencia del peligro. Así, esta teoría busca un equilibrio entre las razones consecuencialistas que justifican el uso de reglas penales y las exigencias de justicia material, evitando tanto la impunidad como el castigo injusto. Nos parece, empero, que los argumentos del autor, aunque coherentes, no superan la doctrina de nuestro Tribunal Constitucional al respecto, resumida inmediatamente *supra* [1092].

### 4.2 **La existencia de un resultado jurídico como daño *sui generis***

La segunda gran postura ha sido defendida particularmente por el alemán Kindhäuser, que, en su monografía, pone de manifiesto una idea que ha acabado asumiendo una parte de la doctrina española [1093]. La idea es, resumidamente, que en los delitos de peligro abstracto lo que se está produciendo es un daño *sui generis*, equiparable al resultado de peligro o a un resultado ordinario, que no es otro que el daño a la confianza que se produce cuando, ante unas expectativas de seguridad de los ciudadanos en determinadas parcelas del derecho, el ciudadano que decide conculcar el tipo penal, lo hace dañando esta confianza.

La postura de Kindhäuser respecto a los delitos de peligro abstracto se fundamenta en una crítica al tradicional «paradigma de la agresión», que entiende que el injusto penal solo se legitima por la lesión directa de un bien jurídico [1094]. Kindhäuser propone un enfoque diferente: considera que los delitos de peligro (tanto concretos como abstractos) constituyen una forma autónoma de injusto material, ya que no solo se lesionan bienes jurídicos cuando se produce un daño real, sino también cuando no se garantiza la seguridad necesaria para disfrutarlos racional y despreocupadamente.

---

[1091] Rusca sostiene que esta presunción es razonable y válida desde un punto de vista probatorio y no vulnera el principio de presunción de inocencia si se interpreta correctamente porque es una inferencia basada en una regularidad empírica sólida: la mayoría de las veces, esas conductas efectivamente son peligrosas. No implica trasladar indebidamente la carga de la prueba al acusado, sino que, una vez probada la conducta, este puede introducir duda razonable sobre el elemento del riesgo. *Ibid.,* p. 120: *«así como resulta legítimo inferir la prueba del dolo en el caso particular de un conjunto de circunstancias genéricas, también puede inferirse la prueba de la peligrosidad ex ante de la conducta en concreto a partir de la realización de una acción que, en general, ocasiona un nivel de riesgo mayor al permitido».*

[1092] N. 1085.

[1093] La obra de referencia de Kindhaüser en este sentido es KINDHÄUSER, U., *Gefährdung als Straftat*, *Rechtstheoretische Untersuchungen zur Dogmatik der abstrakten und konkreten Gefährdungsdelikte*, Vittorio Klostermann, Frankfurt a. M., 1989.

[1094] KINDHÄUSER, U., «Estructura y legitimación de los delitos de peligro del Derecho penal», *InDret*, 1/ 2009, pp. 7 y ss.

Desde su perspectiva, el valor de los bienes jurídicos (entendidos como condiciones que permiten el libre desarrollo del individuo en una sociedad democrática) se ve afectado no solo por la lesión física o directa, sino también por la generación de contextos donde el uso de esos bienes no es seguro. Así, la comisión de una conducta típicamente peligrosa reduce esa seguridad general, menoscabando el valor del bien como medio para el desarrollo personal. En consecuencia, el injusto de estos delitos no radica en la lesión en sí, sino en la afectación de las condiciones objetivas de seguridad que deben rodear al bien jurídico protegido [1095].

Kindhäuser sostiene que esta afectación de la seguridad es un daño *sui generis* que constituye por sí mismo un resultado jurídico reprochable. El ciudadano debe poder confiar racionalmente en que determinadas condiciones de seguridad se mantendrán, por lo que, si alguien las rompe, aunque no cause un daño concreto, está generando un perjuicio punible. Eso sí, advierte que no toda alteración de la seguridad debe penalizarse, solo aquellas que afecten a competencias de protección que el individuo no puede asumir por sí mismo, sino que requieren protección institucional o colectiva. Por ello, la legitimidad de estos delitos depende de que la norma penal garantice estándares de seguridad imprescindibles y que el hecho típico perturbe efectivamente esas condiciones.

En suma, para Kindhäuser, el fundamento del injusto en los delitos de peligro abstracto no es la lesión del bien jurídico en sentido tradicional, sino la alteración de las condiciones de seguridad que permiten su uso tranquilo, lo cual supone una forma autónoma de daño jurídicamente relevante.

Una idea similar asume Corcoy [1096], poniendo el énfasis, empero, en el elemento subjetivo [1097], de tal forma que el ciudadano sufre una inseguridad sobre el disponer pacíficamente de sus bienes.

---

[1095] Las condiciones de seguridad se basan en la expectativa objetivamente fundada de un juzgador racional de poder usar sus bienes despreocupadamente. Kindhäuser, U., *Gefährdung, op. cit*. N. 1093, p. 282. En cambio, en Kindhäuser, U., «Acerca de la legitimidad de los delitos de peligro abstracto en el ámbito del derecho penal económico», *Hacia un derecho penal económico europeo. Jornadas en honor del profesor Klaus Tiedemann*: [Universidad Autónoma de Madrid, 14-17 de octubre de 1992], 1995, p. 449, la define como «la posibilidad de disfrutar sin riesgo de los bienes», dejando de lado la referencia subjetiva a las expectativas.

[1096] Corcoy Bidasolo, M., *Delitos de peligro y protección de bienes jurídico-penales supraindividuales..., op. cit.* N. *932*, p. 208: en los delitos de peligro «*los bienes jurídicos supraindividuales sirven a la persona a través de proteger su seguridad y confianza en el buen funcionamiento y en el nivel de riesgo de las distintas actividades en las que esa persona se ve involucrada […]*». Defiende, no obstante, la autora su postura, afirmando que solo podrá ser exigible esa seguridad en tanto la exigencia sobrepase el riesgo permitido en la actividad de la que se trate.

[1097] Feijoo lo califica, de hecho, de una seguridad en sentido psicológico, trasladando a esta autora todas las críticas que quepan contra el alemán. Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, nota al pie 44.

Las críticas vertidas contra Kindhauser provienen de varios sectores [1098], pero si hay una idea que las vertebra tiene que ver con que es difícil superar, con este argumento, la crítica de formalidad (que aquí se ha venido a llamar el problema del «injusto formal» [1099]), pues, en palabras de Feijoo, aunque Kindhauser ofrece una magnífica obra capaz de sistematizar el injusto administrativo, no se encuentra por ninguna parte por qué la contravención de ciertas cautelas que llevan a la perturbación de la confianza en ciertos bienes debe, en primer lugar, tutelarse por el derecho penal y, en segundo lugar, por qué no está presente entonces está técnica respecto del disfrute no perturbado de cualquier otro bien jurídico [1100]. En un sentido similar, crítica con el formalismo que impregna el razonamiento de Kindhäuser, Mendoza [1101].

---

[1098] Así, tanto Feijoo como Mendoza, que defienden posturas esencialmente distintas, coinciden en su crítica al alemán. Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 328 y Mendoza, por su parte, hablando de «injusto demasiado formalizado». Mendoza Buergo, B., *Límites dogmáticos, op. cit*., N. 1045, p. 140. También Molina, implícitamente, cuando afirma que «*La referencia a bienes como la seguridad supone en último extremo una forma alternativa de expresar la distinción entre tipos de peligro y de lesión. Por ello, cuando se utilizan a la vez ambas categorías se incurre en una cierta contradicción sistemática que consiste en calificar a la vez un tipo como de peligro y como protector del bien jurídico seguridad, cuando en realidad un comportamiento peligroso para un bien entraña una lesión de la seguridad de dicho bien*». Molina Fernández, F., *«Delitos contra la seguridad del tráfico», op. cit*. N. 935, p. 708. Respecto del bien jurídico, Molina, aunque no acepta en su totalidad la tesis del bien jurídico intermedio, sí toma algunos de sus elementos, negando que en los delitos contra la seguridad del tráfico se protejan exclusivamente la vida e integridad física, sino más bien el mantenimiento de la seguridad del tráfico como **presupuesto** de estos otros bienes. (La negrita es nuestra), ibid., p. 709. Una crítica completa de este autor sobre las posturas que intentan buscar una lesividad propia en la tentativa, así como en los delitos de peligro, rompiendo la unidad dogmática del tipo, en Molina Fernández, F., *Antijuridicidad penal y sistema del delito*, BdF, (Montevideo-Buenos Aires), 2021, pp. 813 y ss.

[1099] Muchos menos problemas tiene en este sentido Jakobs, que se conforma con afirmar que, en los casos en que el legislador crea delitos de peligro abstracto para estandarizar determinadas conductas en ciertos ámbitos sociales, el injusto de los delitos de peligro abstracto es eminentemente formal, pues se basa en la desobediencia a las normas. Los injustos de mera desobediencia, por mucha coherencia que puedan encontrar en el sistema del delito de Jakobs, quedan fuera de toda discusión cuando se trata de afirmar, como aquí se hace, la vigencia de un principio de lesividad material. Una explicación de la postura de Jakobs respecto de los delitos de peligro abstracto, en Mendoza Buergo, B., ibid., pp. 258 y ss. Originalmente en Jakobs, G., *Derecho penal: fundamentos y teoría de la imputación*. Parte general, Marcial Pons, Madrid, 2.ª ed., 1997, pp. 212-213 (6/86 y ss.).

[1100] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 328 y ss., afirmando que produce una «insatisfactoria administrativización del injusto». Ello es así, continúa Feijoo, «*ya que KINDHÄUSER no tiene suficientemente en cuenta que en última instancia lo que se quiere proteger son bienes jurídicos como la vida, la salud o el medio ambiente pero variando la técnica de protección; es decir, lo que caracteriza a los delitos de peligro abstracto de los que nos ocupamos aquí no es que tengan un objeto de protección distinto que otros tipos penales como las lesiones o el homicidio sino que la estructura típica es diferente. No se trata simplemente de tematizar la seguridad como bien jurídico u objeto de tutela sino de tematizar el comportamiento que genera inseguridad como injusto penal o como hecho jurídico-penalmente imputable.*»

[1101] Mendoza Buergo, B., *Límites dogmáticos, op. cit*. N. 1045. En este punto, considera Mendoza que las concepciones de Kindhäuser acaban por ser asimiladas por Jakobs cuando este recoge la idea de que la seguridad es condición para poder disfrutar de los bienes como un bien propio. Mendoza Buergo, B., «La configuración del injusto (objetivo) de los delitos de peligro abstracto», *Revista de Derecho Penal*

### 4.3 **Los delitos de peligro como tentativas imprudentes**

Más ampliamente defendida es la postura que equipara el injusto de los delitos de peligro con el de las tentativas imprudentes, en el entendimiento de que cuando se comete un delito de peligro abstracto se produce una vulneración de la norma objetiva de cuidado, sin que se siga un resultado lesivo. La idea central es que los delitos de peligro abstracto pueden ser comprendidos como una forma de conducta imprudente que se sanciona no por el daño causado ni por el riesgo efectivamente producido, sino por la violación de un deber de cuidado.

Se parte del principio de que la conducta típica implica una infracción del deber de cuidado, incluso cuando en la situación concreta el resultado peligroso no se haya producido o no haya sido verificable. De este modo, el injusto se sitúa íntegramente en la acción que el autor realiza, en cuanto transgresora de normas preventivas que buscan proteger bienes jurídicos mediante el control de conductas socialmente riesgosas.

Dentro de este argumento, son variadas las vertientes. Así, los alemanes Horn y Brehm [1102] inician esta discusión, también aceptada por Schünemann. En España encuentra apoyo en autoras como Rodríguez Montañés [1103]. Símbo-

---

*y criminología*, 9, 2002, nota al pie 15. La obra a la que se refiere Mendoza es Jakobs, G., *Sociedad, norma y persona en una teoría de un Derecho penal funcional,* Civitas, Madrid, 1996. Insiste también Mendoza, más allá de resaltar que la teoría de Kindhäuser no puede huir de una necesaria peligrosidad como elemento, en la idea de que no permite distinguir entre sanciones penales o administrativas, en la línea de lo expuesto por Feijoo. Mendoza Buergo, B., *ibid*., p. 52.

[1102] Cabe resaltar, ya incluso antes de explicar la postura de estos autores, que este entendimiento del delito de peligro abstracto como tentativa imprudente sin resultado solo aplica, según la doctrina desarrollada por Brehm, para los casos en que se quieran proteger de forma inmediata bienes jurídicos (entendemos que individuales), mientras que si lo que se quiere proteger son intereses abstractos el tipo solo desempeñaría una función de ordenación y regulación, quedando justificada la sanción por la mera realización de la conducta prohibida. Así se explica en: Pérez-Sauquillo Muñoz, C., *op. cit*. N. 1075, p. 132-133. Una revisión más pormenorizada de los postulados de ambos autores en Mendoza Buergo, B., Limites dogmáticos., *op. cit*. N. 1045, pp. 211-222.

[1103] Rodríguez Montañés, T., *Delitos de peligro*…, *op. cit*. N. 934, pp. 298 y 299. Para esta autora, partiendo de la idea de que la antijuridicidad material exige que, en el caso concreto, la conducta represente un peligro real para bienes jurídicos, descarta la importancia de la producción del resultado, tanto en el plano objetivo como en el subjetivo, destacando que en los delitos de peligro abstracto, a diferencia de los de peligro concreto, «*Lo que interesa es la peligrosidad ex ante, como atributo de la acción y si se da esa peligrosidad ex ante, el que ex post se haya producido o no peligro es algo que ni siquiera ha de ser considerado»* (p. 298), pues, «*Lo relevante es que en el momento de la acción no quedara excluida la potencialidad lesiva de su conducta*» (ídem). Esta construcción implica, en el plano subjetivo, que el dolo y la imprudencia no han de referirse al peligro concreto como resultado de la acción típica, afirmando (de forma que nos parece harto discutible, por la enorme dificultad práctica que conllevará su distinción), que el dolo respecto del peligro abstracto no implica dolo de peligro concreto. La obra dedica unas cuantas páginas, como es normal, a intentar diferenciar, a nuestro juicio de forma muy poco convincente, los supuestos en que hay dolo de peligro concreto y dolo de peligro abstracto. Crítica con la postura de esta autora: Mendoza Buergo, B., «La configuración del injusto…», *op. cit*. N. 1101, p. 59, pues excluye de

lo evidente de que esta interpretación no parece la más adecuada es la excesiva complejidad en la que se ha de sumergir la autora para explicar la diferencia entre la comisión dolosa y la imprudente, que acaba por diluirse y equipararse prácticamente en lo material [1104].

En los orígenes de esta discusión, autores como Horn y Brehm [1105] defendieron que para sancionar una conducta bajo un tipo de peligro abstracto no basta con la mera realización del hecho descrito en la norma, sino que también debe haber una infracción concreta del deber objetivo de cuidado en relación con el bien jurídico. Brehm, además, exige que la acción tenga una aptitud concreta *ex ante* para lesionar, evaluada desde el estándar del «hombre medio» cuidadoso, aunque también considerando posibles conocimientos especiales del autor. Así, estos delitos serían comparables a tentativas imprudentes, en tanto no requieren resultado, pero sí infracción de deber.

Sin embargo, Brehm limita esta visión a los casos en que la norma penal tenga como objetivo directo la protección inmediata de bienes jurídicos. Cuando la norma persigue fines meramente organizativos o regulatorios, el fundamento del delito pasa a ser formal: el mero incumplimiento de la conducta prohibida, prescindiendo del juicio material sobre el riesgo o el bien jurídico.

Schünemann, influido por Brehm, también equipara los delitos de peligro abstracto a tentativas imprudentes, aunque añade que la valoración de la infracción del deber de cuidado debe incluir criterios subjetivos, es decir, debe analizarse si el autor creía que estaba actuando con la debida precaución. Si él mismo sabía que no había adoptado todas las medidas necesarias, entonces hay reproche. No obstante, Schünemann reconoce que, si en el caso concreto

---

su teoría a los tipos que protejan bienes jurídicos supraindividuales inmateriales (a los que interpreta como delitos de lesión del bien intermedio), además de entender que la infracción del cuidado debido se daría con la realización de la conducta descrita en el tipo. En este caso concreto, Rodríguez Montañés, T., ibid., pp. 300 y ss., entiende que no se trata de anticipar la tutela de los bienes esenciales, como de proteger otros bienes cuya peculiar naturaleza exige el empleo de esta técnica (p. 301), motivo por el cual se afirma la «*legítima punición de toda realización típica sin necesidad de constatar la peligrosidad en relación con el bien inmaterial que mediatamente se protege*» (p. 302), con la particularidad respecto del tipo subjetivo de que no es necesario que se refiera al bien jurídico mediatamente protegido, de forma que el dolo o la imprudencia del sujeto solo deben referirse a los elementos típicos (acción típica y resultado, en su caso), sin corrección alguna (p. 303).

[1104] Más directo es en este sentido Romeo Casabona, que acepta que en los delitos de peligro abstracto no es posible diferenciar el dolo de la imprudencia. Romeo Casabona, C. M., «La peligrosidad y el peligro en la estructura del tipo del delito imprudente», en AA. VV., *La ciencia del derecho penal ante el nuevo siglo. Libro homenaje al profesor doctor don José Cerezo Mir*, Tecnos, Madrid, 2002, pp. 955 y ss. Crítico con esta postura, Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 338, arguyendo que la imposibilidad de distinguir entre dolo e imprudencia en los delitos de peligro abstracto va en contra del tenor literal de los preceptos que diferencian ambas modalidades, así como de la definición general de delito del artículo 10 CP.

[1105] Brehm, W., *Zur Dogmatik des abstrakten Gefährdungsdlikts,* Mohr, Tubinga, 1973; y Horn, E., *Konkrete Gefährdungsdelikte,* Otto Schmidt, Colonia, 1973.

la conducta resulta objetivamente inocua (porque el autor adoptó todas las medidas de seguridad) se debería permitir una reducción teleológica del tipo penal, para respetar el principio de ofensividad y evitar el castigo de actos no lesivos [1106].

Roxin, por su parte, y partiendo de las tesis de Schünemann, aunque defiende esta postura, lo hace desde un punto de vista algo diferente, centrándose no solo en el aspecto de vulneración del deber objetivo de cuidado, sino poniendo el énfasis en la vulneración del deber personalmente exigible al sujeto activo [1107].

Ninguna de las vertientes de este argumento puede aquí defenderse pues, una vez más, estas propuestas analizan el delito de peligro abstracto desconociendo las características que lo hacen valioso. No es posible reconducir las referencias del deber de cuidado (ya sea este objetivo o subjetivo), al resultado o la lesión del bien jurídico, pues en este caso estamos perdiendo todas las ventajas que presenta una estructura que persigue prevenir formas particulares de ataque donde existen dificultades para imputar el resultado lesivo final. Lo explica de forma especialmente clara Feijoo cuando afirma que no compagina bien esta interpretación teleológica con la intención técnica de los delitos de peligro abstracto, cuyo objetivo no es otro que estandarizar todo lo posible ciertas conductas, buscando precisamente que no sean los ciudadanos quienes decidan cuáles son las cautelas o precauciones necesarias en cada caso. Añade que el deber de cuidado no puede ir referido a un resultado típico inexistente en estos delitos, pues el objeto del dolo y la imprudencia en los delitos de peligro abstracto no es un resultado lesivo, sino solo los elementos típicos reco-

---

[1106] Resulta innovadora la postura de Schünemann, B., «Consideraciones críticas...», *op cit*. N.407, en un punto, y es que con su idea de los «puntos de conexión colectivos» (p. 201) pone el énfasis en el momento crucial en el que la intervención penal se vuelve necesaria en sociedades modernas, donde los intercambios sociales y económicos están marcados por cadenas de distribución complejas, masivas y anónimas. En estas circunstancias, es difícil identificar causalidades individuales claras. Por eso, se hace necesario adelantar la barrera de protección penal antes de que los cursos causales se vuelvan incontrolables. Cuando una conducta encaja en uno de estos puntos de conexión colectiva (es decir, cuando su peligrosidad está inserta en estas dinámicas colectivas y sistémicas) el adelantamiento de la intervención penal se justifica. Si no se da tal conexión, el tipo penal sería ilegítimo e incluso inconstitucional, por violar el principio de *ultima ratio.* Esta idea nos parece, por tanto, especialmente valiosa en lo que importa a esta investigación.

[1107] Roxin, C., *Derecho penal. Parte general*, *op. cit*. N. 921, § 11/121; que, partiendo de la idea de Horn y Brehm de hacer depender la punibilidad de una infracción del cuidado debido al modo de los delitos imprudentes (peligro sin resultado), añade la concepción de Schünemann, que modifica el punto de partida, añadiendo que la cuestión no depende de la infracción del cuidado debido como del subjetivo, es *decir, «habría que determinar el delito de peligro abstracto como una tentativa imprudente (en su caso inidónea)»,* según la cual la punibilidad sería posible si el autor deja de observar las medidas de precaución subjetivamente requeridas desde su perspectiva, por mucho que las precauciones adoptadas sean objetivamente suficientes para eliminar el peligro.

gidos en el tipo penal de que se trate [1108]. Consideramos oportuno reproducir los términos concretos en los que resume su idea:

*«Desde un punto de vista subjetivo los delitos de peligro des-subjetivizan la referencia al daño, des-subjetivación que aumenta a medida que se abstrae el peligro. En estos delitos no se trata sólo de prescindir de la esclavitud de la relación de causalidad sino también de la esclavitud de tener que exigir la relación subjetiva: la constatación de que el posible resultado fue previsto o podía ser previsto por el autor. Cualquier intento de construir la responsabilidad personal de los delitos de peligro a partir del delito de lesión desconoce la realidad dogmática de las diferentes estructuras típicas de los delitos de peligro y de los delitos de lesión como consecuencia de los diversos fines político-criminales de cada modalidad.»* [1109]

### 4.4 **La traslación de la teoría de la imputación objetiva**

La misma crítica, adelantamos, puede hacerse a la teoría que presenta Mendoza Buergo. Tras un estudio completísimo sobre la evolución del delito de peligro abstracto, la autora propone su propia versión interpretativa, con el claro objetivo de evitar el vaciamiento de lesividad o la formalización del injusto, arriba referidos.

Mendoza considera que todo delito debe responder a una estructura similar y que, de hecho, a todos ellos les es aplicable el juicio de imputación objetiva, por lo que lo esencial para garantizar la lesividad material del caso con-

---

[1108] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p.324, aunque reconociendo, en nota 37, que es dominante esta equiparación en la doctrina española entre delitos de peligro e imprudencia sin resultado que acepta, sin mucho matiz, la idea de que el delito de peligro representa meramente un adelantamiento en las barreras de protección, castigando la tentativa imprudente de manera excepcional. El propio Feijoo dedica un artículo entero a tratar de desmentir esta afirmación en lo relativo a los delitos contra la seguridad colectiva. Feijoo Sánchez, B., *ibid*, defendiendo, en esencia, que los delitos protegidos bajo la rúbrica de delitos contra la seguridad colectiva, aunque no protegen bienes jurídicos nuevos, sí los protegen de forma esencialmente distinta, enfocándose esta técnica en la protección contra riesgos expansivos e incontrolables que pueden afectar a un número indeterminado de personas, acercándose a la lógica del peligro general. La idea es que los delitos que siguen esta lógica no son delitos subsidiarios respecto del delito de resultado, sino que responden a una lógica propia, como es la creación de riesgos para la colectividad sin necesidad de daño efectivo, *passim*. Concretamente en p. 3: «*En definitiva, el hecho que debe ser objeto de referencia (tipo objetivo) del dolo y de la imprudencia (tipo subjetivo) en los delitos contra la seguridad colectiva es distinto al hecho que es objeto de referencia del dolo y de la imprudencia en los tipos puros de resultado. La relación de los delitos contra la seguridad colectiva con el delito imprudente es más bien «funcional» desde un punto de vista político-criminal. Permiten intervenir al Derecho penal en ámbitos en los que sólo podía intervenir mediante los tipos de resultado tradicionales imprudentes, con todas las limitaciones que ello implicaba*».

[1109] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 326.

creto es que el juzgador realice este juicio de subsunción. Solo serán subsumibles en el delito de peligro abstracto aquellas conductas en que se pueda afirmar que se ha creado (o elevado) un riesgo jurídicamente desaprobado. Se trata, así, de analizar todos aquellos factores que normalmente forman parte de este juicio, todo ello con el objetivo de asegurar que, con esta interpretación teleológica, ningún delito formal se castigue, sino solo aquellos que presentan un riesgo jurídicamente desaprobado.

Parte Mendoza de la idea de que la equiparación del injusto del delito de peligro abstracto y el tipo objetivo del delito imprudente es insuficiente, afirmando que *«no se avanza mucho si se dice que el injusto objetivo tiene como contenido material la infracción del cuidado debido que, sin embargo, se afirma respecto de una acción que para cualquier espectador objetivo ex ante con los conocimientos del autor no podía poner en peligro el bien jurídico»* [1110]. Se denuncia que, de esta forma, aunque pretendiendo aportar un criterio material para la delimitación del contenido de injusto, se acaba por interpretar aquel de manera formal, sin llegar hasta el final en la materialización a través de la consideración individual del injusto del hecho concreto. De forma que, con el objetivo de constituir un injusto definitivo, la autora propone realizar un juicio individual sobre la verdadera infracción del deber de cuidado que atienda al juicio de previsibilidad del posible resultado en el caso concreto y determine el incumplimiento o no por parte del autor de las exigencias derivadas del deber de cuidado, siempre en consideración, también, a las circunstancias concretas [1111]. La propuesta de Mendoza pasa por aplicar el primer juicio de imputación, pues a juicio de la autora *«abre una solución más adecuada al problema planteado, de modo que aunque en los delitos de peligro abstracto no se plantee la cuestión de la imputación objetiva de un resultado –el llamado por algunos segundo nivel de imputación– sí debe considerarse necesaria la comprobación individual del primer nivel de imputación, es decir, la imputación de la propia conducta, o imputación del hecho al tipo objetivo, fundada en la creación de un peligro ex ante no permitido abarcado por el fin de protección o dentro del alcance del tipo* [1112]*.»*

---

[1110] Mendoza Buergo, B., «La configuración del injusto…», *op. cit*. N. 1101, p. 60.

[1111] Ibid, p. 60-61.

[1112] *Ibid.,* p. 73. *«La base de esta propuesta parte de la idea de que algunos de los criterios enunciados en la teoría de la imputación objetiva no son aplicables sólo a los delitos de resultado, de modo que las cuestiones de si el autor ha creado un riesgo objetivo desaprobado y cuál es el alcance del tipo no atañen primariamente a la imputación de resultados, sino que circunscriben ya la propia relevancia típica y la antinormatividad de la conducta. En esa medida, el primer nivel de imputación objetiva constituye, en realidad, una teoría sobre la delimitación de la conducta típica y, con ello, del injusto típico; por ello*

Las críticas contra esta forma de entender el injusto del delito de peligro abstracto son, sin embargo, contundentes. Feijoo, en un intento por mostrar los inconvenientes que supondría llevar hasta el final esta interpretación, afirma que *«Si se pretende en serio "igualar" el tratamiento de la determinación y comprobación judicial del injusto objetivo (primer nivel de imputación) de los delitos de peligro abstracto con todos los demás»* [1113], sosteniendo que el primer criterio de imputación deba funcionar de forma similar a como funciona con el delito de lesión o peligro concreto, ello tiene las siguientes consecuencias dogmáticas: «[…] *el delito doloso de peligro abstracto debería entenderse como siendo equivalente a una tentativa (idónea) y, si tras realizar dicha conducta se produce un resultado no imprevisible, la única solución posible sería considerar este hecho un delito de resultado doloso, lo que conduce a una "indeseable ampliación del alcance de los tipos de resultado"*» [1114].

El problema que presenta esta teoría parece evidente, si tenemos en cuenta el objeto de protección que Mendoza tiene en mente. Si el legislador, al tipificar el delito de peligro abstracto de fraude alimentario, por utilizar un ejemplo, estaba pensando en castigar aquellas conductas que, a pesar de tener ínsitas todas las propiedades necesarias para afirmar el juicio de imputación objetiva (al menos en su primer escalón, es decir, el relativo a la creación de un riesgo desaprobado sin resultado), ¿por qué no acudir, entonces, al régimen ordinario de la tentativa? ¿Qué razones existen para castigar, generalmente, de forma privilegiada, estos supuestos, cuando lo que se da no es más que una tentativa de homicidio? [1115] También señala Feijoo los problemas del tipo sub-

*entiendo que resulta generalizable a la comprobación de la tipicidad de toda conducta, sea o no la de un delito de resultado […]», p. 74.*

[1113] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, nota 58.

[1114] *Idem*. En palabras del propio Feijoo, ídem: «*se acaba ahuyentando al diablo con Belcebú*».

[1115] En este sentido, la postura de Doval respecto de los delitos contra la salud pública resulta, creemos, clarificadora, al negar que la estructura de los delitos de peligro abstracto fuera redundante o innecesaria, sino más bien todo lo contrario, pues, aunque los bienes jurídicos finales que el legislador trata de proteger son en esencia los mismos (aquí, salud y vida), los ataques a los que se encuentra expuestos no pueden ser válidamente protegidos ni acudiendo a la estructura de la tentativa ni a la del delito imprudente. Señala el autor que, de hecho, los delitos de peligro se parecen a una y otro; a la tentativa por cuanto ambos dan lugar solo a la puesta en peligro y no a la lesión del bien; y al delito imprudente en tanto en ambos falta la intención lesiva del autor (p. 242), sin que ninguna de las estructuras sea suficiente, aisladamente, para solucionar estos problemas. La tentativa por cuanto falta el elemento intencional que le es propio (pues en los delitos de peligro el dolo del autor se proyecta, cómo máximo, según Doval, sobre la actividad peligrosa, p. 243), y el delito imprudente por cuanto, aunque la conducta de creación de peligro se llevara a cabo de manera imprudente, tradicionalmente se ha considerado que no había forma de castigar estas formas imperfectas de ejecución de delitos de lesión sin resultado (p. 243), seguramente por una cuestión de insignificancia que el autor no explicita. Así, reconoce Doval, aunque, de hecho, a todos ojos la estructura de los delitos de peligro (para la salud) parece ser equivalente a la tentativa imprudente del delito de lesiones, esta construcción encierra algunas contradicciones, pues la tentativa exige la afirmación de una intención que luego es negada con la afirmación de la imprudencia, constituyendo algo así como

jetivo, pues las únicas diferencias entre un delito de peligro abstracto y un delito de resultado serían objetivamente la ausencia de un resultado típico (ya sea lesivo o de peligro concreto) y subjetivamente la ausencia de la volición del resultado, de manera que, en sus palabras, una configuración deficiente del tipo objetivo obliga a acudir al elemento volitivo del dolo como criterio decisivo de imputación. Así, lo que hace Mendoza es negar a los delitos de peligro abstracto una estructura típica propia [1116].

### 4.5 **La resignificación del injusto del delito de peligro abstracto sin referencia a la lesión del bien jurídico**

Debe resaltarse, por ser esencialmente distinta en cuanto a su formulación, la propuesta interpretativa que ofrece Feijoo, que se propone realizar una

---

una imposibilidad lógica (idem, nota 155). Para salvar estas dificultades, por tanto, era necesario que el legislador crease una nueva vía de solución de supuestos «*caracterizados por la indiscriminada puesta en peligro*» pero donde hay una «*ausencia de resolución lesiva en el autor*». Doval Pais, A., *Delitos de fraude alimentario…, op. cit. N.957.*

[1116] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 334, nota 58, apartado 3, especialmente crítico, achacando a esta construcción que no queda del todo claro, entonces, cuál es la contribución concreta de sus referencias a la teoría de la imputación objetiva, ya que las conclusiones alcanzadas no aportan innovaciones significativas respecto a soluciones tradicionales ya conocidas. Si se pretende aplicar realmente un enfoque teleológico de interpretación, es necesario examinar la autonomía conceptual de los delitos de peligro abstracto en relación con las estructuras típicas de los delitos de peligro concreto y de resultado lesivo. Esto implica centrarse en sus elementos distintivos, en lugar de derivarlos directamente de la dogmática de los delitos de resultado. En este sentido, Mendoza no ha aplicado verdaderamente la teoría de la imputación objetiva a los tipos de peligro abstracto, sino que ha trasladado a estos la estructura dogmática propia de los delitos de resultado lesivo. Una de las consecuencias de la crítica de Feijoo es que, al considerar que el injusto de los delitos de peligro es propio y diferenciado del delito de resultado «correlativo» es que, coherentemente, cuando se trate de solucionar cuestiones concursales, los tipos puros de resultado no abarcan todo el desvalor del hecho. Esta observación resulta fundamental porque reconoce que el injusto de los delitos contra la seguridad colectiva no se agota en la lesión efectiva de bienes jurídicos individuales, sino que incluye una dimensión adicional de «peligrosidad abierta» Por el contrario, *«los delitos contra la seguridad colectiva no abarcan el desvalor de la lesividad de la conducta y los tipos puros de resultado no abarcan el desvalor de la 'peligrosidad abierta' que encierran los delitos contra la seguridad colectiva*». Esta complementariedad del desvalor justifica la aplicación del concurso ideal regulado en el artículo 77 del Código Penal. Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 4. En un sentido similar, Doval, ibid., pp. 248 y ss., sosteniendo que en los delitos contra la salud pública (que el autor enmarca como de los delitos de riesgo en general) la particularidad estriba en que los sujetos objeto de ataque no solo son indeterminados *ex ante*, sino incluso también después de la agresión. Esto determina, por tanto, que la forma de castigar estos supuestos no pueda derivarse, tanto solo, de una supuesta anticipación de una tutela individual pues, en todo caso el objeto de ataque es plural y difícilmente quedará abarcado por el delito de resultado ulterior, que no podrá cubrir suficientemente el desvalor generado por el delito de peligro. Esta idea se desarrolla en p 363 y ss., donde Doval individualiza los escenarios concursales siendo, en esencia, su postura favorable al reconocimiento de un concurso ideal en caso de que al delito de peligro se sume un resultado de muerte o lesiones afirmando la imposibilidad de «absorción», que sería la consecuencia lógica de afirmar que aquí no hay más que una tentativa de un delito de resultado.

reconstrucción de la dogmática de los delitos de peligro abstracto atendiendo a sus particularidades, y cuya idea fuerza pasa por rechazar la equivalencia entre el injusto del delito de lesión y el del delito de peligro. Mediante el análisis de los delitos contra la seguridad, aboga por abandonar la discusión del terreno del bien jurídico, afirmando que el problema de estos tipos delictivos no es tanto un problema sobre lo que se protege sino sobre la forma en que estos bienes son normalmente atacados, insertando su discurso dentro del paradigma de la sociedad del riesgo.

Parte, así, de la idea de que los delitos contra la seguridad, si bien están destinados a proteger bienes jurídicos finales tradicionales (vida y salud, destacadamente, medio ambiente, en otros), son algo diferente a meras tentativas de lesión, así como tampoco son meras contravenciones del deber objetivo de cuidado respecto del bien jurídico final. La lógica de ataque, generalizado y difundido, hace que el legislador deba optar por una opción político-criminal diversa, donde lo que se protege mediatamente son estos bienes finales, pero también se protegen las condiciones de seguridad de ciertas instituciones. La protección de la competencia de la Administración para ordenar ciertos sectores riesgosos está en la base, y sirve como condición para la protección de otros bienes finales. La forma de atacar estos bienes finales se aleja de la lógica del ataque doloso dirigido a dañar el bien final, y la forma de concebir el injusto también debe modificarse.

La propuesta de Feijoo se articula en torno al concepto de «organización normativamente insegura». Según esta construcción, los delitos de peligro abstracto tipifican supuestos donde «*no se tipifica la arrogación, sino sólo la desprotección ilegítima de otro ámbito de organización (delitos de peligro concreto) o la incorrecta organización del propio ámbito de organización (delitos de peligro abstracto)*»[1117]. Esta formulación presenta ventajas evidentes sobre las teorías precedentes, especialmente porque supera la dicotomía tradicional entre formalismo puro y administrativización del derecho penal.

En el plano del tipo objetivo, Feijoo establece que *«la conducta no sólo debe reunir todos los elementos típicos de acuerdo con los cuales el tipo define expresamente la conducta como estadísticamente peligrosa, sino que esos elementos deben permitir que se pueda imputar al infractor en el caso concreto una organización insegura del propio ámbito de organización»*[1118]. La precisión es fundamental: no se trata de constatar «la idoneidad de la conducta para lesionar», sino de verificar «desde una perspectiva normativa o jurídica» que

[1117] Feijoo Sánchez, B., «Seguridad colectiva y peligro abstracto…», *op. cit.* N. 951, p. 331.
[1118] *Ibid.,* p. 333.

no se ha «controlado o dominado suficientemente el propio ámbito de organización» [1119].

La construcción del tipo subjetivo es particularmente elegante. El dolo consistiría en la representación de: *(a) «todos los elementos del tipo que definen la conducta como estadísticamente peligrosa» y (b) «la falta de un control o un dominio suficiente de la situación que hace que no se trate de una organización segura para otros potenciales ámbitos de organización»* [1120]. Como señala Feijoo, *«en los delitos de peligro abstracto dolosos no hace falta que el autor realice juicio alguno relacionado con otros ámbitos de organización u otras personas sino solamente enjuicie su propio ámbito de organización como inseguro»* [1121].

La teoría de Feijoo presenta ventajas innegables. En primer lugar, una coherencia sistemática, pues ofrece una construcción que respeta la autonomía dogmática de estos delitos, pero sin llegar a desconectarlos del sistema general. En segundo lugar, una operatividad práctica, pues permite distinguir el injusto penal del administrativo, evitando las críticas de formalismo y administrativización. Finalmente, deja de lado los problemas de constitucionalidad asociados a las teorías basadas en las presunciones. Para completar la teoría de Feijoo (especialmente en lo relativo al elemento de incorrecta organización del propio ámbito) es necesario, no obstante, interpretar su propuesta a la luz de su pensamiento, algo más amplio, en cuanto al injusto de los delitos en general, para así concretar a qué se refiere el autor con el paradigma de la organización insegura [1122]. La norma penal, con la evolución y compleji-

---

[1119] *Ibid.,* p. 334.

[1120] *Ibid.,* 337.

[1121] *Idem.* En todo caso, cabe destacar que ya Schünemann apuntaba, en su defensa del modelo del delito de peligro abstracto, que el deber de cuidado no debe ser interpretado por los tribunales como ocurre con los delitos imprudentes de resultado, sino que el mandato es mucho más determinado y genera más seguridad jurídica., Schünemann, B., «Consideraciones críticas...», *op cit*. N.407, p. 201, «Pues en el delito imprudente de resultado, la norma de comportamiento concreta por regla general no está escrita y es formulada posteriormente por parte de la jurisprudencia, mientras que en el delito de peligro abstracto, el legislador puede y debe crear un tipo mucho más concreto para que este pueda cumplir con el principio *nullum crimen sine lege*».

[1122] Algunas de las ideas que pueden ayudar a resignificar este injusto se pueden encontrar en: Feijoo Sánchez, B., «Imputación objetiva en el Derecho penal económico y empresarial. Esbozo de una teoría general de los delitos económicos», *InDret*, 2/2009, p. 8: «*La pretensión de explicar dogmáticamente todo el Derecho Penal económico a partir de la exclusiva existencia de deberes de no lesionar el patrimonio ajeno o funciones estatales de ordenación de la economía es, en mi opinión, una empresa condenada al fracaso*». Para el autor, en un mundo donde lo permitido y lo no permitido está fuertemente regulado en el orden extrapenal, no será viable ya afirmar que los deberes que vinculan a las personas son simplemente de no lesionar bienes jurídicos, sino que pueden implicar toda una serie de deberes positivos, y esto es precisamente lo que el autor define como hacerse cargo de los ámbitos de organización propios. Para Feijoo, la determinación de la tipicidad penal, especialmente en el derecho penal económico, exige considerar el contexto normativo y las posiciones jurídicas sobre las que recae la conducta. La valoración

zación de las sociedades, no tiene como único objeto la evitación de ciertas conductas lesivas (el paradigma del simple *naeminen ladere* hace tiempo que se ha superado), sino que los sujetos, en función de su posición en la sociedad, tienen una serie de competencias que gestionar, de las que derivan toda una serie de deberes que no solo son de contenido negativo. Es decir, el sujeto obligado por la norma penal no solo debe abstenerse de realizar ciertas conductas, sino que debe organizar su propio ámbito de una manera idónea o, en términos del propio Feijoo, organizada. Cuando esa esfera de organización propia se descontrola tendremos el injusto propio de un delito de peligro abstracto, con independencia de que, más adelante, por factores ajenos a la voluntad del actor, este curso de los acontecimientos no genere un peligro. No es necesario, pues, que haya una afectación a la organización de esferas ajenas (esto correspondería con el paradigma del delito de peligro concreto), sino simplemente que aquello que está encomendado a la incumbencia, competencia o esfera del actor, se mantenga dentro de su dominio, en lo que Feijoo denomina organización segura del ámbito propio. No es necesario, pues, aludir a otros bienes finales distintos de esa esfera propia, previa, cuya organización compete al sujeto, con independencia de sus conocimientos o intenciones sobre el bien final. El delito de peligro abstracto implicará, por tanto, un conocimiento sobre la realización de los elementos del tipo, con conocimiento de que se está produciendo esta desorganización del ámbito que es propio, sin que, por tanto, el dolo deba referirse a más elementos que a aquellos contenidos en el tipo y aquellos que se refieran a la organización cuidadosa. Es, por tanto, teóricamente imaginable y plausible la imprudencia en un terreno como este, pero solo se dará en tanto en cuanto haya un descuido sobre la falta de organización de la esfera o ámbito propio, sin necesidad, insistimos, de referirse esos conocimientos al resultado lesivo final, pues no estamos vinculados a la lesión de bien final alguno.

Es esta una forma de entender el injusto del delito de peligro abstracto, a nuestro modo de ver, absolutamente coherente con su génesis, pues lo hace responder y ser efectivo en aquellas situaciones en que fracasa el tipo tradicional, a saber, el que permite la imputación por el resultado. Traer de nuevo a colación el conocimiento sobre el resultado redundaría en la inutilidad de estas figuras. Y también es coherente con la situación propia de la IA, donde la or-

no puede hacerse en abstracto, sino en función de los deberes (ya sean negativos, de no lesionar, o positivos, de actuar en beneficio de otros) que derivan de esas posiciones. No toda infracción de deberes implica una conducta típica, pues el injusto penal requiere un desvalor cualificado que debe analizarse según las relaciones jurídicas y normas del orden jurídico primario. Así, la idea fuerza es que la teoría del tipo penal no puede ignorar el marco normativo previo, especialmente en ámbitos regulados con deberes especiales.

ganización de ámbitos de competencia propios, altamente especializados y compartimentalizados, es un rasgo innegable del proceso de producción y uso. Es este un ámbito, donde naturalmente tienen sentido las razones por las cuales el injusto propio lo será en cuanto posibilidad de controlar la parcela concreta del proceso productivo sobre el que el sujeto tiene el conocimiento. Conocimiento limitado que, en la configuración de sus deberes de actuación, genera una incumbencia también estrecha, no ya sobre la lesión de los bienes jurídicos finales, sino sobre el proceso en sí mismo. No en vano, son numerosos los autores que acaban por reconocer como excepción al régimen del delito de peligro como tentativa imprudente precisamente los ámbitos de alta organización y formalización de obligaciones.

Destaca Feijoo, no obstante, que no significa su postura una aceptación sin límites del uso de delitos de peligro abstracto puros y, de hecho, aclara que los delitos de aptitud deberían ser la regla, y que se debe intentar fomentar (allí donde sea posible) interpretar los delitos de peligro en el sentido de que exijan una idoneidad lesiva real en el caso concreto, y nunca basar una condena en la presunción. Destaca que, de hecho, contamos con jurisprudencia asentada que, en la interpretación de diferentes delitos del Código Penal, ha ido acuñando un concepto esencialmente correcto del delito de aptitud [1123].

## 5. LA UTILIDAD DEL MODELO DE DELITOS CONTRA LA SEGURIDAD COLECTIVA

Si tratamos de pensar, habida cuenta de lo expuesto hasta el momento, en cuáles son los posibles delitos que más se apegan a los caracteres hasta aquí expuestos la respuesta parece obvia. Si se trata de prevenir hechos que generan riesgos difusos (sobre personas indeterminadas), sobre bienes jurídicos tradicionales (vida, salud, integridad, pero también otros como no discriminación o privacidad), fruto de tecnologías producidas en masa y fácilmente descontrolables, pero con amplios márgenes para su regulación y ordenación por la Administración, parece evidente que el modelo de los delitos contra la seguridad colectiva puede ser un buen punto de partida.

Se adelanta ya que el delito que se propondrá en la parte final de esta investigación beberá de los dos grupos de delitos analizados en este capítulo. En primer lugar, bebe de la prolija discusión sobre la génesis y el sentido de la protección de la salud colectiva en los delitos de fraude alimentario y, en se-

[1123] Nos remitimos aquí a las referencias y citas de la nota 949.

gundo lugar, reproduce la estructura de los delitos contra la seguridad colectiva (replicando esencialmente la estructura de los delitos de riesgo catastrófico) por lo incontrolado y desconocido de sus consecuencias, tanto en lo relativo a los bienes jurídicos finales (mediatamente protegidos), como en lo relativo a la indeterminación de sus víctimas.

Bajo la rúbrica «De los delitos contra la seguridad colectiva», el legislador tipifica un gran número de conductas de naturaleza muy diversa, pero con un denominador común: en la medida en que afectan a la seguridad colectiva se trata de proteger la seguridad material de grandes masas de población ante daños y peligros de gran magnitud y de alguna forma incontrolables [1124]. Incide Muñoz Conde en la idea de que en el fondo de todos estos delitos late una idea común, cual es intervenir penalmente de forma anticipada, castigando conductas peligrosas cuando estas se dan en ámbitos de especial trascendencia, con riesgo tanto para bienes jurídicos personales diversos (vida, integridad física, salud o patrimonio) como para bienes jurídicos sociales o universales (medioambiente, flora y fauna) [1125]. Se configuran, así, en delitos de peligro general, poniéndose el acento tanto en las especiales características de la acción, como en el marco o contexto en que esta se produce. Han de entenderse, por tanto, como delitos de peligro referidos directamente a un bien jurídico de carácter universal (seguridad colectiva), pero indirectamente a bienes jurídicos de carácter individual, lo que les ha valido la denominación de delitos de peligro común.

Resalta Muñoz Conde, asimismo, las dificultades que plantea esta naturaleza doble, tanto a la hora de determinar el concurso respecto de la posible lesión de bienes jurídicos individuales, como la distinción entre delitos de peligro concreto y abstracto, que no aparece de forma clara en algunos supuestos, dificultando, así, la solución concursal. Por lo demás, según señala este autor, en el Título abundan también los delitos de peligro hipotético, en los que la acción exige una aptitud o idoneidad para producir resultados lesivos, más que un concreto peligro.

---

[1124] Muñoz Conde, F., *Derecho Penal. Parte especial, 25.ª Ed, revisada y puesta al día con la colaboración de Carmen López Peregrín*, Tirant Lo Blanch, Valencia, 2023, p. 633. El propio Muñoz Conde pone de manifiesto la posibilidad de interpretar de dos formas esta seguridad colectiva; o bien como garantía de evitación de daños de dimensión supraindividual (tratando, así, todos los tipos bajo esta rúbrica de incriminar conductas que pongan en peligro bienes colectivos o bien por su dimensión cualitativa o bien por el número de personas afectadas), o bien como sinónimo de tranquilidad colectiva, poniéndose el énfasis en la creación de un clima de garantía social. Aunque Muñoz Conde se apresura en señalar que la diversa redacción y significado de los tipos parece conducir a que solo es posible la segunda acepción, no compartimos esta idea, adscribiéndonos, por tanto, a un entendimiento de la seguridad colectiva como forma de proteger bienes jurídicos finales diversos.

[1125] *Ibid.*

El Código Penal español recoge, en el capítulo primero del Título XVII (comprehensivo de los genéricos delitos contra la seguridad colectiva), concretamente en sus artículos 341 a 350, los denominados delitos de riesgo catastrófico [1126]. Nos centraremos en el primero de estos capítulos por encontrar en los delitos aquí tipificados una lógica común con los riesgos tecnológicos arriba descritos, constituyendo ejemplos valiosos de tipificación que pueden servir de inspiración a los nuevos tipos delictivos contra los daños causados por IA.

Este capítulo tipifica delitos contra la seguridad colectiva, entendida como bien jurídico supraindividual autónomo, o bien como la expresión de conductas que, en abstracto, son idóneas para poner en peligro la vida y/o integridad de una pluralidad de personas indeterminadas [1127]. Advierte Carbonell Mateu, empero, que deberá atenderse en la aplicación de los tipos al principio de relevancia para afirmar la tipicidad de las conductas, sin que la mera subsunción sea suficiente para afirmar la tipicidad [1128].

### 5.1 **Delitos relativos a la energía nuclear y a las radiaciones ionizantes [1129]**

El primer tipo de delitos (sección 1.ª) se corresponde con los delitos relativos a la energía nuclear y a las radiaciones ionizantes (arts. 341 a 345), castigándose conductas de diferente índole, todas ellas relacionadas con la puesta en peligro de los bienes personales arriba referidos. Para Muñoz Conde, en estos delitos se presenta una característica adicional, y es que representan un «riesgo global», *«que va más allá del ámbito nacional y afecta a toda la Humanidad en su conjunto e incluso, igual que en el caso del medio ambiente, a las generaciones futuras, por lo que la inter-*

[1126] El resto de los capítulos del título de Delitos contra la seguridad colectiva recogen, respectivamente, los delitos de incendio (Capítulo II), delitos contra la salud pública (Capítulo III), y delitos contra la Seguridad vial (Capítulo IV).

[1127] Cardenal Montraveta, S., «Delitos de riesgo catastrófico: energía nuclear, estrados y otros delitos de riesgo (arts. 341-350)», en Corcoy Bidasolo, M., (Dir.), *Manual de Derecho Penal, Parte especial*. Tomo 1, 3.ª ed. Tirant Lo Blanch, Valencia, 2023, p. 653. La referencia es, a su vez, a pronunciamiento del Tribunal Supremo: STS 538/00, de 25 de abril, FJ1.

[1128] Carbonell Mateu, J. C., «Lección XXXII, Delitos contra la seguridad colectiva (I): riesgos catastróficos», en González Cussac, J. L. (Coord.), *Derecho penal parte especial*, 8.ª ed., Tirant Lo Blanch, Valencia, 2023, p. 635. Esta afirmación, afirmada respecto de otras secciones donde el legislador ha optado por la tipificación de delitos de peligro abstracto puro sin exigencia de indicios de peligrosidad para bienes concretos, puede generar polémica, pero lo cierto es que, como a continuación se verá, en los delitos aquí tipificados se especifica, o bien mediante la alusión a un peligro concreto o bien mediante la alusión a la aptitud para poner en peligro bienes concretos, esta posibilidad de interpretación material que propone Carbonell sin que suponga una fricción con el principio de legalidad.

[1129] Para una profundización en los delitos contra la seguridad nuclear, de obligada referencia: De la Cuesta Aguado, M. P., *Respuesta penal al peligro nuclear,* PPU, Barcelona, 1994.

*vención del Derecho penal se fundamenta sin necesidad de exigir la prueba de un peligro concreto»* [1130].

La energía nuclear supone un magnífico ejemplo de tecnología dual, pues sus beneficios como fuente de energía, aunque no siempre pacíficos, resultan incontrovertidos, pero lo que en ningún caso ha resultado cuestionada es su capacidad destructora. Los ejemplos históricos de desastres de perfil catastrófico son difícilmente desconocidos (los desastres de Chernóbil, Fukushima, Hiroshima y Nagasaki, sin ir más lejos, pero también algunos menos conocidos como el accidente producido en el acelerador de partículas del Hospital clínico de Zaragoza con consecuencias letales para varios pacientes [1131]). El peligro nuclear es indudable, no solo por su potencialidad como arma en el contexto bélico, sino por las consecuencias catastróficas de su empleo negligente.

Entrando a analizar los tipos, se castigan en esta primera sección conductas diversas:

– La liberación de energía nuclear o elementos radioactivos (siempre que se ponga en peligro, con ella, la vida, salud o bienes de personas [1132])

Como explica Muñoz Conde, por la propia configuración de la acción típica (liberación de esta energía), la protección de los bienes individuales adquiere una dimensión necesariamente colectiva, pues la imposibilidad de controlar el peligro creado impide ya, desde la liberación, acotar el riesgo para los bienes de personas concretas. Es evidente, continúa este autor, que en este tipo de situaciones el peligro está implícito en la misma acción, por mucho que luego no se concrete en lesiones determinadas. No se trata, por lo tanto, de un peligro concreto, sino de un peligro general [1133].

Se trata de un delito común, aunque lo más frecuente será que solo determinados sujetos estén en condiciones de provocar estas liberaciones, que serán

---

[1130] Muñoz Conde, F., *Derecho penal. Parte especial*., *op. cit*. N. 1124, p. 634.

[1131] Todos ellos citados por Muñoz Conde, F., ibid., p. 635. Mayores detalles de este accidente nacional y sobre las repercusiones que tuvo a nivel legal, acelerando la trasposición de directivas europeas de protección radiológica y para formalizar la figura del físico médico hospitalario, asegurando un mayor nivel de supervisión en los tratamientos con fuentes de radiación: Arranz, L., «Leçons tirees de l'accident d'un accelerateur lineaire de l'hopital de Saragosse le 5 decembre 1990», *Radioprotection,* vol. 44, 4, 2009, pp. 405-416.

[1132] Art. 341. La pena prevista es pena de prisión de quince a veinte años, e inhabilitación especial para empleo o cargo público, profesión u oficio por tiempo de diez a veinte años. Este artículo se refiere a que la conducta produzca daños contra los «bienes» de las personas, sin que se llegue a especificar a qué tipo de bienes se refiere el legislador. Ha sido criticado por la doctrina, con razón, que se produzca esta equiparación penológica entre bienes eminentemente personales y bienes, que han de entenderse, necesariamente, como patrimoniales, configurando, por tanto, una sanción desproporcionada. En un sentido similar: Muñoz Conde, F., ibid., p. 637.

[1133] *Ibid.,* p. 637.

esencialmente los sujetos de las estructuras empresariales, así como los técnicos de las administraciones encargadas del control administrativo de estas instalaciones [1134].

Especialmente relevante, por su importancia en la práctica, es la tipificación por imprudencia grave (art. 341), pues serán los casos más frecuentes, dado que los casos de accidentes nucleares o escapes radioactivos más comunes se deben o bien a errores humanos o bien a fallos técnicos.

– La perturbación del funcionamiento de instalaciones nucleares o radioactivas y la alteración del desarrollo de las actividades con estos materiales (siempre que se cree una situación de grave peligro para la vida o salud de las personas) [1135]. Estamos, por tanto, también ante un delito de peligro abstracto, con referencia a bienes colectivos, en tanto es imposible acotar la peligrosidad solo en uno o varios sujetos determinados. Sí se produce un cambio significativo respecto del tipo de peligro exigido, (se exige que sea grave) y se suprime la referencia al peligro respecto de los polémicos «bienes». Los sujetos activos son los mismos, así como la tipificación expresa de la imprudencia grave.

– Se castiga, asimismo, la exposición a radiaciones ionizantes [1136] también cuando se produzca fuera de los procedimientos industriales, equiparando en este caso la pena cuando con ello se pone en peligro la vida, integridad, salud o bienes de una o varias personas, con los supuestos en que se ponen en peligro bienes de carácter medioambiental (calidad del aire, suelo, aguas, animales o plantas) [1137]. Una vez más, se trata de un delito común donde se prevé el castigo de la modalidad imprudente.

---

[1134] En este sentido, crítico Muñoz Conde con la falta de referencia a la tipificación de la responsabilidad expresa de los funcionarios. *Ibid.,* p. 638.

[1135] Art. 342, castigado con pena de prisión de cuatro a diez años, e inhabilitación especial para empleo o cargo público, profesión u oficio por tiempo de seis a diez años. La pena es menor respecto del tipo anterior, consistente en la liberación, por ser un tipo residual.

[1136] Cabe aplicar aquí el Reglamento sobre protección sanitaria contra radiaciones ionizantes (RD 783/2001, de 6 de julio). Sin embargo, tal y como señala Muñoz Conde, F., *Derecho penal. Parte especial*., *op. cit*. N.: 1124, el precepto penal no se refiere en modo alguno a las tablas de valores contenidas en dicho Reglamento, *«pero implícitamente el principio de unidad del ordenamiento jurídico fuerza a interpretar que las radiaciones que no superan los límites marcados administrativamente entran en el ámbito del riesgo penalmente permitido»* (p. 641). Sobre las consecuencias que, en materia de riesgo permitido y accesoriedad tiene la regulación administrativa, profundizaremos *infra*, en el capítulo séptimo.

[1137] Art. 343. Las concretas conductas enumeradas se deben entender a título ejemplificativo (vertido, emisión, introducción en el aire, suelo, agua, o exposición por cualquier otro medio). La pena es de prisión de seis a doce años e inhabilitación especial para empleo o cargo público, profesión u oficio por tiempo de seis a diez años. La doctrina (Cardenal Montraveta, S., *op. cit,* N. 1127, p. 654), es crítica con esta equiparación de penas ante la protección de bienes tan dispares como la vida o integridad física de las personas con los bienes de carácter medioambiental.

– Finalmente, el cierre de los delitos nucleares se realiza con el castigo por la posesión, tráfico, gestión, transporte o eliminación de materiales nucleares o radioactivos, siempre que se contravengan las leyes u otras disposiciones de carácter general [1138]. Se requiere que estas conductas causen o puedan causar la muerte o lesiones graves de personas, o bien daños sustanciales a la calidad del aire, suelo, aguas, animales o plantas; así como la producción de estas sustancias o materiales sin autorización [1139]. Según la mayor parte de la doctrina estamos aquí ante un delito de peligro abstracto [1140], de carácter remoto, pues no se exige que llegue a hacerse un uso inadecuado del material radiactivo, sino que basta con que esté siendo manipulado por quien no está autorizado. Señala Maraver que, para que no se acabe convirtiendo en un delito meramente formal y se equipare plenamente a las infracciones de carácter administrativo, se debe comprobar que la conducta realizada supone realmente el riesgo que se pretende evitar en el caso concreto [1141]. Se prevé, asimismo, el castigo de la modalidad imprudente.

El apartado segundo de este tipo aclara la problemática concursal relativa a la posible producción del resultado lesivo, estableciendo que se aplicará la pena más grave en su mitad superior, solución que ha de ser la común con el resto de delitos de energía nuclear, en el entendimiento de que, al configurarse como delitos de peligro contra la seguridad colectiva mantienen su autonomía en caso de que se llegara a producir el resultado lesivo, dando lugar a un concurso ideal, incluso en los casos en que no esté expresamente tipificado [1142]. Solo se podrá imputar el delito de resultado en tanto en cuanto se pueda establecer la relación

---

[1138] Habrá que estar al RD 1464/1999, de 17 de septiembre, sobre actividades de la primera parte del ciclo de combustible nuclear, al RD 97/2014, de 14 de febrero, que regula las operaciones de transporte de mercancías peligrosas por carretera en territorio español, o al RD 102/2014, de 21 de febrero, sobre gestión responsable y segura del combustible nuclear gastado y los residuos radiactivos.

[1139] Art. 345. Pena de prisión de uno a cinco años, multa de seis a dieciocho meses, e inhabilitación especial para profesión u oficio por tiempo de uno a tres años.

[1140] Así, AAP Cádiz de 31 de mayo de 2004, n.º recurso: 32/2004, FJ3.

[1141] Maraver Gómez, M., «Delitos de riesgo catastrófico», *Memento Práctico Penal*, Francis Lefebvre, 2025, nm 14683.

[1142] De otro parecer es la opinión de Maraver Gómez, M., ibid., que parece sugerir que siempre que se dé el resultado lesivo en los casos en que no esté expresamente prevista la solución del concurso ideal, será procedente aplicar la lógica del concurso de leyes, de tal forma que el delito lesivo absorba el de peligro. Así, ibid., nm 14672: *«Este precepto no ha de suponer, sin embargo, un cambio con respecto a la interpretación tradicional, pues cabe entender que el concurso de delitos que expresamente ha decido mencionar aquí el legislador se produce solo si el riesgo afecta a más personas de las que finalmente resultan lesionadas»*. Supone esta una interpretación coherente con la lógica concursal general pero habrá de estarse, en todo caso, al carácter abierto o cerrado del peligro no solo *ex ante*, sino también tras la agresión, en los términos expresados por Doval (como en cita 1115), para atender pertinentemente a las particularidades de estos tipos.

de causalidad y de imputación objetiva, además de la prueba de la contravención de las normas aplicables al ejercicio de estas actividades [1143].

Todas las disposiciones referentes a los delitos relativos a la energía nuclear y radiaciones ionizantes deben complementarse con la remisión a la normativa administrativa que regula estos materiales, a saber: la Ley 25/1964, de 29 de abril, sobre energía nuclear.

## 5.2 **El delito de estragos** [1144]

Según pacífica jurisprudencia, el delito de estragos presenta tres elementos diferenciadores: la potencia destructiva del medio empleado, el objeto contra el que se dirigen [1145], y la peligrosidad para la vida o integridad de las personas [1146]. Lo característico, por tanto, de estos delitos, no son tanto los daños efectivamente causados en los bienes o servicios, sino la gravedad del peligro para la vida o integridad de una pluralidad de personas, derivada precisamente de la presencia de elementos y circunstancias concretas en las que se realiza la conducta [1147]. Es por ello que tiene sentido que, atendiendo a la entidad del

---

[1143] En este sentido, Muñoz Conde, F., *Derecho penal. Parte especial*, *op. cit*. N. 1124, p. 644, trayendo a colación la sentencia de la Audiencia Provincial de Zaragoza de 10 de noviembre de 1993, en la que se condena a los encargados del acelerador de partículas del Hospital General de Zaragoza, cuyo uso incorrecto determinó la muerte de 21 de las 27 personas tratadas. Aunque la sentencia citada por Muñoz Conde no se encuentra en los repositorios, las crónicas de la época dan cuenta de que, en realidad, tras la confirmación parcial del fallo de primera instancia, el único condenado es el técnico de la compañía General Electric, como único responsable del fallo del acelerador, por un delito de imprudencia temeraria con resultado de muerte. En sentido similar, Carbonell Mateu, J. C., *op. cit*. N.1128, p. 636 aunque con algunas dudas, hablando de que la más lógica será la alternativa de la absorción, pero sin descartar la necesidad de contemplar el desvalor del riesgo para los bienes de una generalidad de personas mediante el concurso ideal. A continuación, señala, no obstante, que la magnitud de las penas previstas permite entender que el riesgo para las personas es muy significativo y que, por tanto, el peligro general no puede considerarse absorbido por la producción de un daño personal.

[1144] Art. 346 CP.

[1145] Los objetos son explícitamente enumerados en el artículo: aeropuertos, puertos, estaciones, edificios, locales públicos, depósitos que contengan materiales inflamables o explosivos, vías de comunicación, medios de transporte colectivos, o la inmersión o varamiento de nave, inundación, explosión de una mina o instalación industrial, levantamiento de los carriles de una vía férrea, cambio malicioso de las señales empleadas en el servicio de ésta para la seguridad de los medios de transporte, voladura de puente, destrozo de calzada pública, daño a oleoductos, perturbación grave de cualquier clase o medio de comunicación, perturbación o interrupción del suministro de agua, electricidad, hidrocarburos u otro recurso natural fundamental.

[1146] Así, SSTS 623/22, de 22 de junio, FJ 2: «*En el delito de estragos, desde la configuración proveniente de la LO 15/2003, lo significativo no es tanto la magnitud o la especial trascendencia de los daños causados, sino el peligro para la vida o la integridad de las personas, convertido en el eje central del tipo*», STS 626/12, de 17 de julio; 136/2006, de 15 de febrero. Igual en Cardenal Montraveta, S., *op. cit* N. 1127.

[1147] Algunos de los casos citados por Cardenal Montraveta, ibid., p. 655: SSTS 623/22, de 22 de junio; STS 626/12, de 17 de 7; o STS 349/02, de 22 de febrero (ataque con lanzagranadas contra la Refi-

peligro para las personas, se equiparen las explosiones con otros medios con igual potencialidad destructiva [1148].

El tipo distingue entre tres situaciones de peligro, a las que asigna diferentes penas. Si concurre necesariamente un peligro para la vida o integridad de las personas se impondrá una pena de gravedad muy significativa (prisión de 10 a 20 años); si no concurriese tal peligro, la pena será de 4 a 8 años de prisión, mientras que, resolviendo expresamente la cuestión concursal, si además del peligro concurre una lesión para la vida, integridad o salud de las personas, se castigará separadamente este último delito [1149]. Se castiga tanto la modalidad dolosa [1150] como la imprudente (siempre que concurra imprudencia grave) [1151].

Aunque, como es común en esta sección, la jurisprudencia ha ido oscilando en su calificación como delito de peligro abstracto o concreto, en los casos en que se ha considerado como de peligro abstracto, el argumento utilizado ha sido el de la indeterminación de las personas afectadas [1152].

---

nería Petronor de Vizcaya) o STS 30/01, de 17 de enero (explosión de gas que destruye parcialmente el edificio).

[1148] Así, Cardenal Montraveta, S., *idem,* insiste en que podría considerarse perfectamente típico el supuesto del levantamiento manual de los carriles de una vía férrea, o la manipulación de señales para el tráfico de aquella, insistiendo en la idea de que lo importante no es el medio sino su potencialidad lesiva, primando en todo caso esa idea de riesgo colectivo que vertebra el Título.

[1149] Merece la pena tener en cuenta que, en los arts. 573 y 573 bis.1. 3.ª CP se castigan como delitos de terrorismo los estragos cometidos en las circunstancias allí previstas, dándose lugar en este caso a un concurso de normas.

[1150] El dolo debe darse respecto de los medios y objetos del tipo, es decir, respecto de la destrucción o afectación del bien o servicio, así como respecto de la peligrosidad de la conducta, sin que deba referirse a la producción del resultado, pues, como acertadamente señala Cardenal Montraveta, S., ibid., p. 656, apoyándose en jurisprudencia recaída sobre estos tipos, en caso de producirse un dolo (siquiera eventual) sobre los resultados, cabrá aplicar los delitos de resultado intentados o consumados. Ejemplos en: STS 398/18, de 12 de septiembre, donde se condena por 11 delitos intentados de homicidio, desplazándose el castigo por un delito intentado de estragos, o en la STS 136/06, de 15 de febrero, que castiga por tentativa de homicidio terrorista del artículo 572.1.1.º

[1151] Ejemplos: STS 623/22, 22-6, encender una llama en un local donde se estaba manipulando gas; SAP. Baleares 1.ª 122/00, de 29 de septiembre, por manipulación imprudente de gasolina en contacto con fuego.

[1152] Maraver Gómez, M., «Delitos de riesgo catastrófico», *op. cit*. N. 1141, nm 14707 (STS 1237/1998, de 24 de octubre; auto 812/1998, de 11 de marzo), justificándolo normalmente en la indeterminación de las personas amenazadas (SAN 11/2007, de 28 de marzo). Sin embargo, en reciente sentencia del TS parece justificarse que es un delito de peligro concreto, si bien con una interpretación discutible pues afirma la Sala que «*no era necesario que el peligro amenazara a personas concretas sino que bastaba que el riesgo se ciñera sobre personas indeterminadas*», STS 539/2019, de 28 de noviembre, FJ 4.

### 5.3 **Otros delitos de riesgo**

Finalmente, en la sección 3.ª se prevé el castigo de otros delitos de riesgo provocados por explosivos y otros agentes (arts. 348-350).

Muñoz Conde explica el origen de estos delitos, cuya aparición se ha visto, en ocasiones, condicionada por la sucesión de accidentes con consecuencias lesivas catastróficas, por lo que merece la pena atender al contexto de tipificación para lograr una adecuada interpretación y entendimiento de estos peculiares tipos penales, unidos a contextos muy concretos de producción. El proceso de industrialización, como bien es sabido, ha supuesto desde sus inicios (con la manipulación de maquinaria, sustancias y productos peligrosos), riesgos inevitables para la integridad de las personas, pero el acaecimiento de algunas catástrofes ha puesto de manifiesto el enorme potencial destructor de algunas sustancias cuyo uso y manipulación, aunque sea con fines totalmente lícitos y socialmente valiosos, debe ser escrupulosamente controlado [1153].

Narra Muñoz Conde que en España el problema se ha hecho patente en los últimos años en relación con el transporte de sustancias inflamables, que ha dado lugar a incidentes como la explosión de un camión cisterna en Los Alfaques, originando decenas de víctimas, en los años 70, o la explosión de un coche cargado de fuegos artificiales en Alicante en los años 80, causando varios muertos y heridos. Más allá de los accidentes pirotécnicos en actividades lúdicas, vale la pena resaltar también el origen del actual artículo 348 (incisos relativos a los encargados de la vigilancia de material explosivo), motivado porque, tras el atentado terrorista del 11 de marzo de 2004 en Madrid, se puso de manifiesto la práctica impunidad con que se traficaba con explosivos procedentes de las minas y la falta de vigilancia sobre las instalaciones mineras [1154]. Aunque las conductas básicas ya estaban expresamente tipificadas (manipulación, fabricación, etc.), con la tipificación expresa de la contravención de obligaciones de vigilancia se cierra así la posibilidad de vacíos legales, reforzando la protección con delitos de contenido altamente formal, para cuyo entendimiento resulta útil entender el contexto del que provienen. Lo mismo cabe decir respecto del ámbito de aplicación del artículo 350, relativo a la contravención de medidas de seguridad en determinadas construcciones, fruto de la preocupación por el elevado número de muertes y lesiones en España por accidentes laborales en el ámbito de la construcción, uno de los sectores considerados más peligrosos del ámbito laboral, resultado de la inobservancia de

[1153] Muñoz Conde, F., *Derecho penal. Parte especial*., *op. cit*. N. 1124, p. 646.
[1154] *Ibid.,* p. 646 y ss.

reglas elementales de seguridad. Justifica Muñoz Conde la existencia de estos delitos pues, a pesar de la prolija red de regulación administrativa, las disposiciones son incumplidas continuadamente, con una habitual desidia por parte de los implicados, los empresarios, y la propia Administración [1155]. Una vez explicado el contexto, parece más fácil comprender la concreta regulación de la sección.

El tipo básico del artículo 348 castiga conductas variadas: la fabricación, manipulación, transporte, tenencia o comercialización de explosivos u otras sustancias, capaces de o bien producir estragos o bien comportar grave peligro para la vida, integridad o salud de las personas o medioambiente.

Lo que hace el legislador a través del artículo 348 es fijar una ley penal en blanco, estableciendo un conjunto de normas de seguridad para controlar los riesgos que puedan surgir de todas estas actividades generalmente peligrosas [1156]. En este caso, lo que se exige es que las conductas, además de contravenir las normas de seguridad establecidas, pongan en concreto peligro los bienes referidos. Se castiga en este caso exclusivamente la modalidad dolosa, siendo que el dolo debe referirse tanto al carácter peligro de las materias como al incumplimiento de la normativa de seguridad. Tal y como explica Maraver, en este delito concreto el elemento subjetivo juega un papel primordial para la limitación del tipo pues para evitar que el tipo se convierta en uno de carácter formal, es necesario, no tanto un concreto propósito delictivo del agente, pero sí al menos una conciencia de poseer los explosivos en su condición de instrumentos idóneos para ocasionar los estragos [1157].

La segunda conducta incriminada es el facilitamiento de la pérdida o sustracción de explosivos, y la obstaculización del control administrativo de estos.

---

[1155] *Ibid.,* p. 647.

[1156] Cardenal Montraveta, S., *op. cit* N. 1127, p. 657. Las actividades enumeradas son: la fabricación, manipulación, transporte, tenencia o comercialización de explosivos, sustancias inflamables o corrosivas, tóxicas y asfixiantes, o cualesquiera otras materias; incluyendo expresamente, en el inciso final, las sustancias destructoras del ozono. La referencia al ozono se introduce mediante la LO 5/2010, de 22 de junio. Sin embargo, destaca Muñoz Conde, F., *Derecho penal. Parte especial*., *op. cit*. N. 1124, p. 648, que, a diferencia de lo que sucede en el inciso primero, en el inciso segundo del artículo 348.1 (donde se contiene la referencia al ozono) no se exige una puesta en peligro de bienes jurídicos individuales o del medio ambiente, dándose por supuesta la peligrosidad intrínseca de estas sustancias para el ozono. Pero precisamente por no tener una referencia a la peligrosidad concreta para bienes jurídicos individuales, afirma el autor, y así también lo creemos, este delito hubiera tenido mejor cabida entre los delitos relativos al medio ambiente, por ser éste el bien jurídico directamente afectado por dicha conducta.

[1157] Maraver Gómez, M., «Delitos de riesgo catastrófico», *op. cit*. N. 1141, nm 14730. Cita el autor la sentencia del TS donde se revocó la condena de una persona que acumulaba y manipulaba en su domicilio proyectiles con carga explosiva encontrados en una zona donde se habían librado batallas de la guerra civil española, por considerar que los explosivos no se almacenaban como tales, es decir, como medios con aptitud para poner en peligro el orden público, sino como reliquias bélicas parte de una colección, aplicando el tipo menos grave del artículo 348, y no el del 568 CP. STS 622/2017, de 19 de septiembre.

En este caso, el artículo se aplica a los responsables de la vigilancia, control y utilización de explosivos, siempre que contravengan la normativa [1158].

El inciso cuarto, de cierre, prevé, en forma de delito especial, el castigo para los responsables de fábricas, talleres, medios de transporte, depósitos y establecimientos relacionados con estas sustancias, cuando: obstaculicen la actividad inspectora de la Administración en materia de seguridad, falseen o oculten información a la Administración sobre cumplimiento de medidas de seguridad o bien desobedezcan expresamente sus órdenes encaminadas a subsanar anomalías graves en materia de seguridad. Se trata de un tipo especialmente formal cuya inclusión se explica, como se ha apuntado *supra,* el tráfico clandestino de dinamita proveniente de las minas aprovechando la negligencia de los responsables de su custodia. Como estas conductas, que sí se sancionaban en la normativa administrativa, no encajaban en ningún otro tipo, se procedió a su tipificación expresa como delitos de peligro, sin que se tratase de sancionar por la participación en los actos delictivos concretos que cometían otras personas con el empleo de esos explosivos, pues no se trataba de casos en que esa participación reuniese los caracteres necesarios [1159].

El tipo del 349 castiga, por su parte, a quienes contraviniesen normas o medidas de seguridad al manipular, transportar o en la tenencia de organismos [1160], siempre que concurra un concreto peligro para la vida, integridad física o salud de personas o medioambiente, y el 350, con la misma pena, un ámbito de resultados catastróficos concreto: el de la ejecución de obras [1161], castigando a quienes infrinjan las normas de seguridad correspondientes en

---

[1158] La pena prevista en el inciso primero pena de prisión de seis meses a tres años, multa de doce a veinticuatro meses e inhabilitación especial para empleo o cargo público, profesión u oficio por tiempo de seis a doce años; mientras que la pena del inciso segundo (responsables de vigilancia y control), será de prisión de seis meses a tres años, multa de doce a veinticuatro meses e inhabilitación especial para empleo o cargo público, profesión u oficio de seis a doce años. Se prevé la posibilidad de castigar a las personas jurídicas a tenor de este precepto (apartado 3).

[1159] Comenta Muñoz Conde, F., *Derecho penal. Parte especial*., *op. cit*. N. 1124, p. 650, que es la infracción de deberes administrativos lo que constituye la esencia de los delitos previstos en el apartado 4, aunque en nuestra opinión, tal y como el mismo autor presenta la necesidad de introducción de estos delitos, más que apelar a una teoría esencialmente formalista como la de la infracción de deber, cabe concebir, una vez más, estos delitos como formas de protección adelantada de bienes individuales ante ataques contra bienes variados y personas no individualizables.

[1160] Cabe entender, aunque la remisión no es expresa, que la ley aplicable es la Ley 9/2003, por la que se establece el régimen jurídico de la utilización confinada, liberación voluntaria y comercialización de organismos modificados genéticamente. La STS 668/2019, de 14 de enero, condenó por este delito en el caso de un técnico superior de laboratorio que, de forma deliberada, alteró muestras de pacientes que contenían cepas de tuberculosis enviadas previamente al hospital por la Organización Mundial de la Salud. Como resultado de esta manipulación, se generaron falsos positivos en los diagnósticos de tuberculosis y se puso en riesgo de contagio al personal sanitario.

[1161] Explica Cardenal Montraveta, S., *op. cit* N. 1127, que este precepto se introduce en el Código tras los graves resultados producidos por la rotura de la presa de Tous (STS 492/1997, de 15 de abril).

toda una serie de actividades concretas (apertura de pozos, excavaciones, construcción o demolición de edificios, presas, canalizaciones, o análogas), siempre que la inobservancia de las normas pueda ocasionar «resultados catastróficos» [1162], y pongan en concreto peligro los bienes ya señalados [1163]. Por tanto, la potencialidad lesiva exigida es doble: no basta con el resultado de peligro concreto, sino que, además, la conducta debe tener una potencialidad lesiva para causar resultados catastróficos.

Merece la pena señalar, en este punto, que el resultado catastrófico que puede derivarse de la inobservancia de las normas de seguridad ha sido definido por la jurisprudencia como «*un daño de extraordinaria gravedad e importancia, no por su cuantía patrimonial, sino por su afectación a un gran número, siempre indeterminado de personas, excluyéndose por consiguiente aquellos supuestos en que la infracción de normas de seguridad pueden generar un riesgo sobre personas consideradas en forma individual*» [1164].

Cabe señalar que no es necesario que las conductas enumeradas en esta sección sean realizadas en el marco de actividades empresariales, pues el TS ha tenido ocasión de confirmar la condena por este delito a particulares [1165].

### 5.4 El delito de incendio con peligro para la vida o integridad de las personas

Finalmente, es procedente mencionar, por su importancia histórica como delito de riesgo común, una concreta modalidad del delito de incendio de

---

[1162] Es este un elemento, según afirma Carbonell Mateu, J. C., *op. cit*. N.1128, p. 640, que nos sitúa en un peligro potencial o abstracto.

[1163] Como advierte Muñoz Conde, F., *Derecho penal. Parte especial*., *op. cit*. N. 1124, p. 652, se trata de un precepto derivado del delito contra la seguridad e higiene en el trabajo previsto en el artículo 316, pero con la adición del riesgo para la seguridad colectiva. Por tanto, cuando se produzca, además del riesgo para los trabajadores, un riesgo de resultado catastrófico, se dará el correspondiente concurso ideal entre el delito previsto en el artículo 316 y el previsto en el artículo 350. La contradicción que cabe señalar, empero, es que mientras el delito contra los derechos de los trabajadores prevé la modalidad imprudente, no ocurre así con el delito de riesgos catastróficos.

[1164] Los casos que nombra Cardenal Montraveta, S., *op. cit* N. 1127, p. 659, son: SAP Madrid 5.ª 43/06, de 16 de mayo; SAP Cuenca 55/01, de 26 de mayo, por infracción de las normas de seguridad en derrumbe de edificio, sin riesgo de resultados catastróficos; SAP Asturias 3.ª 504/98, de 10 de noviembre, por obras en edificio realizadas de forma temeraria, pero sin riesgo de resultados catastróficos; SAP Zaragoza 1.ª 117/98, de 11 de marzo, caída de brazo de grúa torre que giraba sobre un colegio, debido a un proceso avanzado de fisuración por fatiga muy difícil de detectar, siendo favorables todas las inspecciones previas.

[1165] Así, STS 828/09, de 13 de julio, confirmando la condena impuesta a quien, tras verter gasolina sobre su propia cabeza y sostener un mechero en la mano, había colocado en el salón cuatro bombonas de butano de gran tamaño, cortado las gomas de dos de ellas, añadido dos bombonas pequeñas de camping gas y, además, había atado a dos de sus hijos a una de las bombonas.

nuestro Código Penal. Como es sabido, los delitos de incendio en el Código Penal español se clasifican en dos grandes grupos: aquellos que suponen un peligro para la vida o integridad física de las personas, y aquellos que no representan este peligro personal, en los que se protegen bienes materiales naturales determinados. Dejaremos fuera del análisis, por tener una relación mucho menos estrecha con el tema que nos ocupa, este segundo grupo de incendios, centrándonos en la (necesariamente breve) exposición del delito de incendio con peligro para la vida o integridad física de las personas del artículo 351 [1166].

Como señala Carbonell Mateu, también en los delitos de incendio nos encontramos con una composición mixta entre la producción de un daño y la creación de un riesgo que normalmente es extremadamente grave [1167].

Llama la atención que, en este caso, el incendio se castiga con pena idéntica a la prevista en el delito de estragos (10 a 20 años de prisión), lo cual nos indica que estamos ante escenarios semejantes. Cabe matizar que el peligro para las personas puede provenir de cualesquiera consecuencias del fuego, no solo del peligro propiamente dicho de morir calcinado o sufrir quemaduras, sino también por consecuencia de todo otro tipo de efectos: asfixia por la inhalación de humo o gases tóxicos, aplastamiento por derrumbe de edificios, etc. [1168] Aunque el delito se consuma con el comienzo del incendio, debe requerirse un peligro inherente al fuego, por lo que si no concurriese este en el caso concreto, no se podría aplicar el tipo, que, en palabras de Carbonell Mateu, requiere la creación de un peligro potencial que califica de concreto, aun cuando no se individualicen los sujetos pasivos, siendo estos todos aquellos que vean incrementado el peligro para su vida o integridad física como consecuencia del incendio [1169].

El tipo requiere, en esencia, que el incendio genere un peligro para un número indeterminado de personas, habiendo sido polémico y objeto de discusión, tradicionalmente, si este peligro ha de ser abstracto o concreto, oscilando tradicionalmente el TS entre ambas interpretaciones. Aunque tanto en la doctrina como en la jurisprudencia se encuentran manifestaciones en ambos sentidos, mientras que en el plano doctrinal es mayoritaria la postura que lo con-

---

[1166] De hecho, son comunes las críticas de la doctrina respecto de la equivocada ubicación sistemática del resto de delitos de incendio, apuntando que solo deberían haberse incluido en este capítulo aquellos incendios generadores de riesgo para la vida o la integridad de las personas, debiendo trasladarse a otros capítulos del Código tales como delitos contra los recursos naturales y el medio ambiente. Así, Maraver Gómez, M., «Delitos de riesgo catastrófico», *op. cit*. N. 1141, nm 14777.

[1167] Carbonell Mateu, J. C., *op. cit*. N.1128, p. 640.

[1168] Cardenal Montraveta, S., *op. cit* N. 1127, p. 663, con cita a STS 611/10, de 21 de junio, delito de incendio cometido por el detenido que prendió fuego al colchón de su celda, produciendo humo que acabó invadiendo el pasillo y otras celdas.

[1169] Carbonell Mateu, J. C., *op. cit*. N.1128, p. 641.

sidera como delito de peligro concreto, en la jurisprudencia es mayor el número de sentencias que se inclinan por considerarlo delito de peligro abstracto [1170]. No obstante, cabe matizar que a menudo se ha definido como delito de peligro hipotético o potencial, explicando que el tipo no contempla la existencia de una situación de peligro (abstracta o concreta), sino la idoneidad del comportamiento que efectivamente se lleva a cabo en el caso concreto para generar ese riesgo, aun cuando éste no se produzca [1171]. También se ha tenido en cuenta, en reciente sentencia del TS, al evaluar la idoneidad de la acción no desde el resultado finalmente obtenido, pues para valorar la consumación del delito de incendio es irrelevante si llegó a materializarse riesgo para la vida o integridad de las personas o si este riesgo, aunque surgiese, decayó tiempo después, pues estas modalidades delictivas de peligro hipotético o potencial no tipifican en sentido propio un resultado concreto de peligro, sino un comportamiento idóneo para producir peligro para el bien jurídico protegido [1172].

Se castiga la comisión dolosa y por imprudencia grave (art. 358), pero en todo caso queda claro, una vez más, que habrá dolo en quien actúa consciente respecto de la provocación del fuego, con consciencia del peligro y aceptación del mismo, sin que sea necesario que el dolo se extienda respecto del resultado de muerte o lesiones [1173] pues, en principio, si el dolo se extiende a estos resultados habrá que castigar por la correspondiente tentativa (o delito consumado) del delito de lesión de que se trate. Esto no obsta a que, en caso de que se produzcan el incendio y el resultado consumado de asesinato, homicidio o lesiones respecto de personas determinadas, deba dejarse sin castigo el peligro general para el número indeterminado de personas, y de hecho el TS no ha tenido problema en apreciar en este caso un concurso de delitos, confirmando la idea de que los bienes protegidos en uno y otro delito no son idénticos, al menos en cuanto al número de personas potencialmente en riesgo [1174].

---

[1170] ATS 457/2021, de 3 de junio; STS 1117/2011, de 31 de octubre, STS 510/2014, de 23 de junio. Calificándolo demo delito de peligro hipotético: STS 1136/2009, de 4 de noviembre; STS 695/2016, de 28 de julio; STS 679/2020, de 11 de diciembre; STS 364/2022, de 8 de abril; STS 844/2022, de 26 de octubre.

[1171] Maraver Gómez, M., «Delitos de riesgo catastrófico», *op. cit*. N. 1141, nm 14798.

[1172] STS 844/2022, de 26 de octubre. Destaca Maraver que *«la jurisprudencia resume su posición, indicando que el riesgo típico desencadenado por el fuego provocado no es el necesario y concreto que se exige en otras figuras delictivas como ocurre con el delito de estragos, sino el potencial o abstracto, a medio camino entre el peligro concreto y el abstracto propio del denominado «delito de aptitud», configurándose como una conducta idónea para poner en peligro el bien jurídico tutelado de la seguridad colectiva y, solo incidentalmente, la propiedad. Lo que requiere la norma penal es que la acción incendiaria comporte un riesgo al menos potencial para la vida o la integridad física de las personas, pero no que ponga en peligro real y efectivo esos valores personalísimos». Ibid.,* nm 14799, con abundantes referencias jurisprudenciales.

[1173] Así: STS 611/10, de 21 de junio; STS 969/04, de 29 de julio; STS 753/02, de 26 de abril.

[1174] STS 569/07, de 29 de junio, analizando la jurisprudencia al respecto en FJ1, explicando que tradicionalmente se ha aplicado el concurso ideal cuando hay dolo directo respecto de la conducta de matar, pero solo eventual respecto del incendio. Cabe reconocer, empero, que la jurisprudencia no es uniforme, y que también son comunes las resoluciones donde solo se castiga por el correspondiente delito de

## 6. CONCLUSIONES: LA DEFENSA DE UN DELITO *A LA MARIE CURIE* EN EL ÁMBITO DEL RIESGO CATASTRÓFICO TECNOLÓGICO GENERADO POR LA IA

Hasta donde se nos alcanza, no es una propuesta extravagante la de optar por el modelo del delito de peligro cuando se trata de regular los riesgos asociados a los *hard-AI crimes*. Empero, las propuestas normativas hasta ahora planteadas en los diferentes ordenamientos y propuestas dogmáticas han sido, quizá, demasiado genéricas [1175].

La tarea verdaderamente pendiente es dar un paso más en la concreción de aquellos supuestos que, hoy por hoy, y con la normativa administrativa en la mano, pueden dar lugar a supuestos de riesgo intolerable, siguiendo la lógica del delito de peligro abstracto aquí esbozado.

Contamos con un inconveniente de partida, que no es otro que el temprano estadio de desarrollo de ciertas tecnologías (veáse, a título de ejemplo, el desconocimiento aún existente sobre los modelos generales de IA, que han dado lugar a una regulación incipiente y a un desarrollo todavía muy tentativo de los mismos). Lo mismo ocurría en los albores del radio cuando los Curie experimentaban con poca cautela y muchas promesas de progreso.

---

resultado. Ejemplos de reconocimiento de concurso ideal entre incendio y el delito de resultado en: STS 63/06, de 31 de enero o STS 653/04, de 24 de mayo, o, incluso, de concurso real en: STS 1007/06, de 10 de octubre; SST 1757/03, de 2 de enero.

[1175] Así, como ejemplo quizá más ilustrativo, y más allá de las genéricas propuestas dogmáticas estudiadas ya en el capítulo tercero, la propuesta de modificación del CP liderada por el Penal Code Review Committee de Singapur, del año 2018, disponible en: <https://www.mha.gov.sg/docs/default-source/media-room-doc/penal-code-review-committee-report3d9709ea6f13421b92d3ef8af69a4ad0.pdf> [Accedido por última vez el 28/01/2026], que en su Sección 3.4 propone una tipificación para los delitos de IA, bajo la rúbrica «*New offences relating to computer programs*», proponiendo la introducción de un nuevo tipo penal para conductas relacionadas con el uso, creación y supervisión de programas informáticos (incluidos sistemas de IA): *(1) Whoever makes, alters or uses a computer program so rashly or negligently as to endanger human life, or to be likely to cause hurt or injury to any other person, or knowingly or negligently omits to take such order with any computer program under his care as is sufficient to guard against any probable danger to human life from such computer program, shall be punished with imprisonment for a term which may extend to one year, or with fine which may extend to $5,000, or with both.* Es decir, planteando el castigo de quien cree, modifique o use un programa de forma temeraria si eso genera un peligro para la vida humana, la probabilidad de causar un daño o lesiones, o bien omita actuar para evitar un riesgo si el programa está bajo su cuidado, sin necesidad de que el daño se materialice. El único daño cubierto es el daño físico. Ha sido criticado que, pese a suponer un avance en cuanto a su propuesta de adelantamiento, por poner el énfasis en la creación de un riesgo sin necesidad de daño, en un intento de modernizar su derecho, no deja de ser un tipo aún demasiado abierto, por cuanto, como acertadamente ha sido señalado por Giannini, no fija los estándares de negligencia o cuidado exigible al usuario o supervisor, haciendo que la fórmula de estar «bajo el cuidado efectivo» de una persona torne estos tipos en delitos con un elemento subjetivo muy débil. Así, Giannini, A., *op. cit*. N. 143, pp. 194 y ss. Cabe destacar, no obstante, que la propuesta, por el momento, no ha prosperado quizá, entre otras cosas, por la oposición que encontró entre la academia dentro de su propio territorio: Singapore Academy of Law's Law Commission, «Report on Criminal Liability, Robotics and AI Systems», febrero de 2021.

Lo que hoy en día sabemos, muchos años después de esos primeros descubrimientos de sustancias y técnicas capaces de causar daños en múltiples bienes personales y en un número indeterminado de personas, es que el derecho penal tiene herramientas de sobra para hacer frente a la prevención de daños descontrolados.

La teoría del bien jurídico resiste, en tanto podemos encontrar un paralelismo claro, en nuestra situación, con los llamados bienes jurídicos intermedios o espiritualizados. No resulta forzado decir que, si de lo que se trata hoy en día es de proteger bienes jurídicos clásicos (vida, salud, integridad, no discriminación o privacidad, estos últimos como nuevos exponentes de la realidad tecnológica), que se ven atacados de una forma normalmente no dolosa, sino como consecuencia de descuidos que pueden dar lugar a ataques indiscriminados, donde sea difícil descartar que los sujetos activos bajo el paraguas de estas tecnologías vayan a sufrir consecuencias lesivas de muy diverso tipo, el derecho penal puede estar legitimado para tipificar delitos.

El delito de peligro abstracto presenta, en su vertiente de peligro hipotético, una enorme utilidad, en tanto es capaz de mantener la justicia material que requiere el principio de lesividad (o, paralelamente, el principio de antijuridicidad material), encontrando justificación en la importancia de los bienes finales protegidos, siempre y cuando se exija una potencialidad lesiva o una aptitud en el caso concreto, tratando de evitar aquellos supuestos en que lo único que se castiga es la mera contravención, alejándonos de argumentos más cercanos a injustos formales. Hemos de acoger, no obstante, una interpretación muy concreta de estos tipos, que ya ha sido defendida por una parte de la doctrina (según aquí se ha explicado, Feijoo, Doval y, para los delitos que protegen bienes instrumentales, Rodríguez Montañés), y que pasa por desvincular el injusto del delito de peligro con el injusto del delito de resultado, pues poca capacidad de rendimiento presentará, entonces, esta estructura. Cuando se trata de bienes jurídicos intermedios, el injusto objetivo se colma con la alusión a la efectiva potencialidad lesiva de una conducta atendidas las circunstancias del caso concreto (excluyéndose, por tanto, el injusto, en aquellos casos en que el autor haya desplegado medidas adicionales que mantengan el riesgo dentro de lo permitido); y, en el plano subjetivo se tendrá que conformar con el conocimiento sobre los elementos del tipo, sin necesidad de que la representación sobre el peligro al que se somete el bien final sea examinada. La contravención dolosa del tipo objetivo, por ser la que implica mayor reproche y necesidad de prevención, satisface por sí misma las exigencias de antijuridicidad en ámbitos de peligro común o general. En estos sectores, infringir las normas de ordenación –que pertenecen a un campo técnico y especializado donde la responsabi-

lidad por el proceso sustituye a la del resultado– supone, en sí mismo, poner en riesgo las condiciones de seguridad de la industria. Sin estas condiciones, medio imprescindible para la protección de los bienes jurídicos que dan sentido final al tipo, no habrá manera de garantizar la protección de los bienes finales. La dinámica de los ataques cometidos con IA encaja, desde luego, en este discurso.

La tarea más difícil será evitar que nos desplacemos a un modelo penal basado en un cumplimiento puro de obligaciones de alto contenido formal remotamente conectadas con la lesividad del caso concreto (como hemos visto ocurre en algunos de los delitos de riesgo catastrófico de nuestra actual regulación).

Merece la pena, por ello, creemos, realizar un trabajo adicional. Si el injusto de estos delitos va a estar unido a la contravención de disposiciones que regulan condiciones de seguridad mínima para bienes jurídico finales diversos, será conveniente marcar una barrera entre las obligaciones altamente formales y aquellas que están estrechamente relacionadas con la protección, ulterior, mediata o instrumentalizada, de bienes finales. El cometido de explorar la normativa extrapenal aplicable en busca de estas obligaciones sustantivas será, pues, necesario. Esta es la tarea que acometeremos, inmediatamente, en el capítulo que sigue.

# CAPÍTULO SÉPTIMO. EL NUEVO ESCENARIO NORMATIVO EUROPEO

> *Ser completamente libre y hallarse al mismo tiempo sujeto al dominio de la ley, he aquí la eterna paradoja de la vida humana, sentida por nosotros a cada momento* [1176].

La paradoja que usa Wilde en su famosa epístola, escrita precisamente desde su encierro carcelario, se torna actual cuando se trata de reflexionar sobre uno de los dilemas esenciales de cualquier sociedad compleja: el de cómo equilibrar la libertad individual con la necesidad de regulación colectiva. Aunque las consideraciones del escritor irlandés trascienden la problemática de la que aquí nos ocupamos [1177], parecen tener sentido como reflexión que da pie a un capítulo dedicado al examen de la regulación cuya mera existencia ha resultado polémica desde su origen.

Las tensiones entre el libre desarrollo más absoluto y el orden social como contrapeso se ponen de manifiesto en un contexto como el que se examina aquí, donde es pertinente reconocer, como así lo hizo Wilde, las dificultades que entraña tanto una libertad excesiva sin limitación, como una regulación desmesurada. El sofisticado pensamiento de Wilde nos muestra, por tanto, un camino, donde la regulación se comprenda en su dimensión compleja, pues puede resultar simultáneamente limitante y habilitante, de forma que la liber-

---

[1176] Wilde, O., *De Profundis*, Trad. Maritza Izquierdo, Verbum, Madrid, 2020, p. 36. En el original, en inglés, *«To be entirely free, and at the same time entirely dominated by law, is the eternal paradox of human life»*.

[1177] En efecto, la extensa carta que Wilde dirige a su amante, Lord Alfred Douglas, desde la cárcel de Reading en 1987 encierra toda una reflexión sobre la naturaleza humana, que toma aún más profundidad si se tiene en cuenta que Wilde experimentaba en sus propias carnes esta tensión entre libertad y legalidad como consecuencia de su encierro.

tad y la regulación no se conciban ya como alternativas mutuamente excluyentes, sino como realidades que han de coexistir de manera dialéctica.

La tensión progreso-regulación se ha presentado, a menudo, como una contraposición insalvable que obligaba a tomar una decisión tajante: o bien la de favorecer el progreso tecnológico y permitir que las nuevas técnicas de la prometedora inteligencia artificial se desarrollaran sin constricción alguna, so pena de frenar los múltiples beneficios que prometía traer a la colectividad; o bien la de procurar su inmediata restricción, prohibiendo cuantas técnicas presentasen un riesgo, siquiera remoto, para la humanidad, por muy lejanos o hipotéticos que estos pudieran ser.

Por fortuna, ni la libertad absoluta ni la prohibición más restrictiva han acabado por instaurarse, o al menos no lo han hecho en lo que se refiere a la regulación de inmediata aplicación para España y para todos los países de la Unión Europea. El objeto del presente capítulo es, pues, explorar cuál es el marco regulatorio actual, esencial en cuanto determinará la aplicación del eventual régimen penal, fijando prohibiciones, permisiones y pautas de interpretación, dotando de seguridad jurídica a una industria en imparable expansión. Se explorarán, también, otros instrumentos regulatorios adicionales y periféricos, con el objeto de complementar el panorama normativo actual, a fin de extraer aquellas obligaciones que pueden ser aplicables en un futuro modelo.

## I. EL REGLAMENTO DE IA DE LA UE COMO PIEDRA ANGULAR

### 1. ESTRUCTURA Y OBJETIVOS GENERALES: EL MODELO DE LA GOBERNANZA CON LEGITIMIDAD DEMOCRÁTICA

Se tratará de desentrañar, por su evidente importancia en cuanto única norma de aplicación general vinculante para los sistemas de IA comercializados dentro de la Unión Europea, el Reglamento de Inteligencia Artificial [1178], en adelante, simplemente el Reglamento o, por sus siglas en español, RIA.

[1178] Nos referimos al Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial o, simplemente, por sus siglas, RIA. El proceso legislativo de este Reglamento ha derivado en varios textos donde han intervenido las distintas instituciones europeas, dando lugar a versiones completas diferenciadas que han permitido seguir las añadiduras que cada institución proponía en las sucesivas fases de negociación. Allá donde no se diga otra cosa, nos referiremos al texto

El modelo regulatorio escogido por la Unión Europea se basa en la ya conocida gobernanza como paradigma, donde podemos afirmar que la responsabilidad proactiva se erige como principio rector de la regulación de la inteligencia artificial [1179]. La idea es que toda organización debe ser capaz de establecer todos los procedimientos necesarios para garantizar el cumplimiento de la normativa aplicable, así como de todas las medidas de seguridad imprescindibles para asegurar el respeto de los derechos fundamentales y la responsabilidad proactiva de la organización y sus órganos de gobierno [1180].

Más allá del papel central de la propia empresa, la regulación se enmarca en el proceso legislativo de la Unión, dando lugar a una norma dotada de una legitimidad democrática innegable. Uno de los temores puestos de manifiesto por la doctrina era, precisamente que, aunque la necesidad de fijar estándares de cuidado y normas vinculantes ante el auge de las nuevas tecnologías parecía evidente, no cualquier entidad estaba dotada de la suficiente legitimidad democrática para llevar a cabo esta tarea.

Así lo expresaba Susanne Beck en un precursor trabajo de 2016, al manifestar que las instituciones encargadas de la estandarización (se refería la autora

finalmente aprobado y en vigor. En el momento de cerrar esta edición (mayo de 2026), no existe todavía una moratoria formalmente vigente que suspenda los plazos del RIA. No obstante, la situación ha evolucionado desde la propuesta inicial: tras presentar la Comisión el paquete Digital Omnibus el 19 de noviembre de 2025, el Parlamento Europeo y el Consejo alcanzaron el 7 de mayo de 2026 un acuerdo político provisional sobre el denominado Digital Omnibus on AI (COM(2025) 836), que modifica el Reglamento (UE) 2024/1689. El acuerdo prevé, entre otras medidas, retrasar la aplicación de las obligaciones de alto riesgo del Anexo III al 2 de diciembre de 2027 y las del Anexo I al 2 de agosto de 2028. El texto está pendiente de adopción formal por ambos colegisladores, prevista antes del 2 de agosto de 2026, fecha en la que de otro modo comenzarían a aplicarse las obligaciones de alto riesgo conforme al RIA vigente.

[1179] Novelli, C./Taddeo, M./ Floridi, L., «Accountability in artificial intelligence: what it is and how it works», *AI & Soc*, 39, 2024, pp. 1871-1882.

[1180] Loza Corera, M., «Datos y gobernanza de datos y conexiones con principios protección de datos en el artículo 10 del Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 571. Señala la autora que, aunque el Reglamento contempla esencialmente el principio de gobernanza con relación a los datos, es evidente que todo el modelo gira en torno a este concepto más amplio. En palabras de Simón Castellano: *«La gobernanza es una actividad transversal que consta de dos elementos principales. El primer elemento se refiere a la gobernanza del propio proceso de gestión de riesgos e incluye la supervisión y revisión, documentación, comunicación y consulta sobre el proceso y sus resultados. El segundo elemento de gobernanza asegura la efectividad del proceso de gestión de riesgos al incorporarlo en la cultura y los procesos de gobernanza más amplios de las organizaciones».* Simón Castellano, P., «Los sistemas de gestión de riesgos como obligación específica para los sistemas de inteligencia artificial de alto riesgo en el artículo 9 del Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 551. Pone de manifiesto este último autor como el paradigma de la gobernanza no es exclusivo de la inteligencia artificial, sino que el tópico de los modelos de gestión de riesgos y supervisión compartida se ha venido usando en campos tan cercanos al derecho penal como la responsabilidad penal corporativa. No en vano, en los albores del hoy ya tan extendido modelo de *compliance* empresarial, explica Bacigalupo cómo, en EEUU, el término surgió asociado a las normas de *soft law* que servían como buenas prácticas relacionadas con la buena gobernanza empresarial. Bacigalupo Sagesse, S., «Compliance», *Eunomía*, 21, 2021, p. 262.

alemana a las normas para determinar los estándares de diligencia en materia de robótica, en aquel momento testimoniales), se enfrentaban al reto de dotar de normas que fueran aceptadas por todos los miembros de la sociedad, en el entendimiento de que, si las normas generadas se basaran simplemente en los intereses económicos particulares de grupos específicos, gozarían de menor legitimidad. Ponía de manifiesto Beck que las instituciones de estandarización (privadas) carecían de la necesaria legitimidad democrática y transparencia, siendo esto un inconveniente si se pretendía trasladar las normas a principios relevantes para el derecho penal. Se apuntaba que sería necesaria una interacción con las instituciones de estandarización para garantizar los procesos normativos [1181].

Esta legitimidad democrática, punto de exigencia común también en lo referido a la concreción del riesgo permitido como institución propia de la sociedad del riesgo, puede considerarse atendida habida cuenta del proceso legislativo que da lugar al actual Reglamento.

Ya en 2018 se inician los trabajos encaminados a la generación de una norma con vocación internacional y con la participación de actores variados, con la creación del Grupo de Expertos de Alto Nivel por parte de la Comisión, que aglutina a profesionales de entornos geográficos y profesionales diversos [1182]. Perfil empresarial, académico y social se aúnan, bajo el encargo de la Comisión, con el objeto de asesorar a la UE en su empeño de obtener una IA ética y robusta, tanto desde el punto de vista jurídico como técnico.

Este primer enfoque, basado más bien en la necesidad de fijar un marco conceptual previo y una estrategia común, que culminó con la publicación del Libro Blanco por parte de la Comisión [1183], dio lugar al inicio de un proceso

---

[1181] Beck., S., «Intelligent agents and criminal law…», *op. cit.* N. 259. La crítica de Beck se basa en la existencia de normas de estandarización de las entonces normas ISO aplicables en materia de robótica, así como del *Deutschen Institut für Normung* (DIN), todo ello de cara a generar estándares de diligencia con aplicación penal. De un modo más general, Paredes Castañón, dando por hecho que la decisión sobre quién establece los riesgos permitidos es una cuestión que remite al proceso de toma de decisiones políticas en materia de riesgos, aceptando que el juez penal acaba por convertirse en un gestor más del riesgo, obligado, en parte, a tomar decisiones que, implícitamente, tienen contenido político. Pareces Castañón, J. M., «El límite entre imprudencia y riesgo permitido en Derecho Penal: ¿es posible determinarlo con criterios utilitarios?, *ADPCP*, vol. XLIX, 1996, p. 914-915.

[1182] El grupo (*HLEG*, por sus siglas en inglés), compuesto por 52 profesionales provenientes del sector empresarial, industrial, académico, y con representantes de la sociedad civil, así como sus objetivos y documentos de trabajo, pueden consultarse en la página de la Comisión Europea: <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai> [Accedido por última vez el 29/01/2026]. Destacan esencialmente sus guías éticas para una IA confiable («Ethics guidelines for trustworthy AI»), publicadas en 2019. Accesible en: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Accedido por última vez el 29/01/2026].

[1183] Comisión Europea, «Libro blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza» [COM(2020) 65 final]. Bruselas: Comisión Europea, disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065> [Accedido por última vez el 29/01/2026]. Un comentario pormenorizado del instrumento desde la perspectiva de la afecta-

legislativo activo propiamente dicho cuando, en abril de 2021, se publica la primera Propuesta de Reglamento [1184]. A lo largo de 2022 los diferentes actores implicados en el proceso legislativo (comisiones del Parlamento Europeo, la propia Comisión, eurodiputados, o el propio Consejo) van presentando sus propuestas sobre el primer texto, para, que, finalmente, en 2023, las tres instituciones adopten su propia posición [1185], dando lugar a un acuerdo provisional en diciembre de ese mismo año. El texto final es aprobado el 21 de mayo de 2024, publicándose el Reglamento en el Diario Oficial de la Unión Europea el 12 de julio de 2024 [1186].

Este proceso legislativo permite afirmar, pues, que el Reglamento de IA goza de legitimidad democrática, habida cuenta de la amplia participación de actores institucionales, técnicos y sociales. El proceso legislativo ordinario de la Unión Europea (con la intervención del Parlamento, el Consejo y la Comisión), sumado a consultas públicas y al trabajo del Grupo de Expertos de Alto Nivel en IA, permitió integrar voces diversas: desde empresas tecnológicas hasta organizaciones de la sociedad civil, expertos académicos y sindicatos. Podemos afirmar, con Silva, que también en el contexto de la IA se ha acabado por imponer, irremediablemente, el modelo democrático-burocrático-tecnocrático, que, si bien presenta los riesgos evidentes de acabar por aceptar las presiones de los lobbies, se presenta ya como inevitable [1187].

---

ción al proceso penal en: De Hoyos Sancho, M., «El libro blanco sobre inteligencia artificial de la Comisión Europea: reflexiones desde las garantías esenciales del proceso penal como «sector de riesgo», *Revista Española de Derecho Europeo*, 76, 2020 pp. 9-44.

[1184] Comisión Europea, «Propuesta de REglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión», Bruselas, 21.4.2021, COM(2021) 206 final, 2021/0106 (COD). Tras la publicación de la propuesta, se abre el trámite de consulta pública, que resulta en más de 300 contribuciones donde, una vez más, actores de diversa procedencia ofrecen sus opiniones. Las propuestas, disponibles públicamente, pueden consultarse aquí: <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback_en?p_id=24212003> [Accedido por última vez el 29/01/2026].

[1185] El 14 de junio de 2023 el Parlamento Europeo adopta su posición negociadora sobre la Propuesta, mientras que el Parlamento y el Consejo adoptan un acuerdo provisional el 9 de diciembre de 2023, tras un periodo de intensos trílogos. Las enmiendas propuestas por el Parlamento: *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence* (Artificial Intelligence Act) and amending certain Union legislative acts [COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)]1.

[1186] Aunque la entrada en vigor del Reglamento se produce de forma escalonada, las primeras partes del texto entran en vigor el 1 de agosto de 2024. El artículo 113 del Reglamento establece el calendario detallado de aplicación gradual del texto, estableciéndose hitos clave cada 6,9,12, 24 y 36 meses, respectivamente.

[1187] Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 81. El punto de partida de Silva es que para afirmar la legitimidad democrática se requiere un acto análogo al del consentimiento presunto por parte de los posibles afectados, y que este cobra especial importancia cuando se trata de permisión de riesgos abstractos. Advierte, asimismo, sobre la posibilidad de que el criterio de legitimación se acerque a

La regulación de la IA no surge, empero, en un contexto aislado, sino que debe entenderse como parte de una política de perspectivas amplias que la Unión Europea viene desarrollando desde hace un tiempo [1188]. Resulta clarificador, a estos efectos, leer a Mantelero, para comprender el contexto geopolítico que imprima el proceso legislativo. La estrategia en la que se enmarca la regulación de la IA surge del diseño europeo para una sociedad digital [1189] que atiende a la necesidad percibida por la Unión de no quedarse atrás en la carrera tecnológica, por lo que, como consecuencia de la intensa actividad legislativa digital, el Reglamento ha de interpretarse de forma sistémica, teniendo en cuenta su relación con otros textos, caracterizados todos ellos por su complejidad técnica. No obstante, Mantelero pone de manifiesto también las debilidades del Reglamento, resaltando que el legislador europeo ha optado por un modelo de gestión de riesgos *ex ante,* basado en una lógica de diseño del producto, dejando algo abandonadas las medidas compensatorias *ex post*, sin que sea posible afirmar que exista un sistema de normas de cierre sobre la responsabilidad adecuado al Reglamento [1190].

---

la tecnocracia (decisiones opacas con un impacto desigual). Pese a que Silva reconoce este riesgo, acaba advirtiendo que, aunque las decisiones de permisión de riesgos no responden plenamente a la lógica democrática, al menos directa, sí permite afirmarse una legitimidad democrática indirecta, y que realmente no cabe otra alternativa realista, sin que exista un modelo alternativo que pueda ser validado de forma general. *Ibid.,* p. 83.

[1188] La estrategia digital europea puede consultarse en el documento publicado por la Comisión «Shaping Europe's Digital Future», de 2020, disponible en: <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/shaping-europes-digital-future_en> [Accedido por última vez el 29/01/2026]. Dicha estrategia se apoya esencialmente en tres pilares: la tecnología centrada en las personas, una economía digital competitiva, y una sociedad abierta y sostenible.

[1189] Mantelero, A., «El Reglamento de inteligencia artificial: la respuesta del legislador europeo a los retos de la inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 54. Dentro de este denominado en ocasiones como «paquete digital» pueden enmarcarse el Reglamento de Mercados Digitales y el Reglamento de Servicios Digitales. Este enfoque basado en la gestión de riesgos pero partiendo de la necesidad de fomentar el desarrollo de las tecnologías en un contexto de debilidad industrial en Europa, marca este periodo, mostrando una diferencia evidente respecto del enfoque mucho más protector del Reglamento Europeo de Protección de datos. *Ibid.,* p. 57. Véanse: Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828 (Reglamento de Mercados Digitales) (Texto pertinente a efectos del EEE), DOUE n.º L 265/1, de 12 de octubre de 2022; y Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento sobre comercio electrónico).

[1190] *Ibid.,* pp. 57-58. Se refiere el autor a las normas de responsabilidad por el producto, señalando que la actualización de la Directiva sobre responsabilidad por el producto no ha sido afortunada, en tanto no ha corrido en paralelo con el Reglamento, por lo que desarrollar un modelo de protección *ex ante* (centrado en el análisis de riesgos), sin desarrollar, paralelamente, un marco para las hipótesis residuales de riesgo en que se produce un daño, minora el impacto global de la intervención normativa del Reglamento, que acaba por ser una «obra inacabada». Destaca también esta idea de obra inacabada o necesitada de

Baste con lo expuesto hasta el momento para entender el contexto y origen en que se inserta el Reglamento, sin perjuicio de que se profundice, en los apartados relativos a sus obligaciones, en las insuficiencias relativas a la protección real que, respecto de los usuarios finales y los derechos fundamentales en juego, supone un sistema basado esencialmente en la autoevaluación como el que finalmente fija el Reglamento.

## 2. LA IMPORTANCIA DEL REGLAMENTO COMO MARCO PARA EL RIESGO PERMITIDO DEL SECTOR

Antes de examinar las obligaciones que establece el Reglamento parece necesario realizar una aclaración sobre cuál es el papel que dicha normativa, en principio extramuros del derecho penal, desempeña para el sector del derecho que nos ocupa, pues aún queda por determinar de qué forma una norma emanada de las autoridades europeas es capaz de influir en el *ius puniendi* estatal.

El punto de vista que aquí se defenderá es que el Reglamento es un marco directamente aplicable para determinar el riesgo permitido en el ámbito que regula (esencialmente, los sistemas de IA que, comercializados dentro de la Unión Europea, generen riesgos para determinados derechos y bienes con relevancia jurídico penal [1191]). No entraremos, por llevarnos demasiado lejos, en la discusión tradicional sobre la ubicación dogmática del riesgo permitido, que daremos aquí, en lo esencial, por conocida, para permitirnos afirmar, con la doctrina mayoritaria, que el hecho de que una conducta esté permitida incide directamente en su tipicidad [1192].

---

complementos, esta vez desde el punto de vista penal, Scollo, haciendo alusión a que, si bien el Reglamento supone un marco del que partir, debería plantearse la opción de aprobar una directiva penal capaz de aportar seguridad jurídica sobre el régimen sancionador. *Vid.*: SCOLLO, L./BRAVO BOLADO, A./PÉREZ VICARIO, A. M./ENSEL, L., «Sketching the Shadows of Criminal Law and Justice in the AI New Era», *European Criminal Law Review*, (en prensa).

[1191] El ámbito objetivo y subjetivo de aplicación se desarrollará inmediatamente *infra*. Baste decir, por el momento, que el propio Silva, en su reciente monografía sobre el riesgo permitido en el derecho penal económico, se refiere precisamente a la inteligencia artificial como uno de los ámbitos más necesitados de regulación, citando directamente la (entonces) propuesta de Reglamento como futura fuente de regulación que, entendemos, fijaría eventualmente el riesgo permitido en el sector. SILVA SÁNCHEZ, J. M., *El riesgo permitido, op. cit.* N. 886, pp. 34-35.

[1192] Adoptamos aquí el punto de vista de Silva en *El riesgo permitido* … ibid, afirmando, en p. 53, que la opción más razonable es ubicar el riesgo permitido en la doctrina del comportamiento desaprobado, de forma que la conducta permitida no puede definirse como típica, aunque nos hacemos cargo de la discusión dogmática tradicional (pp. 45 y ss.).

No obstante, la accesoriedad que esta norma administrativa representa para el derecho penal es limitada, acogiéndose aquí plenamente el concepto de «accesoriedad asimétrica» propuesto por Feijoo [1193].

Aunque Feijoo desarrolla el concepto en lo relativo al derecho penal económico, consideramos plenamente aplicables sus consideraciones en un sector donde las normas se dirigen fundamentalmente a operadores económicos, en un sector altamente profesionalizado.

Así, para este autor, aunque la referencia al contexto regulatorio sea necesaria en determinados sectores, ello no implica, empero, que el injusto penal haya de tener un carácter absolutamente dependiente respecto del orden primario, sino más bien que «el sentido de las conductas depende del contexto normativo que otorga el orden primario». Así, se trata de una accesoriedad legislativa asimétrica, en el sentido de que «*obliga al ordenamiento jurídico-penal a importar las valoraciones básicas del orden primario extrapenal, con independencia de que no dejen de ser valoradas o tenidas en cuenta de acuerdo a las funciones propias del Derecho penal»* [1194]. Ello lleva a afirmar a Feijoo que la infracción del orden primario es requisito necesario e imprescindible para la tipicidad de los hechos [1195].

---

[1193] El término, acuñado por el autor, como él mismo reconoce, basándose en la doctrina alemana, es explicado en Feijoo Sánchez, B., «Los delitos de manipulación del precio de cotización de un instrumento financiero del artículo 284 CP tras la LO 1/2019: retos dogmáticos frente a un desatino legislativo», *La ley digital*, 8176/2019, p. 5 y Feijoo Sánchez, B., «Límites del tipo objetivo en los delitos económicos», en Gómez-Jara Díez, C. (Coord.), *Persuadir y Razonar: Estudios Jurídicos en Homenaje a José Manuel Maza Martín*, Thomson Reuters Aranzadi, 2018, tomo II, p. 161; si bien es desarrollado y explicado en otros trabajos más genéricos. Así, en: Feijoo Sánchez, B., «Imputación objetiva en el Derecho penal económico…», *op. cit*. N. 1122, p. 8 y ss. El reconocimiento de este concepto se encuentra en otros trabajos: Martínez-Buján Pérez, C., *El error en la teoría jurídica del delito*, Tirant Lo Blanch, Valencia, 2017, p. 85. También: Coca Vila, I., «El consentimiento en el derecho penal económico. Un estudio a propósito de los delitos de administración desleal (art. 252 CP) y corrupción en los negocios (art. 286 bis CP)», *Revista General de Derecho penal*, 28, 2017, p. 27; o Bustos Rubio, M., «Insignificancia y Derecho penal», *InDret,* 4/2023, p. 140. Desde luego, no se pretende agotar en absoluto la bibliografía sobre la relación de accesoriedad entre el derecho administrativo y el penal, enumerándose simplemente aquellos autores que han acogido la terminología acuñada por Feijoo. Baste dejar apuntado que Tiedemann fue uno de los autores pioneros en poner de manifiesto que la mayor parte del derecho penal económico es accesoria. Así, Tiedemann, K., *Manual de Derecho penal económico, op. cit.* N. 934, p. 37: *«En relación a todos estos tipos penales y los de Derecho penal accesorio se plantea una cuestión decisiva, consistente en determinar si y en qué medida, deben ser interpretados de modo accesorio al Derecho civil y administrativo o, por el contrario, de un modo autónomo».* Aunque utilizando otra terminología (hablando de accesoriedad negativa y positiva), muy próximo a estas ideas: Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 58.

[1194] Feijoo Sánchez, B., «Los delitos de manipulación…», *ibid.*, p. 5.

[1195] Feijoo Sánchez, B., «Los delitos de manipulación…», *op. cit*. N. 1193, pp. 4-5. Cita el autor en este punto a Nieto Martín, afirmando la importancia de la lectura conjunta del tipo penal y la normativa extrapenal como forma de garantizar la unidad del ordenamiento, sin que quepa afirmar que un comportamiento es delictivo si es un comportamiento permitido por la normativa extrapenal. Nieto Martín, A., en De la Mata Barranco, N./Dopico Gómez-Aller, J./ Lascuraín Sánchez, J. A./Nieto Martín, A., *Derecho Penal económico y de la empresa,* Dykinson, Madrid, 2018, p. 50. En la más reciente edición de

Es posible comprender esta idea en un marco más amplio, entendiendo, de nuevo con Feijoo [1196], que el injusto penal es un injusto cualificado, sin que quepa afirmar, por tanto, que basta la infracción de deberes jurídicos extrapenales, pero siempre teniendo en cuenta que «*el desvalor jurídico-penal de la conducta no puede ser ajeno al contexto normativo o al tipo de relaciones jurídicas establecidas en el subsistema correspondiente*» [1197]. De tal forma que, si el orden primario permite ciertos comportamientos, entendiendo que son conformes con las normas administrativas aplicables, necesariamente habrá que tener en cuenta dicho orden para determinar la atipicidad o la permisión de la conducta [1198].

Resulta también esclarecedora, en este punto, la explicación de Íñigo Corroza, tratando explícitamente la problemática de la accesoriedad administrativa respecto de la responsabilidad penal por el producto. Reconoce la autora, empero, que la problematización se la debemos a Doval, que la formula ya en su trabajo sobre fraude alimentario [1199]. Se encarga Íñigo de dar respuesta al siguiente interrogante: ante una situación permitida, en principio, por la normativa administrativa, ¿es

---

este manual (2.ª ed., año 2024), la cuestión de la accesoriedad es tratada en pp. 64 y ss., distinguiendo entre accesoriedad legislativa (unida al fenómeno de las leyes penales en blanco) y accesoriedad regulatoria (derivada de la autorregulación, el *soft law* y los estándares). Dentro de este segundo apartado se traen a colación los problemas de legitimidad democrática ya citados (p. 73). Como se verá, la actual normativa europea sobre IA puede concebirse como una mezcla de ambas, en cuanto a su contenido directamente vinculante y con rango de ley, pero también con importantes remisiones a estándares técnicos privados.

[1196] Feijoo Sánchez, B., «Imputación objetiva en el Derecho penal económico…», *op. cit*. N. 1122, p. 9.

[1197] *Idem.*

[1198] A continuación, se explicará dónde situar dogmáticamente dicha permisión, pero se adelanta ya, que, en la línea apuntada por Feijoo, *ibid.*, p. 32, con más referencias en nota 76, que una conducta sea conforme con la normativa extrapenal aplicable, en nuestro caso, con el Reglamento de IA, determinará su atipicidad por entenderse que nos movemos en el ámbito del riesgo permitido, situándose este último en el marco de la imputación objetiva. Cabe matizar, no obstante, que Feijoo se aleja explícitamente de la idea de que la infracción de la normativa extrapenal tenga consecuencias por constituir la esencia del delito la infracción del deber. *Ibid.,* p. 32. Nos hacemos cargo, empero, de que la cuestión de la ubicación dogmática del riesgo permitido no es una cuestión pacífica, sino que ha sido objeto de continua discusión entre la doctrina. Resumiendo las posturas enfrentadas: Íñigo Corroza, E., *op. cit.* N 697, p. 126 y especialmente nota 305, explicando las diferentes posturas respecto de la ubicación sistemática del riesgo permitido, especialmente de autores alemanes. También Corcoy Bidasolo, M., *Delitos de peligro y protección de bienes jurídico-penales supraindividuales…, op. cit.* N. 932, pp. 72 y ss., descartando que pueda entenderse como causa de justificación o como causa de exclusión de la culpabilidad, o como causa de exclusión de la imputación del injusto típico, afirmando que el riesgo permitido opera como criterio restrictivo del tipo objetivo (afectando, por tanto, solo al primer nivel de imputación del juicio de imputación objetiva). Finalmente, Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, pp. 50 y ss., partidario de ubicar el problema en sede de tipicidad y solo residualmente en sede de justificación.

[1199] Doval Pais, A., *Delitos de fraude alimentario…, op. cit. N.957,* p. 170. Señala Doval que resulta interesante conocer las normas administrativas relativas a las condiciones que han de reunir los productos destinados al consumo, pues el conocimiento de los ilícitos administrativos será el que permita valorar las relaciones entre el derecho penal y el derecho administrativo alimentario, conocer los criterios de separación entre un régimen y otro, y su coordinación sistemática.

relevante para el derecho penal si persiste un peligro en el caso concreto? Las opciones son dos: o bien no importa si hay peligro sustantivo en el caso concreto, porque lo relevante es si la conducta está prohibida o autorizada administrativamente; o bien entender que es relevante que la conducta autorizada aún sea peligrosa para bienes jurídico-penales [1200]. Íñigo reconoce como inevitable el modelo de la accesoriedad [1201], y acaba por dar solución a todos los supuestos posibles conforme al siguiente esquema: en caso de que el tipo penal exija para su aplicación que no haya autorización administrativa, no basta con una aplicación automática del tipo; es decir, la no autorización será a lo sumo un indicio de criminalidad, pero se ha de exigir, adicionalmente, un peligro para el bien jurídico protegido.

En el caso contrario, la solución es la inversa, es decir, si la conducta presenta riesgo para los bienes jurídicos pero cuenta con autorización administrativa, Íñigo afirma que habrá que distinguir dos situaciones. Si la Administración expide dicha autorización siempre que se cumplan unos requisitos que delimitan la posición jurídica exigible, no puede derivarse responsabilidad penal del fabricante. Si, por el contrario, la posición jurídica exigible no es tan precisa o delimitada, los fabricantes estarán sujetos a normas genéricas de diligencia, por lo que sí cabe plantearse la responsabilidad penal [1202].

La idea de Íñigo, que aquí se comparte, se basa en una concreta concepción sobre los poderes otorgados a la Administración (debiendo entenderse esta última en sentido amplio). Si tras la emisión de una autorización que sirva para amparar la posición jurídica del implicado, aun así la conducta genera un riesgo, que podemos calificar de «residual», será la propia Administración la que deba asumir los problemas generados por una autorización no válida u obsoleta [1203]. Si asumimos un modelo en el que la Administración se arroga una gran cantidad de competencias regulatorias, entendiendo que se está renunciando, con ello, al modelo opuesto que supondría un menor intervencionismo regulatorio público, la propia Administración genera expectativas a la socie-

[1200] Íñigo Corroza, E., *op. cit.* N 697, pp. 207 y ss.

[1201] Curiosamente enmarca esta consecuencia dentro del discurso sociológico de Beck, afirmando, en p. 206, que el haber convertido «el peligro en normalidad», es determinante para la aceptación de los riesgos residuales como permitidos, señalando la autora que esto otorga un poder enorme a la Administración, que puede actuar movida por motivos de todo tipo.

[1202] *Ibid.,* p. 210. Es en este sentido cuando cobra importancia el concepto de asimetría acuñado por Feijoo, plenamente aplicable al contexto del Reglamento de IA, cuyo sistema de certificación se estudiará más adelante. Baste apuntar que, allí donde Íñigo se refiere a la autorización administrativa, puede entenderse el cumplimiento de los detallados estándares de cumplimiento que, para cada sujeto obligado, fija el Reglamento, previendo que en el futuro próximo la conformidad con estos estándares se materialice a través de la expedición de certificados normalizados. El mero incumplimiento del RIA no dará lugar, automáticamente, a la aplicación de una eventual responsabilidad penal (ya sea esta por imprudencia o mediante el sistema de tipificación de peligro aquí propuesto), sino que servirá, a lo sumo, como indicio.

[1203] *Ibid.,* p. 213.

dad (Íñigo se refiere a las expectativas de consumidores y fabricantes [1204], pero aquí serían plenamente aplicables en lo relativo a los usuarios profesionales y no profesionales de IA, así como sus desarrolladores).

Por tanto, la reflexión relativa a la relación entre el derecho administrativo y el penal está indefectiblemente unida a la relativa sobre quién está legitimado para regular y limitar dichas esferas de actividad.

Aceptar un modelo que, a grandes rasgos, presenta una gran legitimidad democrática, y ha demostrado estar informado por expertos pertenecientes a sectores variados de la sociedad, implica, como correlato, aceptar la validez de las precauciones que de él derivan, así como de los riesgos que con él se asumen [1205].

Hemos de puntualizar, no obstante, que la vinculación de las normas técnicas del Reglamento no es absoluta en todo caso, sino que, según ha destacado acertadamente Ibold, se hace necesario distinguir entre las normas que, bajo el paradigma del nuevo modelo regulatorio de la Unión, persiguen un fin de protección directamente dirigido a evitar daños de relevancia penal [1206], de aquellas (se refiere principalmente la autora a las normas dirigidas a los sistemas calificados de bajo riesgo), donde, por su falta de fuerza vinculante, no quedará más remedio que aplicar los criterios genéricos o principios generales para determinar el riesgo permitido. La idea que defiende la autora es que para la aplicación de las normas técnicas del Reglamento con efecto jurídico vinculante en el campo penal se habrán de examinar al menos tres requisitos: que haya una función protectora de bienes jurídicos en la norma técnica de conducta; que la norma técnica pretenda proteger el mismo riesgo que se materializa; y la irrelevancia de las normas técnicas si se despliegan medidas de seguridad comparables (dirigidas a proteger el mismo bien) [1207], tratando de encontrar un vínculo entre el efecto de las normas técnicas y la teoría de la imputación objetiva (principalmente del requisito del fin de protección de la norma).

---

[1204] *Ibid.,* p. 214, hablando de la necesidad de *«asegurar las expectativas normativas que se consideran necesarias para el desarrollo de la vida social».*

[1205] Evidentemente, y esto también se pone de manifiesto por Íñigo, no quiere decir que, como consecuencia de esta legitimidad democrática, exista un consenso social mayoritario al compartir las ponderaciones entre el riesgo y el beneficio que conllevan algunas aplicaciones tecnocientíficas. Íñigo pone el ejemplo de la energía nuclear, y en el campo de la IA algo parecido cabría predicar respecto del modelo finalmente adoptado de regulación de la vigilancia biométrica en tiempo real o la exclusión de las armas autónomas letales. Nada impide, empero, como más adelante se desarrollará, que la misma Administración (sea esta la europea o la nacional), ponga coto a estas aplicaciones en uso de su soberanía. Iñigo Corroza, E., *ibid.* p. 198.

[1206] Ibold V., *op. cit*. N. 104, p. 324, Ibold, empero, solo se centra en estos efectos penales respecto de las normas que están dirigidas a proteger la salud y la vida, dejando de lado otros bienes jurídico-penales respecto de los que, como se ha demostrado a lo largo de la investigación, surgen lagunas de protección.

[1207] *Ibid.,* pp. 313 y ss.

Destaca la autora el carácter de norma jurídica en sentido estricto del Reglamento, con efecto vinculante indiscutible, así como las normas armonizadas dimanantes de los organismos de normalización, que califica de fuente jurídica secundaria, pero con efecto de norma jurídica vinculante en sentido amplio [1208], descartando que los códigos de conducta voluntarios para los sistemas de riesgo bajo tengan un efecto vinculante. Destaca, empero, que el principio de interpretación conforme al derecho de la Unión vincula, sin duda, a los tribunales penales [1209] y que, en principio, aunque no estén obligados, si un fabricante o responsable de un sistema de riesgo bajo cumple con los códigos de conducta voluntarios, deberá considerarse que actúa dentro de los límites del riesgo permitido, sin que el incumplimiento de estos, al ser, precisamente, voluntarios, pueda conllevar un efecto perjudicial para el fabricante.

Parece aceptar Ibold el efecto de accesoriedad asimétrica aquí postulado, pues, respecto de las normas armonizadas [1210], si se cumplen, será la autoridad de vigilancia de mercado quien deba probar que no se han cumplido los requisitos generales de seguridad, mientras que, si se incumplen, no existe presunción inversa de inseguridad, es decir, no se puede deducir que no se cumplan los requisitos para los sistemas de IA de alto riesgo [1211].

## 3. PRINCIPALES OBLIGACIONES IMPUESTAS A LOS AGENTES ECONÓMICOS Y PÚBLICOS

### 3.1 **Introducción general: título competencial y estructura básica**

El Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA supone la propuesta normativa más ambiciosa hasta la fecha frente a los desafíos jurídicos, éticos, sociales y tecnológicos que presenta esta tecnología. Consta de 180 considerandos, 112 artículos y 13 anexos.

Ya en el artículo 1.1 podemos encontrar una referencia al objeto del Reglamento: *«El objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en*

[1208] *Ibid.,* p. 315.
[1209] *Ibid.,* p. 317.
[1210] Pues, como se explicará *infra,* su cumplimiento genera una presunción de conformidad con el RIA.
[1211] Ibold V., *op. cit*. N. 104, p. 320.

*la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA (en lo sucesivo, "sistemas de IA") en la Unión así como prestar apoyo a la innovación.»*

El orden de aparición de los objetivos (dos), no es azaroso. Mercado interior y protección de derechos fundamentales aparecen de la mano ya en la primera disposición legal, pero no en una relación de simetría. El Reglamento fue impulsado por la Dirección General de Mercado Interior, Industria, Emprendimiento y Pymes, y esta opción no es baladí, pues indica que, al fundarse el título competencial en una disposición relacionada con la política de mercado interior de la UE [1212], el enfoque que adopta la normativa tiene como eje vertebrador la aceptación del riesgo y su regulación para dotar de seguridad jurídica a los agentes económicos, consiguiendo que el mercado europeo sea competitivo. Por tanto, el punto de partida necesario para entender la norma es ser conscientes de que el RIA constituye una norma de seguridad industrial, y no, al menos *prima facie*, una norma orientada directamente a la protección de derechos individuales [1213], sin que esta segunda vertiente sea, evidentemente, ajena a la regulación.

Se debe comenzar destacando cuál es la estructura que vertebra el Reglamento, pues condiciona sobremanera la forma de entenderlo. El RIA adopta una estructura escalonada basada en riesgos, de forma que el legislador europeo ha optado por fijar al menos tres niveles de riesgo claros y un cuarto nivel algo más difuso. Así, la norma reconoce un nivel de riesgo inaceptable, para el cual establece prohibiciones de desarrollo y uso; un nivel de riesgo alto, para el cual prevé toda una serie de obligaciones de forzoso cumplimiento; y finalmente, para el resto de sistemas, un riesgo aceptable o limitado, que conlleva que la comercialización y uso estarán permitidas sin necesidad de sujetarse a cautelas específicas. Decimos que existe un cuarto nivel de riesgo, algo más difuso, pues no queda tan clara su posición jurídica, que se corresponde con el riesgo que el Reglamento denomina «sistémico».

---

[1212] Concretamente, el título competencial es el artículo 114 del Tratado de Funcionamiento de la Unión europea, como así se reconoce expresamente en el considerando 3 del RIA: *«[…] es preciso garantizar un nivel elevado y coherente de protección en toda la Unión para lograr una IA fiable, así como evitar las divergencias que obstaculizan la libre circulación, la innovación, el despliegue y la adopción en el mercado interior de los sistemas de IA y los productos y servicios conexos mediante el establecimiento de obligaciones uniformes para los operadores y la garantía de una protección uniforme de los fines imperiosos de interés general y de los derechos de las personas en todo el mercado interior, sobre la base del artículo 114 del Tratado de Funcionamiento de la Unión Europea (TFUE)».*

[1213] Esta realidad es puesta de manifiesto por Mantelero, que pone en contraste esta génesis con la que dio lugar, en su momento, al Reglamento Europeo de Protección de Datos, liderado por la Dirección general de justicia y con un enfoque centrado en la protección de la libertad, la seguridad y la justicia. Mantelero, A., «El Reglamento…», *op. cit*. N. 1189, p. 56.

El Reglamento se articula en trece capítulos[1214] que configuran un marco normativo integral y escalonado. El capítulo I establece las disposiciones generales, incluyendo el objeto, ámbito de aplicación y definiciones fundamentales. El capítulo II tipifica las prácticas prohibidas, identificando aquellos usos de la IA que contravienen los valores fundamentales de la Unión. El capítulo III, eje central del Reglamento, regula en detalle el régimen aplicable a los sistemas de alto riesgo, abarcando requisitos en materia de gestión de riesgos, documentación técnica, trazabilidad, supervisión humana y robustez técnica. El capítulo IV contempla obligaciones específicas de transparencia para determinados sistemas de IA, mientras que el capítulo V introduce requisitos para los modelos de uso general, con especial atención a los modelos fundacionales. El capítulo VI recoge mecanismos de fomento a la innovación, destacando los entornos controlados de prueba *(«regulatory sandboxes»).* El capítulo VII se centra en el modelo de gobernanza, estableciendo las funciones y atribuciones de las autoridades competentes. El capítulo VIII prevé la creación de una base de datos europea de sistemas de alto riesgo, y el capítulo IX regula el régimen de seguimiento y vigilancia post-comercialización. El capítulo X habilita la elaboración de códigos de conducta de adhesión voluntaria para los sistemas de riesgo limitado. Los capítulos XI y XII abordan, respectivamente, la delegación de poderes a la Comisión y el régimen sancionador. Finalmente, el capítulo XIII contiene las disposiciones finales relativas a la entrada en vigor, la aplicabilidad y las disposiciones transitorias.

Partiendo de esta estructura, veremos a continuación sus principales implicaciones para los operadores económicos y públicos.

### 3.2 **Marco de aplicación**

Se analizará a continuación el ámbito de aplicación de la norma desde una triple perspectiva: objetiva, subjetiva, y territorial, respondiendo, respectivamente, a las siguientes preguntas: ¿sobre qué objetos se aplica?, ¿a qué personas se aplica?, ¿en qué territorio se aplica?

[1214] I. Disposiciones generales; II. Prácticas prohibidas; III. Sistemas de IA de alto riesgo; IV. Obligaciones de transparencia de los proveedores y responsables de despliegue de determinados sistemas de IA; V. Modelos de IA de uso general; VI. Medidas de apoyo a la innovación; VII. Gobernanza; VIII. Base de datos de la UE para sistemas de IA de alto riesgo; IX. Vigilancia poscomercialización, intercambio de información y vigilancia del mercado; X. Códigos de conducta y directrices; XI. Delegación de poderes y procedimiento de Comité; XII. Sanciones; y XIII. Disposiciones finales.

a) Ámbito objetivo

El RIA se aplicará a sistemas de IA que funcionen como componentes de productos o que sean productos en sí mismos, que se pretendan comercializar o poner en servicio en el mercado de la UE. Hemos de partir, necesariamente, de la definición básica de la que parte el Reglamento, que no es otra que la del «sistema de IA». El artículo 3, ya en su apartado 1), otorga la siguiente definición: *«un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales».*

Cabe hacer simplemente dos matizaciones respecto de la delimitación conceptual que el Reglamento realiza, pues en todo lo relativo al concepto y las dificultades de delimitación, nos remitimos a lo ya dicho en el capítulo primero. El Reglamento opta por excluir expresamente (considerando 12 [1215]) los sistemas basados en normas definidas únicamente por personas físicas para ejecutar operaciones automáticamente, lo que implica una exclusión de las formas más sencillas o tradicionales de programación [1216].

Del mismo considerando 12 podemos extraer la siguiente consideración de importancia, relativa a la centralidad de la autonomía del sistema, que, si bien puede concretarse en distintos niveles, implica siempre y en todo caso «un cierto grado de independencia», así como capacidades para funcionar sin intervención humana [1217].

---

[1215] *«[…] no debe incluir los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones […]».*

[1216] Así lo destaca Cotino Hueso, L., *«Alcance y delimitación de los sistemas de alto riesgo en el Reglamento de inteligencia artificial»,* en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 117. Se identificarían estas formas con lo que en el capítulo primero se ha denominado sistemas de IA *old-fashioned* o sistemas expertos o predeterminados. Esta exclusión, aunque puede parecer lógica, habida cuenta de los problemas mucho más complejos que plantean los sistemas más sofisticados, no deja de plantear problemas en algunos contextos de aplicación. Sin ir más lejos, y así es destacado por Miró y Santisteban, en lo relativo a los límites aplicables a las tecnologías de policía predictiva o sistemas de predicción del comportamiento criminal, en general, donde los usos más comunes, al menos por el momento, corresponden a esta lógica del software predeterminado. Miró Llinares, F./ Santisteban Galarza, M., «La regulación de los sistemas policiales predictivos en el Reglamento Inteligencia Artificial» en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 335-336.

[1217] Detalla el Considerando que *«La capacidad de adaptación que un sistema de IA podría mostrar tras su despliegue se refiere a las capacidades de autoaprendizaje que permiten al sistema cambiar mientras está en uso».* Se señala, así, la capacidad de autoaprendizaje como elemento capacitador de la autonomía, que debe entenderse, a nuestro juicio, tal y como se explica en profundidad en el capítulo primero, como una autonomía operativa.

b) Ámbito subjetivo

El Reglamento se aplica a diferentes actores de la cadena de valor de la IA, a saber: los proveedores, los responsables de despliegue (denominados en versiones anteriores del Reglamento «implementadores»), los representantes autorizados, los importadores y los distribuidores [1218] de los sistemas de IA, así como a aquellos fabricantes de productos que comercialicen un sistema de IA junto con un producto bajo su marca comercial. En tanto en cuanto el papel de estos operadores económicos sea asumido por administraciones públicas, a ellas se aplicarán, asimismo, todas las disposiciones del Reglamento [1219].

Desarrollaremos, brevemente, el papel de cada uno de ellos y cuál es el marco de obligaciones aplicable, sin perjuicio de que el contenido material de las obligaciones sea desarrollado *infra.*

El proveedor (art. 3.3 RIA) es la persona física, jurídica o autoridad pública que desarrolla un sistema de IA o lo pone en el mercado bajo su propio nombre o marca. Sus obligaciones generales son, además del cumplimiento de los requisitos esenciales (arts. 8-15), realizar una evaluación de conformidad cuando sea pertinente (art. 16), establecer un sistema de gestión de calidad (art. 17), conservar la documentación pertinente (art. 18), así como establecer todas las medidas correctoras exigibles en caso de que se detecte que el sistema no supera la evaluación de conformidad, recuperando el sistema e informando al resto de operadores de la cadena (art. 20). En todo caso tienen deber de cooperar con las autoridades (art. 21) y de investigar los riesgos tras notificaciones de incidentes por parte del responsable de despliegue.

El responsable de despliegue (art. 3.4), es la persona o entidad que usa un sistema de IA bajo su autoridad, salvo que se trate de uso personal no profesional. Sus obligaciones principales (art. 26) consisten en usar el sistema conforme

[1218] Definidos en el artículo 3, apartados 3)-7). Cabe especificar, en este punto, que cuando el RIA se refiere a «representantes autorizados» establece la obligación de que las entidades extranjeras que operen en la UE nombren un representante que se asegure de que se aplique la ley y actúe como contacto con las autoridades europeas.

[1219] Ha sido señalado, no obstante, con acierto, que el Reglamento no está pensado para la aplicación a las administraciones y que esto se traduce en una eventual aplicación imperfecta (amén de privilegiada), para la misma. Especialmente preocupantes son la desmesurada prórroga en la entrada en vigor de las obligaciones, así como la problemática de aplicarles el sistema sancionatorio, no pensado para entidades públicas. En este sentido: Barrachina Navarro, G./Boix Palop, A., «La aplicabilidad del Reglamento de inteligencia artificial al ámbito de la Administración pública y servicios públicos y especialidades respecto de su cumplimiento: Especial atención a Anexo III y actuación administrativa y particularidades cumplimiento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 351-371.

a las instrucciones del proveedor, supervisando que se cumplan, designando y formando, para ello, a personal competente. En todo caso, deben también vigilar el funcionamiento de sus sistemas, notificando al proveedor y a las autoridades ante riesgos o incidentes graves. En este sentido, e interpretando fielmente el texto de la norma, que alude a la necesidad de que los responsables de despliegue informen a los trabajadores de las condiciones de uso del sistema de IA en el entorno laboral, es evidente que el legislador europeo pretende dejar fuera de la aplicación, al menos directa, del RIA, al trabajador que, por muy profesional que sea en su sector, no estará actuando bajo su propia autoridad.

Aunque estos dos son los principales operadores en la cadena de suministro, existen otras tres figuras a las que el RIA impone obligaciones. Así, el representante autorizado (art. 3.5), es la persona que, dentro de la UE, actúa en nombre de un proveedor extranjero, para asegurar el cumplimiento del Reglamento, cuyas obligaciones principales son de cooperación con las autoridades europeas y de conservación de documentación. El importador (art. 3.6), responsable de introducir en el mercado un sistema de IA desarrollado fuera de la UE, verifica que el sistema cumple con los requisitos de evaluación de conformidad, documentación técnica y, en su caso, certificación, con el deber de informar de los riesgos percibidos, asegurando las condiciones de transporte y almacenamiento adecuadas. Finalmente, el distribuidor (art. 3.7), intermediario que pone el sistema de IA en el mercado, verifica el cumplimiento del resto de requisitos (declaración de conformidad e instrucciones de uso), con la obligación de notificar a los implicados en caso de no conformidad.

Es importante resaltar que las obligaciones de cualquiera de los operadores pueden cambiar de sujeto, asumiendo las obligaciones del proveedor desde ese momento, en tres casos: si pone su propio nombre o marca en el sistema de alto riesgo; si lo modifica sustancialmente, o si cambia su finalidad, de forma que pase a considerarse de alto riesgo [1220].

### c) Ámbito territorial

Para la aplicación del Reglamento, los operadores económicos arriba citados deben realizar su actividad profesional en el mercado único de la UE, pero también se aplicará cuando las personas afectadas estén establecidas en el

[1220] En todo caso, el proveedor inicial, que dejará de ser considerado proveedor a efectos del RIA, deberá colaborar estrechamente con el nuevo proveedor para facilitar toda la información necesaria para el cumplimiento de las obligaciones (art. 25.5).

territorio de la Unión. Se prevé así una aplicación del Reglamento para operadores de terceros estados, pero solo cuando se dé un criterio de conexión que permita afirmar que se ha producido esta afectación [1221].

Cabe finalizar el apartado relativo al ámbito de aplicación con la delimitación negativa del objeto regulado por el Reglamento. Quizá la exclusión más polémica tenga que ver con la decisión expresa de dejar extramuros de la regulación europea los sistemas de IA de seguridad nacional, defensa y militares [1222]. Otras exclusiones significativas son: los sistemas o modelos de IA con finalidad de investigación y desarrollo científico; así como la exclusión de las personas físicas que utilicen sistemas de IA en el ejercicio de una actividad puramente personal de carácter no profesional [1223].

### 3.3 **La IA prohibida: el núcleo garantista del Reglamento**

El capítulo dedicado a la regulación de los sistemas de IA prohibidos supone, a nuestro juicio, el reflejo más obvio de la preocupación de la Unión

[1221] El criterio empleado por el Reglamento es el de «vinculación estrecha con la UE», formulación que ha sido criticada por su vaguedad y consecuente inseguridad jurídica. Así, Ortega Giménez, A., «El ámbito de aplicación territorial del Reglamento de inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 127. También en p. 135, hablando del principio de efecto, que implica que solo si un sistema de IA tiene un impacto significativo en los individuos o entidades situados en la UE, se aplicará la ley.

[1222] Se cierra así la puerta, por el momento, a la posibilidad de tener una regulación armonizada en materia de seguridad nacional. Aunque la exclusión de la regulación (y más concretamente, de su prohibición), ha encontrado principalmente críticos entre la doctrina, por entender que era el lugar para prohibir expresamente ciertos usos militares de la IA, cabe encontrar voces de quienes, pese a mostrarse críticos con la exclusión, consideran que la mejor solución hubiera pasado por la permisión y regulación expresa de ciertas aplicaciones. En este último sentido, Gómez de Ágreda, Á., «La exclusión de los sistemas inteligencia artificial de seguridad nacional, defensa y militares del Reglamento y el Derecho aplicable» en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 137-148. En el ámbito internacional, destaca el informe A/79/88 del Secretario General de la ONU, *«Lethal Autonomous Weapons Systems»*, publicado el 1 de julio de 2024, que examina los principales retos de las armas letales autónomas, con observaciones preliminares del Secretario General y recomendaciones de cara a futuras negociaciones, incluyendo las del Comité de Desarme de la Asamblea y la Convención sobre Ciertas Armas Convencionales (CCW), con miras a alcanzar un instrumento legal vinculante antes de 2026. Disponible en: <https://docs-library.unoda.org/General_Assembly_First_Committee-Seventy-Ninth_session_(2024)/A-79-88-LAWS.pdf> [Accedido por última vez el 29/01/2026]. En el ámbito español, sobre esta problemática, por todos: Gil Gil., A., *op. cit*. N. 242, absolutamente favorable a la prohibición en caso de armas con verdadera autonomía operativa.

[1223] Artículo 2, apartados 3,6,8 y 10, respectivamente. El apartado 12 excluye a los sistemas de IA de código abierto, pero no parece tener mucho sentido esta exclusión, pues a continuación se establece la salvedad de que se pongan en servicio como sistemas de alto riesgo o prohibidos, o bajo el ámbito de aplicación del artículo 50 (principalmente aplicables a sistemas de IA generativa). No se entiende, entonces, qué hueco queda para la exclusión en la aplicación. El considerando 89 especifica que «*se debe animar*» a los desarrolladores de estas herramientas a aplicar ciertas cautelas. El considerando 102 define el código abierto como la licencia que *«permita a los usuarios ejecutar, copiar, distribuir, estudiar, modificar y mejorar el software y los datos».*

Europea por la protección de los derechos individuales, constituyendo una plasmación evidente del principio de precaución [1224].

El Reglamento prohíbe, con un alcance muy diverso, hasta ocho prácticas, a saber: 1) la manipulación subliminal o engañosa que altere sustancialmente el comportamiento; 2) la explotación de vulnerabilidades asociadas a la edad, discapacidad, o situación social para influir de forma perjudicial en el comportamiento; 3) la evaluación o clasificación social de las personas (puntuación ciudadana); 4) la valoración del riesgo delictivo basada exclusivamente en perfiles personales; 5) la creación de bases de datos de reconocimiento facial mediante extracción masiva de imágenes; 6) la inferencia de emociones en contextos laborales o educativos; 7) la categorización biométrica sensible para deducir ciertos rasgos protegidos; y 8) la identificación biométrica en tiempo real en espacios públicos [1225].

Se expondrá, a continuación, el régimen aplicable a cada uno, con el objeto de examinar qué intereses esenciales pueden verse conculcados con cada una de las prácticas, siempre con el objetivo fundamental de realizar una interpretación teleológica del Reglamento que nos aporte información sobre la conexión con intereses penales. Comenzaremos con la prohibición menos absoluta de todas, por constituir, más que una práctica prohibida, uno más de los supuestos de sistemas de alto riesgo, por el altísimo número de excepciones a su régimen de prohibición.

### a) Los sistemas de identificación biométrica en tiempo real en espacios de acceso público con fines de garantía de cumplimiento del Derecho (art. 5.1.h)

Debemos entender por datos biométricos todos aquellos que se obtienen del cuerpo de las personas como expresión de algún tipo de estudio o de medida, ya sea que se obtienen con información estática (características anatómico-físicas del cuerpo humano), como dinámicas (destinadas a captar información se-

[1224] Nos remitimos, en este punto, a todo lo afirmado al respecto de este principio y su papel en el capítulo cuarto.

[1225] Cabe señalar que el alcance de la prohibición es el mismo para todas ellas (se prohíbe su introducción en el mercado, su puesta en servicio, y su utilización), salvo para la última, es decir, para los sistemas de identificación biométrica remota en tiempo real la prohibición alcanza solo a la utilización, constituyendo, como veremos inmediatamente, el régimen más detallado y privilegiado de todas las prácticas prohibidas. Aunque en la actualidad el listado es *numerus clausus,* el propio Reglamento prevé (art. 112) que la Comisión evalúe periódicamente la necesidad de actualizar el listado de prácticas prohibidas, así como el de sistemas de alto riesgo.

De hecho, el acuerdo político provisional alcanzado el 7 de mayo de 2026 sobre el Digital Omnibus on AI contempla añadir al art. 5 una novena práctica prohibida: los sistemas de IA destinados a generar o manipular contenido íntimo no consentido o material de abuso sexual infantil (las denominadas apps «nudificadoras»), prohibición pendiente de adopción formal cuya entrada en vigor se prevé para el 2 de diciembre de 2026.

cuencial o parámetros corporales en sentido amplio, como puede ser un patrón a la hora de andar) [1226]. Explica Escajedo que para que exista un sistema de reconocimiento biométrico se debe construir un sistema capaz de examinar el cuerpo de las personas y reconocer en él algún atributo que pueda vincularse con información almacenada previamente [1227]. El RIA define, en su artículo 3, tanto los datos biométricos como la identificación y la verificación biométrica, estableciéndose como diferencia entre una y otra que en la identificación el reconocimiento de características se usa para identificar a una persona física comparando sus datos con otros previamente almacenados, mientras que en la verificación se procede uno-a-uno, es decir, verificando la identidad mediante la comparación con datos individuales, y no uno-a-muchos. Apartándose expresamente de la definición del RGPD, el RIA incluye también no solo los datos con potencial identificador, sino también aquellos usados para la categorización o el reconocimiento emocional, incluso si no permiten la identificación única de la persona (criterio empleado tradicionalmente por la normativa de protección de datos).

La prohibición del uso de estos sistemas con fines de garantía del derecho presenta numerosos inconvenientes, que parten ya de la evidente superposición (y las contradicciones [1228]) entre la regulación del RIA y la prevista en el Reglamento de Protección de Datos. La segunda de las dificultades para la interpretación de esta práctica es que no solo se excepciona la prohibición en numerosos supuestos (regulados de forma bastante exhaustiva en el propio RIA en lo que tiene que ver, esencialmente, con la persecución de determinados delitos [1229]), sino con que, por lo genérico de esta técnica, se acabará entremezclando con muchas otras aplicaciones prohibidas, que no son más que concreciones, en algunos ámbitos, del uso de tecnologías de reconocimiento. Por ello, Escajedo, al analizar este tipo de sistemas, acaba por concluir que sus regímenes regulatorios son múltiples, pudiendo encontrar ejemplos a lo largo del RIA de identificación biométrica prohibida, usos permitidos pero conside-

[1226] Definición propuesta por Escajedo San-Epifanio, L. «El reconocimiento biométrico en el Reglamento de inteligencia artificial: exenciones, prohibiciones y especialidades de alto riesgo», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 191-192. El ejemplo es nuestro.

[1227] *Ibid.,* p. 196.

[1228] Más allá del diferente alcance en cuanto al objeto de protección, (derivado de la diferente definición), la idea vertebradora del RGPD es la de prohibir el tratamiento de datos biométricos salvo excepciones, mientras el RIA permite su uso en casos muy variados.

[1229] Critica Escajedo que quizá aquí el legislador europeo aporta poco, pues las garantías relativas a las autorizaciones para realizar las excepciones se encuentran previstas por los sistemas constitucionales de los Estados miembros. Escajedo San-Epifanio, L., *op. cit.* N. 1226, p. 222, cuestionando, en p. 220, que un Reglamento orientado a evaluar riesgos de mercado sea el lugar idóneo para referirse a estas excepciones, con marcada base constitucional.

rados de alto riesgo, y completamente lícitos, cada uno de ellos con un estatuto jurídico diferente.

Inmediatamente después de anunciar la prohibición de la identificación biométrica, los siguientes apartados del inciso h) del artículo 5 detallan las excepciones. El uso legítimo se limita a alguno de los siguientes objetivos: la búsqueda de víctimas específicas de secuestro, trata o personas desaparecidas; la prevención de amenazas graves e inminentes para la vida o seguridad física; y la identificación de sospechosos de delitos muy graves (especificados en el anexo II [1230], en definitiva, delitos castigados con al menos 4 años de prisión). Solo se podrán usar estos medios para confirmar la identidad de una persona concreta, sin posibilidad de una búsqueda difusa o prospectiva, y atendiendo a parámetros de gravedad (naturaleza del riesgo e impacto sobre derechos fundamentales de las personas afectadas), incluyéndose, por tanto, un juicio de proporcionalidad, previa realización de una evaluación de impacto relativa a los derechos fundamentales (art. 27 RIA), y previa autorización de autoridad judicial o administrativa competente.

Nos centraremos, por el interés que despiertan en cuanto a los riesgos para derechos fundamentales, en los casos de identificación biométrica relacionados directamente con los usos prohibidos. Desde su negociación, las instituciones mostraron posturas diversas, siendo que la Comisión en su posición inicial partía de una postura mucho más restrictiva, apoyada esencialmente por el Parlamento, pero que fue superada cuando el Consejo [1231] se mostró mucho más proclive a proteger intereses propios de la seguridad nacional, dando lugar a un texto final con excepciones amplias, con el correlativo debilitamiento de su planteamiento inicialmente garantista.

Habida cuenta de que la letra h) del artículo 5 establece todo un régimen de excepciones a la prohibición genérica de identificación biométrica en tiempo real con fines de uso público, quedando prácticamente vacía de contenido esta prohibición, merece la pena explorar, con Escajedo, aquellas otras prohibiciones del artículo 5 que puedan tener relación con esta práctica. La autora identifica hasta seis categorías; es decir, prácticamente todo el resto de usos prohibidos, que pueden verse, de una forma u otra, como aplicaciones concretas de esta técnica de vigilancia. Así:

---

[1230] Dichos delitos son los treinta y dos delitos enumerados en la Decisión Marco 2002/584/JAI del Consejo, relativa a la orden europea de detención y entrega.

[1231] Así, especialmente Alemania, Francia e Italia, muy contrarios en sus posiciones a que se limitara de forma completa la posibilidad de identificación con fines de seguridad nacional. ESCAJEDO SAN-EPIFANIO, L., *op. cit.* N. 1226, p. 210.

– El uso de sistemas de evaluación o clasificación social (el llamado *social scoring,* previsto en el artículo 5.1, apartado c); pues algunos sistemas biométricos se usan, de hecho, para evaluar o clasificar a personas o colectivos atendiendo a sus características (sociales o personales), con el consiguiente riesgo de que estas evaluaciones comprometan la dignidad y la autonomía personal de las mismas.

– La predicción de la peligrosidad criminal (apartado d): la prohibición se produce cuando los sistemas, para valorar el riesgo de comisión del delito se basen únicamente en perfiles biométricos, lo que puede dar lugar, según advierte Escajedo, a una «biologización del crimen», o, por decirlo en otros términos, la vuelta a los postulados del positivismo italiano y de la concepción del delincuente como un ser predeterminado al crimen [1232].

– La ampliación de bases de datos faciales (prohibida por el apartado e): se prohíbe en este punto la construcción de bases de datos a partir de imágenes captadas sin consentimiento (por ejemplo, desde CCTV o redes sociales), por el riesgo que presenta la extracción no selectiva.

– La inferencia de emociones en contextos sensibles (apartado f): en este caso, por el uso de estas técnicas, se corre el riesgo de tomar decisiones en los dos ámbitos señalados (laboral y educativo) que afecten directamente a derechos fundamentales.

– Categorización basada en atributos sensibles (apartado g): en este caso la prohibición es usar sistemas que sirvan para deducir raza, religión, orientación sexual, opiniones políticas, o, en definitiva, todos los datos especialmente protegidos por el artículo 9 del RGPD, por lo que la inclusión de esta categoría tiene sentido desde el punto de vista de la no discriminación.

En definitiva, la pretendida prohibición se acaba «fundiendo» en el régimen aplicable al resto de prohibiciones, a pesar de la especial implicación que la aplicación de los diferentes mecanismos y contextos de identificación biométrica puede presentar para derechos relevantes. Escajedo habla del evidente riesgo de hipervigilancia, así como el impacto en el principio democrático, pues en documentos previos ya se había venido expresando con contundencia

[1232] *Ibid.,* p. 212. Ya se había apuntado este riesgo en: Bravo Bolado, A., «Predicción y peligrosidad. ¿El retorno de un nuevo paradigma?», en Lascuraín Sánchez, J. A./ Peñaranda Ramos, E. (Coords.), *Liber Amicorum en homenaje a Julio Díaz-Maroto y Villarejo*, UAM Ediciones, Madrid, 2023, pp. 87-103.

que la vigilancia biométrica impacta de forma significativa sobre una amplia gama de derechos fundamentales y de valores democráticos [1233].

El considerando 32 explicita los riesgos que suponen estas tecnologías, señalando que invaden «*de forma especialmente grave los derechos y las libertades de las personas afectadas, en la medida en que puede afectar a la vida privada de una gran parte de la población, provocar la sensación de estar bajo una vigilancia constante y disuadir indirectamente a los ciudadanos de ejercer su libertad de reunión y otros derechos fundamentales»,* señalando que las imprecisiones técnicas de los sistemas de IA destinados a la identificación biométrica remota de las personas físicas pueden dar lugar a resultados sesgados y tener efectos discriminatorios. Tales posibles resultados sesgados y efectos discriminatorios son especialmente pertinentes por lo que respecta a la edad, la etnia, la raza, el sexo o la discapacidad. Además, se destaca que la inmediatez de las consecuencias y las escasas oportunidades para realizar comprobaciones o correcciones adicionales en relación con el uso de sistemas que operan «en tiempo real» [1234] hacen más evidente el riesgo para los derechos y las libertades de las personas afectadas en el contexto de actividades de garantía del cumplimiento del Derecho.

### b) Las técnicas de manipulación subliminal y explotación de las vulnerabilidades (art. 5.1. a y b)

Estas dos prácticas son, en esencia, técnicas o formas de manipulación mental destinadas a alterar sustancialmente el comportamiento de una persona o grupo, alterando su capacidad de formación de preferencias [1235].

---

[1233] Escajedo San-Epifanio, L., *op. cit.* N. 1226, p. 209, destacando la postura conservadora del Reglamento, así como algunos trabajos de la doctrina, especialmente críticos en cuanto a la desprotección contra la hipervigilancia, considerando que la normativa, en lugar de prohibirla, la legitima: Veale, M./ Zuiderveen Borgesius, F., «Demystifying the Draft EU Artificial Intelligence Act», *Computer Law Review International*, 4/2021, p. 102. En un sentido similar: Barkane, I., «Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance», *Information Polity*, 27, 2, 2022, pp. 147-162.

[1234] El texto diferencia el régimen aplicable a los sistemas que operan en tiempo real y aquellos que lo hacen en diferido (apartados 42 y 43, respectivamente, del artículo 3), entendiendo que la afectación a derechos es de distinta intensidad. Los sistemas que actúan en tiempo real están captando patrones biométricos y, simultáneamente, cotejándolos con otros patrones almacenados en bases de información previas. Sin embargo, en los sistemas en diferido, los datos empleados para la comparación no se intercambiarían instantáneamente, sino que el cotejo se produciría con una demora significativa. No convence esta distinción, empero, a Escajedo San-Epifanio, L., *op. cit.* N. 1226, p. 218, que considera que hubiera sido más garantista exigir que, por ejemplo, para el inicio del cotejo se requiriese una mínima intervención humana.

[1235] González de la Garza, L. M., «El contenido de las llamadas «técnicas subliminales» y las vulnerabilidades de grupo específico de personas en el Reglamento de inteligencia artificial», en Cotino

De todas las prácticas prohibidas por el Reglamento, podría considerarse que la menos actual de todas las preocupaciones es la que se recoge en el presente precepto, que, aunque no define las técnicas concretas de manipulación mental, delinea sus contornos, probablemente con el objeto de no verse desactualizado en un futuro próximo, ofreciendo un listado de posibles usos de estas técnicas, que pueden desplegarse para: lograr que las personas adopten comportamientos no deseados, empujándoles a tomar decisiones que perjudiquen su autonomía, capacidad de decisión y libre elección, o bien comercializando sistemas destinados a distorsionar materialmente el comportamiento humano (con la probabilidad de que se produzcan daños en la salud física o mental o financieros) [1236].

Aunque los ejemplos de sistemas que pueden dar lugar a esta manipulación son numerosos (así, simples estímulos de imagen video o audio, interfaces como gafas de realidad aumentada o virtual), también se hace referencia en el RIA a todas las técnicas, aún en fase más incipiente, de interfaz cerebro-máquina (BIC), pese a su aún escaso desarrollo [1237]. Es precisamente en esta intersección entre la IA y las neurotecnologías donde se puede producir un salto cualitativo en cuanto a la alteración o manipulación del comportamiento de las personas, siendo consciente el legislador de que, gracias a la capacidad de computación de la IA, será posible procesar y «traducir eléctricamente» las señales cerebrales con muchísimo mayor potencial. Pese a la evidente unión de ambos contextos, el legislador europeo no ha llegado a reconocer, no obstante, la afectación de los denominados neuroderechos, tampoco en aplicación de estas tecnologías, perdiendo la oportunidad de dar algún tipo de reconocimiento a su necesidad,

---

Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 251-252. Desarrollando esta cuestión, por todos: Beltrán de Heredia Ruiz, I., «Algoritmos y condicionamiento por debajo del nivel consciente: un análisis crítico de la propuesta de Ley de Inteligencia Artificial de la Unión Europea», *Revista de la Facultad de Derecho de México,* Tomo LXXIII, 286, 2023, pp. 621-650.

[1236] Considerando 29. La alusión al daño financiero, si bien aparece en la versión definitiva del Reglamento, está pensada para combatir fraudes comerciales. En este sentido, como acertadamente destaca González de la Garza, L. M., *ibid.*, con alusión al mismo considerando, habrá que atender a la normativa ya existente relativa a las técnicas de publicidad desleal (incluyéndose la prohibición de la publicidad subliminal), la permisión de la manipulación en el contexto médico debidamente autorizado, así como aceptar las estrategias de neuromarketing legítimas. *Ibid.,* pp. 255-256. Añadiríamos que precisamente por la convivencia de este precepto con normativa sectorial que prohíbe la manipulación en otros contextos, a la hora de dotar de contenido sustantivo a la prohibición, han de buscarse incumplimientos que sean cualitativamente más graves que los previstos, por ejemplo, en la normativa de publicidad.

[1237] *Ibid.,* p. 257, señalando que las técnicas que probablemente ganen auge a lo largo del siglo XXI serán las que creen interfaces cerebro-máquina mediante neuroimplantes y diademas de lectura y modificación de la conducta a través de campos electromagnéticos procesados por IA.

y uniendo el riesgo de estas técnicas con los derechos clásicos [1238]. En todo caso, y aunque autores como González de la Garza se muestran partidarios del reconocimiento de algún neuroderecho de nuevo corte (con la integridad mental como exponente básico [1239]); se pone de manifiesto su estrecha relación con otros derechos de contenido clásico que se ven afectados con estas técnicas, a saber: el libre desarrollo de la personalidad (10.1 CE); la integridad física y moral (15 CE); el no ser obligado a declarar sobre ideología, religión o creencias (16.2 CE); o a la intimidad personal (18.1 CE) [1240].

No nombra el Reglamento, no obstante, entre los riesgos derivados de estas tecnologías, los relacionados con la propaganda computacional inteligente, aprovechando la explotación de sesgos de los electores para diseñar campañas ultra personalizadas [1241], limitándose a nombrar la afectación de la autonomía en la toma de decisiones, la integridad física y mental, así como los intereses financieros.

Aunque habrá que esperar para comprobar cómo se interpretan estas prohibiciones, algunos autores han señalado ya que sería conveniente adoptar un concepto de manipulación que incluya la manipulación sin intención del diseñador, ya sea porque esta se vea favorecida por el objetivo del entrenamiento, o bien porque el modelo aprenda a imitar el comportamiento manipulador en

---

[1238] Para González de la Garza, esto constituye, sin duda, un error. *Ibid.,* p. 256. Considera este autor que la privacidad mental exige nuevos desarrollos técnicos, y que, de hecho, ha sido reconocido por algunas declaraciones de derechos (por ejemplo: principios 6, 7, 8 y 10 del Código de Conducta global de la IA de 30 de octubre de 2023, denominado el proceso de IA de Hiroshima aprobado por el G7.) La propuesta de este autor pasa por afirmar un modelo de patrimonialización del dato. *Ibid.,* p. 269, proponiendo que se cree un mercado de datos en el cual cada ciudadano tenga acceso a la información que cada empresa disponga sobre su persona, recibiendo una parte de los ingresos que con esos datos se generan. La propuesta bebe directamente, y de forma explicitada por el autor, de los postulados del economista Daron Acemoglu, expuestos recientemente en Acemoglu, D./ Johnson, S., *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*, PublicAffairs, Nueva York, 2023 y puede entenderse como una propuesta de superación de los problemas detectados en su momento por Zuboff (explicados *supra,* en capítulo cuarto).

[1239] Imprescindible en este sentido: Pérez Manzano, M., «¿Un estatuto constitucional singular…, *op. cit.* N. 1018. La misma, en: «Justicia penal y neuroderechos», en AAVV., *Estudios político-criminales, jurídico-penales y criminológico. Llibro homenaje al profesor José Luis Díez Ripollés*, Tirant Lo Blanch, Valencia, 2023, pp. 329-348. En lengua inglesa, por todos: Ienca, M./ Andorno, R., «Towards new human rights in the age of neuroscience and neurotechnology», *Life Sciences, Society and Policy*, 5, 2017, pp. 1-27, abogando por una reconceptualización del derecho a la integridad mental (aunque protegido, hoy, mediante el artículo 3 CEDH), de forma que incluya la protección frente a intrusiones tecnológicas no consentidas en el bienestar mental.

[1240] González de la Garza, L. M., *op. cit.* N. 1235, p. 270.

[1241] Las preocupaciones sobre la llamada propaganda cognitiva electoral o computacional son muy numerosas recientemente. Por todos, con un enfoque empírico basado en el examen de casos concretos ya acaecidos: Woolley, S. C./Howard, P. N., *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, Oxford University Press, Nueva York, 2019. Más recientemente, proponiendo una prohibición aún más absoluta de los usos de condicionamiento humano por debajo del nivel de la consciencia, Beltrán de Heredia Ruiz, I., *op. cit*. N. 1235.

sus datos de entrenamiento [1242], pero con el límite que el propio Reglamento (considerando 29) fija, y es que no se podrá presuponer que existe esta intención de alterar el comportamiento de las personas si es resultado de factores externos que escapan al control del proveedor o responsable de despliegue, que no sea lógico prever.

Constituye una variante de esta manipulación la prevista en el apartado b) del artículo 5: la explotación de vulnerabilidades de una persona física o colectivo derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables. Más allá de las críticas que la doctrina ha vertido relativas a la inseguridad que se deriva de la indeterminación de algunos de los conceptos usados por el Reglamento (perjuicio «considerable», o alteración «sustancial»), parece innegable que el RIA pretende proteger a colectivos que, por sus menores capacidades de autoprotección, pueden verse especialmente discriminados e impulsados a tomar decisiones no conscientes, consentidas o queridas. Al daño a la integridad psíquica, moral o económica arriba señalados, debe sumarse la discriminación como potencial daño de estas técnicas.

### c) Los sistemas de clasificación social (5.1.c)

El Considerando 31 del RIA establece que *«los sistemas de IA que permiten a agentes públicos o privados llevar a cabo una puntuación ciudadana de las personas físicas pueden tener resultados discriminatorios y abocar a la exclusión a determinados grupos. Pueden menoscabar el derecho a la dignidad y a la no discriminación y los valores de igualdad y justicia».*

Dichos sistemas de IA, continúa, evalúan o clasifican a las personas físicas o a grupos de estas sobre la base de varios puntos de datos relacionados con su comportamiento social en múltiples contextos o de características personales o de su personalidad conocidas, inferidas o predichas durante determi-

[1242] Así, Fernández Hernández, C., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 195. También en: Carrol, M./Chan, A./Ashton, H/Krueger, D., «Characterizing Manipulation from AI Systems», *Proceedings of Under review*, ACM, NewYork, NY, USA, 2023, argumentando que la manipulación ocurre cuando un sistema de aprendizaje automático afecta intencionadamente el comportamiento o las creencias de las personas de forma que interfiere con su autonomía, especialmente cuando esto sucede sin su conciencia o con información parcial, pero en todo caso sin intención del diseñador, posibles cuando, entre otros escenarios, el sistema se mueve por medio de incentivos para optimizar su objetivo inicial.

nados períodos de tiempo. La puntuación ciudadana resultante puede dar lugar a un trato perjudicial o desfavorable de determinadas personas físicas o colectivos enteros, o a un trato perjudicial desproporcionado o injustificado en relación con la gravedad de su comportamiento social. Por lo tanto, culmina, deben prohibirse los sistemas de IA que impliquen esas prácticas inaceptables de puntuación y den lugar a esos resultados perjudiciales o desfavorables.

Para Presno Linera, la prohibición de este tipo de sistemas conecta directamente con el derecho a no verse discriminados [1243], y responde a una lógica en la que el ser humano se convierte en materia prima aprovechable para prácticas políticas, sociales y comerciales, presididas por una lógica de producción que persigue la modificación conductual. En contraposición a los sistemas nacionales en que este paradigma está ya implantado, Presno considera que en los países de la Unión Europea aún existe un grado significativo de protección y preocupación por la vida privada y los datos personales, así como una preocupación sobre el libre desarrollo de las personas [1244].

La vinculación con el derecho a no verse discriminado parece evidente si se atiende a la génesis de este precepto del Reglamento, donde se exige que, como consecuencia de este tratamiento de datos conducente a la puntuación, se produzca un trato desfavorable, con un matiz importante: que los datos que generen ese trato desfavorable deben haberse obtenido en un contexto diferente de aquel en que se provoca el perjuicio. El ejemplo que pone el autor permite entender la lógica: sí es posible usar datos conseguidos en el seno de una relación laboral para tomar decisiones que afecten a la propia esfera del traba-

---

[1243] Presno Linera, M. A., «La prohibición de sistemas de inteligencia artificial que evalúan y clasifican a las personas a partir de datos que no guardan relación con el contexto donde se generaron y que provocan discriminaciones», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 241, vinculando su dinámica con el sistema de capitalismo de vigilancia dibujado por Zuboff y explicado aquí con detalle en el capítulo cuarto. Ideas similares en Presno Linera, M. A., *Derechos fundamentales... op. cit*. N. 984, obra analizada *supra,* en capítulo sexto.

[1244] Presno alude en este punto a la paradoja de la privacidad, (Presno Linera, M. A., «La prohibición de sistemas…, *ibid.*, p. 243) a la que cada vez cedemos más terreno, y no deja de aceptar que, pese a que nuestro sistema dista mucho de parecerse al chino, cada vez vemos más ejemplos de esta lógica de la vigilancia y el control en espacios especialmente sensibles, como el laboral. Utiliza el ejemplo de la vigilancia de trabajadores por medio de algoritmos con el caso Glovo (explicado *supra,* N. 1013). No obstante, otros autores han puesto de manifiesto la ironía o paradoja de que, aunque los políticos europeos y en general occidentales describen estas prácticas como un problema derivado de regímenes autoritarios, como China, en occidente también existen este tipo de sistemas de clasificación (así, el algoritmo de automatización de decisiones de acceso a la vivienda en Países Bajos, *vid.*: Heikkilä, M., «The AI myth Western lawmakers get wrong, *MIT Technology Review*, 29 de noviembre de 2022, disponible en: <https://www.technologyreview.com/2022/11/29/1063777/the-ai-myth-western-lawmakers-get-wrong/> [Accedido por última vez el 29/01/2026]. También en: Fernández Hernández, C., *op. cit*. N. 1242, p. 199.

jo, pero no sería legítimo usar la menor calificación crediticia de una persona como condicionante para obtener el ascenso en una empresa.

Además de un trato desfavorable, el Reglamento exige, para que entre en juego la prohibición, que dicho trato sea injustificado o desproporcionado con respecto al comportamiento social (de la persona afectada) o la gravedad del mismo. Este plus de desproporción o gravedad, podemos entender, aunque no lo explicita el legislador europeo, indica que estamos ante una situación aún más gravosa para el perjudicado por la decisión, pues se ve sorprendido por unas consecuencias desfavorables sin alcanzar a comprender en qué otros ámbitos de su vida personal pueden verse afectados por una recolección de datos llevaba a cabo en cualquier otra esfera, dándose una situación de desprotección más gravosa respecto de sus derechos [1245].

### d) Sistemas para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de comisión de un delito basándose únicamente en la elaboración de perfiles (5.1.d)

El considerando 42 advierte de que, en consonancia con la presunción de inocencia, las personas físicas siempre deben ser juzgadas con base en su comportamiento real, y nunca con base exclusivamente en la elaboración de perfiles, sin valoración humana, y sin que exista sospecha razonable basada en hechos objetivos comprobables [1246].

Fue esta una de las prohibiciones más discutidas, elevada, finalmente, a prohibición tras las enmiendas introducidas por el Parlamento, si bien exigiendo, como matiz importante, que la evaluación del riesgo se base «únicamente» en el proceso de perfilado. Se incluye, así, una prohibición de algunas técnicas de la llamada policía predictiva, solo en lo relativo a las infracciones penales, aunque aplicable, en principio, a cualquier fase del proceso (desde la propia investigación o *policing*, hasta la fase de sentencia o *sentencing)* [1247].

---

[1245] Los ejemplos de criterios de clasificación crediticia en el ámbito laboral, entre otros muchos, pueden encontrarse en O'Neil, C., *op. cit.* N. 988, pp. 181 y ss.

[1246] Se excluyen, sin embargo, en el mismo considerando, los análisis de riesgos que no estén basados en la elaboración de perfiles de personas o en los rasgos y características de la personalidad de las personas, como los sistemas de IA que utilizan los análisis de riesgos para evaluar la probabilidad de fraude financiero por parte de empresas sobre la base de transacciones sospechosas o las herramientas de análisis de riesgo para predecir la probabilidad de localización de estupefacientes y mercancías ilícitas por parte de las autoridades aduaneras.

[1247] Miró-Llinares, F./ Santisteban Galarza, M, *op. cit*. N. 1216, p. 330. Para una aproximación al ya manido concepto de «policía predictiva», así como sus orígenes, desde un punto de vista crítico, considerando que en las conceptualizaciones más actuales se prefiere hablar de *«data driven policing». Ibid.,* p. 317-318, fijando ya en p. 319 que: *«Más allá de que tenga razón Wilson cuando señala que dentro*

Puede entenderse, con Miró y Santisteban, que la elaboración de estos perfiles con fines de evaluación delictiva comporta un riesgo evidente de que se lleven a cabo correlaciones genéricas que no sean correctas para todas las personas, reduciendo el trato a la persona como miembro de un grupo y no como individuo [1248]. Señalan, además, los riesgos éticos asociados tradicionalmente a estos algoritmos predictivos en el campo penal, distinguiendo con acierto la diferencia sustancial que introduce la IA respecto de los algoritmos clásicos, pues el problema aquí presente no se refiere solo a la automatización de sesgos ya presentes en las decisiones individuales policiales, ni tampoco a la dudosa validez de los pronósticos, sino que, dado que la lógica científica detrás de los algoritmos de IA constituye un cambio de paradigma (basada en el aprendizaje dinámico sobre datos en tiempo real), con una opacidad relevante y con una dificultad intrínseca para explicar el resultado de sus predicciones, ello conduce directamente a la automatización de sesgos que quedan ocultos [1249].

El elemento esencial para que la práctica sea prohibida, criticado por el elemento de inseguridad que introduce, es, entonces, que la decisión sea *únicamente* tomada con base en el perfilado. Esto nos remite, como señalan con acierto Miró y Santisteban, a la interpretación del artículo 22 del RGPD, que establece la necesidad de que las decisiones no sean *completamente* automatizadas, en

---

*del amplio término de policía predictiva nos encontramos con un «popurrí de metodologías y filosofías policiales contemporáneas» que enlazan la digitalización y la tendencia a la automatización en la toma de decisiones con una idea de predicción16, existe acuerdo en destacar dos tipos de técnicas de pronóstico delictivo, los pronósticos delictivos enfocados en el lugar en el que se va a cometer el crimen, y los pronósticos delictivos centrados en las personas que lo van a perpetrar o van a ser víctimas del mismo. Ambas técnicas se pueden llevar a cabo utilizando o no sistemas de IA […]».* En profundidad, una visión realista del uso de algoritmos predictivos policiales, en: Miró-Llinares, F., «Predictive policing: utopia or dystopia? On attitudes towards the use of big data algorithms for law enforcement», *Revista de Internet Derecho y Política*, 30, 2020, pp. 1-18.

[1248] En este sentido, informe de la Agencia Europea de Derechos Fundamentales (FRA), *Guía para prevenir la elaboración ilícita de perfiles en la actualidad y en el futuro*, 2019, disponible en: < https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_es.pdf > [Accedido por última vez el 29/01/2026].

[1249] Miró-Llinares, F./ Santisteban Galarza, M., *op. cit*. N. 1216, pp. 321 y ss. En lo relativo a los riesgos de los algoritmos de predicción policial o judicial clásicos (herramientas de valoración de riesgo), por todos: Martínez Garay, L./ Montes Suay, F., *op. cit.* N. 17. En cuanto al salto cualitativo que supone el *machine learning*, una buena muestra se puede encontrar ya en decisiones judiciales contrarias a permitir las técnicas de IA con aprendizaje automático para la toma de decisiones judiciales. Así, el TJUE en el caso de la utilización de datos del registro de nombres de los pasajeros de aerolíneas para detectar delitos de terrorismo y delitos graves facilitado por la Directiva 2016/681 (Sentencia del Tribunal de Justicia en el asunto C-817/19, *Ligue des droits humains*), o el TC alemán en su prohibición de uso de estos algoritmos por la policía de los Länder de Hesse y Hamburgo: Sentencia del Tribunal Constitucional Alemán de 16 de febrero de 2023, BVerfG, 1 BvR 1547/19, 1 BvR 2634/20, disponible en: <https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2023/02/rs20230216_1bvr154719.html >[Accedida por última vez el 29/01/2026].

un sentido similar al requisito aquí analizado, con la ventaja de haber sido objeto de extensas interpretaciones [1250]. La línea que separa, por tanto, los usos prohibidos y permitidos (aunque de alto riesgo) de la llamada policía predictiva tiene que ver con la existencia de decisiones completamente automatizadas.

Entendemos que, en caso de producirse una decisión de estas características, no solo estaremos ante la invisibilización de sesgos discriminatorios, sino que será inviable saber qué posibles datos han condicionado una decisión restrictiva de derechos fundamentales de primer orden, que pueden alcanzar, en última instancia, a una privación de libertad injusta. La presunción de inocencia y la libertad en sentido estricto se ven, por tanto, en juego con el uso de estas técnicas, cuya prohibición responde a intereses de primer orden para el derecho penal.

### e) Sistemas de IA basados en reconocimiento facial con imágenes extraídas de Internet o televisión (5.1.e), reconocimiento de emociones (5.1.f) y categorización biométrica para clasificación individual cuando deduzcan determinados datos sensibles (5.1.g)

Las técnicas a las que se refiere este inciso se refieren a sistemas biométricos avanzados capaces de reconocer emociones, clasificar a personas, detectar comportamientos, pensamientos e incluso evaluar personalidades, incluyendo los llamados polígrafos inteligentes. Ejemplos de estos sistemas los encontramos en los controles fronterizos, así como en aeropuertos [1251].

Más allá de las técnicas de identificación biométrica más modernas, que se basan en identificadores débiles como la forma de andar, patrones sanguíneos o incluso datos tan sutiles como el patrón en la pulsación de las teclas de

[1250] En este sentido resulta especialmente útil acudir a las directrices del grupo de trabajo del artículo 29, que se ha encargado de interpretar dicho precepto, señalando que el punto central es que no haya participación humana en el proceso de decisión (Comisión Europea, Article 29 Data Protection Working Party, «*Guidelines on Automated individual decisionmaking and Profiling for the purposes of Regulation 2016/679 (WP 251 rev.01)*»,22 de agosto de 2018,). De especial importancia, por su actualidad, y por determinar la aplicación de la prohibición del artículo 22 RGPD en un caso de *scoring crediticio*, STJUE de 7 de diciembre de 2023, asunto C-634/21, caso *Schufa.*

[1251] No solo en EEUU o Reino Unido, donde el debate sobre su uso viene siendo común desde hace tiempo, sino también como proyecto en las fronteras europeas, con la Comisión Europea financiando el proyecto *Intelligent Portable Control System* (iBorderCtrl), basado en supuestos biomarcadores de engaño y microexpresiones faciales. *Vid.*: Sánchez Monedero, J./ Dencik, L., «The Politics of Deceptive Borders: «Biomarkers of Deceit» and the Case of iBorderCtrl», *Information, Communication & Society*, 3, 2020, pp. 413-430, críticos, poniendo de manifiesto su dudosa eficacia técnica, y evidenciando los riesgos de discriminación hacia población vulnerable, con abundantes referencias empíricas que ponen de manifiesto las debilidades técnicas de estos modelos.

un ordenador, la nueva generación de tecnologías da un paso más y comienza a ser viable el reconocimiento de emociones [1252].

En cuanto a la razón de la prohibición de estas técnicas, el considerando 43 expresa, en lo relativo a la extracción no selectiva de imágenes en internet o circuitos cerrados (la también denominada técnica de *scraping),* que se prohíben pues «*agravan el sentimiento de vigilancia masiva y pueden dar lugar a graves violaciones de los derechos fundamentales, incluido el derecho a la intimidad»*. En opinión de Simón Castellano, la prohibición encuentra su fundamentación nuclear, más allá de la privacidad, en la necesidad evidente de prohibir abusos, manipulación, o discriminación, señalando que estas técnicas pueden contravenir la dignidad humana y la autonomía personal, con efectos nocivos en el libre desarrollo de la personalidad [1253].

En lo relativo a la inferencia de emociones, el objeto de prohibición se extiende, tan solo, a los lugares de trabajo y centros educativos (con la excepción de que se realice con fines médicos o de seguridad). El considerando 44 incide en la idea de que estas prácticas suponen una vigilancia masiva y que puede dar lugar a violaciones graves de la intimidad. La explicación del considerando de la prohibición específica del ámbito laboral o educativo tiene que ver con el desequilibrio de poder en sendos contextos, unido al carácter intrusivo de estos sistemas, que podrían dar lugar a un trato perjudicial o desfavorable, en definitiva, discriminatorio, de determinadas personas físicas o colectivos enteros. Simón Castellano hace referencia, de nuevo, a la privacidad y la integridad psicológica, pues la inferencia de emociones, amén de ser una técnica aún falible y propensa a los errores y malinterpretaciones (cuestión reconocida en el propio considerando 44), no solo invade la privacidad, sino que puede manipular su comportamiento y decisiones, con un alto riesgo de error [1254].

Finalmente, en cuanto a la prohibición de la categorización biométrica para la inferencia de datos relativos a la raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual, vuelve a colocarse en el centro de la justificación la preocupación por la privacidad, la igualdad y la no discriminación, dando lugar a un estado de vigilancia masiva,

---

[1252] Simón Castellano, P., «El resto de los sistemas de inteligencia artificial prohibidos o inaceptables en el Reglamento de inteligencia artificial», Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 279.

[1253] *Ibid.,* p. 281.

[1254] *Ibid.,* p. 283.

que conduce, una vez más, a generar un ambiente de desconfianza en la ciudadanía, socavando, finalmente, su libertad individual y autonomía personal [1255].

#### f) Algunas conclusiones sobre los usos prohibidos

Aunque la aplicación e interpretación efectiva de las prohibiciones contenidas en el artículo 5 del RIA será la que confirme la extensión de sus contornos, cabe afirmar que las obligaciones de no hacer que introduce el Reglamento al establecer usos vetados se encaminan directamente, según la propia *ratio legis* expresada por el legislador y reforzada por la doctrina, a la protección de derechos fundamentales del más alto interés. Se ha tratado de bajar a tierra la realidad de estos intereses en cada una de las aplicaciones, que aquí se resumen.

Dejando de lado la dudosa prohibición sobre identificación biométrica, habida cuenta de su caótico régimen de excepciones, susceptible de tratarse más bien como un supuesto de alto riesgo, el resto de las categorías protegen, en esencia: la intimidad, el derecho a no sufrir discriminaciones, la libertad o autonomía, tanto en un sentido amplio (vinculado al libre desarrollo de la personalidad) como estricto (libertad ambulatoria frente a la hipervigilancia y a posibles detenciones o sanciones injustas), la integridad psíquica y, residualmente, los intereses económicos. Sin duda, y sin necesidad de acudir a derechos de nuevo cuño, derechos todos ellos reconocidos como valiosos por nuestro ordenamiento penal.

### 3.4 **Los sistemas de alto riesgo: el verdadero núcleo del riesgo permitido**

Partiendo de la idea (considerando 48) de que los sistemas de alto riesgo (en adelante, *SAR* [1256]), pueden afectar hasta a 21 derechos fundamentales, el Reglamento utiliza dos criterios para determinar que nos encontramos ante uno de estos sistemas: o bien porque se integren en productos amparados por

[1255] *Ibid.,* p. 284. Cabe resaltar que se produce aquí, como así señala la doctrina, una colisión o conflicto con el Reglamento General de Protección de Datos, en tanto el RIA excepciona de esta prohibición el tratamiento de estos datos adquiridos de conformidad con el derecho nacional, cuando sean necesarios para detectar el posible sesgo, lo que supone una posible colisión con el artículo 9 RGPD. Resta comprobar, en este sentido, si esta permisión se considerará un nuevo fin legítimo adicional.

[1256] Cotino Hueso, L., *op. cit.* N. 1216, pp. 291 y ss.

determinada legislación armonizada de la UE [1257] (ya sean productos o componentes de seguridad de esos productos), en cuyo caso se exige, adicionalmente, que la normativa específicamente aplicable al producto en cuestión exija que se someta a una evaluación de conformidad realizada por un organismo independiente para introducirse en el mercado *(inclusión por pertenencia a ámbitos de producción industrial sensibles);* o bien, bajo ciertos requisitos, si están vinculados a alguna de las finalidades o usos enumerados en el Anexo III [1258] *(inclusión por empleo en determinadas áreas de actividad).*

El elemento determinante para que se aplique este Anexo III es la «finalidad prevista» así como la «capacidad de un sistema de IA para alcanzar su finalidad prevista» (art. 3.18 RIA). Pero destaca acertadamente Cotino [1259] cómo se ha ido evolucionando en la redacción de este apartado, pues en las primeras versiones del RIA se aplicaba de forma automática, mientras que ahora es necesario que además de las finalidades se cumplan unos requisitos para ser considerado un sistema de alto riesgo. El sistema ha de generar efectivamente un riesgo y, sobre todo, *«influir sustancialmente en el resultado de la toma de decisiones»* (art. 6.3.º RIA), introduciéndose un cambio sustancial que nos sitúa, una vez más, cerca del artículo 22 RGPD. Sin embargo, el Reglamento concreta en este punto el significado, dando una interpretación original sobre esta influencia «sustancial», pues el considerando 53 explicita que la premisa o punto de partida de los sistemas de alto riesgo es que plantean un *«riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas», por lo que, siguiendo esta lógica, no se considera SAR «un sistema de IA que no afecta al fondo, ni por consiguiente al resultado, de la toma de decisiones, ya sea humana o automatizada» (*cons. 53), pues no planteará, entonces, un riesgo importante para la salud, seguridad o derechos fundamentales. No serán SAR, por ello, los sistemas utilizados para los fines del anexo III, pero con una tarea de *«procedimiento limitado», o para «mejorar […] una actividad*

[1257] Dicha legislación se especifica en el Anexo I del RIA, tratándose de productos con cierto nivel de peligrosidad. Así: máquinas, juguetes, embarcaciones de recreo y motos acuáticas, ascensores, aparatos y sistemas de protección para uso en atmósferas potencialmente explosivas, equipos radioeléctricos, equipos a presión, instalaciones de transporte por cable, equipos de protección individual, aparatos que queman combustibles gaseosos, productos sanitarios, productos sanitarios para diagnóstico in vitro, Aviación civil, vehículos, equipos marinos, y sistemas ferroviarios. Para ver el detalle de cada una de las normas aplicables. *Vid.* Anexo I del RIA.

[1258] A saber: 1. Biometría; 2. Infraestructuras críticas; 3. Educación y formación profesional; 4. Empleo, gestión de los trabajadores y acceso al autoempleo; 5. Acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones; 6. Aplicación de la ley; 7. Migración, asilo y gestión del control fronterizo; 8. Administración de justicia y procesos democráticos. Se explicarán detalladamente *infra,* por encontrarse especialmente vinculados estos usos con los riesgos evidentes para derechos fundamentales y bienes jurídico-penales.

[1259] Cotino Hueso, L., *op. cit.* N. 1216, p. 298.

*humana previamente» realizada; cuando el sistema «no esté destinado a sustituir […] ni a influir»* en la toma de decisiones, o cuando el uso del sistema IA sea claramente accesorio. Con una salvedad (art. 6.4), y es que siempre será alto riesgo el sistema usado para elaborar perfiles de personas físicas [1260].

A pesar de la necesaria aplicación de este segundo requisito sustantivo, ha de puntualizarse que, en principio, ante la concurrencia de una de las finalidades del Anexo III, que enseguida se analizarán, se presumirá de partida que el sistema es de alto riesgo, generándose toda una serie de deberes para los proveedores que deben actuar de forma automática, aplicando todas las previsiones del capítulo III. Asimismo, se debe documentar la cualidad de alto riesgo del sistema e incluirlos en el registro al efecto (art. 6. 4.º).

Analizaremos, entonces, brevemente, cada una de las finalidades que permiten clasificar, bajo los parámetros arriba explicados, un sistema como de alto riesgo, con el objetivo de determinar cuáles son los bienes que pueden verse afectados en cada uno de los sectores.

### a) Relación de ámbitos del Anexo III

#### 1. *Biometría*

Recuérdese que la biometría o reconocimiento biométrico constituía la octava aplicación prohibida en discordia, habida cuenta de las numerosas excepciones, por lo que todo lo dicho *supra* sobre el concepto de biometría y las excepciones a su prohibición son, aquí, plenamente aplicables. Cabe matizar, no obstante, que concurre en su modalidad de sistema de alto riesgo alguna crítica parecida, principalmente relativa a su falta de concreción como un uso o finalidad propia, pues, como acertadamente señala Escajedo [1261], habrá técnicas biométricas en multiplicidad de áreas, prácticamente, en todas aquellas en que los datos recolectados se usen para la elaboración de perfiles. En este sentido, la legislación puede resultar excesivamente reiterativa.

Serán de alto riesgo, con independencia del escenario en que se apliquen: los sistemas de identificación biométrica remota, los sistemas de categorización biométrica basados en atributos o características sensibles, y determina-

[1260] Habida cuenta del complejo sistema de categorización, y de la importancia capital que tiene la consideración de un sistema como de alto riesgo, se explicita que la Comisión debe proporcionar, en el futuro, directrices que especifiquen la aplicación práctica (considerando 53).

[1261] Escajedo San-Epifanio, L., *op. cit.* N. 1226, p. 224.

dos sistemas destinados al reconocimiento biométrico de emociones (punto 1 del Anexo III).

El segundo de los apartados del Anexo III incluye hasta 21 modalidades de IA de alto riesgo, agrupadas en seis escenarios operativos: educación, ámbito laboral, prestaciones esenciales, garantía de cumplimiento del derecho, tránsito transfronterizo y administración de justicia. En estos casos, también es posible que estos sistemas sean de reconocimiento biométrico, por lo que la sistemática resulta bastante confusa, porque, como señala Escajedo, las tecnologías biométricas se pueden usar en multitud de ámbitos operativos, siendo sus objetivos de lo más variados [1262]. En todo caso, en tanto en cuanto el uso de la biometría esté permitido por el derecho de la Unión o el derecho nacional aplicable, su uso será permitido, pero de alto riesgo, con todas las obligaciones y el régimen jurídico que esta calificación conlleva.

### 2. *Infraestructuras críticas*

Son definidas como aquellos *«sistemas de IA destinados a ser utilizados como componentes de seguridad en la gestión y el funcionamiento de las infraestructuras digitales críticas, del tráfico rodado o del suministro de agua, gas, calefacción o electricidad»* (art 6.2). El considerando 33 reconoce que una perturbación de las mismas puede suponer una *«amenaza inminente para la vida o la seguridad física de las personas físicas»,* y el considerando 55 detalla que procede clasificar como del alto riesgo a los sistemas destinados a ser usados como componentes de seguridad de los sectores arriba citados, pues pueden *«poner en peligro la vida y la salud de las personas a gran escala y alterar de manera considerable el desarrollo habitual de las actividades sociales y económicas».*

### 3. *Educación y formación profesional*

El considerando 56, aunque reconoce la potencialidad positiva que pueden tener los sistemas de IA en el ámbito educativo en el fomento de la formación en competencias digitales, considera que se trata de un sector de alto riesgo en tanto *«pueden decidir la trayectoria formativa y profesional de una*

[1262] *Idem.* Téngase en cuenta que se excluye aquí la verificación biométrica con la finalidad única de confirmar que una persona física concreta es la persona que afirma ser.

*persona y, en consecuencia, puede afectar a su capacidad para asegurar su subsistencia. Cuando no se diseñan y utilizan correctamente, estos sistemas pueden invadir especialmente y violar el derecho a la educación y la formación, y el derecho a no sufrir discriminación, además de perpetuar patrones históricos de discriminación».*

En este sentido, se consideran SAR los sistemas diseñados para determinar el acceso o la admisión o «para distribuir a las personas físicas» entre instituciones educativas (3.a); los destinados a evaluar los resultados del aprendizaje (3.b), así como los sistemas que permiten «evaluar el nivel de educación adecuado que recibirá una persona o al que podrá acceder» (3.c), así como los sistemas utilizados «para el seguimiento y la detección de comportamientos prohibidos […] durante los exámenes».

### 4. *Empleo, gestión de los trabajadores y acceso al autoempleo*

El considerando 57 aclara que son considerados SAR en tanto *«dichos sistemas pueden perpetuar patrones históricos de discriminación, por ejemplo contra las mujeres, determinados grupos de edad, las personas con discapacidad o las personas de orígenes raciales o étnicos concretos o con una orientación sexual determinada, durante todo el proceso de contratación y en la evaluación, promoción o retención de personas en las relaciones contractuales de índole laboral. Los sistemas de IA empleados para controlar el rendimiento y el comportamiento de estas personas también pueden socavar sus derechos fundamentales a la protección de los datos personales y a la intimidad».* Se incluyen aquí los sistemas usados para la gestión de trabajadores, acceso al autoempleo, contratación y selección de personal, toma de decisiones de índole laboral, promoción y rescisión de la relación laboral, asignación de tareas y supervisión.

### 5. *Acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones*

El extenso considerando 58 especifica que se trata de servicios *«necesarios para que las personas puedan participar plenamente en la sociedad o mejorar su nivel de vida».* Se destacan como especialmente relevantes algunos ámbitos prestacionales (sanidad, seguridad social, asistencia social, ayuda a la vivienda o protección en situaciones de vulnerabilidad tales como enferme-

dad, vejez o desempleo). En caso de tomarse estas decisiones con ayuda de sistemas de IA, la concesión, denegación o modificaciones de estas prestaciones *«podría tener un efecto considerable en los medios de subsistencia de las personas y vulnerar sus derechos fundamentales, como el derecho a la protección social, a la no discriminación, a la dignidad humana o a la tutela judicial efectiva».* Se incluyen, en el listado del Anexo III, apartado 5, además de los sistemas de asistencia pública, los sistemas usados para evaluar la solvencia crediticia (justifica su importancia el considerando 58 al determinar que implican el posible acceso a recursos y servicios esenciales como la vivienda, la electricidad y los servicios de telecomunicaciones), los destinados a la evaluación de riesgos para la fijación de precios en seguros de vida y salud (pues, según el citado considerando, *«pueden afectar de un modo considerable a los medios de subsistencia de las personas y […] pueden vulnerar sus derechos fundamentales y tener graves consecuencias para la vida y la salud de las personas, como la exclusión financiera y la discriminación»),* y los sistemas diseñados para ser usados en la clasificación de llamadas de emergencia (policía, bomberos y servicios médicos).

Destaca Cotino que se protege el acceso y el disfrute de estos servicios sin distinguir entre los servicios públicos o privados, no obstante, señala que el uso de sistemas de IA por el sector público goza de un, a juicio del autor, desproporcionado plazo de adecuación al Reglamento (6 años), y que, además de ello, se echa en falta la inclusión de algunos sistemas también habituales, que pueden suponer la afectación de derechos fundamentales, como son los sistemas de persecución del fraude [1263].

### 6. *Aplicación de la ley*

La calificación de sistema de alto riesgo se concreta aquí en cinco supuestos: evaluación del riesgo de que una persona sea víctima de delitos, polígrafos o similares, evaluación de la fiabilidad de pruebas durante la investigación o enjuiciamiento de delitos, evaluación del riesgo de que una persona física cometa un delito o reincida, y elaboración de perfiles de personas físicas

[1263] Cotino Hueso, L., *op. cit.* N. 1216. En nota 40 explicita que de hecho los excluye expresamente el apartado 5, b). No obstante, el autor señala que el Considerando 58 no parece que se piense en sistemas públicos de detección de fraudes. Se trataría, sin embargo, de *«sistemas de IA previstos por el Derecho de la Unión con vistas a detectar fraudes en la oferta de servicios financieros y, a efectos prudenciales, para calcular los requisitos de capital de las entidades de crédito y las empresas de seguros no deben considerarse de alto riesgo en virtud del presente Reglamento».*

durante la detección, investigación o enjuiciamiento de delitos. Igual que ocurría con las biometrías, que encontraban su ámbito de aplicación prohibido y el correlativo permitido, lo mismo ocurre ahora con las denominadas técnicas de policía predictiva. Señalan en este punto Miró y Santisteban que resultará harto complejo, en la práctica, delimitar el uso prohibido del de alto riesgo, en tanto lo que hay que examinar es si se adopta un pronóstico de riesgo basado «solamente» en la elaboración de perfiles, destacando ambos autores que esto dependerá esencialmente del papel del operador jurídico en la toma de decisión, y que será determinante el análisis del contexto concreto en que se inserta el sistema, pues *«no solo nos ayudará a determinar si «formalmente» se acepta la decisión del sistema como vinculante, sino también si, de facto, existe un normal proceder en el que el sistema no auxilia al implementador de la IA, sino que sustituye su decisión»* [1264].

Destaca González-Meneses [1265] que todos los supuestos, salvo el relativo al polígrafo, requieren una conexión con la investigación de delitos, quedando extramuros de la clasificación de alto riesgo todos los sistemas de ejecución forzosa, aplicación de la ley o *enforcement* propios de otras ramas del derecho. El autor nombra, como ejemplo, la automatización del cumplimiento de contratos *(smart contracts* [1266]*).*

### 7. *Migración, asilo y gestión del control fronterizo*

Se incluyen aquí los polígrafos o similares para detectar engaños por autoridades públicas, la evaluación de riesgos de entrada de personas en el territorio de un Estado Miembro, la tramitación de solicitudes migratorias, o los sistemas usados en tareas de migración, asilo y control fronterizo para reconocer a personas físicas, salvo que se usen solo para verificar documentos.

### 8. *Administración de justicia y procesos democráticos*

---

[1264] Miró-Llinares, F./ Santisteban Galarza, M, *op. cit*. N. 1216, p. 335.

[1265] González-Meneses García-Valdecasas, M., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 214.

[1266] Puede señalarse, no obstante, que las preocupaciones que se relacionan con el proceso penal han sido objeto de discusión desde la Resolución del Parlamento de 2021 (Resolución del Parlamento Europeo sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI)), donde se ponía ya de relieve la especial importancia de los bienes en juego en estos procesos.

En este último apartado se incluyen dos ámbitos diferenciados y amplios. En lo relativo a la Administración de justicia, se trata de sistemas usados por una autoridad judicial, o en su nombre, tanto para ayudar en una investigación como para el cumplimiento del derecho, o que se aplican para la resolución alternativa de conflictos. Gonzáles-Meneses destaca en este sentido que aquí no hay limitación al proceso penal, ni a asuntos en que estén involucradas personas físicas, pero que sí se debe tratar de una autoridad judicial, no aplicándose a los casos de actuaciones administrativas, que solo serán de alto riesgo en tanto afecten a alguna de las otras áreas específicas del Anexo III. [1267]

El segundo de los ámbitos de este apartado se refiere a los procesos democráticos, es decir, sistemas de IA usados para influir en el resultado de una elección, referéndum o comportamiento electoral. Es importante diferenciar aquí, una vez más, por mucho que la diferencia sea sutil, entre el ámbito de lo prohibido y de lo permitido. Así, queda prohibido un nivel de influencia en la decisión política tal que permita entender que se produce (o sea susceptible de producirse) un perjuicio considerable o bien en la persona manipulada o bien en un colectivo (moviéndonos en el ámbito de aplicación de la prohibición de manipulación subliminal y explotación de vulnerabilidades); mientras que, habrá que entender, *a sensu contrario,* que si ese daño no es considerable, nos moveremos en el ámbito del riesgo permitido, pero limitado (alto riesgo) [1268].

### b) Régimen de obligaciones aplicable a los sistemas de IA de alto riesgo

Como consideración previa, antes de entrar a explicar el contenido de las obligaciones propiamente dichas, merece la pena resaltar la peculiar técnica de cumplimiento empleada por el legislador europeo, que tiene que ver con la aplicación de normas armonizadas. La idea de la que parte el Reglamento es que los poderes públicos fijan un régimen normativo general con unos requisitos esenciales (que a continuación se explicarán), pero con un papel importante para las especificaciones técnicas, que servirán para cumplimentar y concretar los requisitos, y que serán fijadas por los organismos europeos de

[1267] González-Meneses García-Valdecasas, M., *op cit*. N. 1265, p. 215. Cabe destacar en este sentido un imperativo adicional aplicable a las Administraciones, previsto en el artículo 27 RIA, que obliga a las Administraciones que sean responsables de despliegue de sistemas de alto riesgo a realizar una evaluación del impacto en los derechos fundamentales.

[1268] *Idem.* Quedan fuera, no obstante, y de forma expresa, los sistemas cuya información de salida no esté disponible para las personas físicas, tales como herramientas dirigidas a la organización, optimización o estructuración de campañas.

normalización (es habitual hablar de este método como el *Nuevo Enfoque)* [1269]. Los instrumentos que, en el futuro próximo, especificarán el contenido de las obligaciones esenciales, y a los que habrá que acudir para su mejor interpretación serán, por un lado, las *normas armonizadas*, normas técnicas elaboradas por organismos europeos de normalización, a petición y bajo supervisión de la Comisión [1270], y por otro las *especificaciones comunes,* que son documentos técnicos emitidos por la propia Comisión cuando no existen normas armonizadas aplicables o las que existen son insuficientes. Igualmente son normas voluntarias, pero también gozan de presunción de conformidad, ejerciendo la misma presión fáctica sobre los operadores económicos.

Sin embargo, pese a que el Reglamento fija este modelo, aún no se han desarrollado estos instrumentos, y su estado es especialmente incipiente en el ámbito de la UE. Existen normas relativas a la seguridad, la confianza y la ética, con mucho mayor alcance a nivel internacional [1271], pero habrá que espe-

[1269] El modelo se regula en los arts. 40 (normas armonizadas) y 41 del RIA (especificaciones comunes). El enfoque legislativo no es, en absoluto, novedoso, sino que es el enfoque que la UE viene adoptando para la comercialización de productos, combinando una serie de requisitos esenciales obligatorios, una regulación de especificaciones técnicas y, finalmente, mecanismos de evaluación de conformidad. Para un estudio más detallado sobre el modelo: Álvarez García V./ Tahiri Moreno, J., «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque», *Revista General de Derecho Administrativo*, 63, 2023, pp. 1-56.

[1270] Aunque son voluntarias jurídicamente, despliegan un efecto jurídico fuerte, pues se presume la conformidad con los requisitos esenciales del Reglamento si se respetan. Señala Álvarez García, empero, que su transparencia y accesibilidad (y, de alguna forma, por tanto, su legitimidad) es escasa, pues su contenido no se publica en el Diario Oficial de la UE, sino solo su referencia. Álvarez García, V., «La aplicación de las normas armonizadas y de las especificaciones comunes en el ámbito de la inteligencia artificial (artículos 40 y 41 Reglamento)», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 409 y ss. Estas normas armonizadas están reguladas en el Reglamento sobre la normalización europea de 2012, aunque el autor denuncia que su desarrollo es escaso y que, aunque hasta el momento los organismos de normalización han conseguido con éxito plasmar los desarrollos científicos y tecnológicos en un tiempo razonable, esto plantea problemas importantes en el mundo de las tecnologías en general y muy en particular en el de la inteligencia artificial. *Ibid.,* p. 413. También crítico con el escaso poder de control que, de facto, permite el sistema del RIA, Alamillo Domingo, I., «Sujetos y agentes en evaluaciones de conformidad», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 492, sugiriendo que en realidad se ha implantado un modelo de ausencia de control. Para mayor detalle de los desafíos de armonización y concreción de las normas en materia de IA: McFadden, M./ Jones, K./Taylor, E./Osborn, G., «Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation», *Oxford Information Labs*, 2021**.** Una crítica al recurso a organismos privados para la elaboración de estas normas, en lengua inglesa, en: Veale, M./ Zuiderveen Borgesius, F., *op. cit*. N. 1233, p. 105-106, señalando, principalmente, la falta de poder negociador de las organizaciones de consumidores con recursos insuficientes y la necesidad de un escrutinio judicial sobre las normas, que van a estar cargadas de valores que afectan directamente a derechos fundamentales.

[1271] Señala Álvarez García que, por el momento, no se han podido dictar propiamente normas armonizadas específicas para su desarrollo, si bien es cierto que tanto el ETSI como CEN/CENELEC ya han empezado a trabajar sobre esta cuestión desde hace meses, como se demuestra con la consulta de sus agendas normalizadoras, señalando que en el nivel internacional los trabajos más importantes son del

rar aún para observar cómo se van desarrollando estos estándares en el marco europeo.

Dentro de este Nuevo Enfoque regulatorio, el tercero de los pilares, además del cumplimiento de los requisitos esenciales y las especificaciones técnicas, lo constituye el procedimiento de evaluación de conformidad, regulado en el artículo 43, que consiste, básicamente, en un proceso mediante el cual se verifica que un sistema cumple con los requisitos técnicos exigidos por el RIA. El Reglamento contempla dos tipos de evaluación, en función del riesgo asociado al sistema de IA. Así, como primera opción, la autoevaluación (regulada en el Anexo VI), que consiste en que el proveedor verifique internamente que el sistema de gestión de calidad y la documentación técnica cumplen con las previsiones del RIA, aplicable a todos los casos de riesgo derivado de la finalidad, es decir, a los casos de alto riesgo contemplados en el Anexo III (salvo los casos de identificación biométrica); y, como segunda opción, la evaluación por un organismo notificado (contemplados en el Anexo VII), donde un tercero independiente revisa la documentación técnica y el sistema de calidad. Este último método de evaluación externa es obligatorio en ciertos casos, esencialmente cuando se han aplicado normas armonizadas [1272]. Aunque, en principio, es este un mecanismo para asegurar el cumplimiento de los requisitos, directamente dirigido a evitar fallas de seguridad y, por tanto, una protección de los derechos en juego en los sistemas de alto riesgo, lo cierto es que, como destaca Palma Ortigosa, la regla general del RIA acaba siendo la autoevaluación, lo que, amén de demostrar la confianza del legislador en los operadores de la industria, plantea serias dudas sobre su eficacia [1273].

La evaluación de conformidad, cuando deba realizarse por un tercero, corresponderá a los organismos notificados (cuyo régimen está regulado en el ar-

---

conglomerado ISO/IEC43, que cuenta con el *Joint Technical Committee 1* (ISO/IEC JTC 1), destacando dentro de él, a nuestros efectos, el Subcomité *SC 42 on Artificial Intelligence*. Alvárez García, V., *ibid.* p. 415-416. Para consultar las normas técnicas en concreto, nota el pie 44.

[1272] Si el riesgo deriva de que el sistema es parte o componente de un producto ya regulado (Anexo I), entre los que se encuentran, entre otros, los juguetes o productos sanitarios, en principio se aplicará la normativa armonizada de ese sector específico, integrada con los requisitos propios del RIA. No obstante, la norma no señala, por el momento, cómo integrar ambos procedimientos. La idea perseguida (considerando 124) es que el proveedor no deba realizar dos evaluaciones, sino solo la prevista en la legislación aplicable al producto de que se trate, para evitar cargas innecesarias.

[1273] Palma Ortigosa, A., «La evaluación de la conformidad en el diseño y producción de sistemas basados en inteligencia artificial en el contexto del "Nuevo Marco Legislativo"», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 445. No obstante, reconoce también el autor que se otorga un peso importante a los Estados Miembros y la Comisión para asegurar el cumplimiento efectivo, en el caso de los primeros en cuanto a la designación de las autoridades de vigilancia de mercado, y en el caso de la segunda por sus amplios poderes para revisar la autoevaluación exigible actualmente a los sistemas de alto riesgo (art. 43.6).

tículo 31 del RIA). Se trata de organismos constituidos conforme al derecho nacional de los Estados miembros con personalidad jurídica propia, garantizándose su independencia, la ausencia de conflicto de intereses, su imparcialidad, objetividad y confidencialidad [1274]. Su función, en cuanto supervisores del procedimiento de evaluación de conformidad, es verificar que el sistema de IA cumple con los requisitos esenciales (principalmente, los establecidos en la sección 2 del capítulo III, analizados *infra*), aunque evitando, por mandato expreso del Reglamento, cargas innecesarias para los proveedores habida cuenta del tamaño y complejidad de la empresa. Si la evaluación es satisfactoria, será el organismo notificado el encargado de expedir un certificado de conformidad [1275].

### c) El núcleo de las obligaciones de los proveedores e implantadores en sistemas de IA de alto riesgo (Sección 2 del capítulo III)

El punto de partida de esta regulación es que se sigue un enfoque basado en el riesgo, imponiendo cargas solo cuando es probable que un sistema cause riesgos elevados para derechos fundamentales o para la seguridad, de forma que los requisitos establecidos son estrictamente necesarios para mitigar estos riesgos, sin que lleguen a estar cubiertos por otros marcos jurídicos previos [1276].

---

[1274] Cabe resaltar, finalmente, el papel de las autoridades notificantes (reguladas en artículo 30 RIA), que son las designadas por cada Estado Miembro para acreditar organismos de evaluación de conformidad, encargándose de notificar su identidad a las instituciones europeas y otros EEMM, siendo, en esencia, los organismos encargados de supervisar a los organismos notificados. En España, será la Secretaría de Estado de Digitalización e Inteligencia Artificial quien asuma el papel de organismo notificante, mientras que la Agencia Española de Supervisión de la IA (AESIA) se creó con la intención de ejercer de organismo notificado (a pesar de esta primigenia intención, aún no ha sido formalmente designada como autoridad notificante u organismo notificado), no sin críticas, principalmente motivadas por su falta de imparcialidad. En este último sentido: Hernández Peña, J. C., «La gobernanza y vigilancia del Reglamento de inteligencia artificial: autoridades de vigilancia del mercado, Comisión y las diversas entidades» en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 858 y nota al pie 45, tildándola de deficitaria en cuanto al número de miembros con experticia técnica y una mejorable independencia, por estar demasiado vinculada con altos cargos de la Administración General del Estado (siendo el Presidente del Consejo Rector el titular de la Secretaría de Estado de Digitalización e IA).

[1275] Previsto en el Anexo VII del RIA, válido en toda la UE por un periodo de 4 o 5 años según el tipo de sistema, prorrogable tras nuevas evaluaciones.

[1276] Domínguez Álvarez, J. L., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 236. En este sentido, ya el propio artículo 8, a modo de preludio de la sección, establece que, para los casos en que rijan las obligaciones en materia de productos bajo legislación armonizada (Anexo I), los proveedores pueden optar por integrar todos los requisitos del RIA en el procedimiento aplicable en función de la legislación armonizada. El término usado, de vaguedad evidente, «según proceda», deja, sin embargo, mucho lugar a la duda, sin que se haya detallado, por el momento, como se ha comentado *supra,* cómo integrar el cumplimiento simultáneo de sendos ordenamientos.

1. *Implantar un sistema de gestión de riesgos (art. 9)*

El RIA prevé la obligación de implementar un sistema de riesgos como forma de garantizar la confiablidad, seguridad y respeto a los derechos fundamentales [1277], pues, tal y como especifica el apartado 2.a), se tendrán que determinar los riesgos conocidos y previsibles que el sistema pueda plantear para la salud, seguridad o derechos fundamentales, aun cuando el sistema se utilice, y precisamente cuando así sea, según su finalidad prevista (no obstante, también se deben prever y evaluar los riesgos derivados de usos no permitidos pero razonablemente previsibles). Dicho sistema debe configurarse como un proceso continuo e iterativo, abarcando todo el ciclo de vida del sistema, incluyendo su vigilancia post-comercialización. En cuanto al concreto modo de dar cumplimiento a la obligación, las opciones son abiertas, pues estamos ante un campo en constante evolución donde no existe un único modelo de gestión ideal [1278], pero sí debe atenderse al estado actual de la técnica (así, ya en artículo 8.1). En todo caso, cabe resaltar, como acertadamente señala Domínguez Álvarez [1279], que los riesgos a los que se refiere el artículo 9 son solo aquellos que pueden mitigarse o eliminarse razonablemente mediante el desarrollo o el diseño del sistema o bien mediante el suministro de información adecuada, siendo que el riesgo residual habrá de considerarse, a nuestro entender, precisamente el riesgo permitido (tanto el riesgo residual general del sistema como los riesgos residuales particulares asociados a cada uno de los peligros detectados).

El sistema de gestión de riesgos propuesto por el Reglamento no es, una vez más, una invención propia de esta norma, pues el recurso a sistemas de gestión del riesgo viene siendo habitual en diferentes contextos. Así, Simón Castellano define a estos sistemas como un *«conjunto de procesos, políticas, procedimientos y herramientas diseñadas para identificar, evaluar, mitigar y monitorear los riesgos que enfrenta una organización en el logro de sus*

[1277] El Considerando 65 explicita: *«Este proceso debe garantizar que el proveedor determine los riesgos o efectos negativos y aplique medidas de mitigación de los riesgos conocidos y razonablemente previsibles de los sistemas de IA para la salud, la seguridad y los derechos fundamentales, habida cuenta de su finalidad prevista y de su uso indebido razonablemente previsible, incluidos los posibles riesgos derivados de la interacción entre el sistema de IA y el entorno en el que opera. El sistema de gestión de riesgos debe adoptar las medidas de gestión de riesgos más adecuadas a la luz del estado actual de la técnica en materia de IA»*.

[1278] Simón Castellano, P., *op. cit*. N.1180, p. 548 y ss. propone fijarse en modelos de gestión previos, poniendo el ejemplo de la norma ISO 42001, así como el modelo desarrollado por la OCDE: (*High-level AI risk-management interoperability framework*).

[1279] Domínguez Álvarez, J. L., *op. cit*. N. 1276, p. 259.

*objetivos»* [1280]. Por tanto, este requisito debe enmarcarse, para su mejor entendimiento, dentro de la lógica de la gobernanza transversal, donde se combinan tanto los atributos técnicos como los principios éticos y normativos, dotando a estos últimos de verdadera operatividad. Podemos entender que la exigencia de un sistema de gestión de riesgos no es una obligación meramente formal, sino que se instituye en instrumento esencial para la construcción de una IA confiable, responsable y jurídicamente conforme. Solo mediante una integración efectiva de diseño técnico, conocimiento sectorial y cultura organizativa de cumplimiento y gobernanza podrá darse efectivo cumplimiento al mandato, que se instituye como precedente fundamental, condición *sine qua non* para la garantía de los derechos que se tratan de proteger.

Cabe matizar, llegados a este punto, que no debe confundirse el sistema de riesgo explicado en el presente epígrafe, con una obligación similar, contemplada en el artículo 27 del RIA, pero con un régimen mucho más específico. Nos referimos a la evaluación de impacto sobre derechos fundamentales (en adelante, por sus siglas en inglés, FRIA [1281]). Mientras que el análisis de riesgos general del artículo 9, como se acaba de exponer, es obligatorio para el proveedor de todos los sistemas de IA de alto riesgo, las FRIA son evaluaciones adicionales, orientadas al caso de uso concreto, vinculantes para el responsable de despliegue (no el proveedor), y solo exigibles en tres casos: cuando se trate de organismos de derecho público (sea cual sea el sistema de IA de alto riesgo que utilicen), operadores privados que presten servicios públicos, y, finalmente, entidades públicas o privadas que usen sistemas para evaluar solvencia crediticia y clasificación o priorización de llamadas de emergencia [1282].

---

[1280] Simón Castellano, P., *op. cit*. N.1180, p. 535. No en vano, como acertadamente hace Simón Castellano, uno de los ejemplos más claros, con abundante desarrollo en los últimos tiempos, son los sistemas de cumplimiento o *compliance* de las personas jurídicas. Por todos, a título de ejemplo: Nieto Martín, A., «Responsabilidad penal de la persona jurídica y programas de cumplimiento: de la gestión de la legalidad a la legitimidad», *La Ley compliance penal*, 4, 2021, hablando del sentido de los programas de cumplimiento a partir de p. 6.

[1281] *Fundamental Rights Impact Assessments*. Chaveli Donet explica que es una práctica importada de la tradición anglosajona, y que responde a nuestra situación en una sociedad con gran número de riesgos, donde la gestión de los mismos se ha vuelto esencial en muchos sectores, tales como el financiero, el asegurador, el penal, la protección de datos y, finalmente, la IA. Chaveli Donet, E., «La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 497. Para mayor profundización. Simón Castellano, P., *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi, Cizur Menor, 2023. En lengua inglesa: Mantelero, A., «The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template», *Computer Law & Security Review,* 54, 2024, pp. 1-18.

[1282] Las FRIA pueden realizarse conjuntamente con la evaluación de impacto en protección de datos del RGPD, si esta ya cubre parte de las obligaciones.

Hemos de poner de manifiesto la distinción entre estos dos mecanismos cercanos [1283], pues, mientras la evaluación de impacto, centrada en la afectación a colectivos y derechos, tiene una perspectiva sustantiva, el sistema de gestión de riesgos, desde un enfoque más organizacional, sigue la lógica de la prevención, el control y el seguimiento. No por ello, añadimos, consideramos que esta obligación peque de formal, sino que, habida cuenta de su centralidad en la localización de riesgos en muchos casos difícilmente detectables, es condición necesaria para la protección final de los mismos. No en vano, en fase de mitigación de riesgos, monitoreo y revisión, el objetivo es desarrollar e implementar estrategias y medidas para reducir la probabilidad de concurrencia o impacto de los riesgos [1284].

### 2. *Impulsar prácticas de gobernanza y gestión de datos adecuadas (art. 10)*

El artículo 10 del RIA constituye uno de los pilares normativos del régimen, pues establece de forma exhaustiva los requisitos sobre la calidad de los datos y su gobernanza. Se destaca, sin embargo, que la definición de gobernanza de datos resulta difícilmente aprehensible [1285], pues no cuenta con una definición unívoca ni en el propio Reglamento ni entre la doctrina, si bien ha ido adquiriendo relevancia como categoría transversal en la economía digital, evolucionando desde una perspectiva de gobernanza intra-organizacional (micro), a una perspectiva de noción estratégica más amplia [1286]. Serna identifica al menos tres elementos nucleares en este proceso de gobernanza del dato: su valorización como un activo de la organización; la asignación de responsabilidades sobre la toma de decisiones relativa al mismo; y la definición de normas para asegurar la calidad de estos, así como su uso ético [1287]. Loza es partidaria de interpretar este artículo no de manera aislada, sino dentro de una lógica más amplia (macro) que se enmarca en la Estrate-

[1283] En este sentido, en mayor desarrollo: Chaveli Donet, E., *op. cit*. N. 1281, pp. 498 y ss.

[1284] Igual en Simón Castellano, P., *op. cit*. N.1180. p. 545.

[1285] Así, Loza Corera, M., *op. cit*. N. 1180, p. 567.

[1286] La AEPD la define como «la estrategia para la correcta administración y gestión de la política de datos en la organización». AEPD, «Gobernanza y política de protección de datos», 2020, disponible en: <https://www.aepd.es/prensa-y-comunicacion/blog/gobernanza-y-politica-de-proteccion-de-datos> [Accedido por última vez el 30/07/2025].

[1287] Salvador Serna, M., «Inteligencia artificial y gobernanza de datos en las administraciones públicas: relaciones y evidencias para su desarrollo», *Gestión y Análisis de Políticas Pública*, 26, julio 2021, p. 24.

gia Europea de Datos, donde se ha asentado la idea definitiva del dato como activo económico y estratégico [1288].

Ateniéndonos, en cambio, al contenido estricto del artículo 10, se trata de garantizar que los datos usados para el entrenamiento, validación y prueba [1289] de los sistemas de IA sean pertinentes, representativos, libres de errores y completos [1290] (art. 10.3).

Algunos aspectos de la regulación son criticados por Loza, comenzando por la indeterminación sobre quiénes son los sujetos obligados a asumir esta gobernanza de datos en cada una de las distintas fases del ciclo de vida [1291]. Tampoco se alcanza a comprender de qué manera se integrará esta obligación con el resto de las previstas en la sección 2, pues, habida cuenta de la presencia mayoritaria de una autoevaluación como proceso de evaluación de la conformidad con el Reglamento, la trazabilidad del cumplimiento normativo se debilita. Asimismo, critica la autora que se añada en el apartado 4 del artículo 10 una presunción altamente controvertida, según la cual se entenderá que cumplen con todos los requisitos del artículo 10 aquellos sistemas entrenados con datos relativos al entorno geográfico, conductual, contextual y funcional específico. Se trata, a su juicio, de una presunción problemática [1292], en tanto reduce el escrutinio sustantivo sobre la calidad de los datos, sin llegar a exigir una verificación sustantiva, amenazando con convertir esta exigencia en una meramente formal.

---

[1288] Así, la autora señala la necesidad de interpretar el artículo 10 conforme a la normativa íntimamente relacionada, de protección de datos.

[1289] De manera simplificada, dentro del procedimiento de procesamiento de datos podemos distinguir una primera fase de recolección, donde habrá que definir una población para tomar muestras de la misma, identificando, en su caso, cuáles son las características o marcadores que se habrán de tener en cuenta, para dividir ese primer conjunto de datos en dos grupos: los de entrenamiento y los de testeo. El modelo es definido y optimizado con los datos de entrenamiento para, finalmente, realizar las pruebas y validaciones sobre la muestra de datos de testeo. Entre otros: Suresh, H., /Guttag, J.: «A framework for understanding sources of harm throughout the machine learning life cycle», *Equity and access in algorithms, mechanisms, and optimization*, (EAAMO '21), October. 5-9-2021, pp. 2-3.

[1290] No obstante, se añade el inciso, en ese mismo apartado, de que lo serán *«en la mayor medida posible»,* debilitando sobremanera la obligatoriedad del precepto. Señala Loza como en el proceso de negociación Noruega ponía de manifiesto que exigir que los datos fueran completos, sin matices, resultaba utópico. Loza Corera, M., *op. cit*. N. 1180, p. 582.

[1291] Esto es especialmente importante habida cuenta, como señala la autora de que, en muchas ocasiones, debido a los métodos de aprendizaje, serán terceros, ya sean responsables de despliegue o contratistas, quienes estén implicados en procesos de reentrenamiento del sistema (así, en considerando 88). Así lo advirtió en su momento el Supervisor Europeo de Protección de Datos. Loza Corera, M., *ibid.* p. 580 y nota el pie 65, señalando que partiendo de una interpretación conforme del RIA habrá que entender a estas personas como proveedores. Señala cómo otros autores (nota al pie 67) han puesto de manifiesto que se han ignorado otras etapas del *machine learning* que también deberían estar sujetas a criterios de calidad de datos y prácticas de gobernanza.

[1292] *Ibid.,* p.585.

Insiste la citada autora en la idea de que, a tenor de los considerandos 67 y 69, es necesario compatibilizar los postulados del artículo 10 con el cumplimiento del RGPD, sugiriendo que, dado que la evaluación de impacto de derechos fundamentales del RIA es, a todos ojos, escasa, se aplique la evaluación de impacto en protección de datos prevista en el artículo 35 RGPD [1293]. Sin embargo, y aunque es evidente la preocupación que, en materia de privacidad, intimidad, o protección de datos, suscita el tratamiento de grandes cantidades de datos inherente a esta tecnología en cualquiera de sus aplicaciones, el artículo 10 supone un avance por la protección de la que puede dotar a otro de los derechos fundamentales transversalmente en peligro con cualquier sistema de IA: el derecho a no sufrir discriminación.

Así, la problemática del sesgo algorítmico puede ver, precisamente en el artículo 10, una valiosa herramienta de protección *ex ante*. Ya hemos puesto de manifiesto en otros trabajos [1294], así como en el capítulo VI de la presente investigación, que una de las mayores preocupaciones cuando se trata de herramientas de IA es su capacidad para generar resultados sesgados [1295] hacia colectivos típicamente discriminados, con la dificultad añadida, derivada de la

[1293] La evaluación de impacto en protección de datos prevista en dicho artículo es un procedimiento obligatorio para los responsables del tratamiento de datos personales cuando es probable que un tipo de tratamiento, especialmente si emplea nuevas tecnologías, por su naturaleza, alcance, contexto o fines, implique un alto riesgo para los derechos y libertades de las personas físicas. Esta evaluación debe realizarse antes de iniciar el tratamiento. En todo caso, y como acertadamente señala López Calvo, serán de plena aplicación los principios previstos en el artículo 5 RGPD: limitación de la finalidad, exactitud, limitación del plazo de conservación y principio de integridad y confidencialidad. López Calvo, J., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 266 y ss. Si bien explica el autor que el RIA introduce algunos matices relevantes. El matiz más importante es explicado por Loza Corera, M., *op. cit*. N. 1180, p. 578: *«En el párrafo quinto se establece la posibilidad para los proveedores de tratar categorías especiales de datos «en la medida en que sea estrictamente necesario para garantizar la detección y corrección de los sesgos». Indicar que el Parlamento denominaba a dichos sesgos negativos y los definía como aquellos que «crea(n) un efecto discriminatorio directo o indirecto contra una persona física», pero finalmente, el concepto de «sesgo negativo» no se recogió en la versión final. Se establece la posibilidad de establecer una base de legitimación adecuada para poder tratar categorías especiales de datos que, en aplicación de la normativa de protección de datos, no eximirá de la obligación de adoptar las salvaguardias adecuadas para los derechos y libertades de las personas físicas.»*

[1294] Bravo Bolado, A. «Justicia algorítmica: un enfoque sociotécnico», *op. cit. N. 221.*

[1295] Paradigmáticamente, entre nosotros: Soriano Arnaiz, A., «La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de Inteligencia Artificial», en Bonorino Ramírez, P. R./ Fernández Acevedo, R./ Valcárcel Fernández, P. (Dirs.), *Nuevas normatividades. Inteligencia artificial, derecho y género,* Thomson Reuters Aranzadi, Cizur Menor, 2021, pp. 63-88. Concretamente en p. 81 propone la autora como uno de los mecanismos apropiados para garantizar la detección de la discriminación algorítmica que los controladores proporcionen información sobre el sistema, incluyendo datos de entrenamiento, datos de prueba y cómo se definen las etiquetas y la variable objetivo, para que así se pueda experimentar con el sistema *«para determinar los efectos que la pertenencia a un grupo desventajado tiene en los resultados, estableciendo así comparadores hipotéticos mediante la experimentación con el sistema».*

opacidad intrínseca de las técnicas, de detectar su origen. Resulta esencial, pues, una regulación que ponga el foco en el dato como medio imprescindible para atajar los resultados discriminatorios.

Así lo establece el artículo 10 (apartado 2.f), cuando exige que se realice un examen de los datos *«atendiendo a posibles sesgos que puedan afectar a la salud y la seguridad de las personas, afectar negativamente a los derechos fundamentales o dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión, especialmente cuando las salidas de datos influyan en las informaciones de entrada de futuras operaciones».* En este sentido, acogemos la clasificación y ejemplos ofrecidos por López Calvo [1296], distinguiendo entre sesgos estadísticos (ejemplo: IA que realiza un análisis de incidencia de una enfermedad basándose principalmente en datos de población menor de 50 años); sesgos motivados por el suministro de información incorrecta (ejemplo: como consecuencia de las alucinaciones de ciertas IAs, un abogado presenta una demanda con múltiples referencias a sentencias completamente falsas); sesgos sociales (ejemplo: IA usada en época COVD que no consideraba la especial magnitud de la enfermedad en zonas de menor renta); o sesgos de género (ejemplo: IA usada en el contexto médico que no tiene en cuenta los diferentes síntomas que sufren mujeres y hombres ante un ataque al corazón) [1297]. Las consecuencias de estos errores en los datos, son, a todos ojos, lesivas no solo del derecho a no verse discriminado, llegando a repercutir directamente sobre bienes finales de especial relevancia.

A nuestro juicio, que el artículo 10 fije obligaciones respecto de los datos a lo largo del proceso de generación de los sistemas viene a responder a necesidades puestas de manifiesto ya por la doctrina, relativas a que, dado que la IA trabaja por medio de la estandarización de acciones y del aprendizaje automático a partir de datos, hay que controlar las fases donde estos sesgos surgen, pues después se replicarán y pueden acabar arrojando un resultado discriminatorio para individuos y colectivos [1298]. Aunque no

[1296] López Calvo, J., *op. cit*. N. 1293, p. 268-269.

[1297] Las clasificaciones por tipologías y grupos de sesgos son, en este punto, incontables. Baste dejar apuntada, a título de ejemplo, la explicada en Bravo Bolado, A. «Justicia algorítmica: un enfoque sociotécnico», *op. cit. N. 221*., p. 6 y ss. Así, dentro de la fase de recolección y procesamiento de datos: sesgo histórico o preexistente, sesgo en los atributos, sesgo en la etiqueta, sesgo en la muestra o de representación; dentro de la fase de entrenamiento del modelo el sesgo de aprendizaje; y, en fase de evaluación del modelo, el sesgo de evaluación.

[1298] Belloso Martín, N., «La problemática de los sesgos algorítmicos (con especial referencia a los de género)» en Llano Alonso, F. H., (Dir.), *Inteligencia artificial y Filosofía del derecho,* Laborum, Murcia, 2022, p. 46.

contamos con una definición jurídica de «sesgo» (y, dicho sea de paso, el RIA ha perdido una magnífica oportunidad para suplir esta carencia), podemos diferenciar una vertiente humana y una vertiente algorítmica. Así, la humana tiene que ver con *«ideas, predilecciones o prejuicios inconscientes, que se activan la mayoría de las veces de forma automática, porque nuestro cerebro funciona a través de la minimización del esfuerzo cognitivo y los estereotipos permiten tomar decisiones de forma más rápida»* [1299]. Sin embargo, estos sesgos se reproducen también en los algoritmos basados en datos, pudiendo localizar hasta tres fases: en la recolección de datos (porque los datos reflejen sesgos previamente existentes; lo que ha venido a denominarse sesgo histórico); en la fase de preparación de los datos de entrenamiento (es decir, al seleccionar y procesar los atributos y características con las que trabajará el algoritmo); y, finalmente, en el proceso de toma de decisiones o implantación del modelo (donde en función del contexto podemos identificar, por ejemplo, sesgos relativos a la interacción del humano con los resultados del algoritmo, como el sesgo de confirmación) [1300]. Por tanto, cuando hablamos de sesgos o resultados sesgados y discriminatorios, no se trata tanto de que el sistema sea defectuoso como de que el diseño del programa no sea correcto, o bien la selección y tratamiento de datos haya sido incompleta. En definitiva, acogiendo las acertadas palabras de Belloso, *«(e)l problema no son los prejuicios, sino los datos»*. [1301] Los sistemas de IA no discriminan, pero un mal tratamiento de datos puede conducir a un tratamiento discriminatorio que el Reglamento pretende, a toda costa, evitar. Por ello resulta central una obligación

---

[1299] Ramírez, H., «Inteligencia Artificial e igualdad de género», en *Grupo Ático 34*, Madrid (24.11.2021), disponible en: <https://protecciondatos-lopd.com/empresas/inteligencia-artificial-igualdad-genero/> [Accedido por última vez el 29/01/2026]. También, sobre los sesgos cognitivos en el entorno judicial: Muñoz Aranguren, A.*, op. cit*. N. 864 y Alonso Gallo, J., «Errores y sesgos cognitivos en la expansión del Derecho penal», en AA. VV., *Derecho y justicia penal en el siglo XXI: liber amicorum en homenaje al profesor Antonio González-Cuéllar García*, Colex, 2006, pp. 31-48. Según una definición puramente gramatical de la RAE: «*Error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras»,* aunque alude, en este caso, a un significado estadístico.

[1300] Belloso Martín, N., *op. cit*., N. 1298, p. 48. El RIA define, en su artículo 3 (apartados 29-33), lo que debe entenderse por: datos de entrenamiento (usados para entrenar a un sistema de IA mediante el ajuste de sus parámetros entrenables); datos de validación (usados para proporcionar una evaluación del sistema de IA entrenado y adaptar sus parámetros no entrenables y su proceso de aprendizaje para, entre otras cosas, evitar el subajuste o el sobreajuste); conjunto de datos de validación (conjunto de datos independiente o una parte del conjunto de datos de entrenamiento, obtenida mediante una división fija o variable); datos de prueba (datos usados para proporcionar una evaluación independiente del sistema de IA, con el fin de confirmar el funcionamiento previsto de dicho sistema antes de su introducción en el mercado o su puesta en servicio); y datos de entrada (datos proporcionados a un sistema de IA u obtenidos directamente por él a partir de los cuales produce un resultado de salida).

[1301] *Ibid.,* p. 49.

como la impuesta en el artículo 10 pues, de su correcta interpretación y aplicación podremos obtener resultados de salida mucho más justos [1302].

### 3. *Elaborar la documentación técnica (art. 11)*

La documentación técnica, aunque no aparece definida expresamente en el RIA, constituye un componente esencial para comprender el funcionamiento de los sistemas, tanto sus capacidades como sus limitaciones, por lo que puede ser considerada como un componente fundamental para garantizar la seguridad y los estándares éticos con los que ha de contar un sistema de alto riesgo [1303]. Dicha documentación debe elaborarse antes de la introducción del sistema en el mercado, y debe ser actualizada periódicamente. El objetivo es, como determina el propio artículo 11.1, demostrar de manera real y efectiva que el sistema cumple con los requisitos de la presente sección [1304].

El Anexo IV viene a completar el contenido de esta documentación, que, en esencia, comprende: la información general del sistema (aspectos básicos, propósito, funcionamiento y actores involucrados); la documentación estrictamente técnica (incluyendo información detallada sobre los datos utilizados, los algoritmos usados, método de aprendizaje, arquitectura, métodos de entrenamiento y parámetros, así como las métricas de evaluación del sistema, y pruebas de validación); información detallada sobre la supervisión y control del sistema (en los términos exigidos por el artículo 14, explicado i*nfra*), una

[1302] En este sentido, el denominado como «*algorithmic fairness*» constituye uno de los puntos de estudio más provechosos para la disminución de los resultados discriminatorios de los sistemas de IA, donde se pone de manifiesto la necesaria carga ético/filosófica que conlleva la selección de parámetros en la configuración de los algoritmos, pues conceptos como lo justo o lo equitativo resultan imposibles de parametrizar. Selbst, A. D/ Boyd, D./ Friedler, S. A./ Venkatasubramanian, S/Vertesi, J., «Fairness and abstraction in sociotechnical systems», *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 59-68; Binns, R., «Fairness in Machine Learning: Lessons from Political Philosophy», *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 2018, pp. 149-159.

[1303] Puyol Montero, J., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 275.

[1304] El artículo introduce dos salvaguardias que relajan el cumplimiento de los requisitos en dos supuestos: cuando se trate de sistemas de alto riesgo asociados a productos dentro del ámbito de aplicación de los actos de armonización (Anexo I), se elaborará solo un conjunto de documentos técnicos para cumplir con sendas legislaciones. En segundo lugar, cuando se trate de pymes o empresas emergentes, la documentación exigible será simplificada, proporcionando la Comisión un formulario específico simplificado. En este sentido, en España se viene fomentando este impulso de las pequeñas empresas por medio de la Ley 29/2022, de 21 de diciembre, de fomento del ecosistema de las empresas emergentes y la Ley 18/2022, de 28 de septiembre, de creación y crecimiento de empresas.

descripción de los parámetros de rendimiento, y una descripción del sistema de gestión de riesgos (exigido por el artículo 9) [1305].

En definitiva, el objeto de la documentación técnica es permitir a las autoridades de supervisión realizar una labor de control y fiscalización sobre el cumplimiento de las obligaciones del Reglamento, por lo que podríamos afirmar que se trata de una medida que, más que introducir una protección directa o sustantiva de los derechos en juego, busca garantizar que el control sea efectivo. Lo mismo cabe predicar de la obligación, que viene a reforzar la ya citada, de conservar esta documentación durante un determinado periodo de tiempo [1306].

### 4. *Conservar los registros de acontecimientos del ciclo de vida del sistema (art. 12)*

La obligación de conservar los registros constituye un elemento esencial para garantizar la trazabilidad del funcionamiento de los sistemas a lo largo de todo su ciclo de vida, permitiendo detectar situaciones de riesgo, facilitar su control tras la comercialización y posibilitar actuaciones de control o inspección ulteriores. Como destacan Arellano Toledo y Merchán Murillo [1307], la utilidad probatoria de los *logs* o registros radica en su capacidad de demostrar que los sistemas no han sido manipulados por agentes no autorizados, reforzando, así, la integridad de la información y permitiendo la verificación del cumplimiento normativo.

Desde un punto de vista técnico, los registros o *logs* son archivos que almacenan de forma detallada los eventos generados por el núcleo *(kernel)* del sistema, incluyendo marcas temporales precisas y otros datos relevantes sobre el comportamiento del sistema [1308]. Se trata, en definitiva, de una herramienta

---

[1305] Se habrá de incluir, también, dentro de la documentación técnica, una copia de la declaración de conformidad que exige el artículo 47, así como hacer constar el plan de vigilancia post-comercialziación y localizar, en su caso, el listado de normas armonizadas aplicadas y de los posibles cambios que el proveedor haya realizado sobre el sistema a lo largo de su ciclo de vida (Anexo IV, apartados 6-9).

[1306] Hoy, el artículo 16. d) RIA recoge esta obligación, desarrollándose las condiciones (a disposición de las autoridades nacionales competentes) y periodo de conservación (10 años desde la introducción en el mercado) en el artículo 18.

[1307] Arellano Toledo, W./ Merchán Murillo, A., «La obligación de conservar registros de los sistemas de alto riesgo en el Reglamento de Inteligencia Artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 614.

[1308] *Ibid.,* p. 620. Podríamos decir, tomando el símil empleado por los autores, que es algo así como un «cuaderno de bitácora» del sistema de IA. Los registros o logs son archivos donde se almacena información generada por el núcleo del sistema y sus programas asociados, similar a una bitácora. En un sistema de inteligencia artificial de alto riesgo, estos logs guardan datos detallados, como marcas temporales

especialmente valiosa para detectar anomalías y reforzar la seguridad, o, incluso, porque puede llegar a constituir medios de prueba [1309]. El sistema de registros debe estar integrado técnicamente en la estructura del sistema de IA, de forma que esté programado para incluir entre sus funciones este método de *logging* [1310].

El propio Reglamento fundamenta la necesidad de estos registros en su considerando 71, que establece que *«Para permitir la trazabilidad* [1311] *de los sistemas de IA de alto riesgo, verificar si cumplen los requisitos previstos en el presente Reglamento, así como vigilar su funcionamiento y llevar a cabo la vigilancia poscomercialización, resulta esencial disponer de información comprensible sobre el modo en que se han desarrollado y sobre su funcionamiento durante toda su vida útil. A tal fin, es preciso llevar registros y disponer de documentación técnica que contenga la información necesaria para evaluar si el sistema de IA de que se trate cumple los requisitos pertinentes y facilitar la vigilancia poscomercialización».* De esta forma, el propio texto del Reglamento sitúa el mantenimiento de registros y de documentación técnica como medios para lograr esa trazabilidad, es decir, formas de vigilar el funcionamiento del sistema. Resultan obligados a conservar los registros tanto el proveedor como el responsable de despliegue [1312], fijándose asimismo un régimen especialmente garantista cuando se trata de los sistemas de identificación biométrica [1313].

---

precisas, que registran eventos con fecha y hora exacta. Esta información es fundamental porque certifica y prueba la conducta del sistema. En general, los archivos de registro ofrecen datos esenciales sobre las actividades internas y externas tanto de hardware como de software, facilitando la detección de anomalías, problemas y apoyando la seguridad del sistema.

[1309] *Ibid.,* p. 614.

[1310] Campuzano Gómez-Acebo, J., / Sieira Gil, J., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 285.

[1311] Sobre el principio de transparencia, central ya desde la génesis de las primeras discusiones sobre los riesgos éticos planteados por la IA, y principio fundamental en numerosas resoluciones, *vid.* inmediatamente *infra.* En español, véase, por todos: Cotino Hueso, L., «Transparencia y explicabilidad de la inteligencia artificial y «compañía» (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad…). Para qué, para quién y cuánta», en Cotino Hueso, L./ Castellanos Claramunt, J. (Coords.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022, pp. 29-79.

[1312] El artículo 19 completa las condiciones de conservación para los proveedores, estableciendo que los conservarán, en tanto estén bajo su control, durante al menos seis meses, salvo que por aplicación de legislación en materia de protección de datos quepa disponer otra cosa; mientras el régimen aplicable a los responsables de despliegue, en esencia igual, se prevé en el artículo 26.6.

[1313] En el apartado 3 del artículo 12 se establece un régimen especial para los sistemas de alto riesgo del Anexo II, punto 1, letra a) (es decir, sistemas de identificación biométrica remota), acogiendo un sistema de mínimos, estableciendo que el archivo de registros debe incluir, «como mínimo» i) un registro del período de cada uso del sistema, ii) la base de datos de referencia con la que el sistema ha cotejado los datos de entrada, iii) los datos de entrada con los que la búsqueda ha arrojado una correspondencia, y, iv) la identificación de las personas físicas involucradas en la comprobación de los resultados que se citan en

## 5. *Articular medidas de transparencia y comunicación de información (art. 13)*

Para entender la complejidad de la obligación que nos ocupa resulta esencial atender a la definición de su concepto central, que es la transparencia. En opinión de Gutiérrez David [1314], el término se usa de forma completamente polisémica y errática a lo largo del Reglamento, por mucho que sea un concepto que goza de largo recorrido en la literatura especializada en IA. Así, hay quienes describen el fenómeno como una «torre de babel conceptual» [1315], y esta idea se refuerza cuando vemos que el Reglamento utiliza de manera indistinta e inconsistente conceptos como transparencia, explicabilidad o interpretabilidad.

El considerando 27 recoge la idea de que la transparencia implica una trazabilidad y explicabilidad adecuadas, posibilitando el uso correcto del sistema por parte del responsable de despliegue [1316].

---

el artículo 14, apartado 5, que son, en esencia «*dos personas físicas con la competencia, formación y autoridad necesarias*» para re-evaluar ciertas decisiones de supervisión del responsable de despliegue.

[1314] Gutiérrez David, M. E., «Transparencia y comunicación de información a los responsables del despliegue en el artículo 13 Reglamento de inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 629, refiriéndose a la transparencia como un concepto polisémico, variable y modulable en función del dominio (técnico, ético, *soft law,* jurídico), el contexto de aplicación del sistema (sector público o privado), sus impactos (individuales o colectivos), la finalidad de la transparencia (interna o externa) y los sujetos destinarios de la misma. Destaca la autora que ni siquiera dentro del campo científico es posible identificar una taxonomía común. Ha de ponerse de manifiesto que el fenómeno de la IA explicable (XAI), más allá de explorar las implicaciones éticas, se ha convertido en un verdadero campo de investigación técnico. A título de ejemplo: Hacker, P./ Passoth, JH., «Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond», en Holzinger, A., Goebel, R., *et al.* (Eds.) *xxAI - Beyond Explainable AI,* Springer, 2022, pp. 343-376, especialmente a partir de p. 51, reflexionando sobre las implicaciones de la explicabilidad sobre la responsabilidad por el producto, cuestionando si, por ejemplo, un sistema de IA se considerará defectuoso porque un consumidor medio tendrá como expectativa que su funcionamiento sea explicable, planteando que quizá debieran adoptarse otros estándares para atender al carácter defectuoso, más relacionados, al estilo estadounidense, con el análisis riesgo-utilidad (atender a si el coste de una alternativa viable, entendemos que más explicable, aun a pesar de su disminución de utilidad, es inferior a la ganancia en seguridad), planteando el dilema de que la explicabilidad no reduce necesariamente el riesgo general. En español: Ortiz de Zárate, L., «Explicabilidad (de la inteligencia artificial)», *Eunomía. Revista en Cultura de la legalidad,* 22, 2022, pp. 328-344.

[1315] Schneeberger et al., «The Tower of Babel in Explainable Artificial Intelligence (XAI)», en Holzinger, et al. (Eds.), *Machine Learning and Knowledge Extraction.* Springer, 2023, p. 67. Las dudas respecto del macro-concepto incluyen: cuál es su alcance (uso de datos, código fuente, interacción entre ser humano e IA, decisiones automatizadas...), quiénes son los obligados y qué finalidad ha de perseguir (minimización del daño, mejora de la IA, razones jurídicas, fomento de la confianza, principio de democracia), entre otras.

[1316] Cons. 27: *«[…] Por «transparencia» se entiende que los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos. […]»* La base de la que parte este considerando es el reconocimiento de cierto papel a las Directrices éticas para una IA fiable elaboradas por el Grupo de expertos de alto nivel en 2019, donde se reflejan los siete principios éticos no vinculantes para la IA: acción

Aunque la literatura ha intentado desligar el concepto de transparencia del de interpretablidad o explicabilidad, no se puede afirmar que haya consenso y, de hecho, a menudo se acaba solapando con otras propiedades técnicas como la reproducibilidad, la replicabilidad, la verificabilidad, o incluso fiabilidad del sistema [1317]. Por tratar de encontrar un concepto de entendimiento común, se puede acoger la interpretación que propone Gutiérrez [1318], y que pone el foco en el objetivo de desarrollar modelos de IA interpretables, entendibles para el observador humano, a partir de, o bien la información que proporciona el modelo en sí mismo, o bien de técnicas complementarias (generalmente modelos algorítmicos externos) que permiten extraer explicaciones en aquellos casos en que el modelo no es interpretable por un humano, como ocurre, por antonomasia, en los modelos de caja negra [1319]. Así, «mientras la transparencia respondería a la pregunta de cómo *funciona el modelo,* la explicabilidad respondería a qué *información adicional puede extraerse del modelo* (explicaciones) cuando resulta imposible o complejo ver y entender (interpretabilidad)

---

y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas. Señala Gutiérrez David, empero, que las directrices éticas no han llegado a establecer una clara definición de estos conceptos, incurriendo en ocasiones en aproximaciones que tilda de casi-tautológicas en los conceptos de explicabilidad y transparencia, pues, «*mientras el principio ético de explicabilidad («explicability», en su versión en inglés) comprendería, entre otros aspectos, la transparencia de los procesos y la capacidad de poder explicar las decisiones a las partes afectadas; el requisito de transparencia incluiría, a su vez, la "explicabilidad" ("explainability", en su versión en inglés»*. Gutiérrez David, M. E., *op. cit*. N. 1314, nota al pie 30.

[1317] *Ibid.,* p. 646.

[1318] *Ibid.,* p. 647. Dicha interpretación se basa en un extremadamente minucioso estudio de la literatura científica.

[1319] Nos remitimos a lo explicado sobre el *black box problem supra,* concretamente en capítulo quinto, con referencias a: Bathaee, Y., *op cit*. N. 290. Resulta interesante, empero, realizar una reflexión acerca de si puede afirmarse que el Reglamento desincentive, con su regulación sobre transparencia, los modelos de caja negra, destacando la mencionada autora que, en realidad, con el fomento del dominio técnico del XAI, se corre el riesgo de abandonar los modelos interpretables por diseño por aquellos que, aunque no son interpretables en principio, tras el uso de técnicas de extracción de información relevante que generen explicaciones *ex post* (ya sean estas de tipo matemático, estadístico o en lenguaje natural), quizá se acabe por aceptar un modelo contra el que se debería luchar. Gutiérrez David, M. E., *op. cit*. N. 1314, pp. 651 y ss. También, haciendo hincapié en la relación inversa entre interpretabilidad y rendimiento: Edwards, L./ Veale, M., «Slave to the algorithm? Why a «right to an explanation» is probably not the remedy you are looking for», *Duke Law & Technology Review*, 18, 2017, pp. 18-84. Gutiérrez cuestiona, de hecho, la creencia generalizada de que deba existir necesariamente una compensación entre el rendimiento predictivo y la interpretabilidad y, por tanto, la necesidad de implementar modelos complejos de caja negra para obtener el máximo rendimiento predictivo pues, según algunas autoras, cuando se utilizan datos estructurados con atributos representativos y relevantes, pues en tales casos, no suele haber diferencias significativas en el rendimiento entre clasificadores más complejos (redes neuronales profundas, árboles de decisión potenciados, bosques aleatorios) y clasificadores mucho más sencillos (regresión logística, listas de decisión). En este último sentido: Rudin, C., «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead». *Nature Machine Intelligence*, 1, 2019, pp. 206-215.

cómo trabaja éste internamente (caja negra). A su vez, mientras que la interpretabilidad constituye el objetivo final, las explicaciones actúan como instrumentos para lograrla cuando el modelo no resulta interpretable por sí mismo» [1320]. Cabe señalar, no obstante, que las menciones a la explicabilidad en el Reglamento son absolutamente anecdóticas.

En cuanto al contenido puramente normativo del artículo 13, destaca Gutiérrez David al menos tres dimensiones del mismo: una dimensión subjetiva, que establece quién debe garantizar la transparencia -el proveedor-, y quién la recibe –el responsable de despliegue–; una dimensión formal, referida al modo de presentación de la información, que ha de ser concisa, clara, comprensible y accesible; y una dimensión material, referida al contenido mínimo que debe constar en las instrucciones de uso que el proveedor ponga a disposición del responsable de despliegue [1321].

Lo cierto es que, si bien la idea de obligar a comunicar cierta información puede parecerse, en cierto modo, al derecho a la explicación exigido por la normativa de protección de datos (arts. 13-15 del RGPD), de ninguna forma puede afirmarse, tras un examen atento del RIA, que ambas obligaciones tengan un contenido siquiera parecido, pues el artículo 86 del RIA, que reconoce un «derecho a una explicación» (limitado, ya de partida, a los sistemas de alto riesgo del Anexo III, excluyendo los del Anexo I), no prevé verdaderos mecanismos para su aplicación efectiva [1322].

Así, las distintas obligaciones de comunicación de información previstas a lo largo del Reglamento permiten distinguir entre dos modalidades de transparencia, interna y externa, respectivamente [1323]. Por un lado, una transparencia interna, de carácter marcadamente técnico, que comprende la puesta a disposición del responsable del despliegue de instrucciones de uso, la cooperación con

---

[1320] Gutiérrez David, M. E., *ibid.*, p. 657.

[1321] Parece evidente que, habida cuenta de la complejidad técnica que necesariamente conllevará el cumplimiento de estas obligaciones, se ha señalado la conveniencia de acudir a la normalización técnica internacional sobre la materia. Así: ISO/IEC DIS 12792:2024, Inteligencia artificial – Taxonomía de transparencia (ISO/IEC DIS 12792:2024); ISO/IEC 25059:2023, Ingeniería de sistemas y software– Requisitos y evaluación de calidad de sistemas y software (SQuaRE) – Modelo de calidad para sistemas de IA (ISO/IEC 25059:2023); IEEE Std 7001-2021, Norma IEEE para la transparencia de sistemas autónomos (IEEE Std 7001-2021). De hecho, Gutiérrez David, M. E., *ibid.*, señala, en p. 643, que la propia Comisión ha solicitado ya al Comité Europeo de Normalización (CEN) y al Comité Europeo de Normalización Electrotécnica (CENELEC), para que elaboren normas técnicas que concreten el alcance de algunas obligaciones del RIA, entre las que se encuentra la exigencia de una norma técnica sobre transparencia.

[1322] Así, López-Tarruella Martínez, A., «Derecho a presentar una reclamación y derecho a una explicación. Vías de recurso para los particulares en el reglamento de inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 904 y ss.

[1323] Gutiérrez David, M. E., *op. cit*. N. 1314, p. 650.

las autoridades competentes mediante el suministro de información y documentación, así como el acceso a bases de datos vinculadas a sistemas de alto riesgo; y, por otro lado, una transparencia externa, orientada al público en general, que incluye deberes de información hacia las personas físicas expuestas a sistemas de alto riesgo, así como sobre la identidad de los proveedores y responsables del despliegue con los que dichas personas puedan interactuar. En este contexto, la transparencia exigida por el artículo 13 se enmarca claramente en la primera categoría, es decir, constituye una transparencia interna y de naturaleza técnica. En cuanto a su funcionalidad, para el proveedor representa una obligación de transparencia activa cuyo propósito es garantizar el cumplimiento normativo, mientras que, para el responsable del despliegue, se traduce en una transparencia pasiva que, además de facilitar el cumplimiento regulatorio, cumple una función capacitadora, como se desarrollará en los apartados siguientes.

Puede decirse, por tanto, que, habida cuenta de la complejidad conceptual y de la falta de un sistema real que garantice los derechos del usuario final a solicitar una explicación sobre el funcionamiento concreto del sistema, podemos afirmar, con Gutiérrez, que realmente el artículo 13, aunque impone a los proveedores la obligación de garantizar *«un nivel de transparencia suficiente para que los responsables de despliegue interpreten y usen correctamente sus resultados de salida»,* concretando el contenido de las instrucciones de uso para una correcta interpretación del sistema, no puede afirmarse que el Reglamento reconozca la explicabilidad como obligación general [1324]. En definitiva, el artículo 13.1, al exigir un nivel de transparencia adecuado, genera una elevada inseguridad jurídica, pues en ausencia de una definición técnica clara, los obligados difícilmente sabrán cómo dar cumplimiento a esta genérica obligación. La situación no es mejor si atendemos a otro de los artículos del Reglamento donde se regulan obligaciones de transparencia (art. 50), donde la obligación del proveedor consiste, simplemente, en informar a las personas que interactúen directamente con sistemas de IA de que están interactuando con estos, salvo que resulte evidente, estableciéndose, a su vez, excepciones cuando se trate de sistemas para el cumplimiento de la ley [1325].

En suma, tal y como afirma Gutiérrez, estamos ante un enfoque de la transparencia técnico, cerrado y entre expertos, reduciéndose este requisito a una

[1324] De hecho, la única mención, en artículo 13.3.b), iv, exige incluir *«en su caso, las capacidades y características técnicas del sistema de IA de alto riesgo para proporcionar información pertinente para explicar sus resultados de salida».*

[1325] La misma obligación de marcado se establece para los sistemas de IA de uso general, que se explicarán *infra,* en caso de que hayan generado resultados de salida manipulados, así como contenido que constituya lo que el RIA denomina «ultrasuplantación» (art. 50.4).

exigencia instrumental vinculada al cumplimiento normativo y alejada de una concepción sustantiva orientada a la protección de derechos fundamentales [1326]. El único de los resquicios donde, en efecto, cabe entender que existe un mandato con cierta seguridad jurídica orientado a la evitación de perjuicios, sería la obligación, dirigida al responsable de despliegue, de capacitación de su personal para dar un uso adecuado al sistema de conformidad con las instrucciones.

### 6. *Fomentar medidas que permitan una supervisión humana (art. 14)*

El artículo 14 viene a plasmar en forma de obligación uno de los principios éticos que se han considerado pilares de la IA fiable desde el comienzo de las discusiones institucionales. Así, ya el Grupo de Expertos de Alto nivel de la Comisión reconocía la supervisión humana como imprescindible, siguiendo la estela del Libro Blanco sobre IA de 2020, siendo un requisito que formó parte del Reglamento desde sus primeras versiones.

Aunque, una vez más, y debido a lo amplio del concepto, resulta difícil encontrar definiciones unívocas, sí parece haber consenso en torno a su finalidad. Así, Favaretto *et al* establecen que los humanos resultan cruciales no solo para excluir decisiones discriminatorias o sesgadas, sino también para evitar correlaciones indebidas y garantizar la equidad en el análisis de los resultados [1327].

El contenido obligatorio del artículo 14 trata de detallar las obligaciones exigibles, en concreto, a los diferentes actores implicados en la cadena de valor, partiendo de la idea (art. 14.1) de que los sistemas de alto riesgo se diseñarán y desarrollarán para garantizar que puedan ser vigilados de manera efectiva por personas físicas, mientras el apartado segundo del artículo establece la finalidad que persigue la supervisión, que no es otra que prevenir o reducir al

---

[1326] Gutiérrez David, M. E., *op. cit*. N. 1314, p. 664, hablando de *«transparencia de expertos para expertos»*. En todo caso, la valiosa contribución del completísimo trabajo de Gutiérrez es que aporta, tras sus críticas, una posible interpretación del artículo 13, dirigida a dotarlo de contenido sustantivo, entendiendo que, en los casos de modelos no intrínsecamente interpretables (cajas negras) las instrucciones de uso deben incorporar información sobre las técnicas implementadas por el proveedor para generar explicaciones que posibiliten una adecuada interpretación de la información de salida, mientras que los modelos interpretables por defecto deben contener, en sus instrucciones de uso, información suficiente con vistas a garantizar la capacitación del responsable del despliegue para realizar una adecuada interpretación de la información de salida y un uso correcto del sistema. (p. 660). Aunque la utilidad del modelo sustantivo que propone la autora nos parece harto coherente, no podemos afirmar, no obstante, que sea la única de las interpretaciones posibles a la luz de los escuetos postulados del Reglamento, por lo que, por muy deseable que sea una interpretación sustantiva de estos requisitos, lo cierto es que nos encontramos, en este punto, con una regulación con un elevado contenido formal.

[1327] Favaretto, M./de Clercq, E. / Elger, B. S., «Big Data and discrimination: perils, promises and solutions. A systematic review», *Journal of Big Data*, vol. *6*, 1, 2019, p. 16.

mínimo los riesgos para salud, seguridad y derechos fundamentales cuando el sistema se use de una forma razonablemente previsible y, aun así, los riesgos persistan. De esta forma se puede afirmar que la supervisión humana sería algo así como una cláusula de cierre o un complemento al resto de salvaguardas fijadas por el RIA. Si todo lo demás falla, al menos tendremos a una persona supervisando al sistema.

Las exigencias de la supervisión en concreto se desarrollan en los siguientes apartados (14.3 y 4), estableciéndose que la supervisión debe ser proporcional al riesgo, autonomía y contexto de uso del sistema, y que debe implantarse desde la concepción del mismo. Las capacidades que se sobreentienden del supervisor humano (art. 14.4), son las siguientes: deben ser capaces de entender tanto las capacidades como las limitaciones del sistema; deben ser conscientes del sesgo de automatización (es decir, la tendencia humana a dar por válidos los resultados de los sistemas técnicos [1328]), interpretar correctamente los resultados y, en caso de que fuera necesario, rechazar los resultados o detener el sistema, actuando el supervisor como freno de emergencia [1329].

Cabe destacar que, aunque estas obligaciones vinculan al proveedor, pues es el encargado de diseñar, desde un inicio, sistemas donde sea posible implantar estas medidas de supervisión (incluyendo, por imperativo legal, una interfaz humano-máquina), proporcionando instrucciones claras sobre cómo llevar a cabo esta tarea, lo cierto es que será el responsable de despliegue el mayor implicado en esta obligación. A él compete garantizar, en primer término, que el sistema sea usado conforme a las instrucciones proporcionadas, asegurar que los datos de entrada del sistema sean pertinentes, (art. 29.3) y verificar que los supervisores directos estén lo suficientemente cualificados, formados y con recursos y autoridad suficientes para desempeñar su

[1328] También llamado sesgo de complacencia o, en su terminología anglosajona, *automation bias.* De Asís Pulido lo define como el sesgo consistente en ceder la decisión al sistema, al considerarlo más neutral y menos sujeto al fallo, por un miedo a las consecuencias en caso de error o de haberse separado de la respuesta de la máquina. De Asís Pulido, M., *op. cit*. N. 585, p. 301. El fenómeno no es exclusivo de la IA, sino que se presenta en todos los ámbitos donde existen sistemas informatizados de asistencia a la toma de decisiones. Así, ya en estudio empírico llevado a cabo sobre pilotos experimentados y estudiantes en 1999, donde se concluye que el sesgo de automatización no responde únicamente a una preferencia sistemática por la tecnología, sino que refleja una combinación de economía cognitiva, falta de vigilancia y diseño de sistemas que no garantizan la equidad perceptual entre fuentes automatizadas y no automatizadas. Skitka, L. J./Mosier, K. L./Burdick, M., «Does automation bias decisión-making?», *Int. J. Human-Computer Studies*, 51, 1999, pp. 91-1006.

[1329] En este sentido, Laux, crítico con la idea de que se considere al humano como imprescindible y última llamada cuando todo el resto de medidas fallan, y proponiendo enfatizar en la desconfianza institucional de los supervisores con el objeto de crear instituciones de supervisión más fiables. Laux, J., «Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act», *AI & SOCIETY*, 29, 2023, pp. 2853-2866.

tarea (art. 26.2). Hemos de ser conscientes en este punto de que, aunque el Reglamento alude a la posición del responsable de despliegue como único destinatario primigenio de las obligaciones (recuérdese que es solo quien actúa en su propio nombre desplegando un sistema en un contexto de uso), lo cierto es que quienes actúen en el día a día con los sistemas puestos al servicio del público serán los usuarios directos, que serán, en muchas ocasiones, trabajadores a cargo del propio responsable de despliegue y, por tanto, no destinatarios directos de esta norma. El RIA parece ser consciente de esta situación, previendo que los responsables de despliegue no solo capaciten a sus trabajadores y les doten de recursos, sino que también se establece la obligación de que los empleadores informen a los representantes de los trabajadores y a los trabajadores afectados de que van a estar expuestos al uso de sistemas de IA de alto riesgo (art. 26.7), asegurándose de implementar las medidas organizativas que sean adecuadas al efecto (26.1) [1330].

El artículo finaliza con un apartado destinado a reforzar la supervisión en un área especialmente sensible, como es la vigilancia biométrica, donde se exige una doble verificación de los resultados, a realizar por dos personas físicas *«con la competencia, formación y autoridad necesarias»* [1331].

---

[1330] Indudablemente esta necesaria relación de delegación de poderes y de obligaciones afectará a la imputación de las eventuales responsabilidad, por lo que este será otro campo más donde cabrá aplicar toda la dogmática relacionada con la delegación de funciones en estructuras organizadas, con la particularidad, sobre la que se profundizará más adelante, de que cabe poner en duda que, por el momento, existan capacidades reales de supervisión, siendo difícil entender que se trata de una obligación por la cual se pueda afirmar la posibilidad de evitación de resultados lesivos. *Vid. infra*. Ha de tenerse en cuenta, en este contexto, que ya el Comité económico y social europeo, en su Dictamen sobre la Propuesta de Reglamento, advertía de que: *«El requisito de "vigilancia humana" es especialmente relevante en las relaciones laborales, puesto que un trabajador o un grupo de trabajadores se encargarán de ella. El CESE subraya que estos trabajadores deben recibir formación sobre cómo llevar a cabo esta tarea. Además, dado que se espera que estos trabajadores puedan ignorar el resultado del sistema de IA o incluso no utilizarlo, deben establecerse medidas para evitar el miedo a represalias (como la degradación o el despido) en caso de que se tome esa decisión»*. Dictamen del Comité Económico y Social Europeo sobre la «Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión» [COM(2021) 206 final – 2021/106 (COD)] (2021/C 517/09, apartado 4.18. Accesible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX%3A52021AE2482> [Accedido por última vez el 29/01/2026].

[1331] Los motivos, recogidos en considerando 73 RIA, resultan confusos. Compartimos plenamente el cuestionamiento de Lazcoz: «*No deja de ser llamativo que el RIA considere que esta doble verificación por separado es un mecanismo «reforzado», ¿hay razones para pensar que el segundo verificador no incurrirá en el mismo sesgo o error que la primera persona encargada de la supervisión? ¿acaso este mecanismo de supervisión «por separado» puede considerarse más reforzado que un proceso de toma de decisiones «conjunta» que involucre dos supervisores humanos?»*. Lazcoz Moratinos, G., «La vigilancia o supervisión humana en el artículo 14 del Reglamento de inteligencia artificial: ¿un mero requisito obligatorio para los sistemas de alto riesgo?», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi,

Aunque el artículo 14 trata de concretar los contornos de un principio que ha sido objeto de extensas discusiones [1332], lo cierto es que la redacción final del artículo no ha servido para despejar algunas de las dudas que numerosos autores habían puesto sobre la mesa [1333]. El tópico común de discusión, en este sentido, tiene que ver con el cuestionamiento de la eficacia empírica de la supervisión humana. Por ser más claros: ninguna obligación puede ser verdaderamente eficaz si su cumplimiento efectivo es imposible. En este sentido, podemos resumir las críticas de la doctrina en torno a tres ideas fundamentales.

En primer lugar, no existe garantía alguna de que un humano tome mejores decisiones que un sistema automatizado, y, de hecho, en ocasiones se podrá afirmar todo lo contrario, es decir, que habida cuenta de las limi-

---

Madrid, 2024, p. 697. Sin embargo, es común encontrar referencias a las «medidas de doble aseguramiento» al hilo del principio de confianza dentro de la teoría de la imputación objetiva. Así, en Maraver Gómez, M., *El principio de confianza … op. cit.* N. 834, p. 143. Llama la atención el autor, empero, sobre la necesidad de diferenciar este doble aseguramiento en cuanto deberes de cuidado sobre la conducta de terceros (propios de la división horizontal del trabajo), de los casos en que un subordinado tiene un deber de revocación o corrección respecto de indicaciones de un superior, donde dicho subordinado solo debe actuar si reconoce el error, manteniéndose, por tanto, dentro de los límites del principio de confianza (nota 196). Habrá que ver, por tanto, cuál es la configuración final de este doble aseguramiento o supervisión, pues los efectos respecto de la confianza son distintos. En todo caso, las medidas de aseguramiento tradicionalmente exploradas (necesidad de recuento doble del material médico tras una operación, o el deber de mantener una distancia de seguridad respecto del vehículo de delante en caso de que genere una maniobra injustificada), son casos en que el deber superpuesto del supervisor está perfectamente limitado en cuanto a su contenido, por lo que consideramos que la crítica de Lazcoz sigue siendo válida.

[1332] Una revisión completa de lo que significa la supervisión de sistemas de IA, desde la propia génesis del Reglamento, en: Lazcoz Moratinos, G., *Gobernanza y supervisión humana de la toma de decisiones automatizada basada en la elaboración de perfiles*, 2023, tesis doctoral, accesible en: <https://addi.ehu.es/handle/10810/61322> [Accedido por última vez el 29/01/2026]. En España, reflexionando de manera incipiente sobre el paradigma del control humano significativo: Romeo Casabona, C. M., «La atribución de responsabilidad penal…», *op. cit*. N. 179, p. 14 y ss. (igual, el mismo, en: *Estudios político-criminales, jurídico-penales y criminológico. Libro homenaje al profesor José Luis Díez Ripollés,* pp. 721-734). También: Lazcoz Moratinos, G./ Obregón Fernández, A., «La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional Humanitario y el Derecho de la Unión Europea», *Revista electrónica de estudios internacionales*, 42, 2021, pp. 1-29.

[1333] Así, Green, Huq, Laux, Koulu. Green, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, 45, 2022, pp. 1-22, afirmando que las medidas de supervision humanas no se sostienen por pruebas empíricas. Huq, A. Z., «A Right to a Human Decision», *Virginia Law Review*, vol. 106*,* 3, 2020, pp. 611-688, destacando que el hecho de que la decisión del sistema de IA sea errónea no implica necesariamente que un humano hubiera decidido mejor, y que la solución para conseguir la igualdad no debe estar unida al derecho a la decisión o supervisión humana, sino a lo que él llama el derecho a una «decisión automatizada bien calibrada» («*right to a well-calibrated machine decision»,* p. 619), es decir, respetuosa con el debido proceso, la privacidad y la no discriminación*;* Laux, J., *op. cit.* N. 1329 y Koulu, R., «Human control over automation: EU policy and AI ethics», *Journal of Legal Studies*, 1, 2020, pp. 9-46. Un desarrollo de los postulados de estos autores y su influencia para el papel del usuario supervisor, *infra*, en capítulo octavo.

taciones en la cognición y el entendimiento de los factores que dan lugar a una decisión de un sistema de IA, tal vez sea recomendable, acaso, que la decisión más prudente por parte del humano sea devolver el control al sistema [1334].

En segundo lugar, es posible que la supervisión del humano, en tanto no llegue a tener una eficacia real, actúe como mero tranquilizador, aportando una falsa sensación de seguridad [1335], mistificándose las capacidades del supervisor, pero sin considerar las evidentes limitaciones en la cognición [1336].

Esto nos lleva, irremediablemente, a la tercera crítica, y es que corremos el riesgo de transferir la responsabilidad legal a operadores que no tienen, en esencia, un verdadero poder de decisión real en el caso concreto [1337].

Cabría añadir, en este punto, que explicabilidad y supervisión son dos principios que van de la mano. El problema evidente, una vez analizadas las obligaciones concretas impuestas por la normativa, es que aún no ha quedado demostrado que existan deberes concretos capaces de garantizar ni la una ni la otra. La explicabilidad, asociada a la transparencia en un sentido bastante débil, difícilmente podrá permitir una supervisión efectiva, habida cuenta de que el propio Reglamento parece privilegiar (o, al menos, no desincentivar) los sistemas de caja negra, haciendo muy difícil dotar de un contenido verdaderamente material a la supervisión. Evidentemente, una falta de supervisión absoluta redundará en una mayor inseguridad del sistema, pero no puede afirmarse, por el contrario, que una concreta decisión del supervisor individual sea más o menos correcta. Sin estos parámetros de corrección difícilmente podremos hablar de que existe una verdadera concreción del deber de cuidado exigible al supervisor (quizá sí, por su obligación de carácter más formalizado, respecto del responsable de despliegue, pero no tanto

---

[1334] En este sentido, Hilgendorf, haciendo referencia a lo que él llama «*control dilemma*». Hilgendorf, E. «Automated driving and the law…, *op. cit*. N. 833, p. 181-182. Tambien: Wagner, B., «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision Making Systems», *Policy & Internet*, vol. 11, 1, 2019, pp. 104- 122.

[1335] Lazcoz Moratinos, G., «La vigilancia o supervisión humana …, *op. cit*. N. 1331, p. 684.

[1336] Nos remitimos en este punto a lo dicho por Selbst en capítulo V respecto de las repercusiones que, en materia de reconstrucción de los deberes de diligencia esenciales para la imprudencia tienen estas limitaciones en la cognición. Selbst, A., «Negligence …», *op. cit*. N. 718. También Matsumi, H./Solove, D. J., «The Prediction Society: Algorithms and the Problems of Forecasting the Future», *University of Illinois Law Review*, 1, 2025, especialmente en pp. 37-41, afirmando en p. 39.*: «For human involvement to be the answer, the law must set forth exactly how humans would ameliorate the problems with algorithmic predictions in particular cases. Instead, the law just points to a human and says: «Hey, there's a human, so all is fine» even though it remains unclear what the human is to do».*

[1337] Lazcoz habla en este punto del riesgo de *deskilling,* es decir, de que no tengamos las habilidades suficientes para desempeñar una tarea que es encomendada a nivel legal. Es tanto como decir que se exige a alguien aquello que no es posible cumplir. Así, podríamos afirmar, la supervisión es posible, pero no al extremo de poder corregir las decisiones del sistema conforme a parámetros de racionalidad. Lazcoz Moratinos, G., «La vigilancia o supervisión humana …, *op. cit*. N. 1331, p. 698.

respecto del supervisor humano final, obligado a interactuar con un sistema cuyas decisiones son parcialmente explicables y difícilmente corregibles).

La solución que propone Lazcoz a este respecto nos parece adecuada, pero teniendo en cuenta que, una vez más, habremos de esperar a su desarrollo normativo. Afirma este que no basta con imponer una supervisión genérica normativamente, sino que será imprescindible, para hacerla efectiva (exigiendo que los humanos sean capaces de reducir de forma demostrable los riesgos), desarrollar modelos de gobernanza basados en la evidencia [1338], que sean capaces de definir claramente: qué humanos han de supervisar, cuáles son sus funciones reales y qué condiciones (materiales y legales) han de acompañar esta tarea, encontrando un equilibro adecuado entre una supervisión lo suficientemente flexible para cada contexto y el establecimiento de unas normas mínimas comunes [1339].

### 7. *Garantizar un nivel adecuado de precisión, solidez y ciberseguridad (art. 15)*

Podemos entender, con Aba Catoira, que la importancia del artículo 15 del RIA, al establecer los requisitos de precisión, solidez y ciberseguridad, radica en que tales sistemas presentan riesgos potenciales para salud, seguridad y derechos fundamentales [1340], pues, conforme al considerando 75, *«La solidez técnica es un requisito clave para los sistemas de IA de alto riesgo, que deben ser resilientes en relación con los comportamientos perjudiciales o indeseables por otros motivos que puedan derivarse de limitaciones en los sistemas o del entorno en el que estos funcionan (p. ej., errores, fallos, incoherencias o situaciones inesperadas). Por consiguiente, deben adoptarse medidas técnicas y organizativas para garantizar la solidez de los sistemas de IA de alto riesgo, por ejemplo mediante el diseño y desarrollo de soluciones técnicas adecuadas para prevenir o reducir al mínimo ese comportamiento perjudicial o indeseable».*

En lo relativo a la precisión de los sistemas, no existe en absoluto consenso entre la comunidad científica sobre cuáles son los mejores parámetros a utilizar [1341], aunque puede entenderse como la relación entre el propósito previsto del sistema y su desempeño, es decir, la proporción de predicciones co-

[1338] *Ibid.,* p. 695.
[1339] *Ibid.,* p 694.
[1340] Aba Catoira, A., «Precisión y solidez de los sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 702.
[1341] Así, Martínez Garay, L./ Montes Suay, F., *op. cit.* N. 17., p. 16 y ss., explicando con claridad cómo, en el campo de la valoración de riesgo penal, los parámetros para medir la precisión predictiva son

rrectas frente al total de predicciones, debiendo constar en las instrucciones de uso estos datos. Es decir, la precisión indica cuán exactas son las predicciones de un modelo [1342]. El artículo 15 incorpora un apartado que encarga a la Comisión fomentar el desarrollo de metodologías y métricas para medir los niveles exigidos de precisión y robustez. No obstante, se concede flexibilidad a los proveedores para implementar soluciones técnicas que consideren apropiadas.

El artículo 15 se remite, a su vez, al artículo 10, en lo relativo a los datos utilizados para entrenar al sistema, subrayándose la importancia de los controles de calidad desde el diseño y la selección de métricas adecuadas. Esta interrelación nos hace ser conscientes de la importancia fundamental de estos requisitos.

Respecto a la solidez, el artículo 15.4 exige que los sistemas sean resilientes frente a comportamientos perjudiciales, y prevé que los proveedores adopten medidas como planes de respaldo, copias de seguridad, mecanismos de redundancia o planes de actuación para la recuperación de datos. Esta exigencia cobra especial relevancia en los sistemas que continúan aprendiendo tras su comercialización, ya que pueden sufrir degradaciones de precisión y robustez debido a bucles de retroalimentación (positivos o negativos) [1343].

La evaluación del cumplimiento del artículo 15 requiere, por tanto, una verificación y validación del sistema mediante evidencias objetivas, que incluyen pruebas con datos reales y técnicas como la validación cruzada [1344].

Finalmente, en materia de ciberseguridad, atendiendo al considerando 76, resulta *«fundamental para garantizar que los sistemas de IA resistan a las actuaciones de terceros maliciosos que, aprovechando las vulnerabilidades del sistema, traten de alterar su uso, comportamiento o funcionamiento o de poner en peligro sus propiedades de seguridad»*. Una vez más, no cabe duda de que el

---

diversos (sensibilidad, especificidad, área bajo la curva ROC), pudiendo mostrar resultados discrepantes entre ellos, lo que lleva a afirmar la complejidad técnica y contextual de estos parámetros.

[1342] Aba Catoira, A., *op. cit*. N. 1340, p. 708: «*cuando hablamos de precisión referida a un modelo de IA hablamos de la proporción de predicciones correctas realizadas por el modelo en comparación con el total de predicciones realizadas. En otras palabras, la precisión nos indica qué tan exactas son las predicciones de un modelo (accuracy) fundamental dentro del sistema de gestión de calidad junto a la robustez, ciberseguridad, transparencia, gobernanza del dato y la supervisión».*

[1343] Aba Catoira analiza los distintos tipos de errores que afectan a sistemas auto-aprendientes, como la degradación del modelo por cambios en el entorno o el fenómeno del «colapso» en sistemas generativos, causado por la contaminación del conjunto de datos de entrenamiento. *Ibid.,* p. 717.

[1344] No obstante, señala la autora que la calidad del sistema no depende únicamente del rendimiento medido, sino también de su interpretabilidad y equidad. Las métricas de equidad permiten detectar diferencias en el rendimiento entre distintos grupos sociales (sexo, raza, edad) y mitigar posibles discriminaciones, lo cual afecta directamente a la robustez y fiabilidad del sistema. Esta variabilidad no hace más que confirmar, a nuestro juicio, la necesidad de normas técnicas que vengan a concretar las buenas prácticas para cada uno de los supuestos de uso, pero, a pesar de la posibilidad de que varios sistemas encajen dentro de las exigencias de la norma, creemos que no cabe poner en duda su centralidad para logar productos o sistemas seguros, en un sentido directamente vinculado con la evitación de daños sobre bienes esenciales. *Ibid.,* p. 719.

cumplimiento de este requisito eminentemente técnico podrá darse de muy variadas maneras, pero no cabe perder de vista que el artículo 42.2 establece la presunción de conformidad cuando los sistemas cuenten con una certificación o declaración conforme al esquema europeo de ciberseguridad [1345].

En definitiva, aunque el cumplimiento de los requisitos de carácter técnico conlleve la necesidad de mayores especificaciones y la sujeción a normativas de especificación, no cabe dudar de que no sólo garantiza la seguridad y fiabilidad de los sistemas, sino también su legitimidad social y jurídica en contextos de alto riesgo.

### d) El resto de obligaciones y disposiciones del Reglamento

#### 1. *Las obligaciones poscomercialización (art. 72)*

Conforme al artículo 3.25, la vigilancia poscomercialización se define como *«todas las actividades realizadas por los proveedores de sistemas de IA destinadas a recoger y examinar la experiencia obtenida con el uso de sistemas de IA que introducen en el mercado o ponen en servicio, con objeto de detectar la posible necesidad de aplicar inmediatamente cualquier tipo de medida correctora o preventiva que resulte necesaria».* El proceso implica varias fases, desde la recolección de datos del sistema (técnicos y de impacto sobre las personas), pasando por la evaluación de las causas de los posibles problemas, implementación de soluciones y, finalmente, una obligación de comunicación con los agentes involucrados en caso de detectar cualquier indicio de riesgo.

El artículo 72 define las obligaciones concretas, estableciendo que el sistema de vigilancia posterior a la comercialización *«recopilará, documentará y analizará de manera activa y sistemática los datos pertinentes que pueden facilitar los responsables del despliegue o que pueden recopilarse a través de*

[1345] Más allá de la referencia al Reglamento (UE) 2019/881, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación, destaca acertadamente Sánchez Acevedo que habrá que atender a normativa accesoria, nombrando la Directiva (UE) 2022/2555, que establece medidas para un nivel común de ciberseguridad en la Unión, y en España a la Ley 40/2015 para el sector público y el Real Decreto del Esquema Nacional de Seguridad, pues el requisito de ciberseguridad no es aplicable por primera vez a los sistemas de IA, sino que se predica de todos los sistemas o tecnologías de la información, por lo que se puede hablar, en este sentido, de un marco europeo de certificación de ciberseguridad. Sánchez Acevedo, M. E., «Ciberseguridad en sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 721-742.

*otras fuentes sobre el funcionamiento de los sistemas de IA de alto riesgo durante toda su vida útil, y que permiten al proveedor evaluar el cumplimiento permanente de los requisitos establecidos en el capítulo III, sección 2, por parte de los sistemas de IA».* En definitiva, este plan, que debe incluirse dentro de la documentación técnica, tiene como finalidad verificar que se siguen cumpliendo los requisitos esenciales durante todo el ciclo de vida del sistema, fijándose protocolos estandarizados para la detección de anomalías y la reacción frente a ellas [1346].

Es importante tener en cuenta el componente de dinamismo que ofrece esta obligación, pues, habida cuenta de la naturaleza cambiante de los sistemas, será esencial para garantizar la continuidad en las condiciones de seguridad, así como para adaptar los sistemas a los nuevos contextos, preservando la eficacia de las medidas de contención del riesgo. Se trata, por tanto, de una obligación especialmente central en el caso de los sistemas de aprendizaje, pues solo una adecuada vigilancia continua e ininterrumpida tras el lanzamiento del sistema al mercado garantizará que los riesgos de los sistemas que van evolucionando y cambiando conforme van aprendiendo puedan mantenerse dentro del riesgo permitido [1347].

### 2. *Las obligaciones de los modelos de uso general y fundacionales (art. 50)*

La inicial propuesta de Reglamento de 2021 no contemplaba la regulación de los modelos de IA de uso general, pero la fuerza de los acontecimientos tecnológicos (esencialmente, el auge de tecnologías como ChatGPT [1348]), obligó a modificar los postulados del Reglamento para incluir una referencia particular a este tipo de sistemas, mucho más complejos y con riesgos particulares. Así, el artículo 3. 63 los define como aquellos entrenados *«[…]con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas […]».* La IA generativa, estre-

[1346] Así, puede implicar la necesidad de actualizar los algoritmos empleados, modificar los conjuntos de dato que sirven de base al sistema o revisar los procesos de decisión *ex novo*. En este sentido: Salazar, I., / Liébanas, M. A., «Vigilancia poscomercialización en los sistemas de inteligencia artificial de alto riesgo en el Reglamento. Descripción, medidas y casos de uso», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 745, ejemplificando qué tipo de técnicas han de implementarse.

[1347] Así, Ramón Fernández, F., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 683.

[1348] Destacadamente el modelo GPT-4, lanzado al mercado en marzo de 2023.

chamente vinculada a estos modelos, se reconoce como una subcategoría de la IA de uso general, capaz de producir contenido textual, audiovisual o visual, lo que plantea retos específicos en materia de propiedad intelectual [1349] y desinformación (Cons. 99, 105, 110, y 133 RIA; artículo 3.60 RIA).

Se destaca (Cons. 110) que estos modelos de uso general pueden generar lo que se ha venido a denominar «riesgo sistémico», introduciéndose, así, en la última de las versiones del Reglamento, un nuevo escalón de riesgo.

Así, estos modelos pueden generar riesgos sistémicos cuando sus capacidades y su alcance dan lugar a impactos negativos significativos o previsibles, que pueden manifestarse en diversas formas: accidentes graves, afectación a sectores críticos, daños a la salud y seguridad pública, o alteraciones en los procesos democráticos y la estabilidad económica o social, incluyendo la difusión de contenido ilícito, falso o discriminatorio. Su aparición puede darse en cualquier fase del ciclo de vida del modelo y depende de múltiples factores: fiabilidad, equidad, nivel de autonomía, acceso a herramientas avanzadas, estrategias de distribución, y eliminación de salvaguardias. Además, se identifican en el mismo considerando amenazas específicas relacionadas con el uso indebido intencional, la desalineación con la intención humana, y riesgos de carácter químico, biológico, radiológico y nuclear, así como el desarrollo de cibercapacidades ofensivas. También se consideran peligrosas la capacidad de los modelos para autorreplicarse, entrenar otros modelos, interferir en infraestructuras críticas o generar desinformación, sesgos discriminatorios o violaciones de la intimidad. Finalmente, se advierte sobre el riesgo de que un solo incidente pueda desencadenar efectos en cascada que afecten a poblaciones enteras o sectores completos de la sociedad, comprometiendo los valores democráticos y los derechos humanos [1350].

Asimismo, diversos sistemas de inteligencia artificial son capaces de producir grandes volúmenes de contenido sintético que resulta cada vez más indistinguible del creado por personas, lo que plantea serios desafíos para los usuarios. La amplia accesibilidad y el continuo perfeccionamiento de estas tecnologías afectan de manera significativa a la integridad del ecosistema informativo y erosionan la confianza pública, al generar nue-

[1349] «*La Directiva (UE) 2019/790 introdujo excepciones y limitaciones que permiten reproducciones y extracciones de obras y otras prestaciones con fines de prospección de textos y datos en determinadas circunstancias. Con arreglo a estas normas, los titulares de derechos pueden optar por reservarse sus derechos en relación con sus obras u otras prestaciones para evitar la prospección de textos y datos, salvo que su finalidad sea la investigación científica […]*» (Cons. 105). También referencias a la aplicación de la citada Directiva en Considerando 106.

[1350] Resumen del Considerando 110.

vos riesgos asociados a la desinformación masiva, la manipulación de la opinión, el fraude, la suplantación de identidad y el engaño a los consumidores [1351].

Pese al alarmante perfilado de estos novedosos sistemas, sorprende la escasa fuerza normativa de los requisitos a ellos aplicables. Solo en caso de que los modelos de uso general (cuya categorización queda reducida al número de parámetros que usen para su entrenamiento [1352]) presenten los riesgos sistémicos arriba descritos (riesgo, según el artículo 3.65, de *«repercusiones considerables en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor»),* se aplicará el régimen reforzado.

Dicho régimen, previsto en el capítulo quinto (arts. 51 a 55), establece los criterios para que un sistema sea considerado de riesgo sistémico, el procedimiento de calificación como tal, atribuido a la Comisión, y el régimen general de obligaciones reforzadas, que, para los proveedores, implican la obligación de elaboración de documentación técnica, la cooperación con la Oficina de IA (OIA), y la publicación de un resumen del material de entrenamiento, con ciertas excepciones para modelos de código abierto.

El régimen sancionador es, asimismo, particular en estos casos, pues, en lugar de delegarse la potestad sancionadora en las autoridades nacionales, será la Comisión, con asistencia de la Oficina de IA, la encargada de supervisar el cumplimiento, imponer medidas correctoras, e imponer sanciones por infracciones como negarse a facilitar información (art. 91), incumplir medidas correctoras (art. 93), o impedir la evaluación del modelo (art. 92).

---

[1351] Resumen del Considerando 113.

[1352] Al menos mil millones de parámetros. Más allá de ese criterio, otro de los criterios empleados por el RIA para determinar si un modelo de uso general tiene capacidades de gran impacto es el denominado umbral de FLOPS, que es una medida concreta de poder de cómputo, referida a la cantidad acumulada de cálculos usados para el entrenamiento de un modelo de IA de uso general, y que corresponde a la medida de operaciones de coma flotante (un tipo de cálculo matemático usado por los ordenadores para manejar números con decimales, que las IAs hacen por millones o billones). Así, Castillo Parrilla, J. A., «Inteligencia artificial de uso general, modelos fundacionales (y «Chat GPT») en el Reglamento de inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 761. El artículo 51.1.2 dispone, en este sentido: «*Se presumirá que un modelo de IA de uso general tiene capacidades de gran impacto con arreglo al apartado 1, letra a), cuando la cantidad acumulada de cálculo utilizada para su entrenamiento, medida en operaciones de coma flotante, sea superior a 10». En caso de no alcanzarse esta cantidad, habrá que atender a su capacidad de impacto con arreglo a una decisión de la Comisión, teniendo en cuenta los criterios recogidos en el Anexo XIII.»*

Más allá de las obligaciones de contenido técnico, por lo demás muy similares a las exigidas al resto de sistemas, el artículo 50 incluye algunas obligaciones de transparencia que serán aplicables a algunos de estos sistemas. Según el artículo 50.1 RIA, los sistemas que permiten interacción directa con personas (como chatbots o asistentes virtuales) deben informar claramente a los usuarios de que están tratando con una IA. El artículo 50.2 impone obligaciones específicas a los sistemas que generen contenidos sintéticos (texto, audio, imágenes o vídeo), incluidos los modelos de IA de uso general, requiriendo la inclusión de marcas visibles y legibles por máquina que permitan identificar la artificialidad del contenido, mediante soluciones técnicas «eficaces, interoperables, sólidas y fiables».

Resulta llamativo cómo, en este caso, se establecen excepciones para usos en contextos legales (como la investigación de delitos), y en obras creativas o informativas, donde basta con hacer público que el contenido ha sido manipulado o generado artificialmente, sin que ello interfiera con la libertad de prensa. Además, si el contenido ha sido sometido a revisión humana o cuenta con responsabilidad editorial, la transparencia no es exigible, lo cual puede entenderse como una decisión *pro libertate* o una forma de no restringir en demasía otros posibles derechos fundamentales en juego, adoptando un enfoque equilibrado entre prevención de la desinformación y respeto a otros derechos fundamentales, esencialmente la libertad de expresión e información, así como la libertad de creación artística, siempre presente en el debate sobre la restricción de contenidos, donde la tensión entre libertad y control resulta inevitable.

Finalmente, y aunque el artículo advierte también sobre la creciente sofisticación y peligrosidad de las ultrafalsificaciones [1353], que pueden utilizarse para influir en la opinión pública, manipular elecciones, extorsionar o dañar reputaciones, y subraya la necesidad de mecanismos técnicos y regulatorios para identificar estos contenidos, no llega el Reglamento a legislar al respecto, siendo, como es evidente, otro ejemplo de tensión entre derechos fundamentales que queda, empero, lejos de tener una solución jurídica clara, pese a las elevadas repercusiones que puede conllevar. Identificamos aquí un claro ejemplo de vacío de protección, motivado, entendemos, por la dificultad de lograr un equilibrio entre derechos de tan especial importancia.

[1353] Advirtiendo sobre los riesgos penales: Caldwell, M./ Andrews, J. T. A./ Tanay, T./ Griffin, L. D., «*AI-enabled future crime*», *Crime Science,* 9, 2020, pp. 1-13, destacando, entre los riesgos penales más difíciles de detectar, la suplantación por medio de audio y video y la generación de noticias falsas.

### 3. *Los espacios controlados de prueba como salvaguarda jurídica*

El capítulo sexto, bajo la rúbrica «medidas de apoyo a la innovación», regula los espacios controlados de pruebas o *«sandboxes»,* como instrumentos jurídicos destinados a promover la innovación tecnológica por medio de entornos de experimentación [1354]. El artículo 57 RIA establece sus objetivos y fija que cada Estado miembro deberá establecer al menos un *sandbox* nacional dentro de los 24 meses siguientes a la entrada en vigor del RIA (art. 57.1), para evitar una fragmentación normativa [1355].

Estas medidas, además de fomentar el desarrollo controlado de los sistemas, especialmente para asegurar que las pequeñas y medianas empresas, así como las empresas emergentes, no se queden fuera de la carrera tecnológica, tienen sentido en tanto modelo porque fijan un régimen jurídico excepcional, que permite una exención limitada de la responsabilidad en caso de que los participantes de las pruebas actúen de buena fe y conforme al plan aprobado (art. 57.12), aunque sí se generará responsabilidad civil por los daños causados durante las pruebas.

### 4. *El régimen sancionador*

El colofón a la complejísima regulación hasta ahora perfilada se completa con el capítulo XII relativo al régimen sancionador aplicable, objeto de diversas modificaciones a lo largo de las diferentes versiones del texto legal [1356]. Se puede afirmar que la configuración de un sistema sancionador responde al objetivo general del Reglamento, que es garantizar una aplicación coherente y efectiva del marco normativo, si bien el papel de las autoridades europeas es, en este sentido, anecdótico, pues se opta por delegar en cada Estado miembro la concreción del régimen de sanciones (que habrán de ser, en todo caso, usan-

[1354] Definidos en artículo 3.55. Una reciente reflexión sobre el papel de este modelo en: Casado García, M., «La regulación de las neurotecnologías: las «sandboxes» para la innovación», *RJUAM*, 51, 2025, pp. 95-120, especialmente pp. 107 y ss.

[1355] España ha aprobado ya un Real Decreto destinado a regularlos. Se trata del RD 817/2023 pero, según señala Cotino Hueso, L., «Sandbox, espacios controlados y pruebas en condiciones reales de sistemas de inteligencia artificial en el Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 829, que requerirá una reforma, pues es anterior a la aprobación del RIA y tendrá que alinearse con el nuevo marco europeo.

[1356] Un repaso de esta evolución en Sempere, J. F., «El régimen sancionador en el Reglamento de inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 873-892.

do la expresión ya habitual, «efectivas, proporcionadas y disuasorias»), introduciendo un margen de discrecionalidad que podrá dar lugar a efectivas divergencias entre los diversos regímenes nacionales.

En el caso español, aunque se desarrollará con más detalle *infra,* sin perjuicio de que el régimen general aplicable será el previsto en la Ley 40/2015, de Régimen Jurídico del Sector Público (concretamente su capítulo III, relativo a los principios de la potestad sancionadora), se encuentra en tramitación un Anteproyecto de Ley de gobernanza donde se especifica el régimen disciplinario, administrativo, que está por aplicarse [1357].

El artículo 99 RIA, por su parte, fija un sistema escalonado de infracciones, diferenciando esencialmente entre el incumplimiento de las obligaciones del artículo 5 y el resto.

– Así, en primer lugar, el incumplimiento de las prohibiciones absolutas del artículo 5, se sanciona con multas de hasta 35 millones de euros o, si el infractor es una empresa, hasta el 7 % de su volumen de negocios mundial del ejercicio financiero anterior, si esta cuantía fuera superior (art. 99.3).

– En segundo lugar, el incumplimiento de otras obligaciones esenciales, como las de proveedores (art. 16), representantes autorizados (art. 22), importadores (art. 23), distribuidores (art. 24), responsables del despliegue (art. 26), así como de los organismos notificados (arts. 31, 33.1, 33.3, 33.4 y 34). Se asimila aquí el incumplimiento de las obligaciones de transparencia (art. 50), que podrá ser sancionado con multas de hasta 15 millones de euros o el 3 % del volumen de negocios mundial, si fuese mayor (art. 99.4).

– Por último, la más leve de las sanciones, reservada para la presentación de información inexacta, incompleta o engañosa a las autoridades competentes o a los organismos notificados podrá conllevar una multa de hasta 7,5 millones de euros o el 1 % del volumen de negocios mundial, en caso de empresas (art. 99.5).

Se prevé, finalmente, un régimen atenuado para las pymes, ya en el apartado primero del artículo 99, en la línea seguida por toda la normativa.

Quedan, con esta redacción, varias dudas respecto del contenido abierto o cerrado de las infracciones, que en ningún momento se detallan. La alusión a las obligaciones de los distintos operadores no resulta determinante, habida cuenta de la interconexión, como se ha tratado de explicar, de todas las disposiciones del Reglamento. Por poner un ejemplo, por mucho que el artículo 26 regule las

[1357] Nos referimos al Anteproyecto de Ley para el buen uso y la gobernanza de la Inteligencia Artificial, explicado *infra.*

obligaciones de los responsables de despliegue, lo cierto es que habrá que atender a todos los preceptos del RIA para concretar y completar el contenido de dichas obligaciones y, así, a título de ejemplo, el artículo 26.2 se refiere a las labores de supervisión humana encomendadas al mismo, siendo imprescindible para completar el contenido de la obligación atender al artículo 14. Lo mismo es predicable respecto de las obligaciones del resto de operadores.

Resulta también importante resaltar que la sanción por excelencia elegida es la multa, aunque no es la única de las soluciones posibles, pues el artículo 99.1 menciona también las advertencias o medidas no pecuniarias [1358]. Sí se contemplan, en cambio, la retirada y la prohibición de algunos sistemas en casos muy concretos, pero que parecen quedar extramuros del régimen sancionar por su ubicación sistemática [1359].

Más allá de fijar umbrales máximos de sanción, lo cierto es que el Reglamento es, en este punto, absolutamente indeterminado, dejando todo el trabajo de concreción a los Estados Miembros. Esto exige que los Estados, en sus respectivos desarrollos normativos, determinen con claridad qué conductas serán sancionables, y bajo qué criterios de gravedad, algo crucial para garantizar los principios de legalidad y seguridad jurídica. Una de las cuestiones esenciales a dilucidar será, por ejemplo, si el régimen sancionador se aplicará, y, en su caso, de qué forma, respecto del sector público, planteándose la ya problemática cuestión de qué sanción imponer a la Administración [1360], evitando que redunde en perjuicio del administrado. Se ha llegado a afirmar que, en este sentido, la lógica del Reglamento se parece mucho más a la de una Directiva, pues para su aplicación efectiva requiere de una suerte de trasposición o desarrollo normativo expreso [1361].

El artículo 99.6 establece, finalmente, una serie de criterios orientadores para la determinación de la cuantía de las sanciones, entre los que destacan: la

---

[1358] Sempere, J. F., *op. cit*. N. 1356, p. 891, crítico con la no inclusión de otro tipo de medidas sancionatorias (el apercibimiento, la advertencia, medidas de cumplimiento, así como la retirada, prohibición y restricción de un producto de IA).

[1359] Nos referimos a las medidas previstas en el artículo 79, donde se prevé el procedimiento aplicable a escala nacional a los sistemas de IA que presenten un riesgo. Ante indicios fundados de riesgo, la autoridad nacional competente debe evaluar el sistema para verificar su conformidad con el RIA y, en su caso, si se detectan riesgos significativos para derechos fundamentales, y exigir medidas correctoras que, en caso de que cumplirse, pueden dar lugar a la imposición de medidas como la retirada o la recuperación del sistema del mercado.

[1360] Sempere, J. F., *op. cit*. N. 1356, p. 876.

[1361] Gamero Casado, E., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 837. Este autor afirma, de hecho, que no parece posible que se impongan sanciones por el momento, solo por la aplicación del Reglamento, en tanto no se dé un desarrollo normativo interno, siguiendo la estela de lo afirmado respecto de la norma más similar al respecto, esto es, el RGPD. *Ibid.,* p. 839.

naturaleza, gravedad y duración de la infracción; el carácter intencionado o negligente de la conducta; las medidas adoptadas por el infractor para mitigar los efectos del daño; los beneficios obtenidos por la infracción; el grado de cooperación con las autoridades y cualquier reincidencia anterior [1362].

### 5. *¿Vías de recurso de los particulares?*

Una de las carencias más acusadas del texto final del Reglamento es que, pese a su reiterada alusión a la necesidad de construir una IA fiable, respetuosa con los derechos fundamentales de los ciudadanos, las vías para la reclamación de la aplicación efectiva del Reglamento y sus garantías resultan bastante escasas y con un contenido más bien declarativo. Así, la forma de hacer efectivos los pretendidos derechos que el Reglamento protege son, esencialmente, el derecho a presentar una reclamación y el derecho a una explicación.

En primer lugar, el artículo 85 establece el derecho a presentar una reclamación ante una autoridad de vigilancia del mercado, estableciendo como legitimados a toda persona física o jurídica con motivos para considerar que se ha infringido «lo dispuesto en el presente Reglamento», sin especificar a qué concretas obligaciones se refiere ni ante qué concretos órganos reclamar, especificando solamente que dichas reclamaciones se tramitarán conforme a los procedimientos establecidos a tal fin por las autoridades de vigilancia de mercado.

La crítica que una parte de la doctrina formula es que el Reglamento atribuye la función de vigilancia de mercado a varias autoridades, (así, en el caso español, más allá de la AESIA, todas las autoridades que se designan en los instrumentos de implementación de la normativa del Anexo I), y que, además, estas autoridades no

[1362] Ya ha sido apuntado por algunos autores que, aunque materialmente el Reglamento no sea un instrumento de derecho penal, queda aún por determinar cuál es la naturaleza de las sanciones. Así, Scollo, afirmando: «*[…] the AI Act requires States to adopt sanctions that are 'effective, proportionate and dissuasive'. This terminology is now commonplace in EU directives on criminal matters. The second is that the sanctions applicable (up to €35 million or 7% of global annual turnover in the case of prohibited practices; up to €15 million or 3% of global annual turnover in the case of other violations of the AI Act) makes them particularly severe and therefore borderline criminal law provisions, to the extent that it is necessary to ask whether they are themselves punitive sanctions and therefore already criminal provisions […]».* Scollo, L./Bravo Bolado, A., /Pérez Vicario, A. M., /Ensel, L., *op. cit.* N. 1190. En un sentido similar, cuestionando que infracciones de tan alta cuantía sean meramente administrativas solo porque el legislador así las denomina: Błachnio-Parzych, A., «*Criminal Law Protection Against Artificial Intelligence Threats in Finance»* (ponencia presentada en el Congreso Internacional Artifical Intelligence and Criminal Law, Bérgamo, 25 de octubre de 2024).

tienen entre sus funciones conocer reclamaciones de particulares, ni están preparadas orgánicamente o competencialmente para hacerlo [1363]. No existe en el Reglamento ninguna obligación de la autoridad de vigilancia de mercado de informar al reclamante sobre el curso o el resultado de su reclamación, y tampoco obligación expresa de resolver dando una respuesta de fondo [1364]. Por tanto, conforme a la literalidad de la regulación actual, el derecho implica la facultad de cualquier persona de presentar una reclamación ante cualquiera de las autoridades de vigilancia de mercado (en función del tipo de sistema podrá ser la Oficina de IA, la AESIA, la AEPD, la CNMV etc. [1365]), sin que se pueda garantizar uniformidad en cuanto al procedimiento ni en cuanto a los derechos a obtener una resolución sobre el fondo.

En segundo lugar, el artículo 86 contempla el derecho a una explicación en la toma de decisiones individuales, que, por su régimen particular de regulación, no puede afirmarse que contenga un derecho genérico con una efectividad real. El texto del articulado limita este derecho a los sistemas de alto riesgo del Anexo III (excluyendo, por tanto, todos los componentes de seguridad de los productos regulados por actos legislativos del Anexo I), exceptuando a su vez a las infraestructuras críticas de esta protección.

Se establece, como exigencia, que la decisión adoptada tenga un efecto perjudicial para la salud, seguridad, o derechos fundamentales del afectado. Lo único que se especifica en cuanto a la acción debida tras la reclamación es a obtener explicaciones *«claras y significativas acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada»* (art. 86.1).

El apartado 3 del artículo excluye este derecho cuando ya esté previsto de otro modo en el derecho de la UE. Este es precisamente uno de los motivos por los que es posible suponer que en la mayoría de las ocasiones podrá entenderse que el derecho a una explicación previsto por el RGPD hará que se excluya esta obligación adicional [1366].

---

[1363] En este sentido, López-Tarruella Martínez, A., *op. cit*. N. 1322, pp. 897-899, refiriéndose al Reglamento 2019/1020, relativo a la vigilancia del mercado y la conformidad de los productos, destacando que dista mucho de ser un derecho equiparable al derecho a una explicación fijado por el RGPD en su artículo 22.3.

[1364] Sí ha sido, en cambio, establecido así por el TJUE para el caso del derecho a la explicación contemplado en el RGPD. Recientemente, en el caso Schufa *(op cit.* N. 1250). Está por comprobar, por tanto, si estas garantías se trasladan en la aplicación del RIA.

[1365] Se mencionará *infra* el modelo de gobernanza propuesto por el momento por España.

[1366] López-Tarruella Martínez, A., *op. cit*. N. 1322, p. 912 y ss., analizando la relación de este derecho con el RGPD, afirmando que la disposición no aplicaría en prácticamente ningún supuesto, más allá de permitir la reclamación de una explicación a las personas jurídicas, excluidas del derecho a una explicación del RGPD. Los tres requisitos que exige el artículo 22 del RGPD (decisión totalmente automatizada, afectación a datos de carácter personal y efecto legales), serán los que haya que examinar.

Además de la inconcreción respecto de los deberes derivados del derecho, resalta de forma evidente la dificultad de dotar de un contenido sustantivo a esta prerrogativa si tenemos en cuenta lo ya señalado respecto de la explicabilidad genérica de los sistemas de IA. Es decir, existe una dificultad práctica evidente para explicar tanto el rol del sistema usado en la concreta decisión como de los elementos que la han configurado. Debido al carácter opaco de los sistemas, la mera explicación matemática no será suficiente para que la persona afectada por la decisión pueda comprender por qué el sistema ha decidido de la forma en que lo ha hecho [1367].

En suma, tras el examen de sendos artículos, cuyo contenido parece no quedar demasiado claro, cabe destacar que el Reglamento no impide la aplicación del resto de garantías y vías de recurso y reclamación previstas tanto por los ordenamientos nacionales como por el ordenamiento de la Unión. Así, se reconoce expresamente en el artículo 85, *«sin perjuicio de otras vías administrativas o judiciales de recurso»*. En tanto en cuanto el resto de vías de protección quedan disponibles para la salvaguarda de derechos e intereses también protegidos por el Reglamento, en los apartados que siguen realizaremos un breve repaso por la normativa de flanqueo del Reglamento, con el objetivo de acabar de entender el complejo marco de protección, complementario, que se presenta tras la aprobación del RIA.

### e) Recapitulación de los deberes y conclusiones críticas

Una vez examinado el contenido más esencial de la norma de referencia para fijar el riesgo permitido respecto de los sistemas de IA, siempre con el objeto de obtener un mayor conocimiento respecto de las actuaciones adecuadas y del deber objetivo y subjetivo exigible a los distintos operadores, se realizará una recopilación respecto de los deberes y su importancia sustantiva, con el objeto de delimitar aquello que sí puede ser relevante para el derecho penal y aquello que, por su alto contenido formal, debe reservarse a la normativa administrativa, para acabar la reflexión sobre la normativa europea con unas breves conclusiones críticas.

#### 1. *Deberes sustantivos versus deberes formales*

Puede afirmarse sin demasiada complicación que, habida cuenta de la importancia de los diversos derechos en juego en los sistemas de IA de uso

[1367] En este sentido, Agustinoy Guilayn, A., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, p. 775, sugiriendo que este artículo recoge más bien un principio general de interpretabilidad de difícil concreción.

prohibido (que van desde la intimidad, la integridad física y psíquica, la libertad, la presunción de inocencia y la no discriminación), la prohibición de uso de cualquiera de los sistemas de IA de riesgo inaceptable constituye una obligación de contenido indudablemente sustantivo, directamente dirigida a proteger intereses esenciales para los ciudadanos.

En lo relativo a los sistemas de alto riesgo, entre los requisitos esenciales podemos distinguir diferentes tipos de obligaciones.

a. Las obligaciones sustantivas

Así, la obligación de implantación de un sistema de gestión de riesgos capaz de detectar posibles riesgos y reaccionar frente a ellos puede considerarse sustantiva, en tanto es central para una protección *ex ante* de los derechos afectados por los sistemas. Las obligaciones relativas a la gobernanza y calidad de los datos resultan, igualmente, sustantivas, en tanto están directamente orientadas a prevenir discriminaciones, sesgos o errores con efectos nocivos en los sistemas de decisión. Finalmente, también lo serán la obligación de verificar la precisión, robustez y ciberseguridad del sistema, pues se dirige a asegurar la fiabilidad técnica y la protección frente a ataques o fallos que pueden implicar impacto directo en derechos e intereses fundamentales.

b. Las obligaciones formales

Son claramente formales, en cambio, a pesar de constituir requisitos esenciales en el Reglamento, la obligación de mantener la documentación técnica y la de conservación de registros de actividad, en tanto suponen garantías dirigidas a asegurar el cumplimiento de otras obligaciones de carácter sustantivo. Lo mismo puede decirse respecto de las obligaciones dirigidas a cooperar con las autoridades de supervisión, incluidas aquellas que tienen que ver con las (escasas) obligaciones dirigidas a realizar autoevaluaciones de conformidad, cuyo fin primordial es permitir un control o seguimiento del resto de requisitos, tratándose de obligaciones instrumentales [1368].

[1368] Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 122 hace una reflexión valiosa sobre la cooperación con las autoridades, cuestionando si es sustantiva o no. Hablando de las condiciones formales y materiales de permisión administrativa de riesgos, afirma que normalmente las condiciones de permisión administrativa de riesgos tienen en común que contribuyen al control y reducción del riesgo, (ejemplos: medidas de vigilancia de focos de peligro, o recabar el consentimiento de los afectados por un

### c. Las obligaciones intermedias

Dos obligaciones dentro de las enumeradas como esenciales para los sistemas de alto riesgo caerían en una tercera categoría intermedia, pues, debido a su escasa concreción sustantiva, esconden, como se ha explicado, más declaraciones de principios (o máximas de actuación) que reglas cerradas. Es el caso de las obligaciones de transparencia (en todas sus versiones) y las reglas dirigidas a la supervisión humana efectiva. Se ha tratado de argumentar cómo, tanto una como otra, a pesar de tener un contenido mínimo sustantivo, no se puede afirmar que ofrezcan una protección directa o real de los intereses en juego, en tanto su efectividad está puesta en cuestión empíricamente. Por este motivo, es decir, su naturaleza indeterminada que nos permite hablar más de principios que de normas, solo podremos decir que funcionan como obligaciones concretas dirigidas a proteger bienes jurídicos esenciales cuando su vulneración sea absolutamente patente. Así, por poner un ejemplo, será posible afirmar que se afecta de forma sustancial el derecho a la no discriminación si, ante la decisión claramente sesgada de un sistema que deniega sistemáticamente ayudas a un colectivo desfavorecido no existe mecanismo alguno de verificación o supervisión de estas decisiones, pero será difícil afirmar su vulneración en caso de que el supervisor exista, haga su trabajo, pero el resultado resulte ser, finalmente, erróneo [1369]. Lo mismo respecto de las obligaciones que,

---

peligro abstracto), pero, *«Ahora bien, la administración pública también puede condicionar la permisión de un riesgo a la observancia de requisitos que no tengan que ver directamente con la protección de los bienes jurídicos afectados»,* y señala que *«el incumplimiento de condiciones ajenas a la gestión del riesgo excluirá la permisión administrativa. Sin embargo, no obsta a que el juez concluya que la conducta es penalmente atípica»*. Pero esta atipicidad penal no puede basarse en la concurrencia de un riesgo permitido en sentido estricto, sino en la ausencia de una imprudencia (objetiva o subjetiva), del agente. Otro ejemplo de condición similar es la del cumplimiento de deberes de tolerancia pasiva de la inspección de riesgos y la colaboración con esta inspección. En las pp. 123 y 124 no deja clara su postura en cuanto a los deberes de colaborar con las entidades de supervisión (sí menciona que en el caso del delito de negativa a someterse a alcoholemia sí se fundamentó por el STC 161/97 como una cuestión unida a la protección de los bienes; pero que en otros casos de control realizado por entidades privadas puede resultar menos claro). La idea es que habrá que examinar si este deber es sustantivo, es decir, p. 124, *«si un determinado deber de colaboración está orientado a la efectivamente minimización del riesgo generado por la actividad o, por el contrario, tiene un alcance más amplio»*. Aunque se trata de una cuestión discutida aquí se sostiene, de cara a minimizar el papel del derecho penal, que no.

[1369] Por poner un ejemplo acaecido en la realidad. Tras la sugerencia del sistema de predicción del riesgo de violencia contra las mujeres Viogen, el resultado arrojado es de riesgo bajo en un caso donde multiplicidad de factores, no tenidos en cuenta por la herramienta (y siendo esto conocido por el verificador), el supervisor humano decide no corregir la decisión del sistema, con la consecuencia de que el sujeto bajo examen acaba asesinando a su mujer. La referencia de la sentencia por la que se acaba condenando a la Administración a hacer frente a la responsabilidad patrimonial es: SAN 2350/2020, Sala de lo ContenciosoAdministrativo, Sección 5.ª, de 30 de septiembre de 2020. La existencia de un supervisor humano en este caso es, si se quiere, anecdótica o meramente aparente pues, en tanto su labor de confirmación se dirija exclusivamente a confirmar ciegamente la decisión del sistema, no podrá afirmarse que existe supervisión en absoluto. Diferente sería el caso si el policía encargado de verificar este mismo sistema, tras la valoración de los hechos adicionales, hubiera confirmado, a

en materia de supervisión, competen al responsable de despliegue; habrá responsabilidad en caso de apartamiento patente respecto de las instrucciones de uso, así como de falta absoluta de designación de personal supervisor, pero, en tanto estos mínimos se cumplan, será difícil afirmar la vulneración de esta norma.

Mención aparte requieren las obligaciones que tienen que ver con el estadio de poscomercialización de los sistemas donde, habida cuenta de que uno de los rasgos fundamentales de los mismos es su dinamismo y continuo cambio, resultará esencial, y por tanto sustantiva, toda obligación dirigida a mantener el riesgo dentro de los niveles de lo permitido tras la primera fase de comercialización. Así, la obligación de contar con un plan de vigilancia poscomercialización resulta esencial para obtener una protección adecuada frente a los nuevos riesgos, permitiéndose una actualización del riesgo permitido.

Las obligaciones relativas a la actuación tras los incidentes graves tienen, en cambio, una importancia aún mayor, y pueden dar lugar a un régimen de responsabilidad diferente, pues en tanto suponen un aviso en el caso concreto de la existencia, ya demostrada, de un riesgo concreto de dimensiones catastróficas (recuérdese que estos supuestos incluían fallecimientos o graves daños a la salud o derechos fundamentales), activarán un deber mucho más sustantivo, que hará que sea mucho más probable la imputación del resultado final, siempre que se cumpla el resto de requisitos, en comisión por omisión [1370].

---

pesar de todo, la decisión de la máquina, en cuyo caso ningún reproche podría establecerse para el operador obligado (en este caso, la Administración, en calidad de responsable de despliegue).

[1370] El Reglamento prevé un mecanismo de notificación específico en caso de incidente grave (art. 73), que, en virtud del artículo 3.49, se define como: «*un incidente o defecto de funcionamiento de un sistema de IA que, directa o indirectamente, tenga alguna de las siguientes consecuencias: a) el fallecimiento de una persona o un perjuicio grave para su salud; b) una alteración grave e irreversible de la gestión o el funcionamiento de infraestructuras críticas; c) el incumplimiento de obligaciones en virtud del Derecho de la Unión destinadas a proteger los derechos fundamentales; d) daños graves a la propiedad o al medio ambiente*». En caso de que este suceso ocurra, tanto proveedores como responsables de despliegue tienen deber de notificar a las autoridades de vigilancia de mercado en un plazo perentorio de 15 días. Tras el incidente, se prevé la adopción de medidas posteriores, tales como la investigación, la evaluación de los riesgos y de posibles medidas correctoras, así como la obligación de cooperar con las autoridades; todos ellos deberes de contenido muy concreto y directamente dirigidos a evitar que el garante de un riesgo (la fuente de la cual se obtenga la posición puede ser discutida, remitiéndonos a la discusión al respecto del fabricante del producto del capítulo V). Reconociendo la posición de garantes de los operadores económicos del RIA, en calidad de *«garantes de evitación de descontrol del riesgo»,* Roso Cañadillas, R., «Derecho penal adaptado a la IA: entre el derecho penal analógico y el derecho penal digital», *En Letra*, 17, 2023, p. 73. En esta línea, resulta especialmente relevante la propuesta de Roso (desarrollada en nota al pie 68), quien distingue, dentro de la posición de garante, entre el garante defensivo de evitación y el garante agresivo de producción de resultados, con la finalidad expresa de restringir el ámbito de la comisión por omisión. Desde esta perspectiva, no toda infracción omisiva del garante implica la realización del resultado típico, sino que una parte significativa de los supuestos debe reconducirse al ámbito de las omisiones puras cualificadas por la posición de garante. Este esfuerzo por limitar dogmáticamente las posiciones de garante y evitar expansiones automáticas de la responsabilidad penal resulta plenamente compartible. La misma, en una versión más reciente del citado trabajo: «Los retos de la IA para la teoría del delito: ¿un Derecho penal digital?*», Libertas,* Nº. 15, 2026, pp. 90-157,

Podríamos decir que establece así, el Reglamento, el procedimiento específico por el cual puede entenderse que, tras un indicio cualificado de peligrosidad que indica que el riesgo permitido se ha visto superado en un caso concreto, la obligación de actuación será tan esencial que no solo fundamentará un delito de peligro, sino el genuino nacimiento de un deber de evitar. Los deberes de evitación se pueden vincular, directamente, con el procedimiento de actuación tras incidentes graves del RIA, que permite trazar el camino a seguir en cuanto a la notificación y la posibilitación de una investigación efectiva.

### d. Conclusión: escasa garantía de los derechos e indudable necesidad de refuerzo

Aunque el Reglamento constituye un instrumento normativo valioso para la labor de concreción del riesgo permitido en un sector que, hasta su promulgación, se movía en el terreno de la continua incertidumbre jurídica, pudiendo extraerse del mismo contenidos normativos definitivos que permiten su aplicación directa en los ordenamientos nacionales, lo cierto es que deja, como es lógico en un texto con una vocación tan compleja, aun muchas lagunas en cuanto a su concreción, interpretación y aplicación.

Sin perjuicio de que haya que esperar a la labor de concreción realizada, en conjunto, por los tribunales ordinarios encargados de su aplicación, así como por las instituciones europeas, autoridades de vigilancia y organismos privados (en este complejo baile de gestión compartida público-privada), algunas críticas pueden, ya, señalarse, en lo relativo a las lagunas de protección que aún plantea.

Señala la doctrina que uno de los puntos más controvertidos del RIA es, precisamente, la escasa articulación de un régimen robusto de responsabilidad en caso de daños, habiéndose centrado en un enfoque *ex ante* de gestión de riesgos que no se complementa con un marco reparador adecuado, lo cual acaba por minar el impacto de la intervención normativa, que acaba siendo una «obra inacabada» [1371].

A pesar de los esfuerzos paralelos de las instituciones europeas por generar nuevos marcos de responsabilidad civil extracontractual o responsabilidad por el producto, lo cierto es que no se han elaborado en paralelo con el instrumento de aplicación directa más importante. El sistema de sanciones,

---

donde integra y aplica su marco teórico a los problemas específicos derivados de la gobernanza de riesgos tecnológicos.

[1371] Mantelero, A., «El Reglamento…», *op. cit*. N. 1189, p. 58.

finalmente, acaba por ser insuficiente y una evidente tarea pendiente para los Estados miembros, que habrán de manejar este complejo instrumento normativo para su mejor eficacia en la protección de derechos. No cabe duda de que los derechos implicados en los sistemas prohibidos y de alto riesgo que el RIA regula bien merecen una reflexión que se extienda a todos los sectores del ordenamiento.

No falta, de hecho, quien advierte sobre la necesidad de valorar la posibilidad de que sea una Directiva, de contenido penal, la que venga a concretar la aplicación del Reglamento [1372].

## II. OTRAS NORMAS RELEVANTES

Como se ha puesto de manifiesto, el Reglamento no opera en el vacío normativo, por mucho que la cuestión de la que se ocupa sea de una novedad tecnológica apabullante. Resultará recomendable desarrollar, de forma necesariamente breve, cuál es la relación que, a partir de este momento, tienen las normas ya existentes sobre IA, cuando, por diferentes razones, entren en contacto con la nueva norma por excelencia.

### 1. LA IA COMO SISTEMA BASADO EN LA EXPLOTACIÓN DE DATOS (EL REGLAMENTO EUROPEO DE PROTECCIÓN DE DATOS)

Como se ha ido señalando al profundizar en el RIA, si existe una norma directamente vinculada con la inteligencia artificial es la normativa de protección de datos pues, con independencia de su finalidad de uso o su mayor o menor peligrosidad, todo sistema de IA es, en fin, un sistema de procesamiento masivo de datos.

La relación entre RIA y RGPD es, pues, una de las cuestiones clave para entender cómo se protegerá el dato en el nuevo ecosistema digital que surge con la IA. Ambos reglamentos tienen, como se ha puesto de manifiesto, finalidades distintas, pero se hace necesario dotarlos de una interpretación armonizada.

---

[1372] Scollo, L.,., *op. cit.* N. 1190: «*A European directive on criminal matters will have to address and resolve these issues as soon as possible, as they cannot be left to interpretation. This directive will have to make general policy choices in criminal matters, providing a systematic framework for AI concerning the issues of criminality, causality, and culpability*».

Precisamente por la estrechísima semejanza entre uno y otro, será relativamente sencillo fijar su relación: en la medida en que los sistemas de IA tengan capacidad para inferir datos personales (tal y como estos son definidos por la normativa de protección de datos), se activarán todas las garantías del RGPD, sin mayores matices. No es casual, de hecho, que los sistemas de IA a que se refiere el anexo III siempre se considerarán de alto riesgo cuando el sistema de IA efectúe la elaboración de perfiles de personas físicas (art. 6.3, *in fine),* ni que cuando los sistemas de categorización biométrica clasifiquen a las personas sobre la base de sus datos para deducir raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual (categorías de datos especialmente protegidas en la normativa de protección de datos), sean consideradas una de las prácticas prohibidas por el Reglamento (art. 5.1.g).

En todo caso, y sin perjuicio de lo valioso que puede resultar acudir a la interpretación ya recaída sobre determinados elementos de la normativa de protección de datos (esencialmente, como se ha indicado *supra,* la prohibición de tomar decisiones totalmente automatizadas), introduce el RIA una habilitación legal para el tratamiento de categorías especiales [1373] de datos siempre que sea estrictamente necesario para garantizar la detección y corrección de los sesgos asociados a los sistemas de IA de alto riesgo (art. 10.5). Pese a la posible contradicción que, aparentemente, pudiera parecer que surge en este inciso, es aplaudible si se entiende que proviene de una legítima preocupación, expresada repetidamente por la literatura, relativa a la inutilidad de las técnicas de protección contra el sesgo discriminatorio que prohíben usar ciertas categorías de datos, habida cuenta de la innegable capacidad de los sistemas de inferir categorías de datos sin necesidad de que estos sean proporcionados (en lo que se conoce como el problema del *proxy)* [1374].

[1373] Se refiere el RIA, aunque no lo especifique, a los datos sensibles del artículo 9 del RGPD.

[1374] Un *proxy* es un indicador próximo. Se ha hecho patente que el hecho de entrenar a un sistema con un determinado dato que pudiera ser problemático debido a su inclusión dentro de las categorías prohibidas no evita que el sistema acabe infiriéndolo y usándolo. El ejemplo más común en la literatura es el de la raza, que, acaba conociéndose por indicadores próximos (*proxy*) como el código postal del lugar de residencia. Sin perjuicio de la numerosa literatura doctrinal, también la Agencia para la protección de derechos fundamentales puso de manifiesto este problema: European Union Agency for Fundamental Rights (FRA) «*Bias in Algorithms: Artificial Intelligence and Discrimination*», enero de 2023, p. 24, disponible en: <https://fra.europa.eu/en/publication/2022/bias-algorithm> [Accedido por última vez el 29/01/2026].

## 2. NORMATIVA DE CARÁCTER CIVIL: LA DIRECTIVA SOBRE RESPONSABILIDAD POR PRODUCTOS DEFECTUOSOS Y LA PROPUESTA DE DIRECTIVA SOBRE RESPONSABILIDAD CIVIL EXTRACONTRACTUAL EN MATERIA DE IA [1375]

Dos son las iniciativas legislativas que, a nivel comunitario, se presentaron por la Comisión en el año 2022, para regular, desde dos perspectivas, la cuestión de la responsabilidad por daños derivados de inteligencia artificial. Sin embargo, las dos han seguido caminos legislativos bien diferenciados: mientras la propuesta de directiva sobre responsabilidad civil extracontractual se encuentra en un punto muerto, tras un demoledor informe del Parlamento europeo apuntando la equivocada dirección del texto propuesto [1376], la nueva Directiva sobre responsabilidad por productos defectuosos se encuentra ya aprobada (aunque su entrada en vigor no se ha producido por el momento [1377]).

Por este motivo, sin perjuicio de resaltar muy brevemente cuál es el régimen (por lo demás, muy poco novedoso) que proponía la norma sobre responsabilidad extracontractual, que se centraba en exclusiva en aliviar, para determinados casos, la carga de la prueba [1378], estableciendo presunciones de

---

[1375] Las referencias completas a los textos legales son: Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos y por la que se deroga la Directiva 85/374/CEE del Consejo; y Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA).

[1376] El Parlamento Europeo, en su informe de septiembre de 2024 (*Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence: Complementary impact assessment*), propone precisamente avanzar hacia un régimen objetivo, al menos para los sistemas de alto riesgo, definidos conforme al Reglamento de IA. Asimismo, sugiere ampliar el enfoque hacia una directiva general sobre responsabilidad por daños causados por programas informáticos, dada la dificultad de separar conceptualmente la IA de su soporte software y las lagunas existentes en el tratamiento jurídico del software libre y de código abierto. Esta propuesta ha sido abandonada: la Comisión anunció su retirada en su Programa de Trabajo para 2025 (11 de febrero de 2025).

[1377] Estas líneas se escriben a principios del año 2026, mientras la nueva Directiva entra en vigor 24 meses después de su aprobación. Según el artículo 2.1: *«La presente Directiva será aplicable a los productos introducidos en el mercado o puestos en servicio después del 9 de diciembre de 2026»*.

[1378] La Propuesta introducía dos presunciones. En primer lugar, se ordenaba la exhibición de pruebas sobre sistemas de IA de alto riesgo en poder del demandado de forma que, si el demandado no colaborase con dicha orden, se establecía la presunción de incumplimiento del deber de diligencia. En segundo lugar, se establecía en la propuesta una presunción de causalidad del daño cuando el actor acredite tres extremos: que el demandado ha incurrido en culpa; que es razonablemente probable que dicha culpa haya incluido en los resultados generados por el sistema; y que tales resultados hayan causado un daño al demandante. Este régimen pretendía ser aplicable, asimismo, al usuario no profesional del sistema, pero en este caso el juicio sobre la culpa recaerá sobre si el uso se realizó conforme a las instrucciones de uso del proveedor. Cuando se trata de sistemas de IA de alto riesgo, el cumplimiento de las obligaciones que el Reglamento establece deberá probarse como indicio de la culpa. En más profundidad en: Vallespín Pérez, D. «Responsabilidad civil extracontractual en materia de IA: especial referencia a la carga de la prueba y la aplicación de presunciones», *IDP (Revista de Internet y política*), 42, 2025.

causalidad bajo ciertos parámetros, nos centraremos en explicar qué aspectos de la nueva Directiva de responsabilidad por el producto resultan más relevantes, para finalizar con alguna reflexión sobre sus carencias. El diferente enfoque de ambas normas ha sido criticado por una parte de la doctrina, pues, mientras la propuesta de directiva sobre responsabilidad extracontractual estaba presidida por un enfoque de responsabilidad subjetiva de mínimos, es decir, otorgando amplio margen de apreciación a los Estados miembros, la Directiva sobre productos defectuosos parece introducir un régimen de responsabilidad más objetiva con una regulación de máximos o armonizada [1379].

El punto de partida del impulso regulatorio lo podemos situar en la comunicación de la Comisión Europea de 10 de enero de 2017, donde advertía ya sobre la insuficiencia del régimen de responsabilidad objetiva para afrontar los desafíos de los nuevos sistemas de IA. Estas dificultades pivotaban, esencialmente, en torno a la difícil calificación jurídica de los productos inteligentes, cuya autonomía, conectividad y capacidad de aprendizaje dificultaban su encuadramiento dentro del concepto tradicional de «producto», así como la aplicación del estándar de «producto defectuoso» en los contextos de prestación de servicios [1380], así como los obstáculos para probar el nexo causal, habida cuenta de la opacidad y dinamismo de los sistemas. Así, el primer informe del Grupo de expertos de la Comisión propone mantener el principio de responsabilidad objetiva, si bien introduciendo elementos de responsabilidad subjetiva en función del deber de diligencia exigible a los distintos actores del ecosistema digital. Se sugiere, en este sentido, un tratamiento diferenciado según el tipo de operador, el grado de control sobre el sistema y el cumplimiento de obligaciones concretas.

---

[1379] Muñoz García, C., *op. cit. N. 767*, p. 193, denunciando que *«ni se propone el mismo orden normativo, ni iguales reglas de armonización máxima, ni igual responsabilidad objetiva, no los mismos sujetos protegidos, ni los potenciales responsables»*. Pese a las promesas de armonización de la nueva Directiva, cabe vaticinar, que, como venía ocurriendo con la Directiva anterior (84/347), los aplicadores nacionales escapen de la regulación, aplicando preferentemente la responsabilidad por incumplimiento contractual o con dolo, como así ocurre con la aplicación del Reglamento 2023/988 de seguridad general de los productos y la aplicación de los incumplimientos por parte de los operadores de las distintas obligaciones de seguridad. En este sentido: Carrasco Perera, Á., «Análisis de la nueva directiva de responsabilidad por daños causados por defectos de productos. Síntesis y crítica de la regulación más contenida en la Directiva 2024/2853», *Centro de Estudios de Consumo*, 2024, p. 2.

[1380] No obstante, no falta quien asevera que, pese a que la anterior normativa no estuviera pensada para dar cabida a daños causados por estos nuevos productos inteligentes, no parece posible esperar a la entrada en vigor de la nueva Directiva para aplicar el régimen de responsabilidad. Una propuesta completa de interpretación en este sentido en: Herbosa Martínez, I., «Encaje de los sistemas de IA en la definición de producto en la legislación de productos defectuosos Análisis de la legislación vigente con la vista puesta en la Propuesta de Directiva del Parlamento europeo y del Consejo de 28 de septiembre de 2022 (COM/2022/495)», *InDret,* 3/2024, pp. 52-98.

En lo relativo a la consideración del sistema de IA como producto -uno de los primeros obstáculos para la aplicación de la antigua Directiva-, el texto de la nueva Directiva es clarificador, al aclarar en su artículo 4.1 que producto es *«cualquier bien mueble, aun cuando esté incorporado a otro bien mueble o a un bien inmueble o interconectado con estos; incluye la electricidad, los archivos de fabricación digital, las materias primas y los programas informáticos»,* mencionando expresamente la IA en su considerando 13 cuando se ponen ejemplos de programas informáticos, y manifestando que el proveedor de sistemas de IA en el sentido del RIA ha de ser tratado como fabricante [1381]. Sin embargo, hay quien pone en duda que los sistemas de IA que son intangibles puedan entrar a formar parte de este concepto [1382].

En todo caso, ha de tenerse en cuenta que, pese a la posible amplitud del concepto, los daños cubiertos por la Directiva son sustancialmente los mismos que bajo la normativa anterior (daños solo en caso de que el producto no ofrezca la seguridad que una persona tiene derecho a esperar, según el artículo 7.1 [1383]).

---

[1381] Martín Faba, J. M., «La Inteligencia Artificial en la nueva Directiva de responsabilidad por los daños causados por productos defectuosos ¿realidad o expectativa?», *Revista CESCO*, 53, 2025, p. 15. El citado considerando 13 menciona los programas informáticos, los sistemas operativos, los microprogramas, los programas de ordenador, las aplicaciones o los sistemas de IA.

[1382] Así, Martín Faba, J. M., ibid., p. 16, respecto de Chat-GPT, cuya función es generar texto, pero entendiendo que la información no puede considerarse producto, y aunque el archivo digital sí entre en el ámbito de aplicación de la Directiva, el daño cubierto por la norma debe ser causado directamente por el producto defectuoso, y no por la información por él emitida, pues en ella intervienen otros factores intermedios, como la conducta de la propia víctima.

[1383] En este sentido sí se introduce algún cambio, pues la Directiva modifica el enfoque de la evaluación de seguridad, añadiendo nuevas circunstancias a valorar como la presentación que el fabricante hace del producto (con especial atención a las instrucciones de uso), el uso razonablemente previsible, tomando como momento clave el de su introducción en el mercado, pero sin que se acaben aquí las obligaciones del fabricante, acogiéndose las peticiones de la doctrina que aconsejaban tener en cuenta que al tratarse de productos inherentemente dinámicos, el autoaprendizaje (reconociéndose responsabilidad del fabricante por las actualizaciones o mejoras del producto) o la interconectividad son especialmente importantes para fijar el momento en que surgen las obligaciones. Estableciéndose que la seguridad debe evaluarse considerando tanto las expectativas del público como las de los usuarios finales: Atienza Navarro, M. L., «¿Una nueva responsabilidad por productos defectuosos? Notas a la Propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por daños causados por productos defectuosos de 28 de septiembre de 2022 (COM/2022/495)», *InDret* 2.23, p. 13 y ss. Sin embargo, y aunque el legislador sí tiene en cuenta la autonomía y el autoaprendizaje inherentes, no resulta en absoluto claro de qué manera interpretar sus referencias en la Directiva. Explica Atienza (p. 17): *«lo que hace la Propuesta de Directiva es incluir la capacidad de autoaprendizaje de algunos sistemas tecnológicamente avanzados como una circunstancia para valorar su seguridad. Sin embargo, suscita dudas cómo jugará a la hora de apreciar la existencia del defecto. Dicho en otras palabras: no acaba de entenderse que significa esta referencia. ¿Acaso el legislador europeo identifica autoaprendizaje, y la autonomía que lleva consigo, con defecto?; ¿la autonomía en sí misma considerada y, por ende, la imprevisibilidad/impredecibilidad del comportamiento del robot lo convierten en inseguro y por ello en defectuoso?; ¿el autoaprendizaje significa, por sí mismo, que el sistema tiene propiedades peligrosas?; ¿O se considerará defectuoso cuando, fruto del aprendizaje, adopte soluciones erróneas? Dar una respuesta afirmativa a las preguntas anteriores implica un freno al progreso y al desarrollo tecnológico, con lo que no creo que ésa haya sido la intención del legislador europeo»,* apuntando que más que un defecto se trata de «*una de sus características principa-*

Se trata de: daños personales (muerte, lesiones corporales y psíquicas, incluyéndose los daños morales); daños sobre cualquier otro bien que sea distinto del producto; y daños sobre los datos (su destrucción o corrupción) usados con fines no profesionales, pero quedando fuera otro tipo de eventos dañosos relevantes. Así, aclara el considerando 24 que los daños que no están contemplados expresamente en la presente Directiva (como las pérdidas económicas puras, las vulneraciones del derecho a la intimidad o los actos discriminatorios) no deben, por sí solos, generar responsabilidad conforme al régimen establecido en ella, sin que la Directiva limite el derecho a reclamar una indemnización por cualquier otro tipo de daño, incluidos los perjuicios morales, conforme a los regímenes de responsabilidad aplicables fuera de su ámbito.

La nueva Directiva se aplicará exclusivamente a fabricantes o productores en el sentido amplio del Reglamento de IA (es decir, incluyendo al importador, el representante autorizado y el distribuidor, cada uno en su respectivo ámbito [1384]), dejando, por tanto, de lado la responsabilidad del usuario u operador [1385]. Solo pueden reclamar la aplicación de la norma las personas físicas, dejando fuera a las personas jurídicas como posibles legitimadas para la reclamación, cuestión que ha suscitado críticas entre la doctrina [1386].

La carga de la prueba relativa al defecto del producto, al daño sufrido y a la relación de causalidad entre ambos sigue recayendo, como principio general, sobre la parte perjudicada. No obstante, la normativa introduce mecanismos que pueden considerarse una cuasi reasignación de dicha carga probato-

---

*les»,* señalándose la paradoja de que cuanto más autoaprendizaje realicen, más seguros pueden llegar a ser. La reflexión lleva a Atienza a afirmar que quizá, precisamente por esta característica de autonomía y aprendizaje, la normativa más adecuada a aplicar a estos sistemas no sea la de productos defectuosos, dejando ésta reservada a los productos deterministas más sencillos de IA, y aplicando la normativa de responsabilidad civil extracontractual a los verdaderos sistemas autónomos, pp. 18-19.

[1384] La Directiva prevé que también responda en calidad de fabricante quien realice una modificación sustancial del producto (art. 8.2) y reconoce la responsabilidad de un nuevo operador: el prestador de servicios de tramitación de pedidos a distancia (que son aquellos que almacenan, embalan y despachan productos sin ser propietarios). Además de establecer un régimen de responsabilidad en caso de concurrencia de sujetos responsables, se prevé la limitación de responsabilidad en caso de contribución de la víctima.

[1385] Wagner lo valora positivamente. Wagner, G., «Liability rules for the Digital Age -Aiming for the Brussels Effect-», *Journal of European Tort Law,* 13 (3), 2022, p. 196, afirmando que «*Una estrategia dirigida y adaptada al desafío específico que plantean los sistemas digitales autónomos parece más atractiva. El núcleo del desafío que presenta esta nueva tecnología es que el usuario pierde gran parte de su influencia sobre el "comportamiento" del sistema digital*». (traducción propia). No podemos más que estar de acuerdo con esta afirmación. En p. 197: «*Las normas de responsabilidad deben generar los incentivos económicos necesarios para que el fabricante cumpla con sus responsabilidades en cuanto a la seguridad del producto. Esta es la premisa que debe guiar la evaluación de las propuestas para reformar las normas de responsabilidad aplicables a los sistemas digitales.»* (traducción propia).

[1386] Atienza Navarro, M. L., *op. cit*. N. 1383, p. 6. También: Gómez Ligüerre, C., «La Propuesta de Directiva…, *op. cit*. N. 130, p. II.

ria, con el fin de facilitar el acceso efectivo a la tutela judicial en contextos marcados por una especial complejidad técnica o informativa.

En esta línea, el artículo 9 incorpora un procedimiento de exhibición de pruebas, mecanismo que permite al demandante solicitar al demandado la exhibición de información relevante que obre en su poder. Asimismo, el artículo 10 contempla la presunción de la existencia del defecto del producto en determinados supuestos. Esta presunción opera: cuando el demandado no haya cumplido con el deber de exhibición de pruebas; cuando el producto no respete las disposiciones legales o técnicas relativas a seguridad o ciberseguridad [1387]; o bien cuando el daño se haya producido en el marco de un uso previsible o en condiciones normales de funcionamiento. Este último supuesto parece responder a la intención del legislador europeo de recoger una regla general sobre el uso ordinario del producto.

Igualmente, se introduce una presunción del nexo causal cuando el tipo de daño sufrido corresponde típicamente a los efectos esperables de un defecto en un producto de la clase en cuestión. Esta presunción puede activarse también cuando el demandante enfrente dificultades excesivas para acreditar el vínculo causal, atendiendo a la complejidad técnica o a la asimetría de información respecto del productor o proveedor. En todo caso, como es señalado por la doctrina, la normativa adolece de una gran indeterminación que podría abocar en su falta de aplicación [1388].

En términos generales, el artículo 10 refleja una tentativa del legislador europeo de equilibrar la posición procesal entre demandantes y demandados en un contexto de creciente sofisticación tecnológica. No obstante, diversos pasajes del precepto han sido objeto de crítica por su deficiente técnica normativa, lo que genera incertidumbre interpretativa y plantea dudas sobre su adecuada transposición al Derecho nacional [1389]. Se mantiene el ya asentado régi-

---

[1387] En este caso en concreto, la Directiva exige en su artículo 10.2.b) no cumplir con *«los requisitos obligatorios de seguridad del producto* ***establecidos en el Derecho de la Unión o nacional que tienen por objeto proteger contra el riesgo del daño sufrido*** *por la persona perjudicada»* (la negrita es nuestra). Parece evidente, así, la necesidad de discernir, como aquí se ha tratado de hacer *supra,* entre los requisitos de la normativa que cumplen este fin, a efectos, como señala Atienza en el plano civil, de que pueda operar el criterio de imputación objetiva del daño del «fin de protección de la norma violada». Atienza Navarro, M. L., *ibid.,* p. 35.

[1388] Atienza Navarro, M. L., *ibid.* p. 37: «*la propia definición de las «dificultades excesivas» suscita la duda de a quién debe resultarle difícil la prueba, ¿al demandante considerado subjetivamente o, por el contrario, se trata de una referencia objetiva? En la misma línea crítica, la exigencia, para que se aplique la presunción, de que sea probable que el producto fue defectuoso o que la defectuosidad fue la causa «probable» del daño, plantea el interrogante de cómo puede probarse que el defecto fue una causa del daño sin probar su carácter defectuoso, y, sobre todo, provoca el inconveniente de cómo valorar esa probabilidad (¿bastaría demostrar un 51% de probabilidad o debería ser superior?)».*

[1389] Muy crítico: Carrasco Perera, Á., *op cit*. N. 1379, p. 5-6.

men de exoneraciones en lo relativo a la excepción por el estado del conocimiento técnico (los denominados riesgos de desarrollo, arts. 11. e) y 18), con todos los problemas que en materia de armonización ha suscitado, por la diversidad de regímenes de excepción en los ordenamientos nacionales, y se introduce una novedosa causa de excepción a la exoneración, bastante acertada. Teniendo en cuenta que estas tecnologías permiten a los fabricantes ejercer el control más allá de su introducción en el mercado, *«se excluye la posibilidad de que se liberen cuando el defecto aparezca después y sea debido a servicios conexos o programas informáticos que estén bajo su control, ya sea en forma de mejoras o actualizaciones o de algoritmos de aprendizaje automático (art. 10.2.a) y b)».* Lo mismo respecto de la imposibilidad de exoneración de responsabilidad en caso de que el programa no se actualice o mejore [1390].

En definitiva, el marco, aun confuso e incompleto, en el que habrán de convivir los dos regímenes (el régimen de seguridad de los productos propio del ámbito empresarial, de un lado, y el régimen general de responsabilidad por el uso, de otro), presenta carencias importantes y, especialmente, lo hace respecto de uno de los factores más necesitados de concreción en el ámbito civil, como es la determinación del deber de cuidado [1391]. En este punto, no podemos más que mostrarnos de acuerdo con Gómez Ligüerre cuando afirma que *«a pesar de la frenética actividad legislativa en materia de daños causados por el uso de tecnologías emergentes, la iniciativa que todavía queda pendiente es la relativa a la difícil y compleja labor de diseñar unas reglas específicas de responsabilidad para los daños causados por la inteligencia artificial»* [1392].

Compartimos, finalmente, las conclusiones de Atienza, que pone de manifiesto que, pese a los avances que ha supuesto la nueva Directiva (incluir el *software* o sus

---

[1390] Atienza Navarro, M. L., *op. cit*. N. 1383, p. 42.

[1391] Desde luego, no pretendemos afirmar que la determinación del deber de cuidado propia del derecho civil haya de trasladarse, sin más matices, al derecho penal, pero es evidente que la interpretación que respecto del deber objetivo de cuidado fije la normativa civil servirá como punto de partida al derecho penal, y es quizá esta la tarea pendiente de la Directiva, que acaba por aclarar más bien poco sobre la cuestión de fondo aquí planteada. En lo relativo a la relación entre el deber de diligencia civil y penal, concretamente en lo que tiene que ver con el campo en desarrollo de la conducción con vehículos autónomos, nos remitimos a Valerius, que concluye que el futuro desarrollo de la responsabilidad penal por negligencia en la conducción autónoma dependerá en gran medida de cómo el legislador y la jurisprudencia logren equilibrar la protección eficaz de bienes jurídicos fundamentales, como la vida y la integridad física, con la necesidad de mantener la previsibilidad y la seguridad jurídica para los fabricantes, titulares y conductores de vehículos autónomos. La cuestión central será si se opta por una asimilación casi total a los principios de responsabilidad civil, con su catálogo extenso y en parte abierto de deberes de diligencia, o si se establecen límites claros y específicos para la responsabilidad penal, que tengan en cuenta tanto el avance tecnológico como el principio de legalidad penal. Valerius, B, «Sorgfaltspflichten beim autonomen Fahren*», op. cit.* N., p. 15 y ss.

[1392] Gómez Ligüerre, C., «Responsabilidad por daños causados por la inteligencia artificial», *InDret*, 1/2025, p. x.

actualizaciones como producto, ampliar la noción de productor, cambiar el concepto de defecto), los problemas principales se mantienen. Así, el conflicto surge cuando el daño se origina en una decisión tomada por el propio sistema inteligente, como resultado de un proceso de autoaprendizaje, y que solo habría podido evitarse privándolo de esa capacidad y autonomía. Para los operadores económicos, este tipo de situaciones podría considerarse (y previsiblemente será alegado como tal) un supuesto de riesgo de desarrollo. La posibilidad de que estos operadores queden exonerados solo podría eliminarse si se interpreta que la autonomía del sistema constituye en sí misma un defecto, o bien si se suprime en este ámbito la exoneración por riesgos de desarrollo. Ambas opciones conllevan consecuencias negativas similares: si se califica la autonomía como un defecto, se desvaloriza el avance tecnológico; si se elimina la exoneración, se incrementa la carga de responsabilidad sobre los operadores económicos, lo que podría desincentivar la inversión en innovación [1393].

Dicho conflicto, que se encuentra en realidad en la base del problema mismo de la inteligencia artificial, no ha quedado resuelto con los loables intentos de armonización europea.

## 3. *SOFT LAW* Y PRINCIPIOS ÉTICOS. ESPECIAL MENCIÓN AL CONVENIO MARCO DEL CONSEJO DE EUROPA [1394]

El Convenio Marco sobre Inteligencia Artificial del Consejo de Europa establece un marco jurídico vinculante destinado a garantizar que el desarrollo y uso de los sistemas de inteligencia artificial respeten los derechos humanos, la democracia y el Estado de Derecho, constituyendo, por su alcance [1395] y actualidad, uno de los marcos más sólidos de *soft law* [1396], al ser el primer tratado

[1393] Atienza Navarro, M. L., *op. cit*. N. 1383, p. 48.

[1394] La referencia completa es: *Convenio marco sobre la inteligencia artificial, los derechos humanos, la democracia y el Estado de derecho*, en adelante CMIA o, simplemente, Convenio.

[1395] Los 46 Estados miembros del Consejo de Europa, la Unión Europea y 11 Estados no miembros (Argentina, Australia, Canadá, Costa Rica, Estados Unidos de América, Israel, Japón, México, Perú, la Santa Sede y Uruguay) negociaron el tratado, que ha sido firmado finalmente por Andorra, Georgia, Islandia, Liechtenstein, Montenegro, Noruega, Moldavia, San Marino, Suiza, Ucrania, Reino Unido, Canadá, Israel, Japón, Estados Unidos y la Unión Europea.

[1396] Los instrumentos de *soft law* en el campo de la inteligencia artificial resultan, en la actualidad, inabarcables, y consideramos por ello que no merece la pena perderse en este vasto campo *(vid.* a título de ejemplo: Valls Prieto, J., *Inteligencia artificial…, op cit*. N. 46, pp. 47 y ss., explicando el contenido de hasta diez documentos de *soft law*), amén de apuntar que el Convenio que aquí se analiza, aunque vincula en teoría a la Unión Europea por haber sido acogido formalmente por ella, ha sido firmado, también, por países como Estados Unidos o Israel, lo que nos hace vaticinar que, en la línea de lo afirmado por Cotino, habrá aquí muchísimo más «verso» que prosa. Aba Catoira*, et al., op. cit. N. 1025*, p. 79.

internacional vinculante en su ámbito, vinculando tanto a las autoridades públicas como a actores privados.

Establece obligaciones generales para los Estados parte, quienes deben adoptar medidas normativas, institucionales y presupuestarias adecuadas para asegurar la compatibilidad de la IA con dichos principios fundamentales. En particular, se exige prevenir cualquier uso de estos sistemas que comprometa la integridad de las instituciones democráticas o el acceso equitativo al debate público, consagrando principios esenciales, ya consabidos, como la dignidad humana, la transparencia, la responsabilidad, la no discriminación, la protección de datos personales, la fiabilidad y la promoción de una innovación segura y controlada. El Convenio, más allá de su valor ético, representa un instrumento normativo con capacidad de integración sustantiva en los ordenamientos jurídicos nacionales, superando, en cierta medida, el enfoque meramente declarativo del *soft law,* estableciendo también ciertos requisitos de gestión de riesgos (principalmente la obligación de realizar evaluaciones de riesgos e impactos sobre los derechos humanos, la democracia y el Estado de derecho). A diferencia del Reglamento, el Convenio fija un marco de cooperación internacional que, basado en orientaciones flexibles, está llamado a integrar en un ámbito propiamente jurídico lo que, hasta el momento, no eran más que principios éticos.

Finalmente, el texto contempla la articulación con el Derecho de la Unión Europea, permitiendo su aplicación sin perjuicio de las competencias y normas vigentes en el ámbito comunitario, destacando así su potencial como referente normativo en la gobernanza internacional de la inteligencia artificial. [1397]

Habrá que esperar para ver de qué manera el Convenio despliega alguna eficacia, pero, habida cuenta de su capacidad de integración casi constitucional en los ordenamientos de cada uno de los Estados parte, puede resultar interesante por su potencial interpretativo, de forma que sirva a la integración y aplicación de la normativa vinculante en caso de dudas [1398] o bien, como algún autor ha sugerido ya, por entender que establece principios generales que servirán, en el futuro, de marco general para la regulación de la IA [1399]. En todo caso, que el Tribunal Europeo de Derechos humanos tenga ocasión de pronunciarse sobre el particular hará que, previsiblemente, los hasta ahora intermina-

[1397] Así, el Convenio se adopta finalmente por medio de la Decisión (EU) 2024/2218 del Consejo, de 28 de agosto de 2024, relativa a la firma, en nombre de la Unión Europea, del Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho.

[1398] En este sentido, Cotino Hueso, L., en Aba Catoira, *et al., op. cit. N. 1025*, p. 79. También Ziller, J., «El Convenio del Consejo de Europa de inteligencia artificial frente al Reglamento de la Unión Europea: dos instrumentos jurídicos muy diversos», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, p. 70.

[1399] Tudela, J., en Aba Catoira, A. *et al., op. cit. N. 1025*, p. 84.

bles listados de principios éticos con términos no del todo definidos puedan ir acotándose, por lo que, con cautela, puede decirse que este tratado internacional es, en principio, una buena noticia.

## 4. NORMATIVA DE EJECUCIÓN: EL ANTEPROYECTO DE LEY PARA EL BUEN USO Y LA GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL

Para completar el examen de la normativa relevante aplicable, resta mencionar cuál es el sistema de supervisión del Reglamento, tanto a nivel europeo como nacional. El RIA establece un sistema de gobernanza multinivel, en el que los Estados miembros conservan las funciones de supervisión, a excepción de la supervisión relativa a los modelos de propósito general, cuya regulación se centraliza a nivel comunitario.

Podemos caracterizar la estructura de gobernanza como un modelo concéntrico [1400], siendo el primer círculo el Reglamento como norma suprema habilitante; el segundo círculo los organismos notificantes y notificados que verifican la conformidad con los estándares de protección y finalmente un tercer círculo constituido por las autoridades de supervisión y control *a posteriori.*

El papel de los Estados miembros es, en este punto, el de designar a sus autoridades nacionales competentes, tanto notificantes como de vigilancia de mercado, siendo estas últimas las encargadas de aplicar el Reglamento y de supervisar su cumplimiento. Las autoridades notificantes, también nacionales, son las encargadas de designar y evaluar a los organismos notificados.

Con este esquema complejo basado en la cooperación y la supervisión escalonada en mente, pasaremos a examinar muy brevemente la norma española que vendrá a desarrollar, previsiblemente, este sistema de gobernanza, dando cumplimiento a los mandatos del Reglamento. Nos referimos al Anteproyecto de ley para el buen uso y la gobernanza de la inteligencia artificial [1401].

El texto del anteproyecto consta de 37 artículos, distribuidos en cuatro capítulos, y aborda desde las disposiciones generales hasta las sanciones, pa-

---

[1400] Así, Hernández Peña, J. C., *op. cit*. N. 1274, p. 848: «*La gobernanza y vigilancia del Reglamento de inteligencia artificial: autoridades de vigilancia del mercado, Comisión y las diversas entidades*».

[1401] El Anteproyecto fue aprobado por el consejo de ministros el pasado 11 de marzo de 2025. El Reglamento establece como fecha máxima para adoptar un régimen sancionador nacional el 2 de agosto de 2025 (por aplicación del artículo 113.b). A fecha de mayo de 2026, el Anteproyecto sigue en tramitación por la vía de urgencia y no ha sido aún aprobado como ley; su último hito reseñable es el Dictamen 03/2026 del Consejo Económico y Social, de 25 de marzo de 2026.

sando por la gobernanza y la clasificación de las infracciones. Además, se prevé la creación de una base de datos nacional para registrar los sistemas de IA utilizados en infraestructuras críticas.

En lo relativo a la gobernanza, el Anteproyecto designa, en su artículo 4, a la Secretaría de Estado de Digitalización e IA como autoridad notificante. A nivel interno, será la Agencia Española de Supervisión de la IA (AESIA), la encargada de la gobernanza y supervisión nacional. El artículo 6 designa a las autoridades de vigilancia de mercado, especializándose cada una en un ámbito de aplicación. La AESIA será la entidad con potestad sancionadora en la mayor parte de las prácticas prohibidas, así como buena parte de los sistemas de alto riesgo del Anexo III. En sus ámbitos competenciales respectivos: la Agencia española de protección de datos (protección de datos y algunas prácticas prohibidas [1402] y sistemas de alto riesgo directamente relacionados con este aspecto), el Banco de España y la CNMV (sistemas de evaluación de solvencia o calificación crediticia); la Dirección general de seguros y fondos de pensiones (ámbito de seguros de vida y salud); la Dirección de Supervisión y Control de Protección de Datos del Consejo General del Poder Judicial (sistemas de IA en la administración de justicia); y la Junta electoral central (sistemas que afecten a procesos democráticos).

En lo relativo a las prácticas prohibidas, el Anteproyecto se limita a señalar que quedará prohibida la introducción en el mercado, puesta en servicio o uso de un sistema de IA prohibida en virtud del artículo 5 del RIA, desarrollando de forma detallada el régimen de autorización excepcional para el supuesto de identificación biométrica en tiempo real (art. 11) [1403].

Finalmente, el Anteproyecto pretende dar cumplimiento al mandato de determinación del régimen sancionador, estableciéndose en el artículo 13 tres niveles de infracciones administrativas: muy graves, graves y leves. El artículo 12 señala, para despejar cualquier duda, que, en la medida en que las acciones u omisiones no estén tipificadas como delito, pero contravengan el Reglamento, tendrán carácter de infracciones administrativas. La cuantía de las sanciones se especifica, sin separarse del marco establecido en el Reglamento, respe-

[1402] Prácticas prohibidas (art. 5.1 d), g) y h), sistemas de alto riesgo en: Biometría para cumplimiento del derecho o gestión de fronteras, (salvo identificación directa), y sistemas de alto riesgo en: cumplimiento del derecho, migración, asilo y control fronterizo. Para mayor profundización en las implicaciones y un cuadro comparativo con cada uno de los artículos del RIA y el Anteproyecto: Fernández Hernández, C./ Eguiluz Castañeira, J. A., «Análisis del Anteproyecto de Ley para el buen uso y la gobernanza de la Inteligencia Artificial», *Derecho Digital e Innovación*, 23, 2025.

[1403] Se determinan con detalle cuáles son los supuestos que podrán autorizarse, concediendo a los juzgados de lo contencioso-administrativo la potestad de conceder las autorizaciones, e introduciendo un procedimiento administrativo para la solicitud, concesión e informe del uso.

tando los límites máximos y fijando los mínimos, de forma que el margen de la cuantía económica de las multas administrativas a imponer queda configurado como se muestra en la tabla siguiente:

| Tipo de infracción | Importe fijo (€) | Importe alternativo para sociedades (% volumen de negocio global) |
|---|---|---|
| Muy grave (IA prohibida). | 7.500.001-35.000.000 | 2%-7 % |
| Muy grave (IA de alto riesgo). | 7.500.001-15.000.000 | 2%-3 % |
| Grave. | 500.001-7.500.000 | 1%-2 % |
| Leve. | 6.000-500.000 | 0,5%-1 % |

*Tabla 1. Elaboración propia a partir de datos del Anteproyecto*

Sin embargo, la verdadera labor de concreción que viene a realizar este instrumento es la tipificación individualizada de las infracciones de cada clase.

El artículo 14 establece la tipificación de infracciones muy graves, que son, esencialmente, las relacionadas con el incumplimiento de las prácticas prohibidas (en general, el uso de cualquiera de ellas [1404] y el incumplimiento de los requisitos que dan lugar a un uso legítimo de la identificación biométrica en tiempo real [1405]), la ausencia de notificación de un incidente grave, el incumplimiento de las medidas impuestas por la autoridad de vigilancia, y la reiteración en la comisión de infracciones graves.

Los artículos 15 a 20 establecen una extensísima tipificación de infracciones graves, diferenciando en función del operador jurídico, y recogiendo, esencialmente, la contravención de cada una de las obligaciones previstas en el Reglamento.

En el artículo 15 se prevén las infracciones graves aplicables a cualquier operador en general, que consisten en: la resistencia a someterse a actuaciones de evaluación o verificación, el incumplimiento de las medidas provisionales impuestas por las autoridades de mercado, o la resistencia a que se llevan a cabo estas actuaciones de vigilancia, la presentación deficiente de información

[1404] Adviértase, por ser estrictos en cuanto a las conductas objeto de sanción, que para la generalidad de prácticas prohibidas, lo que se prohíbe (tanto en el RIA como en el Anteproyecto) es tanto la introducción o puesta en servicio de estos sistemas en el mercado de la Unión como su uso, mientras que, en el caso concreto de la identificación biométrica remota en tiempo real en espacios de acceso público (letra 5.1.h) del RIA), lo único que se castiga es el «uso».

[1405] Art. 14.1. letras b)-f), que van desde la falta de autorización judicial hasta el no respeto a las limitaciones temporales, geográficas o personales.

requerida por la autoridad de vigilancia y, finalmente, la reiteración de infracciones graves en el plazo de dos años. El artículo 16 establece otro listado de infracciones graves para los proveedores [1406], el artículo 17 hace lo propio para los representantes autorizados [1407], el 18 respecto de los importadores [1408]; el 19 respecto de los distribuidores [1409], y el 20 respecto de los responsables de despliegue [1410]. Finalmente, el artículo 21 establece las infracciones para los organismos notificados.

Los artículos 22 a 26 reproducen el mismo esquema para las infracciones leves. Cabe matizar que aquí, en cuanto a la tipificación general, aplicable a cualquier operador, la delimitación se hace de forma negativa, es decir, todo aquello que suponga un incumplimiento del Reglamento y no previsto como infracción grave o muy grave tendrá la consideración de infracción leve. No merece la pena, pues, ahondar en la prolija enumeración de conductas, pero

[1406] Las infracciones consisten en: No documentar ni registrar un sistema de IA excluido indebidamente como no alto riesgo; No informar a las personas de que interactúan con un sistema de IA; No marcar el contenido sintético como generado por IA; Comercializar sistemas de IA de alto riesgo que no cumplen los requisitos técnicos; No demostrar la conformidad del sistema de IA de alto riesgo cuando lo exige la autoridad; Comercializar sistemas de IA de alto riesgo sin sistema de gestión de calidad; No redactar ni conservar la documentación técnica obligatoria; No conservar los registros generados automáticamente por el sistema; No adoptar medidas correctoras ni facilitar información requerida ante riesgos o incumplimientos; No nombrar representante autorizado o impedirle cumplir su función si el proveedor está fuera de la UE; Comercializar sistemas de IA de alto riesgo sin evaluación de conformidad; No registrar el sistema de IA de alto riesgo en la base de datos nacional; Incumplir las condiciones exigidas para pruebas en condiciones reales; Incumplir las medidas impuestas por la autoridad durante pruebas en condiciones reales.

[1407] No registrar el sistema de IA de alto riesgo en la base de datos nacional cuando están obligados a hacerlo; Incumplir las obligaciones generales establecidas para los representantes autorizados.; No poner fin al mandato ni informar cuando detectan incumplimientos del proveedor.

[1408] No verificar que el sistema de IA cumple con el Reglamento antes de importarlo.; Introducir en el mercado un sistema con dudas fundadas sobre su autenticidad o conformidad; No indicar su nombre, marca registrada o datos de contacto en el sistema; No garantizar que el sistema se mantiene conforme durante el almacenamiento o transporte; No proporcionar a las autoridades la información necesaria para demostrar la conformidad del sistema

[1409] No verificar que el sistema esté marcado correctamente y acompañado de la documentación obligatoria; No actuar ante sospechas de que el sistema no cumple con los requisitos del Reglamento; No asegurar condiciones adecuadas de almacenamiento y transporte que garanticen la conformidad del sistema; No colaborar con las autoridades competentes en las medidas adoptadas respecto al sistema.

[1410] No adoptar medidas técnicas y organizativas para usar el sistema conforme a las instrucciones; No asignar la supervisión humana del sistema a personal cualificado y autorizado; No garantizar que los datos de entrada sean adecuados y representativos para el fin del sistema.; No realizar vigilancia poscomercialización ni comunicar riesgos o incidentes relevantes.; No conservar los registros generados automáticamente por el sistema bajo su control.; No informar a los trabajadores y sus representantes sobre el uso del sistema en el entorno laboral; No usar la información del proveedor al realizar una evaluación de impacto de protección de datos; Usar incorrectamente un sistema de identificación biométrica remota en diferido, por demora o incumplimiento de requisitos; No informar a las personas físicas de que el sistema interviene en decisiones que les afectan; No cooperar con las autoridades competentes en la aplicación del Reglamento; No realizar una evaluación de impacto sobre derechos fundamentales cuando es obligatorio.

como muestra de la efectiva levedad de las mismas, cabe resaltar que se establece como infracción del proveedor no indicar la información de contacto del proveedor en el embalaje del sistema. Los criterios de graduación de las sanciones (art. 27) reproducen los establecidos en el Reglamento, añadiendo simplemente una mención a la reiteración o reincidencia en el caso de las pymes [1411] y estableciendo un criterio relacionado con la existencia de sucesión de empresas o la aplicación de la doctrina del levantamiento del velo en situaciones abusivas o fraudulentas.

En definitiva, una vez examinado el texto provisional que viene a concretar el régimen de supervisión y sanción del Reglamento, puede afirmarse que España continúa siendo uno de los países más atentos a la implementación de las políticas en materia de IA de su entorno. Sin embargo, al mismo tiempo puede afirmarse que, aunque la presente normativa pretende el establecimiento de un sistema efectivo, proporcionado y disuasorio (como así consta ya en la exposición de motivos), el legislador español ha optado por el camino fácil, tipificando cada uno de los incumplimientos del Reglamento en una regulación que, por lo apegada al texto literal del RIA, reproducirá necesariamente sus problemas de interpretación. No es esta, empero, una mala señal, pero parece evidente que la idea del legislador nacional, parcelando la intervención de las autoridades nacionales de supervisión por sectores de especialización y poniendo especial énfasis en los deberes de colaboración con estas autoridades, viene a establecer un régimen sancionatorio altamente formal, donde, sin perjuicio de que los criterios de graduación tengan en cuenta aspectos como la intencionalidad o la gravedad del daño causado, por su contenido absolutamente omnicomprensivo y la inclusión de infracciones de dudosa insignificancia, no solo permite dudar de la proporcionalidad de las mismas, sino que nos lleva a pensar en un sistema sancionador de aplicación automática altamente formal. Lejos queda, por tanto, en este estricto régimen, la protección de derechos sustantivos (tanto de las posibles víctimas de los daños de los sistemas, como de los operadores económicos implicados en las distintas fases).

En definitiva, el régimen sancionador administrativo no constituye un mecanismo suficiente de protección y prevención respecto de los posibles daños con relevancia penal que, en algunos campos concretos, los sistemas puedan causar.

[1411] Uno de los objetivos patentes de la normativa (ya señalados en el RIA) consiste en la aplicación de un régimen privilegiado a las pymes, y así se reafirma en el presente Anteproyecto, que prevé un régimen privilegiado de sanciones para estas empresas.

# CAPÍTULO OCTAVO. UNA PROPUESTA DE *LEGE FERENDA:* DELITOS CONTRA LA SEGURIDAD COLECTIVA TECNOLÓGICA

*Debajo de las multiplicaciones*
*hay una gota de sangre de pato.*
*Debajo de las divisiones*
*hay una gota de sangre de marinero.*
*Debajo de las sumas, un río de sangre tierna;*
*un río que viene cantando*
*por los dormitorios de los arrabales,*
*y es plata, cemento o brisa*
*en el alba mentida de New York* [1412]

El inigualable poeta granadino alzaba con esta magistral fuerza la voz desde su extranjería emocional y cultural ante una ciudad entregada a los ritmos frenéticos de la productividad. En un sentido paralelo se procede aquí, no desde el prisma del poeta, sino desde la visión del jurista, observador atento de los cambios sociales, con la imperiosa necesidad de interrogarnos ante los nuevos lenguajes del poder.

Si en su tiempo las multiplicaciones, divisiones y sumas representaban el orden técnico y capitalista que encubría la violencia estructural, hoy el álgebra de la inteligencia artificial corre el riesgo de esconder un nuevo río de sangre: más invisible, más silencioso, pero no menos real. Debajo de cada línea de código, de cada algoritmo optimizador, puede haber una gota de sangre despersonalizada, la historia de un sujeto reducido a dato, a patrón, a probabilidad. La IA, en su promesa de eficiencia y predicción, amenaza con repetir el gesto que Lorca denunciaba en su tormentoso viaje a la ciudad que nunca

[1412] García Lorca, F., «Nueva York, oficina y denuncia», *Un poeta en Nueva York*, 1940.

duerme: sustituir el rostro por la cifra, el cuerpo por el rendimiento, la comunidad por el control.

Desde nuestro lugar, no tomamos la palabra para denunciar con la fuerza de un poeta, pero sí para proponer de forma cautelosa. El derecho penal no puede asistir pasivamente a esta transformación. Debe erigirse, si no como dique, al menos como brújula. No para sofocar la innovación, sino para proteger al individuo allí donde el sistema (digital esta vez) tiende a disolver su singularidad.

Porque como se ha tratado de exponer a lo largo de la investigación, debajo de las estadísticas predictivas puede haber, como expresaba Lorca, un río de sangre tierna: la dignidad diluida de los vulnerables, los sesgos que repiten la exclusión, las decisiones que se toman sin rostro ni relato. Frente a ello, el papel del jurista no es menor: ser límite frente al cálculo que no conoce compasión ni justicia. El capítulo que culmina esta investigación condensa, así, una propuesta final, directamente encaminada a que el Código Penal español, habida cuenta de lo irrefrenable de la técnica, se modifique para dar cabida a la protección necesaria ante los riesgos hasta ahora dibujados.

## I. INTRODUCCIÓN

A lo largo de la presente investigación se ha tratado de abordar el problema de los *hard AI crimes* desde diferentes perspectivas, explorando tanto las propuestas de otros modelos (principalmente el anglosajón, con una lógica que bebe de las propuestas más objetivadas), como la solución que nuestros ordenamientos continentales han venido dando a los supuestos con ciertos rasgos comunes, para acabar conviniendo que, si hay un modelo óptimo para afrontar las dificultades de imputación de daños propias del esquema tradicional del delito de resultado, es el del delito de peligro abstracto de idoneidad. Hemos señalado, asimismo, que los problemas apuntados tanto por la doctrina angloamericana como por la continental son comunes, y que pueden resumirse en una preocupación por la lesividad de los nuevos delitos, un riesgo de contravención del principio de culpabilidad, y una inquietud porque el castigo resulte desproporcionado.

Es el momento, por fin, de ofrecer un modelo con capacidad de superar todos estos inconvenientes, que, sin romper el aparato instrumental con que el penalista está familiarizado, proteja de manera integral todos los bienes jurídicos en juego en lo que se ha denominado la nueva sociedad del riesgo tecnológico, donde las formas de ataque a los bienes clásicos se caracterizan

por lo difuso de sus víctimas y de los bienes en peligro. Trataremos de ofrecer, en primer lugar, un modelo de tipificación, dando cuenta de sus ventajas e inconvenientes, para luego precisar y justificar algunos de sus aspectos concretos. La propuesta de tipificación se recoge, finalmente, en el Anexo I de este trabajo.

## II. LA PROPUESTA DE *LEGE FERENDA*

El punto de partida de la propuesta es, como se ha indicado, el de tipificar delitos de peligro abstracto basados en la inobservancia de deberes, cuyo anclaje encontramos en la normativa vinculante (el Reglamento de IA). En el capítulo anterior se ha tratado de realizar una delimitación teleológica y penalmente relevante de los deberes impuestos por el Reglamento, distinguiendo entre aquellos deberes cuyo incumplimiento puede (y debe) tener relevancia penal, por estar orientados a la protección de bienes jurídico-penales incontrovertidos, y aquellos otros deberes de tipo meramente formal o procedimental, cuya infracción no debería ser penalizada, porque su función principal es organizativa, documental o técnica, sin que tengan conexión directa (o suficientemente directa) con un bien jurídico penal.

El objetivo es, en todo caso, evitar el riesgo de expansión ilegítima del derecho penal y atender al principio de *ultima ratio.*

La metodología de la que se parte, pues, puede resumirse en tres puntos, que han tratado de responderse en los capítulos anteriores: en primer lugar, la pregunta acerca de la finalidad protectora explícita del bien jurídico penal dentro del Reglamento (es decir, ¿un determinado deber está diseñado para evitar directamente daños a bienes fundamentales?); en segundo lugar, la capacidad del incumplimiento para generar riesgos penalmente relevantes en el caso concreto, incluso aunque no se produzca el daño efectivo (es decir, la relevancia del criterio de idoneidad del delito de peligro abstracto); y, finalmente, la posibilidad de encontrar una analogía con otros delitos del Código penal español (es decir, ¿existen otros tipos que, por su estructura y finalidad, prevean el castigo por el incumplimiento de obligaciones altamente tecnificadas?).

Tras el análisis exhaustivo desarrollado en el capítulo anterior, donde identificábamos deberes sustantivos, formales e intermedios, podemos concluir con la siguiente diferenciación, dirigida a responder la pregunta acerca de si un deber está diseñado para evitar daños a bienes jurídico-penales.

## 1. LOS DEBERES SUSTANTIVOS

Incluimos aquí, los siguientes deberes:

- Deber de no usar o introducir en el mercado sistemas prohibidos.
- Deber de implantación de un sistema de gestión de riesgos.
- Deber de establecer un sistema de gobernanza y calidad de los datos.
- Deber de garantizar la exactitud, robustez y ciberseguridad del sistema.
- Deber de supervisión humana (con matices).
- Deber de transparencia (con matices).
- Obligaciones poscomercialización (con matices).

En los cuatro primeros supuestos (uso de sistemas prohibidos, y deberes sobre el sistema de gestión de riesgos, gobernanza de datos y requisitos de robustez técnica), es innegable que, en la medida en que aplican en sectores de actividad altamente peligrosos (nos remitimos a la delimitación de la categoría de sistemas de alto riesgo del capítulo anterior, así como la *ratio legis* en el caso de los sistemas prohibidos), su incumplimiento afecta directamente a bienes esenciales para el derecho penal. Una ausencia de un sistema de gestión de riesgos puede repercutir en daños a la integridad física o salud de los individuos; una ausencia de control sobre la calidad de los datos puede generar, directamente, discriminaciones indetectables frente a colectivos; y una falta de robustez o seguridad técnica puede dar lugar a accidentes que afecten a la seguridad, intimidad o integridad física de las personas. Recuérdese que los delitos de lesión pertinentes no serán, normalmente, aplicables, pues no habrá daño informático ante una vulnerabilidad no pretendida, ni discriminación punible si el resultado discriminatorio deriva de las propias debilidades del sistema.

Decimos que los tres últimos deberes presentan matices por los motivos que también se han desarrollado *supra*. En cuanto a la supervisión humana y la transparencia, se ha explicado cómo la indeterminación de los deberes exigibles a los operadores obligados es máxima. Se hace difícil saber, pues, en qué consiste la conducta debida en este caso, o más bien, cuál es el resultado esperable tras el cumplimiento de dichos deberes. Por ello consideramos que, en tanto el contenido de las obligaciones de supervisión y transparencia esté por determinar, teniendo estas obligaciones más características de principios que de verdaderas normas, solo se habrá de reputar penalmente relevante el incumplimiento patente y flagrante de estos deberes. Es decir, solo una ausencia absoluta de supervisión y transparencia, allí donde el cumplimiento de la obligación sea posible y el contenido claro, dará lugar a la infracción de un

deber con relevancia penal. Por ser claros, una ausencia absoluta de mecanismos de supervisión será punible, no así el establecimiento de un mecanismo que, tras una actuación de revisión concreta, no sea capaz de evitar una decisión errónea del sistema.

Finalmente, respecto de las obligaciones posteriores a la comercialización, cabe reiterar la idea ya señalada *supra*. Se trata, sin duda, de obligaciones sustantivas, en tanto están dirigidas a prevenir el daño propio de la evolución y autoaprendizaje de los sistemas cambiantes. La ausencia total de estos mecanismos de revisión posteriores a la comercialización tendrá, pues, relevancia penal. No obstante, una vez sucedido un incidente de los calificados como graves por el Reglamento, así como de cualquier incidente que sirva de indicio irrefutable de peligrosidad de un sistema, cualquier actuación no encaminada al restablecimiento de la legalidad abandonará ya el ámbito del riesgo permitido y de la infracción de una obligación dedicada a prevenir un peligro desconocido *ex ante,* para entrar a conformar ya una obligación de evitación en sentido estricto, donde será difícil negar que exista –en el caso de los operadores directamente responsables del sistema, proveedor y responsable de despliegue, respectivamente– una posición de garante. Será este el momento en que entre en juego la aplicación clásica del delito en comisión por omisión, del que no nos ocuparemos, por ser aquí aplicables los postulados tradicionales.

## 2. LOS DEBERES FORMALES

Podrá decirse, en general, que el resto de obligaciones del Reglamento, si bien esenciales desde el punto de vista administrativo, por su lejanía respecto de la protección del bien jurídico penal, habrán de quedar extramuros de los nuevos tipos. Así:

- Obligación de inscripción de los sistemas en bases de datos de la UE.
- Obligación de conservación de documentación técnica.
- Obligación de registro de actividad del sistema.
- Deberes de notificación y cooperación con organismos.

En los supuestos aquí enumerados se persiguen finalidades centradas en la posibilitación del control por parte de la Administración. Evidentemente, el fin último de todas ellas consiste en asegurar la integridad de los bienes jurídicos finales, pero se considera que, habida cuenta del alejamiento evidente respecto del bien final, su tratamiento como deberes con contenido penal supondría la tipificación de

una amenaza de contenido demasiado formal, sin que sea evidente, por otra parte, que el mero incumplimiento de estos requisitos haya de poner en riesgo ningún bien jurídico más allá de la capacidad de ordenación de la Administración [1413]. Un sistema sin control sobre el dato será un sistema directamente expuesto al sesgo discriminatorio, pero un sistema que no haya conservado debidamente sus registros durante el tiempo estipulado puede ser perfectamente inocuo.

Partiendo, por lo tanto, de esta delimitación clara de los deberes fijados extrapenalmente, podemos avanzar en la propuesta. El segundo de los puntos a tener en cuenta es, como se ha indicado, la búsqueda de un criterio que condicione la idoneidad de la intervención penal.

En este sentido, son posibles diferentes modelos, una vez que partimos de la exclusión del delito imprudente de resultado por los motivos explicados a lo largo de la investigación y esencialmente en el capítulo quinto. Si la idea es plantear delitos cuya conducta típica consista en la infracción de las obligaciones sustantivas arriba expuestas, lo siguiente será establecer cuál será el marcador para delimitar la intervención penal, que no puede ser otro que la existencia de un riesgo especialmente grave o cualificado.

Tres alternativas son posibles, aquí, en abstracto: que el marcador de peligro sea un daño concreto que actúe como condición objetiva de punibilidad; que el daño sea un resultado de peligro concreto (tal y como lo interpreta la doctrina mayoritaria, recuérdese, implica que el autor haya podido representarse el daño, que actúa como resultado del delito); o bien, finalmente, en forma de condición de idoneidad de un delito de peligro abstracto-concreto. Se descarta por completo la opción del delito de peligro abstracto presunto por su formalidad, en atención a los motivos expuestos en capítulo sexto.

## 3. LA EXCLUSIÓN DE LA CONDICIÓN OBJETIVA DE PUNIBILIDAD

Ya Hilgendorf apuntaba, en un escrito incipiente sobre sistemas autónomos [1414], la posibilidad de acudir al modelo del delito de peligro abstracto para

[1413] En este punto se toma partido por una concreta opción político-criminal, pero siendo conscientes de que no es la única de las posibles y que, en ocasiones, teniendo en cuenta la importancia de los bienes en juego y, seguramente, lo habitual de su puesta en peligro, se ha optado por tipificar penalmente este deber de obediencia. Así, paradigmáticamente, y con aval constitucional, en el delito de negativa a someterse a las pruebas de alcoholemia. Por ello, en caso de que este tipo de hechos diera lugar habitualmente, por ejemplo, a un incremento en los accidentes de sistemas sin marcado de conformidad, podría plantearse otra situación. El punto de partida que aquí se usa es que, en principio, la menor protección penal será la más respetuosa con los principios constitucionales. *Vid.:* STC 161/97, de 2 de octubre.

[1414] Hilgendorf, E., «Autonome Systeme, künstliche Intelligenz..., *op. cit*. N. 872, p. 111.

los sistemas autónomos que causaren daños, en un intento de eludir la responsabilidad objetiva que introducía el modelo de la ley de productos defectuosos. Sin embargo, en su propósito por poner límite a la intervención penal, proponía que este delito exigiera la concurrencia de un daño como condición objetiva de punibilidad [1415].

Ha de tenerse en cuenta, en este punto, que las condiciones objetivas de punibilidad son definidas como elementos ajenos al contenido del injusto penal específico propio de la conducta típica [1416]. Lo definitivo de las mismas es que no deben ser abarcadas por el dolo. Por tanto, el problema que plantean estas condiciones es el siguiente: la concurrencia de un perjuicio o daño improbable no debería tener incidencia para incrementar la punibilidad del hecho. Y por ello, si las condiciones objetivas de punibilidad tienen como efecto un castigo con una pena no merecida o excesiva, serán ilegítimas [1417]. Queda un hueco, no obstante, para la legitimidad de las condiciones objetivas de punibilidad, que serán aquellas, y solo aquellas, cuya concurrencia impida o limite la imposición de la pena, hablando aquí la doctrina de condiciones propias o auténticas, siendo solo las impropias o falsas las que plantean problemas de constitucionalidad por vulneración del principio de imputación subjetiva [1418] y, añadimos, de culpabilidad en sentido amplio. Siguiendo esta sencilla clasificación nos tememos, por tanto, que la condición que plantea Hilgendorf adolecería de las críticas aquí expuestas.

Habida cuenta de lo expresado *supra* (capítulo tercero), al explorar los argumentos opuestos por la doctrina anglosajona respecto de la *strict liability,* Spender y Pedain afirmaban que, en el contexto continental, en nuestro intento por mantener intacto el principio de culpabilidad, quizá acabásemos filtrando «por la puerta de atrás», una reminiscencia del *versari in re illicita* [1419]*,* que tan

---

[1415] *Idem*: «*eines abstrakten Gefährdungsdelikts des Inverkehrbringens gefährlicher Produkte ohne hinreichende Sicherung, eventuell in Kombination mit dem Eintritt eines Schadens als objektiver Bedingun der Strafbarkeit*».

[1416] Silva Sánchez, J. M., *Derecho Penal. Parte General, op cit*. N. 521, 49/27.

[1417] *Ibid.,* 51/27 *«[…] el ámbito del merecimiento del agente se circunscribe a su comportamiento en tanto que lesivo de la norma de conducta e imputado subjetivamente».* En este sentido, rechazando la propuesta de Hilgendorf del delito de peligro con condición objetiva de punibilidad por considerarla contraria al principio de responsabilidad subjetiva: Ibold V., *op. cit*. N. 104, p. 413.

[1418] *Idem.* Silva pone como ejemplo de condiciones auténticas y, por tanto, legítimas, las condiciones de perseguibilidad de determinados delitos, como es el caso de las insolvencias punibles, o la condición de reciprocidad del artículo 606. 2 CP para castigar el delito de atentado contra un jefe de estado extranjero en España.

[1419] Spencer, J. R., / Pedain, A., *op. cit*. N. 388. En un sentido similar, específicamente sobre los delitos de manipulación de mercado y la tradicional discusión sobre la condición de resultado/condición objetiva de ciertos elementos exigidos en el tipo: Feijoo Sánchez, B., «Los delitos de manipulación …», *op. cit*. N. 1193, p. 14: «*La doctrina considera con buen criterio que la regla debe ser que los elementos del tipo no son sólo elementos del tipo legal, sino también del tipo de injusto (49). Aquí se comparte esta*

mala fama se ha granjeado en nuestro entorno. No se acoge aquí, pues, la propuesta de Hilgendorf, aun teniendo en cuenta su mayor valor en cuanto a la concreción de la peligrosidad, por dos motivos: primero, por ser una manera «fácil» de superar las dificultades de imputación del resultado, simplemente aludiendo al mismo (entendemos que con el riesgo, por cierto, de que dicho resultado condicione una pena superior, aunque ni siquiera se trate de imputarlo al autor), y porque nos parece que en los contextos descritos de riesgos de ataques indeterminados a víctimas colectivas y bienes variados, quizá esperar al acaecimiento de un resultado reste eficacia preventiva al tipo penal.

## 4. LA EXCLUSIÓN DEL DELITO DE PELIGRO CONCRETO

Hemos de partir de la clasificación y caracterización que se realizaba en el capítulo sexto para diferenciar uno y otro tipo de delitos de peligro, según la doctrina mayoritaria. Si el delito de peligro concreto exige que se haya dado una situación de peligro que es altamente probable que se materialice, generando una situación que permite afirmar que dicho resultado es un resultado de peligro, nos encontraremos ante un escenario que exige, al menos, un fuerte indicio de concreción del peligro en un resultado de lesión. Indicio, por cierto, que ha de ser, según la doctrina mayoritaria, conocido, asumido o querido por el sujeto [1420].

El primer y principal inconveniente de este tipo de delitos deriva, pues, de su pacífica condición como delito de resultado, lo cual nos llevará a todos los problemas que, en sede de imputación del resultado lesivo, se han puesto de manifiesto en el capítulo quinto. Es decir, si el problema de la imprevisibilidad intrínseca unida a la autonomía y opacidad del sistema impide afirmar la relación de riesgo, porque no es descartable que haya sido otro (u otros) factores ajenos al primigenio incremento de riesgo el que contribuye finalmente al que se materialice el resultado (en este caso el resultado de peligro), imputar dicho resultado supondrá una presunción contra reo inadecuada e ilegítima.

---

*tesis por ser la más compatible con el principio de responsabilidad subjetiva consagrado en el artículo 5 CP. Las condiciones objetivas de punibilidad dan lugar a una especie de responsabilidad estricta u objetiva de tal manera que el sujeto puede ser responsable –por un delito doloso– de consecuencias o hechos de muy difícil previsión».*

[1420] En función de la concepción de dolo que se sostenga la representación del «resultado de peligro» habrá de reunir más o menos elementos pero, en definitiva, y como mínimo, ha de ser conocido de forma evidente por el sujeto. *Vid.,* en detalle, con un excelente repaso por la polémica del elemento subjetivo en el delito de peligro concreto, Rodríguez Montañés, T., *Delitos de peligro…*, *op. cit*. N. 934, pp. 141 y ss.

Un segundo argumento que señala los inconvenientes del delito de peligro concreto lo apunta Silva, con el que coincidimos, si bien con algunos matices [1421]. Cuando estamos ante actividades reguladas administrativamente, esa regulación extrapenal delimita el riesgo permitido en el ámbito penal. En tales casos, el legislador ha de ser cauto al elegir delitos de peligro concreto. Si se produce el indicio de peligrosidad concreta para uno de los bienes jurídicos que Silva llama «intrínsecos» –que, atendida su posición global, hemos de entender como eminentemente personales–, ese indicio podría bastar para activar deberes reforzados de evitación en otros contextos. Sin embargo, al permanecer dentro del margen del riesgo permitido, podrán darse situaciones de impunidad que resulten intolerables.

Lo explica Silva con un ejemplo sencillo. Si la normativa administrativa de seguridad de los trabajadores quedase obsoleta, ello impediría aplicar el delito de peligro concreto previsto por el legislador, que parte de la base de que se han de incumplir los deberes previstos por la normativa laboral. Es decir, ante un indicio evidente y grave de peligrosidad (para la vida o integridad física) en el puesto de trabajo, el empleador podrá eludir la responsabilidad penal (en aplicación ajustada a derecho del tipo), pues no contraviene la normativa, por mucho que sea consciente de su inadecuación para el caso concreto, ante un peligro inminente y grave. Lo mismo podría predicarse, y quizá en muchísimos más casos, con la normativa relativa a la inteligencia artificial, que se prevé, por sus características, vaya a quedar obsoleta o inadecuada con el paso del tiempo.

La no aplicación de los tipos penales, por ceñirse la conducta al (obsoleto) riesgo permitido, se dará también si los tipos existentes son de peligro abstracto, pero en este caso no se habrá producido ningún indicio de peligrosidad cualificada, por lo que no habrá un indicador concreto de riesgo que debiera activar obligaciones más cualificadas de evitación, siendo en este caso la falta de respuesta penal menos preocupante.

En definitiva, exigir un peligro concreto en situaciones donde se trata de proteger, siquiera mediatamente, como es el caso, bienes jurídico-penales de importancia primordial (sean estos o no intrínsecos o, en terminología de Silva, que dan lugar a delitos *«mala prohibita»),* bloquea, *de facto,* la posibilidad de exigir responsabilidad por la lesión, por muy palmario que sea el indicio de peligrosidad. No nos parece esta la mejor respuesta político-criminal.

---

[1421] Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, pp. 150 y ss. Nuestro matiz se deriva, únicamente, de que no compartimos la separación de la que parte Silva entre delitos *«mala in se»* y *«mala quia prohibita»,* que parece estar detrás de esta obligación positiva del Estado de actuar castigando allá donde los bienes sean merecedores de una protección casi sacra.

## 5. EL MODELO DEL DELITO DE IDONEIDAD

El modelo elegido será, pues, el de la cláusula de idoneidad, identificado tradicionalmente con el modelo de delito abstracto-concreto o hipotético *(supra,* capítulo sexto).

Se trata de buscar cláusulas de idoneidad adecuadas, para dejar claro que, en todo caso, estos tipos penales no se aplicarán con la mera infracción formal de la normativa administrativa, sino que este incumplimiento ha de ser idóneo para comprometer los diversos bienes implicados. En esencia: salud, libertad, igualdad e intimidad, e integridad física, psíquica y moral.

El segundo de los criterios de delimitación de la gravedad será la exclusiva previsión de la modalidad dolosa. Así, no solamente se exigirá que se hayan omitido los deberes más esenciales previstos en la normativa, sino que los sujetos obligados y capacitados para su cumplimiento lo hayan hecho a sabiendas de su obligación. Es necesario aclarar aquí que el dolo ha de referirse no al resultado el resultado sobre los bienes finales, sino sobre las condiciones de seguridad establecidas en la normativa [1422]. Aunque lo más normal será que el incumplimiento, además de ser doloso, dé lugar a una conducta objetivamente idónea para causar un daño a los bienes finales, este indicio de idoneidad lesiva podrá ser refutado, alegando, por ejemplo, el cumplimiento de requisitos adicionales no previstos directamente, dirigidos a mantener limitado el riesgo a un nivel equivalente al permitido por el Reglamento. Piénsese, por ejemplo, en una empresa australiana que decide vender sus sistemas de IA en el mercado europeo y, aunque no cuente con un sistema de gobernanza de datos que cumpla los requisitos del artículo 10 del RIA, es capaz de demostrar que su sistema de tratamiento es suficiente para prevenir los sesgos más habituales en sistemas equivalentes porque, por ejemplo, evite que sus decisiones sean adoptadas sin consentimiento del sometido a la decisión.

Recuérdese que, ya de partida, la gravedad de los daños viene delimitada por el ámbito propio de actuación de los sistemas de alto riesgo y los sistemas prohibidos, que solo operan en contextos especialmente sensibles. Nada impide, no obstante, que el legislador, en uso de su soberanía nacional, decida incluir

[1422] En este punto solo podemos añadir, además de lo explicado en capítulo sexto, que el sujeto no ha de representarse el daño finalmente producido, que en muchas ocasiones será si no absolutamente imprevisible sí difícilmente asumible en su concreta configuración, por ser dependiente de factores completamente desconocidos, por la interacción de diversos factores y agentes (humanos y no humanos). Lo que debe representarse es, simplemente, el incumplimiento de obligaciones, cuya existencia es condición esencial e imprescindible para la protección de los bienes finales, y que suponen, en los términos explicados *supra,* un descontrol o desorganización sobre su concreta esfera de incumbencia o ámbito propio de responsabilidad.

dentro de los usos prohibidos nuevos ámbitos de aplicación, siempre y cuando se puedan mantener los criterios aquí expuestos que justifiquen su inclusión en el derecho penal. Una opción muy común en la doctrina es la relativa a incluir una prohibición (penal) de uso de armas autónomas. Nada impediría, trasladando los postulados aquí mantenidos, incluir dicha prohibición penal.

## 6. UBICACIÓN SISTEMÁTICA

Como se ha venido adelantando, la propuesta aquí formulada trata de proteger bienes jurídico penales muy variados (que incluyen la vida, integridad física y psíquica, salud, intimidad, igualdad, libertad o patrimonio), pero lo hace desde una lógica muy específica, que no es otra que la necesidad de prevención sistemática del riesgo generado por tecnologías potencialmente peligrosas para personas indeterminadas y bienes indeterminados.

Esta estructura, como se ha apuntado *supra,* responde a la lógica de los delitos contra la seguridad colectiva, donde es muy común la estructura del delito de peligro abstracto basado en el incumplimiento de deberes técnicos. Así, la propuesta que aquí se presenta cabría sin problemas dentro del Título XVII, quizá pasando a crearse una nueva sección o capítulo dedicado a «Delitos contra la seguridad tecnológica e inteligencia artificial».

Esta forma de proceder presenta, somos conscientes, ventajas e inconvenientes. Parece evidente que los tipos penales que castiguen los incumplimientos de deberes sustantivos relativos a sistemas de alto riesgo encontrarán un buen acomodo entre los delitos contra la seguridad colectiva, habida cuenta de que responden a la dinámica de tipificación expresa del incumplimiento de requisitos técnicos, muy propia de este Capítulo.

Mayores dudas ofrece, en cambio, el acomodo de los tipos penales que tipifiquen el uso de un sistema de IA clasificada como prohibida, pues aquí, como se ha tratado de explicar, los bienes jurídicos implicados serán muy diferentes en cada contexto de uso. Así, y a grandes rasgos, podría defenderse como coherente la introducción de nuevos tipos penales en las siguientes secciones:

– Dentro de los delitos contra la integridad moral (Título VII) podría tener lógica cabida un nuevo tipo que castigase la manipulación subliminal con idoneidad para causar un daño físico o psicológico grave [1423].

[1423] No cabría aplicar ninguno de los delitos del Título, por cuanto el artículo 173.1 CP, protegiendo la forma de ataque a la integridad moral más básica, exige en todo caso un trato doloso. No es necesario

– Dentro de los delitos contra la Constitución (Título XXI, Capítulo IV), tendrían perfecta cabida los delitos de explotación de vulnerabilidades de grupo o la puntuación social generalizada, así como el perfilado con fines penales, todos ellos por el especial riesgo de discriminación en sectores o contra colectivos especialmente sensibles [1424].

– Dentro de los delitos contra la intimidad (Título X [1425]), podrían tener cabida los casos de uso de sistemas de reconocimiento emocional en el trabajo o la educación, la creación de bases de datos biométricos mediante imágenes sin consentimiento y la identificación biométrica en tiempo real.

Como enseguida habrá comprobado el lector, resulta difícil encajar estas nuevas formas de ataque en un solo bien jurídico, y, aunque en algunos casos resulta más claro que en otros, la dispersión normativa podría resultar un inconveniente.

Es posible encontrar, sin embargo, la lógica propia de la seguridad colectiva una vez más. No existen en estos supuestos víctimas determinadas, sino una colectividad afectada o potencialmente afectada (piénsese, por ejemplo, en los ciudadanos expuestos a vigilancia masiva). Los bienes jurídicos afectados son múltiples (al igual que ocurre en los delitos contra la seguridad colectiva, que protegen salud, integridad física y, en algunos casos, incluso medioambiente o patrimonio). Por tanto, no es necesario para este tipo de delitos que el bien jurídico afectado sea único. En todos los casos, será difícil detectar una

---

que el proveedor persiga infligir este trato, sino que, como consecuencia de la habitual monetización de estas conductas, este daño colateral no perseguido dolosamente se produzca, quedando impune.

[1424] Siguiendo la misma lógica que en la nota anterior, el Título XXI contempla, en su Capítulo IV, delitos relativos al ejercicio de los derechos fundamentales y libertades públicas, entre los que se encuentran los delitos de denegación de servicios por motivos discriminatorios en el sector público y privado (arts. 511 y 512, respectivamente). Ninguno de ellos, empero, sería suficiente para dar cabida a supuestos de discriminación propios de la lógica de los sistemas de IA, explicados *supra,* donde la tipificación de la modalidad imprudente sería, en todo caso, insuficiente para evitar los efectos secundarios no deseados ni representados por los operadores.

[1425] Una vez más, las disposiciones del Título, con los delitos contra la intimidad concebidos respecto de los ataques generalmente realizados, dejan fuera el ámbito de aplicación propio de las conexiones difícilmente aprehensibles de los sistemas de IA. La inferencia de emociones, la extracción masiva de datos no necesariamente sensibles pero que den lugar al conocimiento ulterior (por medio de la lógica del cálculo masivo y el *proxy)* de datos protegidos, amén de dar lugar a vulneraciones a la intimidad, pueden afectar a otros derechos de forma indirecta, si tenemos en cuenta que el paradigma de la hipervigilancia puede convertirse, *de facto,* en un obstáculo a la libertad de movimientos en sentido amplio. La sensación de sentirse sometido a una vigilancia invisible respecto de todo tipo de datos conlleva, en no pocas ocasiones, presupuesto de la comisión de delitos contra la libertad como el conocido como *stalking (*172 ter CP, exigiendo la alteración del normal desarrollo de la vida cotidiana). No debemos dejar de lado el potencial de estas conductas que, si bien no encajan con un solo tipo existente, se parecen, si no en la dinámica comisiva, por no perseguirse este hostigamiento de forma directa sino, una vez más, como «daño colateral», sí en el resultado.

lesión concreta del bien jurídico afectado *ex post,* por lo que la lógica de la anticipación de la respuesta tiene pleno sentido.

La diferencia clave respecto de este nuevo título dentro del capítulo de delitos contra la seguridad colectiva es que los bienes jurídicos afectados son nuevos, de corte más individual en su concepción tradicional; así, la intimidad, la libertad y la igualdad. Recordemos que, tal y como ocurrió en su momento con la categoría de los delitos contra la salud pública o salud alimentaria, el bien jurídico que se protege, en sentido amplio, y como bien jurídico mediato, es la «seguridad en el entorno algorítmico», que solo quedará garantizada en tanto se cumplan las condiciones esenciales para generar esta especie de confianza pública en el entorno digital automatizado, que integra, entre otras condiciones clásicas como la no causación de daños físicos, la ausencia de discriminaciones, la ausencia de vigilancia indiscriminada o la falta de manipulación conductual.

La ventaja de la coherencia sistemática implica que será más fácil a un eventual aplicador interpretar que la lógica propia de estos delitos es la misma: la del peligro abstracto colectivo con idoneidad lesiva, adaptada a los bienes jurídicos estructuralmente afectados por las tecnologías emergentes, conservando así la coherencia técnica, aunque se renuncie a la sistematicidad clásica, poniendo el foco en la lógica del riesgo sistémico y la necesidad de protección *ex ante* por las especiales características de los ataques.

La propuesta final del articulado, habida cuenta de las particularidades aquí previstas, se adjunta en el Anexo I de la presente obra.

## 7. LAS DIFICULTADES DEL MODELO DE LEY PENAL EN BLANCO CON REMISIÓN ABSOLUTA

Si atendemos a los tipos finalmente propuestos parece innegable su consideración de tipos penales en blanco, figura, por otra parte, muy frecuente cuando se trata de regular delitos de peligro abstracto. Como bien es sabido, las únicas restricciones fuertes de este tipo de normas, ya asentadas de forma pacífica por la doctrina constitucional, es que, en cumplimiento del principio de legalidad, contengan el núcleo esencial de la prohibición, estén justificadas en razón del bien jurídico protegido y presenten una claridad sobre el núcleo de la conducta prohibida [1426]. Una vez superados estos, empero, hemos de tener

[1426] En este sentido, por todas, la STC 127/1990, FJ3: *«[…] que el reenvío normativo sea expreso y esté justificado en razón del bien jurídico protegido por la norma penal; que la ley, además de señalar la pena, contenga el núcleo esencial de la prohibición y sea satisfecha la exigencia de certeza […]».*

en cuenta cuáles son los efectos más sutiles que despliega esta concreta técnica legislativa, y que han sido puestos de manifiesto, entre nosotros, por autores tanto alemanes como españoles.

En relación con el modelo de «multinormatividad» introducido por el legislador europeo en la regulación de la seguridad del producto, Ibold pone de manifiesto cómo se ha de valorar la afectación de los principios constitucionales, especialmente el principio de legalidad [1427]. Mientras que en el modelo más tradicional de determinación del estándar de actuación debida caso por caso por medio de la jurisprudencia sufre irremediablemente el mandato de determinación del citado principio, si se adopta el modelo de la norma penal en blanco se verá resentido el mandato de reserva de ley. A la hora de poner en una balanza ambas soluciones se habrá de considerar, pues, cuánto estamos dispuestos a sacrificar de cada uno de ellos, teniendo en cuenta que la seguridad jurídica y el principio democrático están, en cada uno de los casos, en juego.

Por su parte, Silva destaca que cuando el legislador opta por remitirse al ordenamiento extrapenal de modo expreso, se reduce la posibilidad de que el juez se separe de las disposiciones administrativas, lo cual, en la terminología empleada por este autor, conduce a un efecto bloqueo [1428]. Dicho efecto bloqueo tiene, por supuesto, sus ventajas, principalmente en cuanto a la seguridad jurídica e igualdad en la aplicación uniforme de la ley, resultando mucho más favorecedor que la técnica de los delitos imprudentes con elementos normativos, donde la determinación del cuidado exigible o del riesgo permitido queda, completamente, en manos de los diferentes aplicadores, con muy pocas pautas para su formulación uniforme [1429].

Sin embargo, y a pesar de las ventajas que en materia de seguridad jurídica e igualdad presenta este modelo, los inconvenientes han de hacerse notar también, pues no tratamos de esconder las vicisitudes aplicativas que puede conllevar una propuesta como la que aquí se formula. Señala Silva dos variantes de este bloqueo condicionado por la remisión en bloque a la normativa extrapenal. En primer lugar, un bloqueo prohibitivo, en tanto el aplicador habrá de asumir las disposiciones administrativas que prohíben una determinada actividad (es el caso, por excelencia, de la prohibición de los sistemas de riesgo inaceptable). No le preocupa demasiado a Silva este efecto, pues, si bien la prohibición administrativa habrá encontrado su fundamento en razones, en principio, ajenas al ordenamiento penal (ya hemos examinado, cómo, en defi-

[1427] Ibold, V., *op. cit*. N. 104, p. 127 y ss.

[1428] Silva Sánchez, J. M., *El riesgo permitido, op. cit.,* N. 886, p. 146.

[1429] Silva hace, en este sentido, una interesante propuesta de concreción, que bautiza de *«modelo de actuación para el juez penal». Ibid.,* p. 137 y ss.

nitiva, son un reflejo del legítimo principio de precaución), la exigencia de ulteriores elementos propiamente penales para la imputación del delito (la idoneidad lesiva para bienes jurídico penales [1430] y el dolo, esencialmente) logran impregnar de una lógica algo más acorde con los postulados del derecho penal a estas prohibiciones.

Sin embargo, preocupa más a Silva la variante permisiva del bloqueo, que es la que tiene que ver con las disposiciones administrativas que, bajo ciertas condiciones, permiten la actividad. Es aquí donde se desarrolla propiamente la eficacia del riesgo permitido, y donde se produce el riesgo de que la permisión de una conducta por parte de la normativa a la que el tipo se remite bloquee la tipicidad de un hecho que, en realidad, sigue siendo peligroso. Sin embargo, llegados a este punto no podemos más que afirmar nuestro compromiso con lo afirmado *supra* respecto de la accesoriedad en el capítulo séptimo, pues, y en la lógica de lo también defendido por Silva [1431], el delito de peligro abstracto convive con esta tensión propia del riesgo permitido (hay restricciones de libertad, en este caso a los operadores económicos de un sector productivo, a cambio de dejar un margen de libertad dotado de seguridad jurídica en el que operar sin miedo a la sanción penal). Es la necesaria cesión a los empresarios en un modelo que opta por cercenar, en cierta medida, su libertad de empresa.

Es aquí donde puede apreciarse, precisamente, la diferencia entre el modelo de peligro concreto y abstracto. Piénsese en un supuesto en el que el responsable de despliegue de un sistema de IA de alto riesgo (por ejemplo, usado en el contexto médico) desarrolla todo un mecanismo de supervisión de la información de salida de su sistema, cumpliendo escrupulosamente con los requisitos impuestos por el Reglamento pero, a pesar de ello, se produce un diagnóstico del sistema erróneo, de forma que el resultado que devuelve diagnostica de forma incorrecta un cáncer de piel, ignorando su prevalencia entre mujeres afroamericanas.

Con un delito de peligro abstracto de idoneidad, parece evidente que dicho operador no podrá, en ningún caso, responder de estos daños, por mucho que sea evidente, *ex post,* que el sistema es idóneo para causar un daño, pues

---

[1430] En este sentido, esencialmente idéntico, SILVA SÁNCHEZ, J. M., *ibid.*, p. 146, aludiendo a que el juez puede resolver este «bloqueo prohibitivo» realizando una verdadera interpretación teleológica del tipo. Más aun teniendo en cuenta que la remisión no es, aquí, total a la normativa, sino que la redacción elegida matiza no solo las conductas penadas, sino los efectos para bienes jurídicos que deben esperarse. Así, en nota 208, con alusión a la STC 24/2004, de 24 de febrero. Añade Silva, en p. 148, que es necesario, en primer lugar, identificar el bien jurídico protegido por la ley penal y exigir que la conducta de la que se trate comporte un riesgo penalmente relevante para el bien jurídico, ya que en los casos en los que no sea posible apreciar un peligro en medida relevante la conducta debe reputarse atípica.

[1431] *Ibid.,* p. 150 y ss.

dicho riesgo, aún remoto o difícilmente previsible *ex ante,* estaba dentro de lo permitido, dando lugar a una impunidad que no parece demasiado problemática. La diferencia con respecto a un eventual delito de peligro concreto es que, con este último, a pesar de la existencia de un peligro concreto y grave para la salud de las personas, se debería afirmar la impunidad. Aunque estemos de acuerdo en que la diferencia entre peligro concreto y abstracto esté en el conocimiento sobre la probabilidad de acaecimiento del resultado lesivo, parece evidente que, en caso de producirse varios incidentes en los que varias mujeres afroamericanas son incorrectamente diagnosticadas, con un riesgo evidente para su salud, será mucho más problemático sostener que, como se trata de una actuación conforme a la normativa administrativa, el hecho es completamente lícito, por mucho que exista un riesgo concreto y patente para los bienes jurídicos finales [1432].

Finaliza Silva su reflexión en este sentido poniendo de manifiesto que la doctrina ha tratado, en varias ocasiones, de revisar los presupuestos del nexo de imputación objetiva entre conducta y resultado en delitos de homicidio y lesiones dolosas o imprudentes, con propuestas relacionadas, por ejemplo, con introducir presunciones de causalidad [1433]. Es precisamente este el camino que se ha tratado de evitar en esta investigación al que, creemos, nos veríamos irremediablemente abocados si no optásemos por la tipificación propia del peligro abstracto con las limitaciones de idoneidad hasta ahora expuestas.

## 8. LA PROPORCIONALIDAD DE LA PENA

Aunque el modelo que aquí se propone no pretende más que ser una primera aproximación a la tipificación de unos eventuales nuevos delitos, merece la pena explicar, por su importancia, por qué se opta por una pena cumulativa y con triple naturaleza: privativa de libertad, pecuniaria e interdictiva.

La propuesta bebe del modelo ya implantado por el legislador en los delitos contra la seguridad colectiva que recogen la tipificación de los delitos más similares en un triple sentido: corresponden a delitos de peligro abstracto con cláusula de idoneidad; protegen diferentes bienes jurídicos importantes pero

[1432] Considera Silva que la existencia de delitos de peligro concreto con remisión total a leyes permisivas, en el caso de peligro concreto y grave para bienes que Silva califica de intrínsecos (especialmente la vida y la integridad física), da lugar a la vulneración de deberes de protección del Estado. *Ibid.,* pp. 153 y ss., con referencias en nota 216 a la sentencia del caso Lüth (comentada *supra,* capítulo sexto). Sobre los derechos positivos de protección del Estado: Tomás-Valiente Lanuza, C., «Deberes positivos del Estado y Derecho penal en la jurisprudencia del TEDH», *InDret* 3/2016, pp. 1-72.

[1433] Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 159.

no necesariamente con riesgo inmediato; y representan una intervención penal que se distancia de la mera infracción administrativa (especialmente por la mayor fuerza preventiva de la prisión y de la pena interdictiva), pero sin llegar a los rangos de castigo más elevados del catálogo del Código, que se dan en los supuestos donde se pone en concreto peligro la vida de numerosas personas (así, explosiones, incendios o estragos).

En lo relativo a la pena, el modelo más cercano a los delitos aquí propuestos es, tal vez, el previsto en los arts. 359 y 361 [1434]; respectivamente, el delito de elaboración y expedición de sustancias nocivas o productos sin autorización que sean idóneos para causar estragos sin reunir las condiciones legalmente requeridas, y el delito de fabricación o distribución de medicamentos peligrosos, idóneos para perjudicar la salud. Sendos tipos pueden considerarse de peligro abstracto hipotético, sin exigencia de resultado lesivo y con referencia a bienes plurales (la referencia, si bien indirecta, la encontramos en la alusión a la idoneidad para causar estragos, de forma que la vida, integridad, o salud quedan mediatamente protegidos); y, lo más importante, con referencia a incumplimientos de deberes técnicos o normativos a los que se remite el tipo, orientados en última instancia a la protección de bienes jurídico-penales.

Esta triple pena encaja perfectamente con nuestra propuesta, que se diferencia así del régimen meramente administrativo otorgando un plus de prevención, pero operando con coherencia sistemática respecto de otros tipos de naturaleza similar.

La pena de prisión propuesta no es elevada, pues no se exige la puesta en peligro concreto ni un daño como condición de punibilidad, y plantea un marco que puede llevar a la entrada efectiva en prisión en los casos más graves, pero evitando excesos de punición, lo que la hace acorde con el principio de proporcionalidad.

La pena de multa, por su parte, no pretende tener la capacidad aflictiva que genera el monto ya previsto en la normativa administrativa. Si nos atenemos a las cuantías previstas en el CP, en aplicación del artículo 50.4, la multa no podría

[1434] Sobre estos delitos: Rueda Martín, M. Á., «El bien jurídico protegido en los delitos relativos a productos de consumo masivo», *Estudios Penales y Criminológicos*, vol. XXX, 2010, pp. 395-445. La autora defiende aquí una postura favorable a la consideración del bien jurídico colectivo, atendiendo a la convergencia de intereses de los distintos actores implicados en las relaciones sociales fabricante-distribuidor-comerciante-consumidor (p. 430), caracterizado por desempeñar una función negativa de contención de riesgos para bienes jurídicos individuales, inserto en el funcionamiento del sistema concreto, y una función positiva de delimitación de condiciones necesarias para otros bienes involucrados (pp. 433-444). De este carácter colectivo unido a la función de contención de riesgos deriva Rueda la no necesidad de acreditar que la acción infractora en la situación concreta pusiera en peligro la salud del consumidor individual (p. 437).

superar, para las personas físicas, en el escenario más gravoso, los 288.000 € (habida cuenta de que se propone una multa que pueda llegar hasta los 24 meses), estando, por tanto, muy lejos de alcanzar las cuantías previstas en la norma administrativa. La multa proporcional, aunque permitiera en su caso obtener cuantías más elevadas, amén de quedarse también lejos de la cuantía prevista en la normativa administrativa, ya suficientemente elevada, no es la habitual cuando se trata de castigar delitos relacionados con bienes eminentemente personales, como los que aquí se proponen, no encajando tanto la lógica de atender al daño causado, el valor del objeto del delito o el beneficio reportado por el mismo.

No es una cuestión novedosa, en este aspecto, que la capacidad aflictiva de las multas penales quede absolutamente en entredicho en aquellos delitos donde el beneficio económico detrás de las actividades objeto de tipificación es elevado. La literatura penal es pacífica sobre la necesidad de repensar la efectividad de una pena de carácter pecuniario que, amén de resultar mayoritaria en el sistema, adolece de una capacidad preventiva real [1435]. No podemos, en este punto, desarrollar mucho más las ideas que se han puesto de manifiesto en otros lugares [1436], pero somos conscientes de que, si bien la pena de multa parece irrenunciable por la tradición con la que cuenta y el respaldo penal, no es desdeñable plantear, de una vez por todas, su escasísima utilidad y, de paso, cuestionar su ubicuidad allí donde deja de tener sentido [1437].

En el caso de las personas jurídicas, tomando, una vez más, el modelo del Código penal para los delitos contra la seguridad colectiva, e imponiendo, esta vez, el máximo previsto, la multa resultante, aplicando nuestra regulación (art. 50 CP) consistiría en el pago de una suma resultante de multiplicar una cuota diaria, determinada entre 30 y 5.000 euros conforme a la capacidad económica de la entidad (art. 50 CP), por el número de días que establezca la sentencia, cuya duración máxima aquí propuesta asciende a cinco años (por tanto, un máximo de 1.800 días computados a razón de 360 días por año). De esta forma, la cuantía máxima que puede alcanzar la sanción pecuniaria para una persona jurídica se situaría en los 9.000.000 euros (art. 50.4 CP). Una vez

[1435] En este sentido: Blay, E., «A vueltas con la multa. De la relativa invisibilidad de la pena de multa y de las razones para que se investigue sobre ella», *InDret* 4/2022, Editorial. También: Coca Vila, I., «La pena de multa en serio. Reflexiones sobre su dimensión y aseguramiento aflictivos a través del delito de quebrantamiento de condena (art. 468 CP)», *InDret*, 3/2021, pp. 69-99.

[1436] Así, en: Bravo Bolado, A., «La protagonista olvidada del sistema penal», blog de la Facultad de Derecho, 24 de mayo de 2023, disponible en: <https://www.blog.fder.uam.es/la-protagonista-olvidada-del-sistema-penal/>[Accedido por última vez el 30/01/2026].

[1437] Recientemente, entre nosotros: Rodríguez Horcajo, D., «La pena de multa como imposible eje del sistema de consecuencias jurídicas del delito de la persona física», en Rodríguez Horcajo, D./Basso, G. (Dirs.), *Repensando las consecuencias jurídicas del delito,* Tirant lo Blanch, Valencia, 2025, pp. 219-253.

más, lejos de representar el montante previsto en la normativa administrativa, pero ciertamente, con mayor capacidad aflictiva que en el caso de las personas físicas.

Finalmente, la pena interdictiva adquiere aquí un especial sentido por su finalidad preventiva, tal y como indica la doctrina española que se ha encargado de su estudio.

La propuesta que aquí se formula (pena de inhabilitación especial para actividades relacionadas con inteligencia artificial) se enmarcaría en las penas privativas de derechos que el Código Penal contempla, concretamente las penas de inhabilitación especial para empleo o cargo público, profesión, oficio, industria o comercio, u otras actividades (art. 39.b). Se trata de una pena centrada en determinadas actividades laborales que han de concretarse de forma expresa y motivada, donde la autoridad judicial podrá restringir la inhabilitación, en principio, a determinadas actividades o funciones dentro de la profesión, oficio o actividad, permitiendo, tal vez, que se sigan ejerciendo funciones no directamente relacionadas con el delito [1438]. En todo caso, debido a la variedad de formulaciones en el Código, que a veces alude a profesión u oficio, otras a industria o comercio, o combinaciones variadas de todas ellas, es evidente que el legislador no quiere que ninguna clase de ocupación laboral (exija o no permiso, habilitación o licencia para su ejercicio, por poco común o inusual que sea), quede fuera de la aplicación de la pena [1439], por lo que no entrañaría ninguna dificultad una alusión a las profesiones relacionadas con el sector de la inteligencia artificial.

En ocasiones el Código especifica, de hecho, el tipo de profesión afectada (sanitaria, educativa, docente, deportiva, etc.) pero, por imperativo del artículo 45, será el juez el que, en sentencia, de forma expresa y motivada, concrete sobre qué tipo de actividades recae la inhabilitación, expresando la relación de la prohibición de la actividad o actividades con el delito cometido [1440].

---

[1438] Boldova Pasamar, M. A., «Penas privativas de derechos», en Boldova Pasamar, M. A./ Alastuey Dobón, C. (Coords.), *Tratado de las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023, p. 215.

[1439] *Idem,* nota al pie 39: «*[…], lo que se castiga es la actividad en sí misma considerada, por mucho, poco o nada reglamentada que esté en la esfera administrativa».*

[1440] Este aspecto, aunque puede derivarse de una lectura teleológica del artículo 45 del CP, ha sigo exigido de forma reiterada por la jurisprudencia respecto de la redacción del anterior CP. Así, las SSTS de 23 de noviembre de 1970, de 13 de febrero de 1987, de 14 de septiembre de 1990, 2521/1992, de 24 de noviembre y 2063/1994, de 28 de noviembre. Referencias en nota al pie 42 de: Boldova Pasamar, M. A, *ibid.* No obstante, también existen sentencias que no exigen que se especifique la especialidad (STS 2139/2001, de 15 de noviembre). En todo caso, queda sin resolver la duda de si el artículo 45 y sus limitaciones están pensadas solo para los casos en que esta pena se prevé como accesoria, en el sentido de no estar expresamente prevista en el CP como consecuencia jurídica aparejada a un determinado delito. Así, Ramos Vázquez, aclarando que no es posible plantear limitaciones a la imposición en aquellos delitos en que esté expresamente prevista

Resulta plenamente atendible, en este contexto, el llamamiento que hace Gutiérrez sobre la conveniencia de usar este tipo de penas con mayor protagonismo en los delitos característicos de la sociedad tecnológica y del riesgo [1441]. En lo relativo a la finalidad primordial que cabe otorgarles, merece la pena transcribir sus palabras:

> *«[…] es necesario articular un sistema de penas privativas de derechos construido sobre una concepción estrictamente preventiva que, tras despojarlas de su tradicional significado infamante, adecue toda su regulación a los principios propios del Derecho penal democrático, algunos de los cuales resultan claramente vulnerados como consecuencia de la existencia de penas privativas de derechos de aplicación automática e indiscriminada y, por tanto, dotadas de un fundamento y un fin infamantes. Pero esta labor no puede desarrollarse de espaldas a las especiales características de estas penas, pues su eficacia preventivo general y especial depende en gran medid [sic] de una adecuada comprensión de los rasgos característicos de estas penas y de su correcta incorporación a su regulación positiva. Así, debe tenerse en cuenta que nos encontramos ante «penas especiales», que recaen sobre derechos que, aunque en abstracto son predicables de cualquier persona, en la práctica sólo afectan a determinados grupos de ciudadanos que los ejercen de manera efectiva y que, en consecuencia, se verán gravemente afectados por su privación.»* [1442]

Se destaca también su versatilidad, pues su contenido puede adoptar distintas formas en lo referido al derecho, cargo o profesión objeto de la pena, así como en cuanto a la intensidad de sus efectos prohibitivos, apareciendo como penas idóneas para cumplir de forma tanto o más eficaz que otras los fines de prevención general y especial, por tanto, sin producir efectos desocializadores de tanta intensidad. Además, se trata de penas dotadas de suficiente contenido aflictivo, desplegando la fuerza intimidatoria que se exige en general a la pena, pero siendo lo suficientemente volubles como para poder contribuir, incluso, a la reeducación o resocialización del penado, amén de su capacidad para prevenir la reiteración delictiva por la privación ínsita de continuación en el ejerci-

---

la pena, siendo solo de aplicación en los casos en que se prevé como pena potestativa, que, por imperativo del artículo 56 CP, son solo respecto de penas accesorias que sean inferiores a diez años de prisión (p. 120). Ramos Vázquez, J. A., «La pena de inhabilitación especial para profesión, oficio, industria o comercio» en Faraldo Cabana, P./ Puente Aba, L. M. (Dirs.), *Las penas privativas de derechos y otras alternativas a la privación de libertad,* Tirant Lo Blanch, Valencia, 2013, p. 120.

[1441] Gutiérrez Castañeda, A., «¿Las grandes olvidadas de la reforma penal? Reflexiones sobre el papel de las penas privativas de derechos políticos y profesionales en la modernización del sistema de penas» en Pérez Álvarez, F., (Ed.), *Moderno discurso penal y nuevas tecnologías*, Memorias del III Congreso Internacional de Jóvenes investigadores en ciencias penales, 17,18 y 19 de junio de 2013, Ediciones Universidad de Salamanca, 2014, p. 411.

[1442] *Ibid.,* p. 412.

cio de la actividad en cuyo desarrollo se cometió el delito [1443]. La idea de Gutiérrez va, sin embargo, un paso más allá, pues propugna que se supere la consideración tradicional de las mismas como penas accesorias, para pasar a ocupar la categoría de penas principales [1444].

Finalmente, en cuanto a su contenido, cabe matizar, con Pozuelo, que no se trata de privar a alguien de sus conocimientos profesionales [1445], ni se trata en este caso de la privación del estatuto o profesión (a diferencia de lo que puede ocurrir con las inhabilitaciones para cargo público), sino que se priva de las facultades derivadas del cargo. Es común, por otra parte, que en el ámbito de los delitos socioeconómicos sí se determine el concreto ámbito de actividad en que recaería la inhabilitación, y, en este sentido, Fernández Teruelo aboga por usar como criterio decisivo el del bien jurídico afectado, de forma que si el bien jurídico no es de naturaleza económico-patrimonial la inhabilitación ha de recaer sobre el sector concreto sobre el que opera la empresa o negocio en cuestión, y sobre el que, de forma precisa, se ha cometido el delito (por ejemplo, el sector de la construcción o el de productos alimenticios peligrosos) [1446]. Nos parece este, en estos términos, el modelo que habría que seguirse en los delitos aquí propuestos.

---

[1443] *Ibid.,* p. 413.

[1444] Reconociendo, no obstante, y en la línea de lo ya apuntado inmediatamente *supra,* que se trata, en el caso de la pena de prohibición de actividades, oficio, industria o comercio, de una pena cuya aplicación no es automática, sino que requiere que se constate una relación directa entre su contenido y el delito cometido. Y esto, explica la autora, no hace más que reflejar la idea de que en estas penas el fundamento está en la incapacidad del condenado para un adecuado ejercicio de los derechos propios de una profesión, oficio o cargo, incapacidad ya manifestada con la comisión del delito. Niega la autora (p. 416) que deba entenderse que los fines son los de prevención general e inocuización del delincuente, pues dicha prevención general se ve opacada si se impone junto con una pena de prisión que acaba por «eclipsarla» y que, en la práctica, siempre la acompaña. Lo mismo desde el punto de vista de la incapacitación del delincuente, conseguida, también y principalmente, con la entrada en prisión. Estos argumentos son puestos de manifiesto por Gutiérrez para aclarar (p. 417) que la solución de un sistema de penas moderno debe ser optar por la supresión de la accesoriedad de estas penas, consagrándolas como penas autónomas en cuanto a su aplicación y duración.

[1445] Pozuelo Pérez, L., *Las penas privativas de derechos en el Código penal*, Colex, Madrid, 1998, p. 47.

[1446] Fernández Teruelo, J. G., «La inhabilitación profesional del delincuente económico-empresarial», en Carbonell Mateu, J. C./Rosal Blasco, B. del/Morillas Cueva, L./Orts Berenguer, E./Quintanar Ruiz, M. (Coords.), *Estudios en homenaje al profesor Cobo del Rosal*, Dykinson, Madrid, 2005, p. 355. Destaca de la propuesta de Fernández Teruelo sus precauciones sobre la eficacia real que esta pena puede tener cuando las funciones propias del cargo del sujeto, en el entorno de delitos socioeconómicos, tiene que ver con las labores de administración de una sociedad, pues, *de facto,* y a pesar de la pena, será difícil evitar que se siga realizando esta tarea de diseño de políticas y toma de decisiones puesto que, una vez más, el pensamiento no se puede inhabilitar. Lo original de su propuesta, que merecería la pena explorar para una aplicación realmente material en estos casos, pasa por la creación de un modelo específico de inhabilitación profesional dentro de la sociedad. *Ibid.* p. 361.

Cabe afirmar, finalmente, que la opción por los delitos de peligro en los términos aquí expuestos responde, como así se ha tratado de argumentar en los capítulos precedentes, a una cuestión de oportunidad, y no de legitimidad. El legislador europeo ha avanzado ya en su configuración del modelo a seguir, siendo que la apuesta no es solo por la reparación, sino por la prevención general, pues el mandato a los Estados exige un modelo de sanciones disuasorias. En un sentido esencialmente igual, los argumentos que Ibold utiliza para apostar por un modelo de intervención [1447] pasan por reconocer que la prevención no debe realizarse exclusivamente sobre la base de la responsabilidad civil en aplicación del modelo del producto defectuoso, pues esto implicaría una «reprivatización del conflicto social» [1448], haciendo que la ejecución de estas sanciones dependa solo de la actuación de la parte perjudicada, entendiendo que no es este el espíritu que guía al legislador europeo y su decisión consciente de no optar, en ninguno de los casos, por un modelo de responsabilidad objetiva.

## III. EPÍLOGO: LOS SUJETOS OBLIGADOS. EL PAPEL DEL USUARIO SUPERVISOR NO PROFESIONAL

Para finalizar la justificación de los nuevos tipos, es necesario explicar a qué sujetos están dirigidos, cuestión en absoluto baladí, pues determina el alcance fundamental de la tipificación y modula la fuerza expansiva del derecho penal.

El modelo que se propone es dual, pues los sujetos obligados por los dos grupos de delitos son distintos. Así, en la sección 1.ª (relativa a los usos prohibidos de sistemas de IA), se especifica que podrá cometer las infracciones penales *«el que desarrollare, comercializare, pusiere en servicio o utilizare»* dichos sistemas prohibidos, de forma dolosa. Es decir, nos encontramos ante un delito común, que puede ser realizado, siempre con conocimiento sobre el contenido del sistema [1449], en principio, por cualquier persona.

---

[1447] Ibold, V., *op. cit*. N. 104, desarrollado en p. 414 y ss.

[1448] *Ibid.,* p. 418: «*Reprivatisierung des sozialen Konflikts*».

[1449] Nos remitimos, en este punto, a lo establecido respecto del dolo para los delitos de riesgo catastrófico que aquí pueden tomarse como inspiración. Así, Maraver Gómez, M., «Delitos de riesgo catastrófico», *op. cit*. N. 1141, nm. 14667: *«Respecto a la modalidad dolosa [hablando, en este caso, del delito de exposición a radiaciones ionizantes del artículo 341], se entiende que es suficiente el dolo eventual y que debe abarcar tanto la conducta relativa a la exposición a radiaciones ionizantes como el riesgo que tal exposición puede generar sobre la vida, la salud o los bienes de las personas, o sobre la calidad del aire, del suelo o de las aguas, o sobre animales y plantas.»* De forma paralela, el dolo de nuestros nuevos delitos debe abarcar, como mínimo, el conocimiento sobre el desarrollo de un sistema de riesgo inaceptable, así como el riesgo que tal sistema puede generar sobre los bienes en juego, en cada caso.

Seguimos, de esta forma, el esquema del Código cuando se trata de delitos de riesgo catastrófico donde, aunque lo normal será que las personas que efectivamente manipulen materiales, productos o sistemas que generen un peligro elevado sean profesionales, no se desea cerrar la puerta al castigo en caso de que cualquier otra persona, no profesional, pusiere en riesgo a la colectividad.

Se trata, en este caso, de incriminar conductas activas de despliegue o uso de sistemas que permita captar las distintas fases de su ciclo de vida, pues, aunque en este caso el deber extrapenal vincula solo a determinados operadores sujetos al Reglamento, la finalidad que ha de perseguir la norma penal no es tanto reforzar un deber normativo específico de los proveedores técnicos, como evitar el despliegue de un determinado tipo de sistemas que, siempre que se den las condiciones de peligro previstas en los tipos, causarán un riesgo no permitido con incumbencia penal. La normativa de referencia que permitirá realizar una interpretación teleológica adecuada (e integrar el riesgo permitido, mucho más escaso, pero aún posible, en estos sectores), habrá de completar el tipo, por mucho que la remisión a la norma sea implícita.

Sin embargo, en el segundo grupo de delitos (aquí conformantes de la sección 2.ª, sobre el incumplimiento de obligaciones para los sistemas de IA de alto riesgo), la tipificación que se propone es distinta, abarcando el tipo al que *«en el desarrollo, comercialización, puesta en servicio o utilización profesional de un sistema»,* contravenga ciertas obligaciones determinadas en la normativa extrapenal. En este caso, se toma la decisión de dejar fuera al usuario no profesional, convirtiendo a estos delitos en delitos especiales propios.

Debe realizarse, no obstante, una matización importante. Tal y como se ha desarrollado en el capítulo séptimo, los sujetos obligados por el Reglamento comprenden toda una serie de operadores económicos que, en esencia, y a salvo del representante autorizado, actúan siempre en su propio nombre. Cuando en la propuesta de *lege ferenda* se habla de sujetos que «desarrollen, comercialicen, o pongan en servicio de forma profesional», en la medida en que las obligaciones que les incumben son las del Reglamento, han de interpretarse de una manera conforme a esta norma. Y esto es especialmente relevante respecto del usuario profesional, que se identificaría con el responsable de despliegue del sistema, es decir, la persona que opera y usa el sistema, pero siempre, tal y como indica el artículo 3.4, «bajo su propia autoridad». Esta delimitación implica que los usuarios profesionales pero que no actúen bajo su propia autoridad (el ejemplo emblemático es el de los trabajadores), quedan fuera de los sujetos obligados del Reglamento y, por tanto, haciendo una interpretación conforme del tipo penal, también fuera del escrutinio penal, por las razones que a continuación se expondrán.

Aquí, los tipos penales no castigan la generación de un producto peligroso, sino en la infracción grave de obligaciones jurídicas en el desempeño de una actividad fuertemente regulada. Somos conscientes, claro está, de las dificultades que conlleva la introducción de tipos penales especiales, pues, a pesar de que cabe la punición de las conductas de partícipes no cualificados, entrarán en juego las limitaciones propias del principio de accesoriedad [1450]. El modelo a seguir podría ser, en este caso, ya no el del delito de riesgo catastrófico, más propio de los delitos de la primera sección, sino, una vez más, el del artículo 361 CP *(«el que fabrique, importe, exporte, suministre, intermedie, comercialice, ofrezca o ponga en el mercado, o almacene… medicamentos o productos sanitarios…»).*

No obstante, la decisión de optar por una tipificación de tales características obedece a unas razones de fondo, de carácter político-criminal, que a continuación se explicarán, y no tiene que ver con la idea de que los tipos aquí propuestos son de los bautizados por un sector de la doctrina como tipos de infracción de deber [1451].

Dos son las razones esenciales que nos llevan a propugnar este círculo de sujetos activos, una de carácter jurídico y otra de carácter político-criminal.

## 1. EXCLUSIÓN DEL USUARIO NO PROFESIONAL POR MOTIVOS JURÍDICOS

Al proponer la tipificación de un delito que se produce cuando un sujeto actúa más allá de lo delimitado por el riesgo permitido, merece la pena distinguir a quién se dirigen las normas que delimitan estos ámbitos de actividad socialmente permitidos. En este sentido, Silva menciona la diferencia entre el sujeto que está llamado a organizar o gestionar el riesgo permitido (siendo él y solo él quien recibe el permiso para gestionar esta organización peligrosa y siendo él, también, el único obligado a minimizar los riesgos propios de ese

[1450] En este sentido, imprescindible: Peñaranda Ramos, E., *La participación en el delito y el principio de accesoriedad,* BdeF, Buenos Aires, 2015.

[1451] En un sentido crítico: Peñaranda Ramos, E., «Sobre el alcance del artículo 65.3 del Código Penal. Al mismo tiempo una contribución a la crítica de los delitos de infracción de deber», en AA. VV., *Estudios penales en homenaje a Enrique Gimbernat*, Edisofer, Madrid, 2008, Tomo II, pp. 1419-1452. También: Alcácer Guirao, R., *¿Lesión del bien jurídico o lesión de deber?: Apuntes sobre el concepto material del delito*, Atelier, Barcelona, 2003, haciendo un repaso por el pensamiento de los diferentes autores partidarios de la noción de la lesión de deber en pp. 33 y ss., ofreciendo una visión integradora de la teoría del bien jurídico (a la que otorga, en todo caso, prioridad) y las teorías apegadas a la confirmación de la vigencia de la norma.

espacio de actividad), del sujeto que, aunque actúa en un espacio de riesgo permitido organizado por un tercero, no gestiona ese riesgo [1452]. Mientras que para el gestor del riesgo rigen todo tipo de deberes fuertes (que incluyen, en última instancia, deberes de evitación de resultados lesivos), para los sujetos que actúan dentro de su esfera de organización, les competen, en palabras de Silva, solo deberes generales de solidaridad, de forma que solo cuando se deleguen en ellos cautelas concretas asumirán deberes de evitación concretos. Pero, en principio, para quienes simplemente obren en el espacio de riesgo, debe regir un principio de confianza en la correcta actuación del gestor del espacio de riesgo permitido, a cuyas instrucciones quedarán sujetos [1453].

En definitiva, estos sujetos (profesionales o no, pero, en todo caso, no sujetos a las obligaciones del Reglamento, pues no se trata de personas que actúen en su propio nombre, residiendo ahí la clave de la distinción), más allá de cumplir con diligencia los deberes en que se concreten las condiciones de permisión, por parte de sus superiores, no deberán adoptar ninguna prevención más respecto de la evitación de ulteriores resultados [1454] y, añadimos, respecto de la gestión ordenada del riesgo para bienes jurídicos finales. En otras palabras, también manejadas a lo largo de la presente investigación, a cada cual incumbe un ámbito de organización propio claramente delimitado.

No es tanto una cuestión, insistimos, de sujeción al rol o de infracción de un deber asignado de forma personal, como un motivo de fondo bastante evidente, relativo a la capacidad real de adoptar decisiones relevantes sobre el diseño, comercialización o puesta en servicio del sistema. Para tener un papel significativo, por tanto, en la generación de riesgos penalmente relevantes (o en su minimización), no basta con ser profesional en sentido amplio, es decir, tener algún tipo de formación especializada, sino que es esencial tener capacidad de modificación real de los estándares de seguridad o protocolos de supervisión, por poner solo dos ejemplos.

Probablemente se entienda mejor con un caso ilustrativo. Un programador que implementa instrucciones sin diseñar la arquitectura del sistema ni aprobar su salida al mercado no tendrá ninguna autonomía funcional ni capacidad real de cumplimiento de los deberes impuestos en el Reglamento, sin

---

[1452] Silva, Sánchez, J. M., *op. cit*. N. 1421, p. 121.

[1453] *Ibid.,* p. 122. Añade Silva que, en virtud de este principio de confianza, cabe que incluso se den situaciones de error invencible sobre el carácter permitido del riesgo. Piénsese, por concretar esta idea, en el trabajador cuya labor consiste, simplemente, en la tarea de limpieza de datos, cuyo sistema viene dado de antemano, y que no goza de capacidad real para decidir cuáles son los datos que se usan para entrenar un determinado modelo.

[1454] *Ibid.,* p. 125.

que se le pueda atribuir función de garante del cumplimiento normativo, por lo que su actuación, a lo sumo, podrá ser valorada en sede de participación.

## 2. EXCLUSIÓN DEL USUARIO NO PROFESIONAL POR MOTIVOS POLÍTICO-CRIMINALES

Parece evidente que una gran parte de las obligaciones previstas en el Reglamento en forma de cautelas técnicas no será aplicable, con carácter general, al usuario profesional, que no estará en condición ni posición apta para cumplirlas (así, esencialmente, las obligaciones de establecer un sistema de gestión, de prever un sistema de gobernanza de datos o de prever un sistema de vigilancia poscomercialización). Sin embargo, una de las obligaciones es precisamente la que, por excelencia, va a acabar recayendo irremediablemente en el último eslabón del ciclo de vida de la IA, que no es otra que la supervisión efectiva y final del desempeño de los sistemas. En este punto, tanto el trabajador (profesional, pero no obligado por el Reglamento), como el usuario final, en absoluto profesional, habrán de verse excluidos de la aplicación del tipo que nos ocupa, ya no en alusión a su posición de obligados por una norma jurídica, sino, yendo más al fondo de la cuestión material, por la imposibilidad de evitación fáctica debida al conocido como el dilema del control.

El dilema del control, al que se ha aludido en algunas ocasiones en esta investigación, se ha puesto de manifiesto ahí donde se ha discutido sobre el alcance del principio de control significativo o de supervisión. Control, supervisión, o *human in the loop* son términos comunes en la literatura sobre IA desde sus inicios. También la ubicua alusión a una IA donde el ser humano esté en el centro. El ser humano como último eslabón de la cadena de control de los complejos sistemas de IA actúa, así, como alivio inmediato en situaciones donde la automatización completa no es una opción.

Como se ha explicado, el propio Reglamento reconoce que la supervisión humana es un requisito de los calificados como esenciales para los sistemas de alto riesgo. Se ha argumentado, también, que este requisito adolece no solo de concreción, sino que, de forma ínsita, quizá adolece de sentido práctico. Cabe resaltar, no obstante, y como punto de partida del resto de reflexiones que a continuación se desarrollarán, que no toda forma de control es interpretada aquí como un imposible, pues, de ser así, la presente investigación carecería de sentido. Por utilizar la clasificación dicotómica de Ibold, se partirá de la consideración de que el control humano es posible no externamente (es decir, no se podrán controlar los casos o resultados concretos, pues precisamente los

sistemas han sido creados para ello), pero sí será posible un control interno, basado esencialmente en un control sobre sus fines y posibilidades de comportamiento [1455]. Este matiz resulta esencial para poner el énfasis en que hay una fase donde el control es, no solo posible, sino deseable.

Las consideraciones que siguen se realizan, pues, sobre el control el sentido externo, esto es, el control sobre el resultado o *output* del sistema, sobre el que pivotan las obligaciones de supervisión del Reglamento.

Diversos autores han puesto de manifiesto la paradoja del control humano sobre herramientas cuyo significado no solo no se llega a comprender del todo (lógica del *black box),* sino sobre decisiones que están pensadas para ser más correctas que las que tomaría un humano. No se trata de reproducir los argumentos expresados *supra* respecto de la capacidad de supervisión, sino de reflexionar sobre cuál ha de ser el control del supervisor en su condición de usuario no profesional o trabajador por cuenta ajena. En una sociedad donde la última palabra sobre la decisión de los sistemas la tiene una persona (cualificada o no), será pertinente saber qué responsabilidad queremos otorgar al eslabón final y, probablemente más débil, de la cadena.

Ya lo ponía de manifiesto Hilgendorf al hablar del *«control dilemma»* [1456], refiriéndose en este caso a la conducción autónoma, ámbito donde en mayor medida han surgido estos cuestionamientos. Afirmaba el autor alemán que lo que afrontamos es un dilema sobre el control donde la ley está quedándose obsoleta, pues dice que es técnicamente posible que un conductor humano ceda el control a un vehículo autónomo (al menos autónomo en alguna de sus partes), pero esta opción no es legalmente posible, o al menos no completamente, en la medida en que el conductor siempre mantiene el deber de controlarlo. Para que un conductor sea capaz de adaptar el nivel de su conducción al nivel requerido, primero tiene que saber reconocer, por ejemplo, en cada instante, el nivel de autonomía que le ofrece el vehículo, dando a entender que, por la complejidad de los vehículos actualmente en circulación será posible una combinación de automatización y control en diferentes niveles y difícil de percibir por el conductor [1457]. La idea es que el conductor, para cumplir con el

---

[1455] Ibold, V., *op. cit*. N. 104, p. 361: «*Technische Autonomie führt dazu, dass ein KI-System zwar nicht* ***äußerlich kontrollierbar*** *ist – weil es im Einzelfall nicht steuerbar ist bzw. die Entwickler die Funktionsweise nicht im Detail vorgeben (was konzeptionell ja gerade so sein soll); es ist aber sehr wohl* ***innerlich kontrollierbar****, weil alle Zwecke und Verhaltensmöglichkeiten durch die Entwickler gesetzt wurden und die Nutzer über ihren Einsatz entscheiden*». La negrita es nuestra.

[1456] Hilgendorf, E., «Automated driving and the law…, *op. cit*. N. 833, pp. 181-182.

[1457] Así, *ibid.*, p. 182, explicando que el vehículo cuenta con diferentes subsistemas que funcionan de forma independiente y que, por ejemplo, puede ser que el sistema de alumbrado se encuentra completamente automatizado, pero el subsistema de aparcamiento autónomo lo sea solo parcialmente.

deber de cuidado exigible, debe prestar atención a todos los sub-sistemas que, dentro del vehículo, pueden contribuir en diferente medida a un eventual daño, generándose, a medida que los vehículos (y los sistemas en general) son más complejos, situaciones donde el humano tendrá serias dificultades, de partida, para reconocer hasta dónde llega su margen de control o evitación.

Más allá del caso de la conducción autónoma, esta paradoja ha sido puesta de manifiesto por varios autores, aludiendo no ya a la complejidad de los sistemas sino a la posibilidad real de control o evitación del humano, negando, de hecho, su mejor desempeño en algunas ocasiones y poniendo de manifiesto que quizá esta capacidad de control esté empíricamente infundada [1458].

Koulu ha criticado el discurso oficial de la Unión Europea sobre el «control humano», señalando la disonancia entre el marco ético y la complejidad real de las interacciones humano-máquina, denunciando que las propuestas regulatorias ignoran la vasta literatura de interacción humano-máquina que evidencia las serias limitaciones de dicho control en la práctica [1459].

Desde el punto de vista de la responsabilidad, Wagner ha puesto de manifiesto que se corre el riesgo evidente de que la agencia humana se acabe instrumentalizando para obtener una fuente de responsabilidad, pero sin que exista un poder real de intervención o control [1460], manifestando que el humano se convierte en una instancia formal de validación *(«rubber-stamping»* [1461]), siendo su capacidad de control muy superficial y sirviendo, por tanto, a una deficiente protección de los derechos humanos [1462].

Desde un punto de vista más apegado a la normativa, las críticas de Green son también fundamentales en este aspecto, pues la evidencia empírica demuestra que los humanos no pueden desempeñar de manera fiable las tareas de supervisión previstas, y, además, este tipo de políticas que alientan una supervisión de alguna forma vacía incentivan una falsa sensación de seguridad que facilita la elusión de respon-

---

[1458] Green, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, 45, 2022 p. 15: *«human oversight is pre- sented as a safeguard against algorithmic harms despite minimal empirical evidence that it actually provides reliable protections»*.

[1459] Koulu, R., *op. cit*. N. 1333. La idea de esta autora es que el discurso oficial refuerza dicotomías problemáticas humano-máquina, amén de omitir la complejidad de estos sistemas, trasladando riesgos y responsabilidades a operadores humanos, pero sin garantizar, a cambio, una protección efectiva. Este fenómeno ocurre no solo cuando se trata de monitorizar sistemas de IA, y la autora así lo expresa, dando ejemplos de otros campos de automatización. Así, en nota el pie 28: «*There is much evidence that people are not good monitors of automation' for various reasons, including boredom that ensues from monotonous tasks*», citando estudios del campo de la aeronáutica o la medicina.

[1460] Wagner, B., «Liable, but Not in Control?», *op. cit*. N. 1334.

[1461] *Ibid., passim.*

[1462] El autor analiza escenarios reales de aplicación (ibid., pp. 109 y ss.), examinando vehículos autónomos, con estudios sobre la capacidad y velocidad de reacción; verificaciones policiales basadas en los datos PNR de pasajeros; y sistemas automáticos de moderación de contenidos.

sabilidades por parte de los proveedores [1463]. Así, la supervisión, lejos de actuar como salvaguarda efectiva frente a los riesgos de la automatización, funciona como un mecanismo de legitimación y distribución descendente de responsabilidad. En la misma línea, Beck y Burri analizan cómo la legislación ha pasado de usar la denominación «control humano» a «supervisión humana», afrontando esta última obstáculos insalvables inherentes a la psicología humana, insistiendo en la idea de que la actual apuesta por la supervisión (entendida como capacidad de comprender, evaluar y eventualmente intervenir para anular el funcionamiento de un sistema), acaba por conducir al supervisor humano a «zonas de deformación moral» [1464].

La evidencia científica no es, en este caso, esperanzadora, sino todo lo contrario. Sin ánimo de agotar la literatura científica en este campo, pueden citarse algunos estudios que respaldan las siguientes afirmaciones:

### 2.1 **Los humanos tienden a validar las decisiones de las máquinas (sesgo de automatización)** [1465]

Numerosos trabajos muestran que tanto legos como profesionales experimentados pueden desarrollar lo que se conoce como «sesgo de automatiza-

---

[1463] Green, B., *op. cit*., N. 1458, p. 7: *«First, drawing on recent empirical evidence about how people interact with algorithms in government and other settings, I consider whether people can provide the types of oversight that policies call for. Despite the hopes of policymakers, the vast majority of evidence suggests that people cannot provide the envisioned protections against algorithmic errors, biases, and in- flexibility. Second, I consider the implications of these limits to human oversight. Ungrounded assumptions of effective hu- man oversight promote a false sense of security in adopting algorithms and reduce the accountability that vendors and policymakers face for algorithmic harms».*

[1464] Beck, J./ Burri, T., «From "Human Control" in International Law to "Human Oversight" in the New EU Act on Artificial Intelligence», en Mecacci J. *et al, Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*, Elgar, 2024, p. 120, haciendo referencia a la expresión «*moral crumple zone»,* describiéndola como aquella situación en que la responsabilidad por un resultado concreto se atribuye, por error, a un actor humano con un control limitado sobre el sistema autónomo, de forma que el humano, en las interacciones humano-máquina altamente complejas, se puede convertir en mera ancla para la atribución de responsabilidad en caso de mal funcionamiento.

[1465] Un repaso por la evidencia científica que ha demostrado el peso de este sesgo en: Green, B., *op. cit.* N. 1458, p. 8 y ss., con citas a numerosa evidencia empírica. En resumen, se puede afirmar que numerosos estudios han demostrado que las personas (incluyendo también a los expertos) son susceptibles a este sesgo, lo que implica que tienden a confiar en los sistemas automatizados, lo cual entraña una reducción del nivel de análisis en la toma de decisiones, conduciendo tanto a errores por omisión como por comisión. Además, se ha demostrado que automatizar solo ciertas partes de las tareas encomendadas al humano puede hacer que el resto de partes (las no automatizadas) se tornen más complejas, dando lugar a distintos tipos de errores en lugar de reducir el número total. Asimismo, la automatización disminuye la sensación de control, de responsabilidad y de agencia moral de los operadores humanos, de tal forma que los juicios sobre cuándo y cómo el humano debe desviarse de la recomendación algorítmica suele ser incorrecto. Se nombran también estudios (especialmente en el ámbito policial y judicial) donde se muestra que el juicio humano agrava las disparidades raciales, facilitando la introducción de nuevas formas de inconsistencia y sesgo en las decisiones.

ción»: una inclinación a confiar excesivamente en las recomendaciones de sistemas automáticos, lo que disminuye la capacidad crítica o el análisis autónomo en la toma de decisiones [1466]. Este sesgo puede dar lugar a errores por omisión (cuando no se interviene debido a la ausencia de alertas automatizadas), así como a errores por comisión (cuando se siguen indicaciones automatizadas erróneas, incluso ante evidencias contrarias). Asimismo, la automatización de ciertas tareas puede dificultar la correcta ejecución de las tareas restantes y provocar un deterioro progresivo de las competencias humanas necesarias para su supervisión [1467]. En consecuencia, lejos de eliminar errores, los sistemas automatizados suelen transformarlos en errores de otra índole o naturaleza, generando una sensación de disminución de la sensación de control, responsabilidad y agencia moral [1468].

### 2.2 **Los humanos acaban por interiorizar inconscientemente el sesgo de los sistemas**

En un reciente estudio Glickman y Sharot [1469] investigan, mediante una serie de experimentos con más de 1.400 participantes, cómo la interacción repetida entre humanos y sistemas de inteligencia artificial puede amplificar los sesgos en los juicios humanos, produciendo un efecto de retroalimentación. Los autores demuestran que incluso sesgos menores presentes en los algoritmos (ya sea por los datos con los que fueron entrenados o por sus propias características) tienden a intensificarse no solo en la máquina, sino que, tras el contacto, son internalizados y amplificados por las personas. Este fenómeno se documenta en tareas perceptivas, emocionales y sociales, y es más acusado que el sesgo transmitido

[1466] Parasuraman, R./Manzey, D. H., «Complacency and Bias in Human Use of Automation: An Attentional Integration», *Human Factors: The Journal of the Human Factors and Ergonomics Society* vol. 52, 3, 2010, pp. 381-410; Skitka, L. J./ Mosier, K. L./, Burdick, M. /Rosenblatt, N., «Automation bias and errors: Are crews better than individuals?», The *International Journal of Aviation Psychology,* 10, 1, 2000, pp. 85-97. Recientemente, un estudio sobre las decisiones de concesión de libertad previa al juicio en EEUU: Green, B./ Chen Y., «Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments», *Proceedings of the Conference on Fairness, Accountability, and* Transparency, 2019.

[1467] Bainbridge, L., «Ironies of automation», *Automatica*, 6, 1983, pp. 775-779.

[1468] Berberian, B./ Sarrazin, J-C./ Le Blaye, P./Haggard, P., «Automation Technology and Sense of Control: A Window on Human Agency», *PLoS ONE*, 7, 2012, desarrollando un estudio basado en un simulador de vuelo con diferentes niveles de automatización, llegando a la conclusión de que, a mayor automatización, menor sensación de agencia. También: Cummings, M. L., «Automation and accountability in decision support system interface design» *J. Technol. Stud*, 32, 2006, pp. 23-31, utilizando el término «amortiguador moral», p. 26, *(«moral buffer»)* para referirse a la forma de distanciamiento de las personas respecto del resultado de los sistemas**.**

[1469] Glickman, M./Sharot, T., *op. cit.* N. 854, pp. 345-359.

únicamente entre humanos. Los experimentos muestran, además, que los participantes a menudo subestiman el grado en que las recomendaciones del sistema afectan su juicio. Los resultados se extienden a contextos reales, mostrando cómo, por ejemplo, exponerse brevemente a imágenes generadas por IA sesgadas influye en las percepciones sociales y de género de los individuos. El trabajo concluye que la interacción humano-IA puede desencadenar una espiral de amplificación de sesgos difícil de detectar [1470].

### 2.3 **Los humanos acaban por perder capacidades cuando interactúan con sistemas que están diseñados para desempeñar tareas que les son propias**

Ligeramente relacionado con el sesgo de automatización, puede hablarse del efecto de *«deskilling»* o desprofesionalización, que puede darse en humanos sometidos al uso continuo de tecnologías, y que provoca o bien un desgaste en sus habilidades ya existentes, o bien la falta de desarrollo de competencias que normalmente formarían parte del proceso de adquisición de experticia [1471].

También en este sentido puede nombrarse el reciente trabajo de Sutton, Arnold y Holt [1472], que analizan el impacto de la automatización de los siste-

---

[1470] Un interesante dato que los investigadores ponen de manifiesto es que esta «absorción» del sesgo se produce incluso cuando los humanos no son conscientes de que interactúan con sistemas de IA, aunque en menor medida. *Ibid. ,* p. 354: *«Specifically, we observed that when participants were told they were interacting with a human when in fact they were interacting with an AI, they learned the AI's bias to a lesser extent than when they believed they were interacting with an AI (although they did still significantly learn the bias)».*

[1471] En un primer momento, desarrollando la «Teoría de la Dominancia Tecnológica»: Arnold, V./ Sutton, S. G., «The theory of technology dominance: Understanding the impact of intelligent decision aids on decision makers' judgments», *Advances in Accounting Behavioral Research* 1, 1998, pp. 175-194. Postulan estos autores que un tomador de decisiones puede volverse dependiente de una ayuda inteligente para la toma de decisiones bajo dos condiciones: que tenga poca experiencia, y que presente niveles altos en todos los factores influyentes (experiencia en la tarea, complejidad de la misma, familiaridad con la ayuda y ajuste cognitivo), generando un efecto de pérdida de habilidades a largo plazo en el usuario y obstaculizando su crecimiento cognitivo en su área. Más recientemente, actualizando su teoría, teniendo ya en cuenta las complejidades introducidas por la IA, con el objetivo de comprender cómo, por qué y cuándo ocurren estos efectos: Sutton, S. G./ Arnold, V./Holt, M., «An extension of the theory of technology dominance: Capturing the underlying causal complexity», International Journal of Accounting Information Systems, 50, 2023, pp. 1-18.

[1472] Sutton, S. G./ Arnold, V./ Holt, M., «How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work», *Journal of Emerging Technologies in Accounting*, vol. 15, 2, 2018, pp. 15-25. La reflexión final, en p. 22, relativa al necesario cambio sobre el concepto de competencia humana *(expertise)* nos parece interesante. La reproducimos traducida al español: «Probablemente, la competencia no deba evaluarse en términos del cerebro humano operando de forma aislada, sino en función de si el cerebro humano puede importar eficientemente hechos, información y reglas de procesamiento para

mas de inteligencia artificial en el trabajo profesional basado en el conocimiento, especialmente en el ámbito de la contabilidad. Los autores advierten sobre el fenómeno de «tecnología dominante» y el «sesgo de automatización», que pueden llevar a una dependencia excesiva de los sistemas automatizados y a la consiguiente desprofesionalización o pérdida de habilidades *(«deskilling»)* entre los trabajadores del conocimiento. Argumentan que la incorporación de sistemas inteligentes puede obstaculizar tanto la adquisición como la retención de saber experto, reduciendo la autonomía y la creatividad profesional, y sugieren que esta transformación tiene profundas implicaciones éticas y sociales.

Explican estos autores que cada vez más las organizaciones ofrecen a sus empleados pocas alternativas más allá de emplear los sistemas de inteligencia artificial disponibles. Si, como muestran las evidencias empíricas, los usuarios no desarrollan las competencias y conocimientos que habitualmente se adquieren a través de la experiencia, con el tiempo todos tienden a convertirse en usuarios con bajo nivel de experiencia o escasa pericia, y los demás factores pierden relevancia, terminando por depender inevitablemente del sistema [1473].

### 2.4 Los humanos tienen una capacidad de reacción limitada en contextos que requieren respuestas rápidas

Así lo pone de manifiesto Wagner [1474], aludiendo a diversos estudios, en uno de los campos más explorados, como es el de los vehículos autónomos, explicando que el tiempo necesario para que una persona transite de un «estado automatizado» a un «estado manual» en el que el conductor retoma plenamente el control sigue siendo relativamente elevado. Normalmente, se requieren alrededor de 10 segundos para una reasunción básica del control [1475] y entre 30 y 40 segundos para «reanudar un control adecuado» [1476]. Al mismo tiempo, hay consenso en que exigir a los conductores humanos que retomen el

tomar decisiones efectivas en dominios de decisión complejos». Los cambios relativos a la competencia técnica habrán de afectar, inevitablemente, a nuestros modelos penales.

[1473] *Ibid.,* p. 17.

[1474] Wagner, B., «Liable, but Not in Control?», *op. cit*. N. 1334, p. 109.

[1475] Inicialmente, planteando el problema: Kyriakidis et al., «A human factors perspective on automated driving», *Theoretical Issues in Ergonomics Science*, 20, 3, 2017, pp. 223-249. Una investigación completa con conclusions basadas en la observación en: Lu, Z./ Coste, X./ de Winter, J., «How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving», *Applied Ergonomics: human factors in technology and society,* 60, 2017, pp. 293-304.

[1476] Merat, N./ Jamson, A. H./ Lai, F. C. H./ Daly, M./ Carsten, O. M. J., «Transition to manual: Driver behaviour when resuming control from a highly automated vehicle» *Transportation Research Part*, 27, 2014, p. 281.

control del vehículo no constituye una solución adecuada, ya que la desconexión de la automatización no es un método particularmente eficaz para mantener a los conductores involucrados en el proceso [1477].

En definitiva, y tras el análisis de algunos de los estudios empíricos más indiscutidos, en lo que aquí nos interesa, al derecho penal le han de importar las evidencias empíricas, si de ellas se desprenden deberes de muy difícil o casi imposible cumplimiento, deberes para los cuales deberíamos desplegar capacidades de las que, aunque se nos suponen, carecemos. Además de esta innegable evidencia, cabe concluir con un argumento más, de tipo ético, y que tiene que ver con el papel en que esta asunción de responsabilidad imposible asume cada persona. Sobre esto, compartimos plenamente la reflexión de Lazcoz, que nos invita a repensar qué valores humanos queremos preservar en nuestra forma de interactuar con la tecnología. Concretamente, cuando se trata de la supervisión humana, esto implica, valga la redundancia, «humanizar» al supervisor, es decir, tratar al humano como un fin en sí mismo [1478]. Este trato humano pasa por una decisión del legislador (en este caso, creemos, eximiendo de responsabilidad penal), pero también de los operadores económicos que ejercen su autoridad sobre los supervisores, a los que se encomienda esta ardua labor en su puesto de trabajo, poniendo en juego sus derechos como trabajadores y colocando sobre ellos, en definitiva, una espada de Damocles. No debemos olvidar que supervisores de la IA acabaremos por ser todos, sea en nuestro papel de conductores, funcionarios públicos, trabajadores, jueces, policías, banqueros o médicos.

La dicotomía surgida como consecuencia de esta espada de Damocles no queda, empero, ahí, sino que acaba por salpicar a toda la dogmática penal. Con la ubicua automatización de decisiones, la permisión de riesgos acaba por alcanzar a ámbitos insospechados y de lo más variados, y ello implica, por la afectación de los bienes en juego, que allá donde se automaticen decisiones habrá que resolver situaciones dilemáticas [1479]. El mecanismo en este caso es común a todas las áreas donde un sistema de IA pueda asistir en la toma de decisiones. El Estado ha de permitir que, a pesar de generarse riesgos para ciertos bienes jurídicos (pueden ser, evidentemente, de lo más variados), los sistemas decidan con arreglo a una determinada programación. Esta autoriza-

---

[1477] *Ibid.,* p. 274.

[1478] Lazcoz Moratinos, G., «La vigilancia o supervisión humana …, *op. cit*. N. 1331, p. 698.

[1479] Así, Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 105. Destaca Silva, como es habitual, el manido problema de los dilemas en casos de vehículos autónomos, pero cabe mencionar que no es, ni de lejos, el único. El trabajo de referencia en este campo es Coca Vila, I., «Coches autopilotados …», *op. cit*. N. 220. Igual, en inglés: Coca Vila, I., «Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law», *Criminal Law and Philosophy*, 12, 2018, pp. 59-82.

ción no es baladí, pues, como acertadamente manifiesta Silva, comporta la permisión *ex ante* de modelos de decisión automatizada que, por definición, van a poner en peligro bienes jurídico-penales, convirtiendo lo que normalmente podían ser conductas típicas y antijurídicas en conductas amparadas por el riesgo permitido [1480]. La autorización ya no depende de que, en una situación conflictiva concreta, la conducta sea conforme a derecho. Esa decisión individualizada –propia de las causas de justificación– no será adoptada en muchos casos por un humano, sino por el sistema. Ahora bien, configurar esa decisión resulta problemático, porque la realidad ofrece infinidad de supuestos difíciles de programar. Esto es aún más evidente si tenemos en cuenta que la metodología actual no se basa en programar soluciones caso por caso, sino en la experimentación y en el aprendizaje dinámico del propio sistema para cada situación particular. La verdadera situación dilemática se traslada, por tanto, a un momento anterior: el momento en el que hay que decidir si el sistema (con toda la automatización de la que es capaz), ha de usarse en un determinado contexto o no [1481]. En palabras de Silva, la declaración de conformidad a derecho de la puesta en peligro de ciertos bienes jurídicos intrínsecos reduce los niveles sistemáticos de exención de responsabilidad a su mínima expresión: o riesgo permitido o injusto culpable [1482]. Por decirlo con nuestras propias palabras, el momento relevante para la consideración de todos los factores en juego no puede ser uno en el que sea la máquina, y no el humano, quien haga la valoración del caso concreto, lo que nos obliga, una vez más, a poner toda nuestra atención en las fases previas, de control interno.

## 3. REFLEXIÓN FINAL

La pregunta central en este capítulo y, en general, la que vertebra toda la investigación, tiene que ver con quién ha de soportar los riesgos del progreso tecnológico.

No se ha tratado de negar que la tecnología pueda conllevar, en algunos sectores, ventajas y mejoras que repercutan positivamente en el bienestar de la ciudadanía. Esta valoración, de carácter eminentemente político, se ha ido to-

[1480] Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 111.

[1481] Así, destacadamente, Selbst, A., «Negligence …», *op. cit*. N. 718, p. 1339, afirmando que la decisión sobre la «justificación» del uso no dependerá del caso concreto, donde el control escapa en gran medida al usuario, sino que se deberá tomar antes, solo a la hora de valorar si el uso de un determinado sistema en un determinado contexto genera o no un riesgo permitido. *« […]the reasonableness of an action in individual cases must be tied to the decision to use AI in the first place»*.

[1482] Silva Sánchez, J. M., *El riesgo permitido, op. cit.* N. 886, p. 111.

mando a medida que se iban descubriendo las limitaciones, riesgos y bondades de la prometedora inteligencia artificial.

En un momento en el que la incertidumbre sobre los deberes va reduciéndose paulatinamente, es papel del legislador penal, y también de la doctrina, poner coto a la inseguridad jurídica extrema que ha reinado hasta el momento.

La propuesta de tipificación que aquí se mantiene pretende, ante todo, dotar de un papel mínimo al derecho penal, siendo conscientes de su potencial preventivo, pero también de los necesarios límites al poder punitivo, imprescindibles aún en épocas de catastrofismo tecnológico. La opción con la que se concluye esta propuesta es, por tanto, limitada, y puede ser objeto de críticas tanto por exceso como por defecto, pero se considera que puede ser una buena forma de colmar las lagunas que podrían generarse si atendemos solo al modelo hoy establecido en el Código penal.

Quedan, desde luego, muchos interrogantes por responder, porque, aunque se ha tratado de poner coto a la ilimitada libertad de empresa de los poderosos operadores económicos del sector tecnológico, en un intento de evitar los riesgos del «capitalismo de vigilancia» basado en el paradigma de la libertad absoluta de mercado, la opción por no castigar, o al menos, no castigar más allá de los claros casos en que se pueda afirmar una imputación de resultado lesivo conforme a los parámetros más clásicos, responde a una decisión meditada que llama a la parsimonia en el castigo de quien no está llamado a lucrarse de esta industria sino a quien, por ser hijo de su época, «pasaba por allí». Retomamos, así, la imagen de un Lorca sumido en la preocupación al ver en qué se estaba convirtiendo ese mundo industrializado y despersonalizado, pero con la intención de no hacer pagar al paseante, que atiende, llevado por la masa, al imparable cambio, con una responsabilidad que le pilla por sorpresa, sobre la que aún no existen límites claros y que ha de pensarse, por tanto, aún con mucha más calma, si no queremos que la recurrente expansión del capitalismo salvaje acabe por repercutir, solo, a quienes atienden como espectadores a la evolución de la enésima sociedad del riesgo.

Retomando, así, el lenguaje lorquiano, es necesario entender que en un mundo cada vez más atravesado por procesos automatizados, donde la velocidad de la innovación tecnológica supera la capacidad de comprensión y decisión del ciudadano común, se impone una reflexión sobre la distribución de riesgos y responsabilidades. La promesa de la inteligencia artificial (como en época de Lorca lo fue la del cemento y la electricidad) se instala no solo como progreso, sino como imperativo no consultado, donde el individuo no elige participar. Y, sin embargo, al tiempo que se le (nos) impone esta exposición forzada, se le (nos) atribuye la carga moral y jurídica de resistir, consentir o

adaptarse. El ciudadano es, así, un espectador pasivo en la «última fiesta de los taladros», donde las decisiones las toman otros (las grandes plataformas que Zuboff caracteriza en su obra), pero las consecuencias acaban por recaer sobre los más vulnerables. No podemos consentir que el ser humano, ya afectado por los riesgos estructurales de esta nueva arquitectura de poder, sea además convertido en el chivo expiatorio de sus efectos. La responsabilidad y el beneficio han de ir, pues, de la mano. Si el lucro es de unos pocos, la carga no puede ser de todos.

*Yo denuncio la conjura*
*De estas desiertas oficinas*
*Que no radian las agonías*

Oficinas de acero y papel o, en su versión actual, oficinas de silicio y de datos, pero igual de impasibles ante las consecuencias que provocan. Nuestro papel como juristas pasa por abrir los canales necesarios para que la justicia no sea cancelada en nombre del progreso.

## CONCLUSIONES

1. Para hallar un concepto de «inteligencia artificial» capaz de explicar los desafíos jurídicos que plantea, es necesario atender a la evolución de la técnica, distinguiendo claramente entre la IA simbólica y la IA basada en aprendizaje automático, que funciona de manera experimental y sin indicaciones explícitas sobre cómo actuar en cada escenario. La definición más completa hasta la fecha es la del Grupo de Expertos de alto Nivel de 2018, pues destaca los rasgos esenciales: sistemas con objetivos definidos por humanos, alimentados por datos masivos, capaces de procesar información para resolver problemas humanos. Una definición completa debe incorporar el elemento de autonomía, recogido en el Reglamento de IA. Dicha autonomía debe entenderse en sentido operativo y no filosófico: no implica que la IA genere sus propios fines, sino que decide cómo alcanzar los objetivos fijados por humanos. Esta característica genera fenómenos de «caja negra», donde las decisiones del sistema resultan difíciles de explicar y el control humano se ve limitado. La combinación de autonomía y complejidad en los sistemas de IA avanzada supone un salto cualitativo respecto de la IA simbólica, planteando desafíos para el derecho penal, que se fundamenta en el control y la responsabilidad.

2. Existe un riesgo de que se produzcan vacíos de responsabilidad, ya que la autonomía creciente de los sistemas puede debilitar el vínculo entre la acción humana y sus consecuencias, dificultando la atribución de responsabilidad. La IA incide especialmente en *la responsabilidad causal* –por la dificultad de trazar el nexo entre hecho y causa humana– y en *la adscripción de responsabilidad,* al complicar la identificación de los hechos como propios de un individuo. Los problemas jurídicos se presentan de dos maneras: por un lado, la culpa penal basada en la intención, el conocimiento o el control se ve comprometida, dado que los sistemas autónomos pueden generar excusas legíti-

mas para los errores; por otro, existe el riesgo de castigos injustos, pues para evitar la impunidad podría buscarse responsabilizar a humanos como «chivos expiatorios», imponiendo sanciones desproporcionadas e injustas, atribuyendo toda la gravedad de un daño. Aunque algunos autores cuestionan la existencia de lagunas, pues la responsabilidad penal depende del cumplimiento de deberes de diligencia, el verdadero problema radica en definir esos deberes en el contexto de las nuevas tecnologías y los riesgos que estas generan.

3. En la búsqueda de modelos de castigo válidos, se descarta el castigo directo de la IA, pues los sistemas no poseen capacidad cognitiva, moral ni de control suficientes para ser considerados responsables penalmente, partiendo de que el castigo penal requiere agencia humana, capacidad de culpabilidad y capacidad de verse afligidos por una pena. Tampoco resulta suficiente el modelo de responsabilidad de la empresa, en tanto los modelos de autorresponsabilidad redundan en los problemas de capacidad de culpabilidad y los de heterorresponsabilidad resultan insatisfactorios porque acaban proponiendo castigos basados en la mera asunción de riesgos. La propuesta más viable es la del delito de peligro, ya analizada por diversos autores angloamericanos que proponen crear tipos penales por el riesgo o el uso indebido de sistemas de IA. Este enfoque enfrenta tres desafíos clásicos: (i) la reducción del principio de lesividad, lo que obliga a establecer umbrales de daño y mecanismos de cooperación entre expertos ante la incertidumbre tecnológica; (ii) la objetivación de la culpabilidad, advirtiendo la tendencia a matizar las exigencias subjetivas, incluso en el ámbito continental, y (iii) la proporcionalidad de la pena.

4. Se observa una transformación social en la que la sociedad contemporánea (caracterizada como sociedad del riesgo algorítmico), ha transitado del paradigma del riesgo industrial al paradigma del riesgo algorítmico, en el que los riesgos son digitales, estructurales y globales, generados por tecnologías como la IA, *el big* data y la vigilancia masiva. Emergen nuevas formas de poder y deshumanización, asociadas a fenómenos como el dataísmo, el tecnofeudalismo y el capitalismo de la hipervigilancia, que afectan a la autonomía individual, la privacidad y la capacidad crítica, concentrando decisiones que afectan a derechos fundamentales en estructuras opacas. El derecho tradicional debe adaptarse a estos riesgos no tradicionales, anticipados y despersonalizados. En este contexto, el principio de precaución, entendido en su vertiente proactiva, se presenta como una herramienta útil destinada a guiar la labor del legislador en la adopción de medidas de cuidado y regulación frente a tecnologías potencialmente lesivas, sin reemplazar los criterios de imputación penal basados en la previsibili*dad ex* ante del riesgo ni la presunción de inocencia. Asimismo, este principio permite decidir colectivamente, sobre la

base de evidencia científica, acerca del nivel de riesgo tolerable, fomentando normas de protección ante riesgos catastróficos o irreversibles.

5. El modelo de delito de resultado inspirado en el paradigma de la responsabilidad penal por el producto defectuoso presenta diversos inconvenientes. En primer lugar, la IA, debido a su autonomía operativa, dinamismo y opacidad, no encaja en la lógica del producto defectuoso, no tanto por su condición de producto, sino porque el juicio sobre su defectuosidad resulta complejo y dependiente de su desempeño y contexto. A ello se suman problemas de causalidad derivados de la lógica de caja negra de los sistemas: la multiplicidad de factores y la opacidad interna dificulta la reconstrucción de la cadena causal. No obstante, un concepto mínimo de causalidad basado en evidencias probabilísticas suficientemente sólidas puede permitir establecer el vínculo fáctico entre acción del sistema y daño.

6. Más difícil resulta, empero, la atribución del resultado lesivo habida cuenta de la imprevisibilidad. Aunque será posible identificar que hubo una contribución causal del sistema («sí»), mucho más complejo será reconstruir la forma concreta en que se produjo el daño («cómo»). Los problemas de previsibilidad sobre el resultado afectan en dos planos: (i) en la imputación subjetiva, donde surge la pregunta sobre si es posible imputar un resultado cuya previsión era difícilmente representable para el operador– que se presenta como una suma de pequeñas contribuciones individuales mediadas por la tecnología, y (ii) en la imputación objetiva, puesto que se plantea si es legítimo considerar que el riesgo generado se ha concretado en el resultado si no podemos saber cuáles son los factores concurrentes que han determinado el concreto curso causal.

7. En la imputación subjetiva, el dilema que presenta la imprevisibilidad intrínseca del sistema es, pues, que o bien todo evento lesivo de la IA será imputable, en cuanto es técnicamente posible, o bien ninguno lo será. La interacción humano-IA obliga a ajustar los deberes de cuidado, pero siempre considerando la posibilidad real de controlar los resultados y evitar los daños. La complejidad de los sistemas conducirá a la especialización y la compartimentación de roles, de modo que cada persona en el ciclo de vida del sistema poseerá conocimientos muy limitados, generando posibles lagunas de responsabilidad respecto del resultado final, obligando a descartar el modelo de hombre medio ideal como base para el deber de cuidado en la imprudencia.

8. La IA no requiere la creación de nuevos bienes jurídicos, sino más bien la adaptación de los ya existentes. Lo particular de los ataques perpetrados mediante sistemas de IA es la forma de ataque respecto de los bienes tradicionales, especialmente la privacidad y la no discriminación, que pueden verse afectados

de manera masiva e indirecta. La protección penal debe centrarse en la defensa de estos bienes frente a ataques difusos o colectivos, sin alterar su naturaleza individual. Se propone como modelo idóneo el de protección de las condiciones necesarias para el disfrute de estos bienes individuales, con la protección de un bien jurídico intermedio llamado *«seguridad colectiva algorítmica o digital»,* como forma de protección mediata de los bienes finales.

9. *El delito de peligro abstracto de idone*idad es el modelo adecuado para la regulación penal de la IA, pues permite intervenir antes de que se produzca un daño concreto, protegiendo las condiciones de seguridad de bienes en sectores regulados. Se fundamenta en la idea de desorganización del ámbito propio del actor como forma de responsabilidad por el proceso. El modelo del delito de idoneidad –evitando el recurso a un delito de peligro abstracto puro– impide una aplicación formalista o automática, garantizando que el juez verifique que la conducta sea potencialmente lesiva según las circunstancias del caso concreto. Subjetivamente, se propone tipificar la modalidad dolosa, con el fin de limitar la intervención penal.

10. El modelo más adecuado para esta propuesta es el de los delitos contra la seguridad colectiva, que protege a sujetos indeterminados frente a riesgos difusos, combinando la defensa de bienes individuales como la vida, salud, privacidad, y no discriminación. La normativa debe identificar *obligaciones sustantivas de seguridad,* diferenciándolas de las obligaciones formales, para garantizar la efectividad del control y la protección de los bienes finales. La contravención dolosa de normas de organización segura dentro del ámbito propio colma los requisitos de antijuridicidad y justifica la intervención penal. Se evita así trasladar la responsabilidad al cumplimiento formal sin conexión con el riesgo real.

11. El Reglamento de Inteligencia Artificial define el riesgo permitido en el sector, creando un marco de referencia extrapenal cuya conformidad exime de responsabilidad siempre que los operadores ajusten su conducta a los estándares de seguridad y diligencia requeridos. El modelo a seguir es el de la accesoriedad asimétrica, según el cual las conductas conformes al Reglamento no generan responsabilidad penal, y los incumplimientos normativos solo son indicios de delito, pero no determinantes. Las obligaciones penalmente relevantes del Reglamento se concentran en fabricantes y responsables de despliegue, quedando excluidos los usuarios no profesionales. Los usos prohibidos protegen bien jurídicos fundamentales como la autonomía, la dignidad, la libertad, la intimidad, la no discriminación y la presunción de inocencia; siendo, por ello, obligaciones sustantivas a efectos penales. Dentro de las obligaciones aplicables a sistemas de alto riesgo podemos diferenciar entre: (i) deberes sus-

tantivos, que incluyen obligaciones de implantación de sistemas de riesgo y cumplimiento de estándares de calidad de datos, así como garantizar niveles de solidez y ciberseguridad medibles que eviten errores, (ii) las obligaciones formales, entre las que se encuentra la obligación de mantener documentación técnica y registros de actividad, cuya relevancia penal ha de ser excluida por su carácter meramente instrumental; y (iii) las obligaciones intermedias, es decir, los deberes de transparencia y supervisión, que, dada su escasa aplicabilidad práctica y su carácter de principios y no tanto de normas, solo podrán tener relevancia penal en caso de vulneraciones patentes y ausencia absoluta de requisitos mínimos.

12. El régimen de vigilancia poscomercialización fijado por el Reglamento será clave para integrar los deberes del fabricante y operadores afines, pudiendo dar lugar a la responsabilidad por el resultado en omisión en caso de no ejecución de medidas correctoras ante incidentes graves, partiendo de la posición del fabricante como garante de una fuente de riesgo.

13. Respecto al resto de normativa complementaria, el Reglamento deberá interpretarse junto con el RGPD y la nueva Directiva sobre responsabilidad por productos. Aunque esta última incluye en su ámbito los sistemas de IA, no resuelve los problemas de incertidumbre relacionados con los riesgos de desarrollo que exoneran de responsabilidad, ni las dudas sobre la defectuosidad de un producto dinámico. Su única ventaja, por tanto, será el alivio de la carga de la prueba, aunque bajo parámetros aún relativamente inciertos. Por último, el Convenio marco del Consejo de Europa podría servir en el futuro para integrar los principios éticos que orienten la aplicación del Reglamento.

14. El Reglamento deja pendiente de desarrollo nacional el establecimiento del concreto régimen sancionador, limitándose a exigir que las sanciones sean «proporcionadas, eficaces y disuasorias». En la búsqueda de una sanción de estas características, se apuesta por una propuesta de tipificació*n de lege ferenda* que, además de castigar con pena de multa, y prisión en los casos extremos, incluya la prohibición de ejercer actividades relacionadas con la IA, partiendo de la insuficiente capacidad aflictiva y preventiva de la multa.

15. La propuesta de tipificación consiste en la introducción de un nuevo capítulo en el Código Penal titu*lado «De los delitos contra la seguridad digital en sistemas de Inteligencia artificial»*, ubicado sistemáticamente dentro de los delitos contra la seguridad colectiva. Su objetivo es tipificar exclusivamente obligaciones directamente orientadas a proteger bienes fundamentales para el derecho penal frente a nuevas formas de ataque, bajo la premisa de la idoneidad. Mientras los delitos por usos prohibidos son comunes, dirigidos al público en general, los delitos por incumplimiento de obligaciones sustantivas en

sistemas de alto riesgo son especiales, limitados a los operadores obligados por el Reglamento, en la idea de individualización de roles y limitación de responsabilidades. La propuesta se dirige a exigir que la responsabilidad penal recaiga solo en quienes tienen verdadero poder de control sobre tecnologías de alto riesgo, evitando que trabajadores o usuarios terminen pagando los riesgos derivados de decisiones empresariales.

# ANEXO I. PROPUESTA DE REGULACIÓN

## TÍTULO XVII

## De los delitos contra la seguridad colectiva

## CAPÍTULO V

## De los delitos contra la seguridad digital en sistemas de Inteligencia artificial

### *Sección 1.ª De los usos prohibidos*

**Artículo X.1** *Uso de sistemas de IA para manipulación.*

El que desarrollare, comercializare, pusiera en servicio o utilizare un sistema de inteligencia artificial diseñado para aplicar técnicas subliminales más allá de la conciencia de la persona, o que explotare vulnerabilidades de personas debidas a su edad, discapacidad o situación social para distorsionar de manera significativa su comportamiento cuando ello resulte objetivamente idóneo para poner en grave peligro la libertad, igualdad, integridad psíquica, integridad moral o salud, será castigado con la pena de prisión de seis meses a tres años, multa de seis a veinticuatro meses e inhabilitación especial para actividades relacionadas con sistemas de inteligencia artificial por tiempo de uno a cinco años.

**Artículo X.2** *Uso de sistemas de categorización social prohibida.*

Con la misma pena será castigado el que, de manera objetivamente idónea para generar discriminaciones injustificadas, restricciones indebidas de derechos o daños graves a la libertad, igualdad, integridad psíquica o moral de las personas desarrollare, comercializare, pusiera en servicio o utilizare un sistema de inteligencia artificial:

1. Para establecer puntuaciones sociales de personas físicas por una autoridad pública u operador autorizado.
2. Para realizar inferencias o evaluaciones totalmente automatizadas sobre la probabilidad de que una persona cometa delitos futuros, basándose en su perfil personal, características individuales o conductas pasadas.
3. Para elaborar bases de datos de reconocimiento facial mediante la recolección indiscriminada de imágenes de personas en espacios accesibles al público, en condiciones contrarias al Reglamento de IA.
4. Para generar un sistema de reconocimiento de emociones en contextos laborales, educativos o formativos, cuando ello sea contrario a las prohibiciones legales.
5. Para elaborar un sistema de categorización biométrica que infiera datos sensibles como origen étnico, convicciones religiosas, orientación sexual o afiliación política, en contravención del Derecho de la Unión Europea o nacional.
6. Que permita, sin habilitación legal expresa y fuera de los supuestos autorizados por el Derecho de la Unión Europea o nacional, la identificación biométrica remota en tiempo real en espacios accesibles al público.

*Sección 2.ª Del incumplimiento de obligaciones para los sistemas de IA de alto riesgo*

**Artículo Y.1** *Infracción de deberes de control en sistemas de inteligencia artificial de alto riesgo.*

Será castigado con la pena de prisión de seis meses a tres años, multa de seis a veinticuatro meses e inhabilitación especial para actividades relacionadas con sistemas de inteligencia artificial por tiempo de uno a cinco años, el que, en el desarrollo, comercialización, puesta en servicio o utilización profesional de un sistema de inteligencia artificial calificado como de alto riesgo conforme al Derecho de la Unión Europea, y de forma objetivamente idónea

para generar un daño grave a la vida, integridad física o psíquica, salud, intimidad, o una grave discriminación:

1. Incumpliere las obligaciones de establecimiento, aplicación, documentación y actualización del sistema de gestión de riesgos previsto legalmente.
2. Incumpliere los requisitos legales relativos a la gobernanza de los datos y a la calidad, representatividad, exactitud y ausencia de sesgo de los conjuntos de datos empleados en el entrenamiento o validación del sistema.
3. Incumpliere los requisitos legales relativos a la precisión, solidez técnica, resiliencia a errores y protección frente a accesos no autorizados.
4. Incumpliere las obligaciones legales relativas a la integración de mecanismos eficaces de supervisión humana, cuando ello permita razonablemente prever que el sistema operare en condiciones de absoluta autonomía técnica sin ningún mecanismo de control humano.
5. Incumpliere las obligaciones legales de transparencia que permitan a los usuarios profesionales interpretar y operar adecuadamente el sistema.

**Disposición común.** *Responsabilidad penal de las personas jurídicas.*

**Artículo Z.** *Responsabilidad de las personas jurídicas por infracciones relativas a sistemas de inteligencia artificial.*

Cuando, de acuerdo con lo establecido en el artículo 31 bis, una persona jurídica sea responsable de los delitos recogidos en esta sección, se le impondrá la pena de multa de dos a cinco años.

# ANEXO II. BIBLIOGRAFÍA


Aba Catoira, A., «Precisión y solidez de los sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 701-720.

Aba Catoira, A./Balaguer Callejón, F./Cotino Hueso, L./Hernández Ramos, M./Presno Linera, M./Rebollo Delgado, L./Tudela Aranda, J., «Encuesta sobre Inteligencia artificial y Derechos fundamentales», *Teoría y Realidad Constitucional*, 55, 2025.

Abadías Selma, A., «La protección penal de los trabajadores frente a la inteligencia artificial en el ámbito del delito de discriminación laboral», *Revista General de Derecho Penal*, 39, 2023, pp. 1-61.

Abbott, R./Sarch, A., «Punishing artificial intelligence: legal fiction or science fiction», *UC Davis Law Review*, 53(1), 2019, pp. 323-384.

Acale Sánchez, M., «Los delitos de mera actividad», *Revista de Derecho penal y Criminología*, 10, 2002, pp. 11-45.

Acemoglu, D./Johnson, S., *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*, PublicAffairs, Nueva York, 2023.

Adams-Prassl, J./Binns, R.,/Kelly-Lyth, A., «Directly Discriminatory Algorithms, *The Modern Law Review,* 2022*,* pp. 1-32,

Agustina, J. R., «Nuevos retos dogmáticos ante la cibercriminalidad», *Estudios penales y criminológicos*, 41, 2021, pp. 705-777.

Agustinoy Guilayn, A., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024.

Aires de Sousa, S., *Responsabilidade Criminal pelo Produto e o Topos Causal em Direito Penal. Contributo para uma protecção de interesses do consumidor*, Coimbra Editora, Coimbra, 2014.

AIRES DE SOUSA, S., «Não fui eu, foi a máquina: teoria do crime, responsabilidade e inteligência artificial», en Miranda Rodrigues, A. (Coord.), *A Inteligência Artificial no Direito Penal*, Almedina, Coimbra, 2020, pp. 189-215.

— «Portuguese report on traditional criminal law categories and AI», *RIDP*, Vol. 94 issue 1, 2023, pp. 319-330.

— «The importance of AI for financial crimes», *Revue Internationale de Droit Pénal*, 95/1, 2024, pp. 163-180.

AIRES DE SOUSA, S./MONIZ, H., «Manifestações do princípio da precaução no direito português», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 357-380.

ALAMILLO DOMINGO, I., «Sujetos y agentes en evaluaciones de conformidad», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 437-492.

ALCÁCER GUIRAO, R., *¿Lesión del bien jurídico o lesión de deber?: Apuntes sobre el concepto material del delito*, Atelier, Barcelona, 2003.

ALEXANDER, L./KIMBERLY KESSLER, F., «Beyond the Special Part», en Duff, R./Green, S. (Eds.), *Philosophical Foundations of Criminal Law*, Oxford University Press, Oxford, 2001, pp. 253-278.

ALONSO ÁLAMO, M., Bien jurídico penal: más allá del constitucionalismo de los derechos», *Estudios penales y criminológicos*, 29, 2009, pp. 61-105.

— *Bien jurídico penal y derecho penal mínimo de los derechos humanos*, Ediciones Universidad de Valladolid, Valladolid, 2014.

— «¿Riesgos no permitidos? Observaciones sobre la incidencia del principio de precaución en el Derecho penal», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 50-80.

— «Fundamentación pre-positiva de los bienes jurídico-penales y derecho penal mínimo de los derechos humanos», *RGDP*, 15, 2021, pp. 1-33.

ALONSO GALLO, J., «Errores y sesgos cognitivos en la expansión del Derecho penal», en AA. VV., *Derecho y justicia penal en el siglo XXI: liber amicorum en homenaje al profesor Antonio González-Cuéllar García*, Colex, 2006, pp. 31-48.

ÁLVAREZ GARCÍA, V., «La aplicación de las normas armonizadas y de las especificaciones comunes en el ámbito de la inteligencia artificial (artículos 40 y 41 Reglamento)», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 409-414.

ÁLVAREZ GARCÍA, V./TAHIRI MORENO, J., «La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque», *Revista General de Derecho Administrativo*, 63, 2023, pp. 1-56.

AMADOR HIDALGO, L., *Inteligencia artificial y sistemas expertos*, Universidad de Córdoba Servicio de Publicaciones, Córdoba, 1996.

AMELUNG, K., *Rechtsgüterschutz und Schutz der Gesellschaft. Untersuchungen zum Inhalt und zum Anwendungsbereich eines Strafrechtsprinzips auf dogmengeschichtlicher Grundlage. Zugleich ein Beitrag zur Lehre von der «Sozialschädlichkeit» des Verbrechens*, Athenäum, Frankfurt am Main, 1972.

— «El concepto de «bien jurídico» en la teoría de la protección penal de bienes jurídicos», en HEFENDEHL, R./VON HIRSCH, A./WOHLERS, W. (Eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?*, Marcial Pons, Madrid, 2016, pp. 227-264.

AMIUNE, J. M., «Varoufakis, Yanis. Tecno-feudalismo. El sigiloso sucesor del capitalismo. Ciudad Autónoma de Buenos Aires», *Ciclos en la historia, la economía y la sociedad*, 62, 2024, pp. 193-194.

AMORE, N./ROSSERO, E., *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità. Una ricerca sociologica e giuridico-penale*, Il Mulino, Bolonia, 2024.

AMOROSO, D./TAMBURRINI, G., «What makes human control over weapon systems «meaningful»?», *ICRAC Working Paper Series*, 4, 2019, pp- 1-21.

ANARTE BORRALLO, E., «Incidencia de las nuevas tecnologías en el sistema penal: aproximación al Derecho penal en la sociedad de la información», *Derecho y conocimiento: anuario jurídico sobre la sociedad de la información y del conocimiento*, 1, 2001, pp. 191-257.

ANDORNO, R., «Validez del principio de precaución como instrumento jurídico para la prevención y la gestión de riesgos», en ROMEO CASABONA, C. M. (Ed.), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, pp. 17-33.

ANDORNO, R./IENCA, M., «Towards new human rights in the age of neuroscience and neurotechnology», *Life Sciences, Society and Policy*, 5, 2017, pp. 1-27.

ANGWIN, J./LARSON, J./MATTU, S./KIRCHNER, L., «Machine Bias. There's software used across the country to predict future criminal. And it´s biased against blacks», *ProPublica*, 2016.

ANGWIN, J./MATTU, S., «Amazon Says It Puts Customers First. But Its Pricing Algorithm Doesn't», *ProPublica*, 2016.

ARELLANO TOLEDO, W./MERCHÁN MURILLO, A., «La obligación de conservar registros de los sistemas de alto riesgo en el Reglamento de Inteligencia Artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 613-626.

ARNOLD, V./SUTTON, S. G., «The theory of technology dominance: Understanding the impact of intelligent decision aids on decision makers' judgments», *Advances in Accounting Behavioral Research*, 1, 1998, pp. 175-194.

ARRANZ, L., «Leçons tirees de l'accident d'un accelerateur lineaire de l'hopital de Saragosse le 5 decembre 1990», *Radioprotection*, 44/2009, pp. 405-416.

ASARO, P., «A Body To Kick, but Still No Soul to Damn: Legal Perspectives on Robotics», en LIN, E. P./ ABNEY, P. K./BEKEY G. A., (Eds.), *Robot Ethics. The Ethical and Social Implications of Robotics*, Cambridge, 2012, pp. 169-186.

— «The liability problem for autonomous artificial agents», Association for the Advancement of Artificial Intelligence, *Papers from the 2016 AAAI Spring Symposium*, 2016, pp. 190-194.

ASHWORTH, A., «Should strict criminal liability be removed from all imprisonable offences?» *Irish Jurist,* 45, 2010, pp. 1-21.

ASIMOV, I., «El Círculo Vicioso (Runaround)», *Astounding Science Fiction*, 1942.

ATIENZA NAVARRO, M. L., «¿Una nueva responsabilidad por productos defectuosos? Notas a la Propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por daños causados por productos defectuosos de 28 de septiembre de 2022 (COM/2022/495)» *InDret* 2.23, pp. 1-53.

BACIGALUPO SAGESSE, S., «Compliance», *Eunomía*, 21, 2021, pp. 260-276.

BAINBRIDGE, L., «Ironies of automation», *Automatica*, 6, 1983, pp. 775-779.

BALKIN, J. B., «The Path of Robotics Law», *California Law Review*, 6, 2015, pp. 45-60.

BANDURA, A., «Moral Disengagement in the Perpetration of Inhumanities», *Personality and Social Psychology Review*, 3(3), 1999, pp. 193-209.

BANTEKA, N., «Artificially intelligent persons», *Houston Law Review*, 58/2021, p. 537-596.

BARATTA, A., «Principios de Derecho Penal Mínimo (Para una teoría de los derechos humanos como objeto y límite de la ley penal)», en *Criminología y Sistema Penal (Compilación in memoriam)*, BdeF, Buenos Aires, 2004.

BARKANE, I., «Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance», *Information Polity*, 27, 2, 2022, pp. 147-162.

BAROCAS, S./HARDT, M./NARAYANAN, A., *Fairness and machine learning. Limitations and opportunities*, MIT Press, Cambridge, 2023.

BAROCAS, S./SELBST, A., «Big Data 's Disparate Impact», *Cal. Law Review*, 104, 2016, pp. 671-732.

BARONA VILAR, S., «De la sociedad líquida a la justicia penal líquida. Una lectura desde la obra de Zygmunt Bauman», *Ius Puniendi. Política Criminal y Criminología*, 2017, pp. 371-397.

— «Justicia Penal líquida (desde la marida de Bauman)», *Teoría y Derecho*, 22, 2017, pp. 64-91.

— «Justicia penal desde la globalización y la postmoderidad hasta la neomodernidad», *Revista boliviana de Derecho*, 27, 2019, pp. 20-53.

BARRACHINA NAVARRO, G./BOIX PALOP, A., «La aplicabilidad del Reglamento de inteligencia artificial al ámbito de la Administración pública y servicios públicos y especialidades respecto de su cumplimiento: Especial atención a Anexo III y actuación administrativa y particularidades cumplimiento», en COTINO HUESO,

L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 351-371.

Barresi, O., «Italian report on traditional criminal law categories and AI», *RIDP*, vol. 94, 1, pp. 269-302.

Bartholomew, R., «'Navegar é preciso; viver não é preciso': The Impossible Journeys of Kierkegaard and Pessoa», en Falcato, A.,/ Cardiello, A. (Eds.), *Philosophy in the Condition of Modernism*, Palgrave Macmillan, Cham, 2018, pp. 385-414.

Bathaee, Y., «The artificial intelligence black box and the failure of intent and causation», *Harvard Journal of Law & Technology*, 31(2), 2018, pp. 889-938.

Bauman, Z., *Miedo Líquido: La sociedad contemporánea y sus temores*, Paidós, Barcelona, 2007.

Beck, J./Burri, T., «From 'Human Control' in International Law to 'Human Oversight» in the New EU Act on Artificial Intelligence», en Mecacci, J. *et al.* (Coords.), *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*, Elgar, 2024, pp. 104-130.

Beck, S., «Roboter und Cyborgs – erobern sie unsere Welt», en Beck, S. (Coord.), *Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Ed. Nomos, Baden-Baden, 2012, pp. 9-22.

— «Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood», *Robotics and Autonomous Systems*, 86, 2016, pp. 138-143.

— «Robotics and Criminal Law. Negligence, Diffusion of Liability and Electronic Personhood», en Hilgendorf, E./Feldle, J. (eds.), *Digitization and the Law*, Nomos, Baden Baden, 2018, pp. 41-56.

Beck, S./Gerndt, S., «Google-Cars, Software-Agents, Autonome Waffensysteme - neue Herausforderungen für das Strafrecht?», en Beck, S./Meier, B.-D./Momsen, C. (Eds.), *Cybercrime und Cyberinvestigation*, Nomos, Baden-Baden, 2015, 14, pp. 9-33.

— «German report on traditional criminal law categories and AI», *RIDP*, 195, 2023, pp. 202-204.

Beck, U., *Risikogesellschaft. Aufdem Weg in eine andere Moderne*, Suhrkamp Verlag, Frankfurt am Maine, 1986.

— *Gegengifte: Die organisierte Unverantwortlichkeit*, Suhrkamp Verlag, Frankfurt am Maine, 1988.

— *Risk Society: Towards a New Modernity*, Sage, Londres, 1992.

— «De la sociedad industrial a la sociedad del riesgo» (trad. Del Río Herrmann), *Revista de Occidente*, 150/1993, pp. 19-40.

Beck, U./Giddens, A./Lash, S., *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*, Stanford University Press, Stanford, 1995.

Beckers, A./Teubner, G., *Three Liability Regimes for Artificial Intelligence. Algorithmic Actants, Hybrids, Crowds*, Hart, Oxford; New York, 2021.

BELLOSO MARTÍN, N., «La problemática de los sesgos algorítmicos (con especial referencia a los de género)», en LLANO ALONSO, F. H., (Dir.), *Inteligencia artificial y Filosofía del derecho*, Laborum, Murcia, 2022, pp. 45-78.

BELTRÁN DE HEREDIA RUIZ, I., «Algoritmos y condicionamiento por debajo del nivel consciente: un análisis crítico de la propuesta de Ley de Inteligencia Artificial de la Unión Europea», *Revista de la Facultad de Derecho de México*, Tomo LXXIII, 286, 2023, pp. 621-650.

BENÍTEZ ORTUZAR, I. F., «Responsabilidad por el hecho cometido por el robot dotado de inteligencia artificial. ¿Existe un paralelismo con la responsabilidad de la persona jurídica? Breves reflexiones», en AA. VV, *Estudios político criminales, jurídicos penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, Tirant Lo Blanch, Valencia, 2023, 2.ª ed., pp. 503-515.

BERBERIAN, B./SARRAZIN, J-C./LE BLAYE, P./HAGGARD, P., «Automation Technology and Sense of Control: A Window on Human Agency», *PLoS ONE*, 7, 2012.

BERISTAIN PIÑA, A., «Resultado y delitos de peligro», *Revista de la Facultad de Derecho de la Universidad de Madrid*, XIII, 1969, pp. 457-465.

BERLANGA SILVENTE, V., RUBIO HURTADO, M. J., /VILÀ BAÑOS, R., «Cómo aplicar árboles de decisión en SPSS», *REIRE Revista d'Innovació i Recerca en Educació*, 6(2), 1, 2013, pp. 65-66.

BERTOLESI, R., *Intelligenza artificiale e responsabilità penale per danno da prodotto*, 2020, tesis doctoral en abierto, disponible en: <https://air.unimi.it/handle/2434/711618?mode=complete > [Accedido por última vez el 02/02/2026].

BERTOLINI, A., «Robots as products: the case for a realistic analysis of robotic applications and liability rules», *Law, innovation and technology*, 5(2), 2013, pp. 214-247.

BIALAKOWSKY, A., «Alienación, aceleración, resonancia y buena vida. Entrevista a Hartmut Rosa», *Revista Colombiana de Sociología*, 41/2, 2018, pp. 249-259.

BINDING, K., *Die Normen und ihre Übertretung, Eine Untersuchung über die rechtmäßige Handlung und die Arten des Delikts*, vol. 4: *Die Fahrlässigkeit*, Scientia, Aalen, 1965.

BINNS, R., «Fairness in Machine Learning: Lessons from Political Philosophy», *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 2018, pp. 149-159.

BIRD, K. R., «Natural and Probable Consequences Doctrine: Your Acts Are My Acts», *W. St. UL Rev.*, 43, 2006, pp. 43-66.

BŁACHNIO-PARZYCH, A., «Criminal Law Protection Against Artificial Intelligence Threats in Finance» (ponencia presentada en el Congreso Internacional Artifical Intelligence and Criminal Law, Bérgamo, 25 de octubre de 2024).

BLANCO CORDERO, I., «Un derecho penal de humanos para entidades artificiales inteligentes. ¿Machina delinquere et puniri potest?», en DOVAL PAIS, A./GUTIÉRREZ PÉREZ, E. (Dirs.), *Manifestaciones de desigualdad en el sistema de justicia penal. Volumen I*, Aranzadi, Cicur Menor, 2022, pp. 371-391.

BLAY, E., «A vueltas con la multa. De la relativa invisibilidad de la pena de multa y de las razones para que se investigue sobre ella», *InDret*, 4/2022, Editorial.

BOLDOVA PASAMAR, M. A., «Penas privativas de derechos», en BOLDOVA PASAMAR, M. A./ALASTUEY DOBÓN, C. (Coords.), *Tratado de las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023.

BONIS SANZ, A., «Implicaciones jurídicas del uso de técnicas de Big Data y Sistemas de Inteligencia Artificial por las Administraciones Tributarias», *Revista de la Asociación de Abogados del Estado*, 57, mayo 2022, pp. 2-19.

BORGHETTI, J-S., «Civil Liability for Artificial Intelligence: What Should its Basis Be?», *La Revue des Juristes de Sciences Po*, 17/2019, pp. 94-102.

BORUM, R./LODEWIJKS, H./BARTEL, P. A./FORTH, A. E., «Structured Assessment of Violence Risk in Youth (SAVRY)», en Otto, R. K./Douglas, K. S. (Eds.), *Handbook of violence risk assessment*, Routledge/Taylor & Francis Group, 2010, pp. 63-79.

BOSTROM, N., «A history of transhumanist thought», *Journal of evolution and technology*, 14(1), 2005, pp. 1-25.

BOSTROM, N., *Die Zukunft der Menschheit: Aufsätze (suhrkamp taschenbuch wissenschaft)*, trad. Jan-Erik Strasser, Suhrkamp Verlag, Frankfurt a. M., 2018.

BOULANIN, V./VERBRUGGEN, M., *Mapping the Development of Autonomy in Weapon Systems*, Stockholm International Peace Research Institute, SIPRI, Stockholm, 2017.

BRADSHAW, J./HOFFMAN, R./JOHNSON, M./WOODS, D., «The seven deadly myths of "autonomous systems"», *IEEE Intelligent Systems*, 2013, pp. 54-61.

BRANDARIZ, J. A., *El modelo gerencial-actuarial de penalidad: Eficiencia, riesgo y Sistema penal*, Dykinson, Madrid, 2016.

BRAVO BOLADO, A., «El guardián de Gotham. De software predictivos y derechos fundamentales: ¿el fin de la inteligencia artificial al servicio del crimen? (Comentario a la Sentencia del Tribunal Constitucional Alemán de 16 de febrero de 2023)», *La ley penal: revista de derecho penal, procesal y penitenciario*, 165(1), 2023, pp. 1-11.

— «Igualdad e inteligencia artificial: los modernos ataques a la igualdad en el sistema de justicia penal», en DOVAL PAIS, A./GUTIÉRREZ P., E. (Dirs.), *Manifestaciones de desigualdad en el sistema de justicia penal*, Aranzadi, Navarra, 2023, vol. 2, *(online).*

— *Justicia algorítmica. Un enfoque sociotécnico*, Estudios Penales y Criminológicos, 44, 2023, pp. 1-43.

— «La protagonista olvidada del sistema penal», Blog de la Facultad de Derecho, 24 de mayo de 2023.

— «Predicción y peligrosidad. ¿El retorno de un nuevo paradigma?», en LASCURAÍN SÁNCHEZ, J. A./PEÑARANDA RAMOS, E. (Coords.), *Liber Amicorum en homenaje a Julio Díaz-Maroto y Villarejo*, UAM Ediciones, Madrid, 2023, pp. 87-103.

Bravo Bolado, A.,«De las cosas a las no-cosas. Una transformación de los sujetos, los paradigmas de responsabilidad y las sanciones en la era del crimen tecnológico. Desdibujando las fronteras de lo conocido en tres movimientos», en Rodríguez Horcajo, D./Basso, G. J. (Dirs.), *Repensando las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2025, pp. 161-184.

— «Negociando en la sombra: Inteligencia artificial, alta frecuencia y desafíos para el derecho penal económico», en Feijoo Sánchez, B/Peñaranda Ramos, E. (Dirs.), *Teoría y práctica de los Delitos Económicos. Hacia una racionalización de la protección penal del orden socioeconómico*, BdeF, Buenos Aires, 2025, pp. 243-288.

— «El sesgo discriminatorio y el derecho a la igualdad en los sistemas de IA», Gregoraci Fernández, B.,/Castillo Abela, J./Velasco Caballero, F., (Dirs.), *Derecho de la neurotecnología, Libro Blanco para el desarrollo de la Carta de Derechos Digitales en el ámbito de la neurotecnología,* Marcial Pons, 2026, pp. 847-866..

Bravo Bolado, A./Ensel, L./Scollo, L./Vicario Pérez, A. M., «Sketching the Shadows of Criminal Law and Justice in the AI New Era», *European Criminal Law Review*, (en prensa).

Brehm, W., *Zur Dogmatik des abstrakten Gefährdungsdlikts*, Mohr, Tubinga, 1973.

Bricola, F., «Responsabilità penale per il tipo e il modo di produzione», en *La Responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Giuffrè, Milán, 1978, pp. 75-90.

Brooks, D., «The Philosophy of Data», *The New York Times*, 4 de febrero de 2013.

Burchard, C., «Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft», *Normative Orders Working Paper*, 02/2019, pp. 3-31.

— «Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft», Joerden, J. C./Schuhr, J. C. (Coords.), *Jahrbuch für Recht und Ethik. Annual Review of Law and Ethics*, 27, Duncker & Humblot, Berlin, 2019, pp. 527-555.

Busato, C., «De máquinas y seres vivos: ¿Quién actúa en los resultados delictivos derivados de decisiones cibernéticas?», en Demetrio Crespo, E. (Dir.), *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant Lo Blanch, Valencia, 2022, pp. 349-356.

Bustos Ramírez, J., «Los bienes jurídicos colectivos (Repercusiones de la labor legislativa de Jiménez de Asúa en el Código penal de 1932)», *Revista de la Facultad de Derecho de la Universidad Complutense*, 11, 1986.

— «Los bienes jurídicos colectivos», *Revista de Derecho penal*, 27, 2019, pp. 472-473.

Bustos Rubio, M., «Insignificancia y Derecho penal», InDret, 4/2023, pp. 110-170.

Caldwell, M./Andrews, J. T. A./Tanay, T./Griffin, L. D., «AI-enabled future crime», *Crime Science*, 9, 2020, pp. 1-13.

CALO, R., «Robotics and the Lessons of Cyberlaw», *California Law Review*, 103(3), 2015, pp. 513-563.

CAMPUZANO GÓMEZ-ACEBO, J./SIEIRA GIL, J., en BARRIO ANDRÉS, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 283-292.

CANCIO MELIÁ, M., «Caso Contergan», en SÁNCHEZ-OSTIZ GUTIÉRREZ, P. (Coord.), *Casos que hicieron doctrina en derecho penal*, La Ley, Madrid, 2011, pp. 207-222.

CANTARD, N., «Acerca de la función del disvalor de resultado en derecho penal», *Revista del centro de investigaciones de la Facultad de ciencias jurídicas y sociales de la Universidad Nacional del Litoral*, 16, 2015, pp. 9-30.

CAPPELLINI, A., «Reati colposi e tecnologie dell'intelligenza artificiale», en BALBI, G./DE SIMONE, F./ESPOSITO, A./MANACORDA, S. (Coords.), *Diritto penale e inteligenza artificiale. «Nuovi scenari»*, Giappichelli Editore, Turin, 2022, pp. 19-32.

CARBONELL MATEU, J. C., «Lección XXXII, Delitos contra la seguridad colectiva (I): riesgos catastróficos», en GONZÁLEZ CUSSAC, J. L. (Coord.), *Derecho penal parte especial*, 8.ª ed., Tirant Lo Blanch, Valencia, 2023.

CARDENAL MONTRAVETA, S., «Delitos de riesgo catastrófico: energía nuclear, estragos y otros delitos de riesgo (arts. 341-350)», en CORCOY BIDASOLO, M., (Dir.), *Manual de Derecho Penal, Parte especial. Tomo 1*, 3.ª ed., Tirant Lo Blanch, Valencia, 2023.

CARRASCO PERERA, Á., «Análisis de la nueva directiva de responsabilidad por daños causados por defectos de productos. Síntesis y crítica de la regulación más contenida en la Directiva 2024/2853», *Centro de Estudios de Consumo*, 2024, pp. 1-7.

CARROL, M./CHAN, A./ASHTON, H./KRUEGER, D., «Characterizing Manipulation from AI Systems», Proceedings of Under review, ACM, NewYork, NY, USA, 2023.

CARVER, R., *De qué hablamos cuando hablamos de amor*, Anagrama, Barcelona, 2006.

CASADO GARCÍA, M., «La regulación de las neurotecnologías: las "sandboxes" para la innovación», *RJUAM*, 51, 2025, pp. 95-120.

CASTILLO PARRILLA, J. A., «Algunos retos éticos, y sobre todo jurídicos, de la inteligencia artificial», *Anales de la Cátedra Francisco Suárez*, 57, 2023, p. 315-321.

— «Inteligencia artificial de uso general, modelos fundacionales (y «Chat GPT») en el Reglamento de inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, p. 757-777.

CEREZO MIR, J., «Los delitos de peligro abstracto en el ámbito del derecho penal del riesgo», *Revista de Derecho Penal y Criminología*, 10, 2002, pp. 47-72.

CERVANTES, M. DE, *Don Quijote de la Mancha*, dirigida por Francisco Rico, Edición del Instituto Cervantes, 1998.

CHAGAL-FEFERKORN, K. A., «Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers», *Stanford Law & Policy Review*, 30(1)/2019, pp. 62-114.

CHALMERS, J./LEVERICK, F., «Fair labelling in criminal law», *The Modern Law Review*, 71(2), 2008, pp. 217-246.

CHAVELI DONET, E., «La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 495-533.

CHERÑAVSKY, N. A./RIQUERT, M. A./SUEIRO, C. C., «Argentinian National report on criminalisation of AI-related offences», *RIDP*, mayo de 2024.

CHESTERMAN, S., *We, the robots?*, Cambridge University Press, Cambridge, 2021.

CHOPRA, S./WHITE, L. F., *A legal theory for Autonomous Artificial Agents*, University of Michigan Press, Michigan, 2011.

CHRISTMAN, J., «Autonomy in Moral and Political Philosophy», en ZALTA, E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition).

CLARK, A./CHALMERS, D., «The extended mind», *Analysis*, 58/1998, pp. 7-19.

CLARKE, S., «Future technologies, dystopic futures and the precautionary principle», *Ethics and Information Technology*, 7(3), 2005, pp. 121-126.

CLARKSON, C., «Aggravated endangerment offences», *Current Legal Problems*, *60* (1), 2007, pp. 278-295.

COCA VILA, I., «Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal», *Cuadernos de Política Criminal*, 122, 2017, pp. 235-275.

— «El consentimiento en el derecho penal económico. Un estudio a propósito de los delitos de administración desleal (art. 252 CP) y corrupción en los negocios (art. 286 bis CP)», Revista General de Derecho penal, 28, 2017, pp. 1-39.

— «Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law», *Criminal Law and Philosophy*, 12, 2018, pp. 59-82.

— «La pena de multa en serio. Reflexiones sobre su dimensión y aseguramiento aflictivos a través del delito de quebrantamiento de condena (art. 468 CP)», *InDret*, 3/2021, pp. 69-99.

COECKELBERGH, M., «Artificial intelligence, responsibility attribution, and a relational justification of explainability», *Science and engineering ethics*, 26(4), 2020, pp. 2051-2068.

— *Green Leviathan or the poetics of political liberty: navigating freedom in the age of climate change and artificial intelligence*, Routledge, Nueva York, 2021.

— *La filosofía política de la inteligencia artificial,* Ediciones Cátedra, Madrid, 2023.

COFFEE, J., C., JR., «No Soul to Damn: No Body to Kick»: An Unscandalised Inquiry into the Problem of Corporate Punishment», 79 *MICH. L. REV.*, 386, 1981, pp. 386-459.

CONSULICH, F., «Il rastro di möbius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato», *Banca Borsa Titoli di credito*, 2, 2018, pp. 195-234.

— «Flash offenders. Le prospettive di accountability penale nel contrasto alle Intelligenze Artificiali devianti», *Rivista italiana di diritto e procedura penale*, 2022, p. 1015-1055.

CONTRERAS CHAIMOVIC, L., «La responsabilidad penal del fabricante por la infracción de sus deberes de vigilancia, advertencia y retirada», *Polít. crim*., *10*(19), 2015, pp. 266-296.

CORCOY BIDASOLO, M., «Responsabilidad penal derivada del producto: En particular la regulación legal en el código penal español: delitos de peligro», en MIR PUIG, S./LUZÓN PEÑA, D. M. (Coords.), *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, J. M. Bosch, Barcelona, 1996, pp. 247-262.

— *Delitos de peligro y protección de bienes jurídico-penales supraindividuales: nuevas formas de delincuencia y reinterpretación de tipos penales clásicos*, Tirant Lo Blanch, Valencia, 1999.

— «La lección de la experiencia española en materia de responsabilidad por el producto del caso de la colza a nuestros días», en FOFFANI, L. (Coord.), *La sicurezza agroalimentare nella prospettiva europea. Precauzione, prevenzione, repressione*, Ed. Giufre, Milán, 2014, pp. 433-456.

— «Delitos contra la salud de los consumidores», en CORCOY BIDASOLO, M./ GÓMEZ MARTÍN, V. (Coords.), *Manual de derecho penal, económico y de empresa, parte general y parte especial: (adaptado a las LLOO 1/2015 y 2/2015 de Reforma del Código Penal): doctrina y jurisprudencia con casos solucionados*, tomo 1, Tirant lo Blanch, Valencia, 2016, pp. 565-594.

— «Principio de culpabilidad ¿responsabilidad objetiva en determinados delitos?», en AA. VV, *Estudios político criminales, jurídicos penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, Tirant Lo Blanch, Valencia, 2023, 2.ª ed., 2023, pp. 569-590.

— *Delitos de peligro y protección de bienes jurídico-penales supraindividuales: nuevas formas de delincuencia y reinterpretación de tipos penales clásicos*, BdeF, Madrid, 2024.

CORN, E., «El tímido presente y el futuro incierto del principio de precaución en la parte especial del derecho penal italiano: organismos genéticamente modificados y otras aplicaciones», en GÓMEZ TOMILLO, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 277-230.

CORTINA, A., «Fundamentos filosóficos del principio de precaución», en Romeo Casabona, C. M. (Ed.), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, pp. 3-16.

COTINO HUESO, L., «Nuevo paradigma en la garantía de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial», en COTINO HUESO, L. (Dir.), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, pp. 69-105.

— «Transparencia y explicabilidad de la inteligencia artificial y «compañía» (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad…). Para qué, para quién y cuánta», en COTINO HUESO, L./CASTELLANOS CLARAMUNT, J. (Coords.), *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo Blanch, Valencia, 2022, pp. 29-79.

— «Los requisitos del Tribunal Constitucional alemán para el análisis automatizado de datos o con inteligencia artificial, que no se cumplen en España», blog de la Revista Catalana de dret público, 12 de abril de 2023.

— «Alcance y delimitación de los sistemas de alto riesgo en el Reglamento de inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 291-313.

— «Sandbox, espacios controlados y pruebas en condiciones reales de sistemas de inteligencia artificial en el Reglamento», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 817-844.

COVARRUBIAS, S., /NOYDENS, R./ RIQUER, M. DE, *Tesoro de la lengua castellana o española: según la impresión de 1611*, *con las adiciones de Benito Remigio Noydens publicadas en la de 1674*, Horta, Barcelona, 1943.

CUESTA PASTOR, P., *Delitos obstáculo. Tensión entre política criminal y teoría del bien jurídico*, Comares, Granada, 2002.

CUMMINGS, M. L., «Automation and accountability in decision support system interface design», *J. Technol. Stud*, 32, 2006, pp. 23-31.

DANAHER, J., «Robots, Law and Retribution Gap», *Ethics and Information Technology*, 4, 2016, pp. 299-309.

— *Automation and Utopia: Human Flourishing in a World without Work*, Harvard University Press, 2019.

DE ASÍS PULIDO, M., «La justicia predictiva: tres posibles usos en la práctica jurídica», en LLANO ALONSO, F. H., (Dir.), *Inteligencia artificial y filosofía del derecho*, Laborum Ediciones, Murcia, 2022, pp. 285-312.

DE LA CUESTA AGUADO, M. P., *Respuesta penal al peligro nuclear*, PPU, Barcelona, 1994.

— «Reflexiones a propósito de «La robótica ante el derecho penal: el vacío de respuestas jurídica(s) a las desviaciones incontroladas», de Gonzalo Quintero Olivares», *Revista Electrónica de Estudios Penales y de la Seguridad*, 2017, pp. 1-5.

— «Inteligencia artificial y responsabilidad penal», *Revista Penal México*, 9(16-17), 2020, pp. 51-62.

De la Cuesta Aguado, M. P.,«El derecho penal frente al Reglamento de Inteligencia Artificial», *Ius Criminale*, 1, octubre de 2024, pp. 83-107.

De la Fuente Cardona, F., *La teoría del bien jurídico y su función de crítica legislativa. Una contribución a la discusión sobre la evaluación constitucional de las normas penales,* Marcial Pons, 2025.

De La Mata Barranco, N./Pérez González, S., «Inteligencia artificial autónoma y responsabilidad penal de las personas jurídicas», blog Almacén de Derecho, 12 de septiembre de 2013.

De Siles, E. L. «AI, on the Law of the Elephant: Toward Understanding Artificial Intelligence», *Buffalo Law Review*, 69(5), 2021, pp. 1389-1470.

Del Rosal Blasco, B., «¿Hacia el Derecho penal de la Postmodernidad?», *RECPC*, 11-08, 2009, pp. 1-64.

— «¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la inteligencia artificial?», *Revista electrónica de responsabilidad penal de personas jurídicas y compliance*, 2, 2023, pp. 1-49.

Demetrio Crespo, E., «Del «Derecho penal liberal» al «Derecho penal del enemigo»«, *RDPC*, 14/2004, pp. 87-115.

— «Libertad de voluntad, investigación sobre el cerebro y responsabilidad penal», *InDret*, 2/2011, pp. 1-38.

— «Fundamento de la responsabilidad en comisión por omisión de los directivos de las empresas», *Revista Penal México*, núm., 3, 2012.

— *Responsabilidad penal por omisión del empresario,* Centro Mexicano de Estudios en lo Penal Tributario, 2017.

Descartes, R., *Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences*, Jean Marie, Leiden, 1637.

Diamantis, M. E., «The extended corporate mind: When corporations use AI to break the law», *NCL Rev.*, 98/2019, p. 893-932.

— «Employed algorithms: labor model of corporate liability for ai», *Duke Law Journal*, 72(4), 2023, pp. 797-859.

Díaz-Maroto y Villarejo, J., *El Derecho Penal ante los Fraudes Alimentarios. Responsabilidad por el Producto en la moderna sociedad del riesgo*, Civitas/ Thomson Reuters, Cizur Menor, 2010.

— Sociedad del riesgo y fraudes alimentarios», en AA. VV., *Un modelo integral de Derecho penal. Libro homenaje a la profesora Mirentxu Corcoy Bidasolo*, BOE, Madrid, 2022, pp. 1099-1109.

Díez Ripollés, J. L., «De la sociedad del riesgo a la seguridad ciudadana: Un debate desenfocado», *Nuevo Foro penal*, 69/2006, pp. 198-248.

Domínguez Álvarez, J. L., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 227-243.

Domínguez Peco, E. M., «Los robots en el derecho penal», en Barrio Andrés, M. (Dir.), *Derecho de los robots*, Wolters Kluwer, Madrid, 2018, pp. 131-150.

DOPICO GÓMEZ-ALLER, J., «¿Dogmática o probatoria? La cuestión de la causalidad y su prueba en las intoxicaciones masivas con agentes tóxicos desconocidos», *La ley Penal*, 169, 2024, pp. 1-24.

DOVAL PAIS, A., *Fraudes alimentarios. Evolución histórica y elementos esenciales del sistema de protección penal*, Valencia, 1995.

— *Delitos de fraude alimentario. Análisis de sus elementos esenciales*, Aranzadi, Pamplona, 1996.

DSOUZA, M., Don't panic: Artificial intelligence and Criminal Law, en BAKER, D. J./ ROBINSON, P. H. (Dirs.), *Artificial Intelligence and the Law. Cibercrime and Criminal Liability*, Routledge, Oxon, 2021, 2020, pp. 247-264.

— «Why Autonomous AI cannot be a Criminal Law agent», 29 de noviembre de 2024, Criminal Justice Theory Blog.

DUFF, R. A., «Crime, Prohibition, and Punishment,» *Journal of Applied Philosophy*, 19, 2002, pp. 97-102.

— «Criminalizing endangerment», *La. L. Rev.*, 65, 2004, pp. 941-965.

— «Who is responsible, for what and to whom?», *Ohio State Journal Of Criminal Law*, 2/2005, pp. 441-461.

— «Legal and moral responsibility», *Philosophy Compass*, 4/2009, pp. 978-986.

— «Strict Responsibility, Moral and Criminal», *Journal of Value Inquiry*, 43, 2009, pp. 295-313.

EDWARDS, L./VEALE, M., «Slave to the algorithm? Why a «right to an explanation» is probably not the remedy you are looking for», *Duke Law & Technology Review*, 18, 2017, pp. 18-84.

EKELHOF, M., «Moving beyond semantics on autonomous weapons: Meaningful human control in operation», *Global Policy*, 10(3), 2019, pp. 343-348.

ESCAJEDO SAN-EPIFANIO, L., «El reconocimiento biométrico en el Reglamento de inteligencia artificial: exenciones, prohibiciones y especialidades de alto riesgo», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 183-235.

ESCOTT, E., «What Are the 3 Types of AI? A Guide to Narrow, General and Super Artificial Intelligence», Codebots, 24 de octubre de 2017.

ESCRIVÁ GREGORI, J. M., *La puesta en peligro de bienes jurídicos en Derecho penal*, Bosch, Barcelona, 1976.

EUBANKS, V., *Automating Inequality: how high-tech tools profile police and punish the poor*, St. Martin's Press, Nueva York, 2018.

FARMER, L., *Making the modern criminal law: Criminalization and civil order*, Oxford University Press, 2016.

FARQUHAR, S. *et al.,* «Detecting hallucinations in large language models using semantic entropy», *Nature*, 630 (8017), 2024, pp. 625-630.

FATEH-MOGHADAM, B., «Innovationsverantwortung im Strafrecht: Zwischen strict liability, Fahrlässigkeit und erlaubtem Risiko – Zugleich ein Beitrag zur Digitalisierung des Strafrechts», *ZStW*, 131/2019, pp. 863-887.

FAVA, G., «L'ordinanza del Tribunale di Bologna in merito alla possibile discriminatorietà dell'algoritmo utilizzato da Deliveroo», *Lavoro Diritti Europa*, 1/2021, pp. 1-4.

FAVARETTO, M./DE CLERCQ, E./ELGER, B. S., «Big Data and discrimination: perils, promises and solutions. A systematic review», *Journal of Big Data*, 6, 1, 2019, pp. 1-27.

FEELEY, M./SIMON, J., «Actuarial Justice: The Emerging New Criminal Law», en NELKEN, D. (Ed.), *The Futures of Criminology*, Sage Publisher, London, 1994, pp. 173-201.

FEIJOO SÁNCHEZ, B., «Delitos contra la seguridad colectiva: ¿tentativas imprudentes?», *La Ley*, 22432/2001, 1997, pp. 1-9.

— «Seguridad colectiva y peligro abstracto. Sobre la normativización del peligro», en AA. VV., *Homenaje al profesor Dr. Gonzalo Rodríguez Mourullo*, Civitas, Madrid, 2005, pp. 307-342.

— «Sobre la «administrativización» del Derecho penal en la «sociedad de riesgo»: un apunte sobre la política criminal a principios del siglo XXI», en AA. VV., *Derecho y justicia penal en el siglo XXI: liber amicorum en homenaje al profesor Antonio González-Cuéllar García*, Colex, 2006, pp. 137-176.

— *Normativización del derecho penal y realidad social*, Univ. del Externado, Bogotá, 2007.

— «Imputación objetiva en el Derecho penal económico y empresarial. Esbozo de una teoría general de los delitos económicos», *InDret*, 2/2009, pp. 1-74.

— «Límites del tipo objetivo en los delitos económicos», en Gómez-Jara Díez, C., *Persuadir y Razonar: Estudios Jurídicos en Homenaje a José Manuel Maza Martín*, Thomson Reuters Aranzadi, 2018, tomo II, pp. 159-175.

— «Los delitos de manipulación del precio de cotización de un instrumento financiero del art. 284 CP tras la LO 1/2019: retos dogmáticos frente a un desatino legislativo», *La ley digital*, 8176/2019, pp. 1-38.

— «El Derecho penal de la culpabilidad ante el neurodeterminismo», en AA. V.., *Un modelo integral de Derecho penal: libro homenaje a la profesora Mirentxu Corcoy Bidasolo*, BOE, Madrid, 2022, pp. 601-613.

— «Neurociencias y Derecho Penal de un sistema de libertades», en DEMETRIO CRESPO, E., (Dir.), *Derecho penal y comportamiento humano: avances desde la neurociencia y la inteligencia artificial*, Tirant Lo Blanch, Valencia, 2022, pp. 127-146.

— «La función de la responsabilidad penal de las personas jurídicas en el derecho penal español», *REDEPEC: Revista Electrónica de Responsabilidad Penal de Personas Jurídicas y Compliance*, 1/1, 2023, pp. 1-121.

FEINBERG, J., «Action and Responsibility», en FEINBERG, J., *Doing and deserving. Essays in the theory of responsibility*, Princeton University Press, Princeton, Nueva Jersey, 1970.

— *Harm to others*, Vol. I, Oxford University Press, 1984.

FERNÁNDEZ HERNÁNDEZ, C., en BARRIO ANDRÉS, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 180-204.

FERNÁNDEZ HERNÁNDEZ, C./EGUILUZ CASTAÑEIRA, J. A., «Análisis del Anteproyecto de Ley para el buen uso y la gobernanza de la Inteligencia Artificial», *Derecho Digital e Innovación*, 23/2025.

FERNÁNDEZ TERUELO, J. G., «La inhabilitación profesional del delincuente económico-empresarial», en CARBONELL MATEU, J. C./ROSAL BLASCO, B. DEL/MORILLAS CUEVA, L./ORTS BERENGUER, E./QUINTANAR RUIZ, M. (Coords.), *Estudios en homenaje al profesor Cobo del Rosal*, Dykinson, Madrid, 2005, pp. 351-362.

FERRAJOLI, L., *Derecho y razón: teoría del garantismo penal*, Trotta, Madrid, 1995.

FIGUEROA RUBIO, S., *Adscripción y reacción: Responsabilidad jurídica y moral desde una perspectiva interpersonal*, Marcial Pons, Ediciones Jurídicas y Sociales, Madrid, 2019.

FLETCHER, G. G. S., «Deterring Algorithmic Manipulation», *Vanderbilt Law Review*, 74/2, 2021, pp. 259-325.

FLORES ANARTE, L., «Sesgos de género en la Inteligencia Artificial: El Estado de derecho frente a la discriminación algorítmica por razón de sexo», *Revista Internacional De Pensamiento Político*, 18/2023, pp. 95-120.

FLORIDI, L., «Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions», *Royal Society's Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, 374, 2016, pp. 1-22.

FLORIDI, L./KING, T. C./AGGARWAL, N./TADDEO, M., «Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions», *Science and Engineering Ethics*, 26, 2020, pp. 89-120.

FLORIDI, L./SANDERS, J. W., «On the morality of artificial agents», *Minds and machines*, 14, 2004, pp. 349-379.

FORNASARI, G., «El principio de precaución en la experiencia legislativa, jurisprudencial y doctrinal italianas. Aspectos de parte general», en GÓMEZ TOMILLO, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 149-184.

FRAGASSO, B., «La responsabilità penale del produttore di sistemi di intelligenza artificiale», *Diritto Penale Contemporaneo*, 1/2023, pp. 26-45.

FREITAS, P. M./ANDRADE, F./NOVAIS, P., «Criminal liability of autonomous agents: From the unthinkable to the plausible», en CASANOVAS, P./PAGALLO, U./PALMIRANI, M./SARTOR, G. (Eds.), *AI approaches to the complexity of legal systems. AICOL 2013. Lecture notes in computer science*, Vol. 8929, Springer, Berlin, 2014, pp. 145-156.

FRIEDMAN, B./NISSEMBAUM, H., «Bias in computer systems», *ACM Transactions on Information Systems*, 14/3, 1996, pp. 330-347.

FRONZA, E., «Code is Law'. Note a margine del volume di Antoine Garapon e Jean Lassègue, justice digitale. Révolution graphique et rupture anthropologique, Puf, Paris, 2018», *Diritto penale contemporaneo*, dic. 2018, pp. 1-5. «Justice digitale: révolution graphique et rupture anthropologique» De Antoine Garapon y Jean Lassègue», *En letra: Derecho penal*, 9/2019, pp. 180-191.

GALÁN MUÑOZ, A., *Libertad de expresión y responsabilidad penal por contenidos ajenos en internet. Un estudio sobre la incidencia penal de la ley 34/2002 de Servicios de la Sociedad de la Información y el Comercio Electrónico*, Tirant lo Blanch, Valencia, 2010.

— «La problemática utilización del principio de precaución como referente de la política criminal del moderno derecho penal: ¿hacia un derecho penal del miedo a lo desconocido o hacia uno realmente preventivo?», *Revista General de Derecho penal*, 23, 2015, pp. 1-55.

GALLAS, W., «Abstrakte und konkrete Gefährdung», en AA. VV., *Festschrift für Ernst Heinitz zum 70. Geburtstag*, Walter de Gruyter, Berlín, Nueva York, 1972, pp. 171-184.

GALLEGO ARRIBAS, D., «La prohibición de acceso a las redes sociales como consecuencia jurídica del delito. A propósito de la STS 547/2022, de 2 junio», *Diario La ley*, 10332, 2023, pp. 1-11.

GAMERO CASADO, E., en BARRIO ANDRÉS, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 835-837.

GARAPON, A., *La despazializzacione della giustizia*, Mimesis, Milán, 2021.

GARAPON, A./LASSÈGUE, J., *Justice Digitale: révolution graphique et rupture anthropologique*, Presses Universitaires de France, Paris, 2018.

GARCÍA LORCA, F., «Nueva York, oficina y denuncia», *Un poeta en Nueva York*, 1940.

GARLAND, D., *The culture of control. Crime and Social Order in Contemporary Society*, University of Chicago Press, 2001.

GARROCHO SALCEDO, A., «Delitos de denegación de una prestación por motivos discriminatorios de los artículos 511 y 512 CP», *Eunomía. Revista en Cultura de la Legalidad*, 5, 2013, pp. 240-247.

GIANNINI, A., *Criminal behavior and accountability of artificial intelligence systems*, [Doctoral Thesis, Maastricht University, University of Florence], Eleven Publishers, 2023.

GIANNINI, A./KWIK, J., «Negligence failures and negligence fixes. A comparative analysis of criminal regulation of AI and autonomous vehicles», *Criminal law fórum*, 34/1, 2023, pp. 43-85.

GIL GIL, A., «Sistemas de armas autónomos letales. Comentario a la posición española, con especial atención a algunos problemas jurídico-penales», *Revista General de Derecho penal*, 38, 2022, pp. 1-50.

GIMBERNAT ORDEIG, E., *Delitos cualificados por el resultado y causalidad*, Reus, Madrid, 1966.
— «La omisión impropia de la dogmática penal alemana: Una exposición», *ADPCP*, L, 1997, pp. 5-26.
— *Ensayos Penales*, Tecnos, Madrid, 1999.
— «¿Las exigencias dogmáticas fundamentales hasta ahora vigentes de una parte general son idóneas para satisfacer la actual situación de la criminalidad, de la medición de la pena y del sistema de sanciones?», *ADPCP*, LII, 1999, pp. 51-72.
— «Omisión propia e incremento del riesgo en el Derecho penal de empresa», *ADPCP,* LIV, 2001, pp. 5-112.
— «Fin de protección de la norma e imputación objetiva», *ADPCP*, 2008.
GLESS, S., «Von der Verantwortung einer E-Person. Würdigung eines unveröffentlichten Regelungsentwurfs von Professor Dr. Dr. h.c. Kurt Seelmann anlässlich seines 70. Geburtstags», *Goltdammer's Archiv für Strafrecht (GA)*, 6, 164. Jahrgang, 2017, pp. 324-329.
GLESS, S./SILVERMAN, E./WEIGEND, T., «If robots cause harm, who is to blame? Self-driving cars and criminal liability», *New Criminal Law Review*, 19(3), 2016, p. 412-436.
GLESS, S./WEIGEND, T., «Intelligente Agenten und das Strafrecht (Intelligent Agents and Criminal Law)», *ZSTW*, 126/3, 2014, pp. 561-591.
GLICKMAN, M./SHAROT, T., «How human–AI feedback loops alter human perceptual, emotional and social judgements», *Nature Human Behavior*, 9/2024, pp. 345-359.
GOBERT, J./PUNCH, M., *Rethinking corporate crime*, Cambridge University Press, 2003.
GOLLA, S. J., «Lernfähige Systeme, lernfähiges Polizeirecht: Regulierung von künstlicher Intelligenz am Beispiel von Videoüberwachung und Datenabgleich», *Kriminologisches Journal*, 52(2), 2020, pp. 149-161.
GÓMEZ BENÍTEZ, J. M., *Causalidad, imputación y cualificación por el resultado*, Ministerio de Justicia Centro de publicaciones, Madrid, 1988.
GÓMEZ COLOMER, J. L., *El juez-robot (La independencia judicial en peligro)*, Tirant Lo Blanch, Valencia, 2023.
GÓMEZ DE ÁGREDA, Á., «La exclusión de los sistemas inteligencia artificial de seguridad nacional, defensa y militares del Reglamento y el Derecho aplicable», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 137-148.
GÓMEZ LIGÜERRE, C., «La Propuesta de Directiva sobre responsabilidad por daños causados por productos defectuosos», *InDret*, 4/2022, editorial.
— «Responsabilidad por daños causados por la inteligencia artificial», *InDret*, 1/2025.

Gómez Pereira, *Antoniana Margarita,* Servicio de Publicaciones Universidade de Santiago de Compostela, 2000.

Gómez Tomillo, M., «El principio de precaución en el Derecho penal. Contribución político-criminal a los delitos de riesgo normativo», en Gómez Tomillo, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 81-122.

Gómez-Jara Díez, C., «La responsabilidad penal de las empresas en los EE. UU», Ed. Universitaria Ramón Areces, Madrid, 2006.

González de la Garza, L. M., «El contenido de las llamadas «técnicas subliminales» y las vulnerabilidades de grupo específico de personas en el Reglamento de inteligencia artificial», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 251-274.

González Granado, J., *De la persona a la personalidad algorítmica*, Publicacions i Edicions de la Universitat de Barcelona, 2020.

González Rus, J. J., *Bien jurídico y Constitución (Bases para una teoría)*, Fundación Juan March, Madrid, 1983.

— «Recensión del libro de Javier Valls Prieto, Inteligencia artificial, Derechos Humanos y bienes jurídicos», *Revista electrónica de ciencia penal y criminología*, 24, 2022, pp. 1-10.

González-Meneses García-Valdecasas, M., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 205-222.

Gracia Martín, L., *Prolegómenos para la lucha por la modernización del Derecho penal Y para la crítica del discurso de resistencia. A la vez una hipótesis de trabajo sobre el concepto de Derecho penal moderno en el materialismo histórico del orden del discurso de criminalidad*, Tirant lo Blanch, Valencia, 2003.

— «Contribución al esclarecimiento de los fundamentos de legitimidad de la protección penal de bienes jurídicos por el Estado social y democrático de Derecho», en Demetrio Crespo, E. (Dir.), Crisis financiera y derecho penal económico, Edisofer, Madrid, 2014, pp. 339-380.

— «Concepto categorial teleológico y sistema dogmático del moderno Derecho penal económico y empresarial de los poderosos», *InDret*, 3/2016, pp. 1-130.

— «Crítica de las modernas construcciones de una mal llamada responsabilidad penal de la persona jurídica», *Revista Electrónica de Ciencia Penal y Criminología* (en línea). 2016, 18-05, pp. 1-95.

Grandi, C., «Positive Obligations (Garantestellung) Grounding Criminal Responsibility for not having Avoided an Illegal Result Connected to the AI Functioning», 2023, 1 *RIDP*, 59, pp. 67-76.

Greco, L., «Gibt es Kriterien zur Postulierung eines kollektiven Rechtsguts?», en Heinrich, M. (Coord.), *Strafrecht als Scientia Universalis: Festschrift für Claus Roxin zum 80. Geburtstag*, Walter Mouton de Gruyter, vol. 1, 2011, pp. 199-214

GRECO, L.,«Vehículos de motor autónomos y situaciones de colisión», en BASSO, G. / CANCIO MELIÁ, M., /MARAVER GÓMEZ, M. /FAKHOURI, C., (Coords.), *Libro Homenaje al prof. Dr. Agustín Jorge Barreiro*, Ed. Universidad Autónoma de Madrid, 2019.

— *Poder de julgar sem responsabilidade de julgador*, Marcial Pons, São Paulo, 2020.

GREEN, B., «The flaws of policies requiring human oversight of government algorithms», *Computer Law & Security Review*, 45, 2022, pp. 1-22.

GREEN, B./CHEN, Y., «Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments», *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019.

GREEN, S. P., «Six senses of strict liability: a plea for formalism», en SIMESTER, A. (ed.), *Appraising Strict Liability*, Oxford University Press, 2005.

GUERINI, T., «Disinformation, Fake news and Deep fakes», en MIRÓ-LLINARES, F./ DUVAC, C./TOADER, T./SANTISTEBAN GALARZA, M. (EDS.), «Criminalisation of AI-related offences», *RIDP*, 95, 2024, pp. 111-138.

GUIDOTTI, R. *et al.,* «A Survey of Methods for Explaining Black Box Models», *ACM Computing Surveys*, 51, 5, Article 93, 2018, pp. 1-42.

GUNKEL, D. J., «Mind the gap: responsible robotics and the problem of responsibility», *Ethics and Information Technology*, 22, 2020, pp. 307-320.

GUTIÉRREZ CASTAÑEDA, A., «¿Las grandes olvidadas de la reforma penal? Reflexiones sobre el papel de las penas privativas de derechos políticos y profesionales en la modernización del sistema de penas», en PÉREZ ÁLVAREZ, F. (Ed.), *Moderno discurso penal y nuevas tecnologías, Memorias del III Congreso Internacional de Jóvenes investigadores en ciencias penales, 17,18 y 19 de junio de 2013*, Ediciones Universidad de Salamanca, 2014, pp. 409-424.

GUTIÉRREZ DAVID, M. E., «Transparencia y comunicación de información a los responsables del despliegue en el artículo 13 Reglamento de inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 627-680.

GUTIÉRREZ-RUBÍ, A., «Tecnopolítica y los algoritmos», en SABARIEGO, J./JOBIM DO AMARAL, A./CARVALHO SALLES, E. B. (Orgs.), *Algoritarismos*, Tirant lo Blanch, Sao Paulo, 2020, pp. 102-112.

HACKER, P., «Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law», *Common Market Law Review*, 55 (4), 2018, pp. 1143-1186.

HACKER, P./PASSOTH, J. H., «Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond», en HOLZINGER, A./GOEBEL, R., *et al.* (Eds.), *xxAI - Beyond Explainable AI*, Springer, 2022, pp. 343-376.

HALLEVY, G., «I, Robot – I, Criminal – When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offences», *Syracuse Sci. & Tech. L. Rep.,* 2010, pp. 1-37.
— «The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control», *Akron Intellectual Property Journal*, 4/2, 2010, pp. 171-201.
— «Virtual Criminal Responsibility», *Original Law Review,* 6 (1), 2011, pp. 6-27.
— «Unmanned Vehicles. Subordination to Criminal Law Under the Modern Concept of Criminal Liability», *Journal of Law, Information and Science,* Vol 21, 2012, pp. 200-211.
— *When Robots Kill. Artificial Intelligence under Criminal Law*, Northeastern University Press, Boston, 2013.
— *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015.
— «Dangerous Robots – Artificial Intelligence vs. Human Intelligence», 2018.
— «The Basic Models of Criminal Liability of AI Systems and Outer Circles», 2019.
HAN, B. C., *Müdigkeitsgesellschaft*, Matthes & Seitz, Berlín, 2010.
— *Transparenzgesellschaft*, Matthes & Seitz, Berlín, 2012.
— *Im Schwarm: Ansichten des Digitalen*, Matthes & Seitz, Berlín, 2013.
— *La sociedad de la transparencia* (trad. R. P. Gabás), Herder, Barcelona, 2013.
— *Psychopolitik: Neoliberalismus und die neuen Machttechniken*, Matthes & Seitz, Berlín, 2014.
— *La sociedad del cansancio* (trad. A. Saratxaga), Herder, Barcelona, 2018, 2.ª ed.
— *Infokratie: Digitalisierung und die Krise der Demokratie*, Matthes & Seitz, Berlín, 2021.
— *No-cosas: Quiebras del mundo de hoy* (trad. J. Chamorro), Taurus, Madrid, 2021.
— *Undinge: Umbrüche der Lebenswelt*, Ullstein, Berlín, 2021.
— *Infocracia. La digitalización y la crisis de la democracia,* Taurus, Madrid, 2022.
— *Vida contemplativa: Elogio de la inactividad* (trad. M. Alberti), Taurus, Madrid, 2023.
HAO, K., «What is machine learning?», *MIT Technology Review*, 2022.
HARARI, Y. N., *Homo Deus: Breve historia del mañana*, (trad. R. Joan Domènec), Debate, Barcelona, 2017.
HARAWAY, D. J., *Manifiesto Ciborg*, Kaótica Libros, 2020.
HARBERS, M., PEETERS, M. M., /NEERINCX, M. A., «Perceived autonomy of robots: Effects of appearance and context», en ALDINHAS FERREIRA, M. I., SILVA SEQUEIRA, J., TOKHI, M. O., KADAR, E., / VIRK, G. S. (Eds.), *A World with Robots*, Springer, Cham, 2017.
HARCOURT, B., *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age*, University of Chicago Press, Chicago, 2007.

HART, H. L. A., *Punishment and Responsibility, Essays in the Philosophy of Law*, Clarendon Press, Oxford, 1970.
— *Punishment and Responsibility. Essays in the Philosophy of Law* (with an Introduction by John Gardner), Oxford University Press, Oxford, 2.ª ed., 2008.
HASHMI, M./KOKER, L. D./CASANOVAS, P., «Legal compliance through design: preliminary results of a literature survey», *TERECOM2018@ JURIX, Technologies for Regulatory Compliance*, 2309, 2018.
HASSEMER, W., *Theorie und Soziologie des Verbrechens: Ansätze zu einer praxisorientierten Rechtsgutslehre*, Scriptor-Verlag, Kronberg im Taunus, 1973.
— «Grundlinien einer personalen Rechtsgutslehre», en SCHOLLER, H./PHILIPPS, L. (Coords.), *Jenseits des Funktionalismus: Arthur Kaufmann zum 65. Geburtstag*, Decker & Müller, Heidelberg, 1989.
— «Kennzeichen und Krisen des modernen Strafrechts», *Zeitschrift für Rechtspolitik*, 25, 1992, pp. 378-383.
— «Rasgos y crisis del Derecho Penal moderno», *ADPCP*, 45, 1992, pp. 235-250.
— «Viejo y nuevo derecho penal», en HASSEMER, W., *Persona, mundo y responsabilidad*, Tirant lo Blanch, Valencia, 1999.
HASSEMER, W./MUÑOZ CONDE, F., *La responsabilidad por el producto en derecho penal*, Tirant Lo Blanch, Valencia, 1995.
HEFENDEHL, R., «¿Debe ocuparse el Derecho penal de riesgos futuros? Bienes jurídicos colectivos y delitos de peligro abstracto», *RECPC*, 04-14, 2002, passim.
— *Kollektive Rechtsgüter im Strafrecht*, Heymans, Köln, 2002.
HEIKKILÄ, M., «The AI myth Western lawmakers get wrong», *MIT Technology Review*.
HEINE, G., «Accesoriedad administrativa en el Derecho Penal del Medio Ambiente», *ADPCP*, 46/1, 1993, pp. 289-316.
HERBOSA MARTÍNEZ, I., «Encaje de los sistemas de IA en la definición de producto en la legislación de productos defectuosos Análisis de la legislación vigente con la vista puesta en la Propuesta de Directiva del Parlamento europeo y del Consejo de 28 de septiembre de 2022 (COM/2022/495)», *InDret*, 3/2024, pp. 52-98.
HERMAN MELVILLE, H., *Moby Dick,* Editorial Bruguera, 5.ª edición, Colección Historias n.º 7, Barcelona, 1973.
HERNÁNDEZ PEÑA, J. C., «La gobernanza y vigilancia del Reglamento de inteligencia artificial: autoridades de vigilancia del mercado, Comisión y las diversas entidades» en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 845-872.
HERNÁNDEZ RAMOS, M., *El impacto de la inteligencia artificial en el derecho constitucional, Centro de Estudios Políticos y Constitucionales,* Madrid, 2025.
HERZOG, F., *Gesellschaftliche Unsicherheit and strafrechtliche Daseinsvorsorge. Studien zur Vorlegung des Strafrechctschutzes in den GefÄhrdungsbereich*, R. V. Deckers, Heidelberg, 1991.

HERZOG, F.,«Límites al control penal de los riesgos sociales (Una perspectiva crítica ante el derecho penal del peligro)», trad. Larrauri Pijoan y Pérez Álvarez, *Nuevo Foro Penal,* 53, 1991, pp. 303-312.

— «Sociedad del riesgo, derecho penal del riesgo, regulación del riesgo», en ARROYO ZAPATERO, L./NEUMANN, U./NIETO MARTÍN, A. (Coords.), *Crítica y justificación del derecho penal en el cambio de siglo. El análisis crítico de la Escuela de Frankfurt*, Ediciones de la Universidad de Castilla-La Mancha, Cuenca, 2003, pp. 249-258.

HILDEBRANDT, M., «Ambient intelligence, criminal liability and democracy», *Criminal Law and Philosophy*, 2, 2008, pp. 163-180.

HILGENDORF, E., *Strafrechtliche Produzentenhaftung in der „Risikogesellschaft»*, Duncker & Humblot, Berlin, 1993.

— «Relación de causalidad e imputación objetiva a través del ejemplo de la responsabilidad penal por el producto», *ADPCP*, LV, 2002, pp. 91-108.

— Bayerischer Landtag, Ausschuss für Wirtschaft und Medien, Infrastruktur, Bau und Verkehr, Energie und Technologie, «Anhörung zum Thema: „Autonomes Fahren» (Audiencia parlamentaria en el parlamento de Baviera)», 29 de octubre de 2015.

— «Automated driving and the law», en HILGENDORF, E./SEIDEL, U. (Eds.), *Robotics, Autonomics, and the Law*, Nomos, Baden-Baden, 2017, pp. 171-194.

— «Autonome Systeme, künstliche Intelligenz und Roboter», en AA. VV., *Festschrift für Thomas Fischer*, Beck, Múnich, 2018, pp. 99-114.

— «The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems», en HILGENDORF, E./FELDLE, J. (Eds.), *Digitization and the Law*, Nomos, 2018, pp. 57-90.

HOEFER, C., «Causal Determinism», *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.).

HOMERO, *La Iliada*, Trad. Emeterio Fuentes, Verbum, Madrid, 2020.

HORN, E., *Konkrete Gefährdungsdelikte*, Otto Schmidt, Colonia, 1973.

HÖRNLE, T., «La protección de sentimientos en el StGB», en HEFENDEHL, R./VON HIRSCH, A./WOHLERS, W. (Eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?*, ed. en español a cargo de Rafael Alcácer Guirao, María Martín Lorenzo e Íñigo Ortiz de Urbina, Marcial Pons, 2016, pp. 375-392.

— «The Trolley Problem Reloaded, ¿Cómo deben programarse los vehículos autónomos para los dilemas de ¨vida contra vida"?», (trad. por José Béguelin), en HÖRNLE, T., *Criminalización, castigo, y dilemas morales en la obra de Tatjana Hörnle,* Buenos Aires, Editores del Sur, 2022.

HOROWITZ, M. C./SCHARRE, P., «Meaningful human control in weapons systems: A primer», *Center for a New American Security*, 2015.

HOYER, A., *Die Eignungsdelikte*, Duncker & Humboldt, Berlín, 1987.

HU, YING, «Robot criminals», *University of Michigan Journal of Law Reform*, 52(2), 2019, pp. 487-532.

HUERTA TOCILDO, S., «El derecho fundamental a la legalidad penal», *Revista Española de derecho constitucional*, 13(39), 1993, pp. 81-113.

HUNT, E., «Tay, Microsoft's Al Chatbot, Gets a Crash Course in Racism from Twitter», *The Guardian*, 24 de marzo de 2016.

HUQ, A. Z., «A Right to a Human Decision», *Virginia Law Review*, 106, 3, 2020, pp. 611-688.

HUSAK, D., *Overcriminalization: The limits of the criminal law*, Oxford University Press, 2008.

IBOLD, V., *Künstliche Intelligenz und Strafrecht. Zur strafrechtlichen Produktverantwortung in der Innovationsgesellschaft*, Nomos, Munich, 2024.

IMAI, T., «Japanese report on traditional criminal law categories and AI», *RIDP,* 94, 1, 2023.

ÍÑIGO CORROZA, E., *La responsabilidad penal del fabricante por defectos de sus productos*, J. M. Bosch Editor, Barcelona, 2001.

IRELAND, J. D., (Trad), *The Udāna & The Itivuttaka: Inspired Utterances of the Buddha & The Buddha's sayings*, Pariyatti Publishing, 2018.

JAIN, N., «Autonomous weapons systems: new frameworks for individual responsibility», en BHUTA, N. *et al.* (Eds.), *Autonomous weapons systems, Law, ethics, policy*, Cambridge University Press, Cambridge, 2016, pp. 303-324.

JAKOBS, G., *La imputación penal de la acción y de la omisión en derecho*, Universidad Externado de Colombia, 1996.

— *Sociedad, norma y persona en una teoría de un Derecho penal funcional*, Civitas, Madrid, 1996.

— *Derecho penal: fundamentos y teoría de la imputación. Parte general*, Marcial Pons, Madrid, 2.ª ed., 1997.

— «La Ciencia del Derecho Penal ante las exigencias del presente», en *Estudios de Derecho judicial,* 20, 1999, pp. 119-146.

— *La pena estatal: significado y finalidad* (traducción y estudio preliminar de M. Cancio Meliá y B. Feijóo Sánchez), Thomson-Civitas, Madrid, 2006.

— «¿Daño social? Anotaciones sobre un problema teórico fundamental en el Derecho Penal», *Cuadernos de Política Criminal*, 100, 2010, pp. 29-45.

JANUÁRIO, T. F. X., «Inteligencia artificial y responsabilidad penal de personas jurídicas: un análisis de sus aspectos materiales y procesales», *Estudios Penales y Criminológicos*, 44 (Especial «Inteligencia artificial y sistema penal»), 2023, pp. 1-39.

JI, Z. *et al.,* «Survey of Hallucination in Natural Language Generation», *ACM Computing Surveys*, 55, 2022, pp. 1-59.

JIN, H., «Tesla recalls almost half a million electric cars over safety issues», blog Reuters, 31 de diciembre de 2021.

JUANATEY DORADO, C., «Responsabilidad penal omisiva del fabricante o productor por los daños a la salud derivados de productos introducidos correctamente en el mercado», *ADPCP*, LVII, 2004, pp. 19-52.

KAHNEMAN, D., *Pensar rápido, pensar despacio*, Debate, Barcelona, 2012.

KAIAFA GBANDI, M./SACHOULIDOU, A./LIMA, D., «Greek report on traditional criminal law categories and AI», *RIDP*, 94, 1, 2023, pp. 223-252.

KANT, I., *Grounding for the Metaphysics of Morals*, en KANT, I. *Ethical Philosophy*, Ellington, J. W. (trad.), Hackett Publishing Co, Indianapolis, 1983.

KAPLAN, A./HAENLEIN, M., «Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence», *Business horizons*, 62/2019, pp. 15-25.

KARNOW, C. E., «Liability for distributed Artificial Intelligences», *Berkeley Technology Law Journal*, 11, 1996, pp. 147-204.

KAUFMANN, A., «Tatbestandsmäßigkeit und Verursachung im Contergan-Verfahren. Folgerungen für das geltende Recht und für die Gesetzgebung», *JZ*, 1971, pp. 569-576.

KIM, P. T., «Data-Driven Discrimination at Work», *Wm. & Mary L. Rev*, 58, 2016, pp. 857-936.

KINDHÄUSER, U., *Gefährdung als Straftat. Rechtstheoretische Untersuchungen zur Dogmatik der abstrakten und konkreten Gefährdungsdelikte*, Vittorio Klostermann, Frankfurt a. M., 1989.

— «Acerca de la legitimidad de los delitos de peligro abstracto en el ámbito del derecho penal económico», en *Hacia un derecho penal económico europeo. Jornadas en honor del profesor Klaus Tiedemann: Universidad Autónoma de Madrid, 14-17 de octubre de 1992*, 1995, pp. 441-452.

— «Estructura y legitimación de los delitos de peligro del Derecho penal», *InDret*, 1/2009, pp. 1-19.

KLEINBERG, J./MULLAINATHAN, S./RAGHAVAN, M., «Inherent trade-offs in the fair determination of risk scores», *PROC. INNOVATIONS THEORETICAL COMPUTER SCI.*, 2017.

KÖHLER, S./ROUGHLEY, N./SAUER, H., «Technologically blurred accountability? Technology, responsibility gaps and the robustness of our everyday conceptual scheme», en ULBERT, C./FINKENBUSCH, P./SONDERMANN, E./DEBIEL, T. (Eds.), *Moral Agency and the Politics of Responsibility*, Routledge, 1.ª ed., 2017, pp. 51-67.

KOOPS, B. J., HILDEBRANDT, M./JAQUET-CHIFFELLE, D. O., «Bridging the accountability gap: Rights for new entities in the information society?», *Minn. JL Sci. & Tech.*, 11, 497, 2010, pp. 497-561.

KOULU, R., «Human control over automation: EU policy and AI ethics», *Journal of Legal Studies*, 1, 2020, pp. 9-46.

KOWERT, W., «The Foreseeability of Human - Artificial Intelligence Interactions», *Texas Law Review*, 96, 2017.

KUHLEN, L., *Fragen einer strafrechtlichen Produkthaftung*, Müller C. F., Heidelberg, 1989.
— «Strafhaftung beim unterlassenem Rückruf gesundheitsgefährdender Produkte», *NStZ,* 1990, pp. 566-570.
— «Grundfragen der strafrechtlichen Produkthaftung», *JZ*, 49(23), 1994, pp. 1142-1147.
— «Zum Strafrecht der Risikogesellschaft», *GA*, 1994, pp. 347-367.
— «Cuestiones fundamentales de la responsabilidad penal por el producto», en MIR PUIG, S./LUZÓN PEÑA, D. (Coords.), *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Bosch, Barcelona, 1996, pp. 231-246.
— «Necesidad y límites de la responsabilidad penal por el producto», *ADPCP*, 55 (1), 2002, pp. 67-90.
KYRIAKIDIS, M. *et al.,* «A human factors perspective on automated driving», *Theoretical Issues in Ergonomics Science*, 20, 3, 2017, pp. 223-249.
LACAZE, M./WALTHER, J., «French report on traditional criminal law categories and AI», RIDP, 94, 1, 2023, pp. 153-194.
LACRUZ MANTECÓN, M., «Potencialidades de los robots y capacidades de las personas», en ROGEL VIDE, C. (Coord.), *Los robots y el Derecho*, Reus, Madrid, 2020, pp. 25-77.
LAGIOIA, F./SARTOR, G., «Ai systems under criminal law: a legal analysis and a regulatory perspective», *Philosophy & Technology*, 33/2020, pp. 433-465.
LAHTI, R; KOISTINEN, J AND LAHTI, L., «*Finnish national* report on criminalisation of AI-related offences», *RIDP*, 95, pp. 300-301.
LASCURAÍN, J. A., «Fundamento y límites del deber de garantía del empresario» en AAVV., *Hacia un derecho penal económico europeo. Jornadas en honor al profesor Klaus Tiedemann,* BOE*,* Madrid*,* 1995, pp. 209-227.
— «Los delitos contra los derechos de los trabajadores: lo que sobra y lo que falta», *ADPCP*, VOL. LVII, 2004, pp. 19-52.
— «Bien jurídico y objeto protegible», *ADPCP*, LX, 2007, pp. 119-163.
— *Sólo penas legales, precisas y previas. El derecho a la legalidad penal en la jurisprudencia constitucional*, Thomson Reuters Aranzadi, 2009.
— «Compliance, debido control y unos refrescos», en ARROYO ZAPATERO, L./NIETO MARTÍN, A. (dirs.), *El derecho penal económico en la era compliance*, Tirant Lo Blanch, Valencia, 2013, pp. 111-136.
— «La responsabilidad penal individual en los delitos de empresa», en DE LA MATA BARRANCO, N./DOPICO GÓMEZ-ALLER, J./LASCURAÍN SÁNCHEZ, J. A./ NIETO MARTÍN, A., *Derecho penal económico y de la empresa*, Dykinson, Madrid, 2018, pp. 87-128.
LAURENZO COPELLO, P., *El resultado en Derecho Penal*, Tirant Lo Blanch, Valencia, 1992.

LAUX, J., «Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act», *AI & SOCIETY*, 29, 2023, pp. 2853-2866.

LAZCOZ MORATINOS, G., *Gobernanza y supervisión humana de la toma de decisiones automatizada basada en la elaboración de perfiles*, tesis doctoral, 2023.

— «La vigilancia o supervisión humana en el artículo 14 del Reglamento de inteligencia artificial: ¿un mero requisito obligatorio para los sistemas de alto riesgo?», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 681-699.

LEHMAN-WILZIG, S. N., «Frankenstein unbound: Towards a legal definition of artificial intelligence», *Futures*, 13(6), 1981, pp. 442-457.

LIGTHART, S. *et al.,* «Forensic Brain-Readind and Mental Privacy in European Human Rights Law: Foundations and Challenges», *Neuroethics*, 14, 2021, pp. 191-203.

LIMA, D., «Could AI agents be held criminally liable: artificial intelligence and the challenges for criminal law», *SCL Rev.*, 69/2018, pp. 667-696.

LIU, H. Y./ZAWIESKA, C., «From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence», *Ethics and Information Technology*, 22/2020, pp. 321-333.

LLEDÓ BENITO, I., *El derecho penal, robots, IA y cibercriminalidad: desafíos éticos y jurídicos: ¿hacia una distopía?*, Dykinson, Madrid, 2022.

LLONÍN BLASCO, L., «Acerca de la relación entre inteligencia artificial y responsabilidad penal empresarial», *Revista Sistema Penal Crítico*, 3, 2022, pp. 27-48.

LOHMAN, A., *Strafrecht im Zeitalter con Künstlicher Intelligenz. Der Einfluss von autonomen Systemen und KI auf die tradierten strafrechtlichen Verantwortungsstrukturen*, Nomos, Baden-Baden, 2021.

LÓPEZ CALVO, J., en BARRIO ANDRÉS, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 264-272.

LÓPEZ GOROSTIDI, J., *Ciberdelincuencia: proporcionalidad y bienes jurídicos protegidos*, Comares, Granada, 2023.

— «Sobre la posibilidad de comisión de ciberdelitos contra la vida a la luz de la teoría de la imputación objetiva», *RGDP*, 43, 2025, pp. 1-43.

LÓPEZ-TARRUELLA MARTÍNEZ, A., «Derecho a presentar una reclamación y derecho a una explicación. Vías de recurso para los particulares en el reglamento de inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 893-918.

LORD GRIFFITHS/ DE VAL, P./ DORMER, R. J., «Developments in English Product Liability Law: A Comparison with the American System», *Tulane Law Review*, 353, 2021.

LOZA CORERA, M., «Datos y gobernanza de datos y conexiones con principios protección de datos en el artículo 10 del Reglamento», en COTINO HUESO, L./SIMÓN

CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 565-593.

LU, Z./COSTE, X./DE WINTER, J., «How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving», *Applied Ergonomics: human factors in technology and society*, 60, 2017, pp. 293-304.

LUHMANN, N., *Soziologie des Risikos,* Walter de Grurter, Berlín, 1991.

— *Sociología del riesgo*, (trad. Por Silvia Pappe, Brunhilde Erker, Luis Felipe Segura, Javier Torres Nafarrete), Universidad Iberoamericana, Jalisco, 1992.

LUZÓN PEÑA, D. M., *Estudios penales*, PPU, Barcelona, 1991.

MALDONADO FUENTES, F., «Reflexiones sobre las técnicas de tipificación de los llamados «delitos de peligro» en el moderno derecho penal», *Revista de Estudios de la Justicia*, 7, 2006, pp. 23-63.

MANES, V., «L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia», *disCrimen*, 2020, pp. 1-22.

— «L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia», en RUFFOLO, U. (ed.), *Intelligenza Artificiale – Il diritto, i diritti, l'etica*, Giuffré, Milano, 2020, pp. 547-567.

MANTELERO, A., «Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection», *Computer Law & Security Review*, 32, 2016, pp. 238-255.

— «El Reglamento de inteligencia artificial: la respuesta del legislador europeo a los retos de la inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 53-66.

— «The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template», *Computer Law & Security Review*, 54, 2024, pp. 1-18.

MARAVER GÓMEZ, M., *El principio de confianza en derecho penal, Un estudio sobre la aplicación del principio de autorresponsabilidad en la teoría de la imputación objetiva*, Civitas, Cizur Menor (Navarra), 2009.

— «Delitos de riesgo catastrófico», *Memento Práctico Penal*, Francis Lefebvre, 2025.

MARINUCCI, G./DOLCINI, E., «Derecho penal 'mínimo' y nuevas formas de criminalidad», *Revista de Derecho Penal y Criminología*, 9, 2002, pp. 147-167.

MARTÍN FABA, J. M., «La Inteligencia Artificial en la nueva Directiva de responsabilidad por los daños causados por productos defectuosos ¿realidad o expectativa?», *Revista CESCO*, 53, 2025, pp. 14-19.

MARTÍNEZ GARAY, L./MONTES SUAY, F., «El uso de valoraciones del riesgo de violencia en Derecho Penal: algunas cautelas necesarias», *InDret*, 2/2018, pp. 1-47.

MARTÍNEZ GARCÍA, J. I., «Pensar el riesgo. En diálogo con Luhmann», en PÉREZ ALONSO, E./ARENA GARCÍA, E./MERCADO PACHECO, P./SERRANO MORENO, J. L. (Coords.), *Derecho, globalización, riesgo y medio ambiente*, Tirant Lo Blanch, Valencia, 2012, pp. 323-354.

MARTÍNEZ PÉREZ, E. J., «El principio de cautela en la práctica internacional y europea: concepto, naturaleza jurídica y contenido», en GÓMEZ TOMILLO, M. (Dir.), *Principio de precaución y derecho punitivo del Estado*, Tirant Lo Blanch, Valencia, 2014, pp. 17-49.

MARTÍNEZ-BUJÁN PÉREZ, C., *El error en la teoría jurídica del delito*, Tirant Lo Blanch, Valencia, 2017.

— *Derecho penal económico y de la empresa. Parte general*, Tirant Lo Blanch, Valencia, 6.ª ed., 2022.

MASSARO, A., «Principio di precauzione e diritto penale: nihil novi sub sole?», *Diritto Penale Contemporaneo*, 9 mayo 2011, pp. 1-23.

MATA Y MARTÍN, R. M., *Bienes jurídicos intermedios y delitos de peligro*, Comares, Granada, 1997.

MATSUMI, H./SOLOVE, D. J., «The Prediction Society: Algorithms and the Problems of Forecasting the Future», *University of Illinois Law Review*, 1, 2025, pp. 1-62.

MATTHIAS, A., «The Responsibility Gap. Ascribing Responsibility for the Actions of Learning Automata», *Ethics and Information Technology*, 4, 2004, pp. 175-183.

— *Automaten als Träger von Rechten, Plädoyer für eine Gesetzänderung*, Humboldt Universität, Berlin, Logos Verlag, 2007.

MAURACH, R./SCHRÖDER, F-C., *Strafrecht. Besonderer Teil*, 6.ª edición, 1981.

MCALLISTER, A., «Stranger than science fiction: The rise of AI interrogation in the dawn of autonomous robots and the need for an additional protocol to the UN convention against torture», *Minnesota Law Review*, 101, 2017, pp. 2527-2573.

MCCARTHY, J./MINSKY, M. L./ROCHESTER, N./SHANNON, C. E., «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence», 31 de Agosto de 1955.

MCFADDEN, M./JONES, K./TAYLOR, E./OSBORN, G., «Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation», *Oxford Information Labs*, 2021.

MECACCI, G./SANTONI DE SIO, F., «Meaningful human control as reason-responsiveness: The case of dual-mode vehicles», *Ethics and Information Technology*, 22, 2020, pp. 103-115.

MENDOZA BUERGO, B., «Exigencias de la moderna política criminal y principios limitadores del Derecho penal», *ADPCP*, 52, 1999, pp. 279-322.

— *El derecho penal en la sociedad del riesgo*, Civitas, Madrid, 2001.

— *Límites dogmáticos y político-criminales de los delitos de peligro abstracto*, Comares, Granada, 2001.

MENDOZA BUERGO, B.,«La configuración del injusto (objetivo) de los delitos de peligro abstracto», *Revista de Derecho Penal y criminología*, 9, 2002, pp. 39-82.

— «Principio de precaución, derecho penal del riesgo y delitos de peligro», en ROMEO CASABONA, C. M., (Coord.), *Principio de precaución, biotecnología y derecho*, Comares, Bilbao-Granada, 2004, pp. 436-442.

MERAT, N./JAMSON, A. H./LAI, F. C. H./DALY, M./CARSTEN, O. M. J., «Transition to manual: Driver behaviour when resuming control from a highly automated vehicle», *Transportation Research Part*, 27, 2014, pp. 274-282.

MINSKY, M. L., *Semantic information processing*, MIT Press, Cambridge, 1968.

MIR PUIG, S., «Sobre lo objetivo y lo subjetivo en el injusto», *ADPCP,* vol. 41, 1988, pp. 661-684.

MIRÓ LLINARES, F., «El «moderno» Derecho penal vial y la penalización de la conducción sin permiso», *InDret*, 3/2009.

— *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*, Marcial Pons, Madrid, 2012.

— «La criminalización de conductas «ofensivas». A propósito del debate anglosajón sobre los «límites morales» del Derecho penal», *RECPC*, 17-23, 2015, pp. 1-65.

— «Inteligencia artificial y Justicia Penal. Más allá de los resultados lesivos causados por robots», *Revista de Derecho Penal y Criminología*, 20, 2018, pp. 87-130.

— «Lo que queda (y debe quedar) del bien jurídico en Derecho penal», en GONZÁLEZ CUSSAC, J. L./LEÓN ALAPONT, J. (eds.), *Estudios jurídicos en memoria de la Profesora Doctora Elena Górriz Royo*, Tirant lo Blanch, Valencia, 2020, pp. 599-620.

— «Predictive policing: utopia or dystopia? On attitudes towards the use of big data algorithms for law enforcement», *Revista de Internet Derecho y Política*, 30, 2020, pp. 1-18.

— «General report», en MIRÓ-LLINARES, F./DUVAC, C./TOADER, T./SANTISTEBAN GALARZA, M. (Eds.), *Criminalisation of AI-related offences*, *RIDP*, 95/2024, pp. 19-90.

— "Derecho penal y desafíos de la inteligencia artificial (en el contexto del nuevo marco regulatorio europeo)" en CASAS BAAMONDE, M. E., (Dir.), *Derecho y tecnologías*, Fundación Ramón Areces, 2025.

MIRÓ-LLINARES, F./CASTRO TOLEDO, F. M., «¿Correlación no implica causalidad? El valor de las predicciones algorítmicas en el sistema penal a propósito del debate epistemológico sobre «el fin de la teoría», en DEMETRIO CRESPO, E./DE LA CUERDA MARTÍN, M./GARCÍA DE LA TORRE, F. (Coords.), *Derecho penal y comportamiento humano. Avances desde la neurociencia y la inteligencia artificial*, Tirant Lo Blanch, Valencia, 2022, pp. 507-530.

MIRÓ-LLINARES, F./SANTISTEBAN GALARZA, M., «La regulación de los sistemas policiales predictivos en el Reglamento Inteligencia Artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 315-336.

MOLINA FERNÁNDEZ, F., «Delitos contra la seguridad del tráfico», en BAJO FERNÁNDEZ, M. (Dir.), *Compendio de Derecho Penal (Parte Especial)*, vol. II, Ceura, Madrid, 1998.

— «Presupuestos de la responsabilidad jurídica (análisis de la relación entre libertad y responsabilidad)», *ADFUAM,* 4, 2000, pp. 169-284.
— *Antijuridicidad penal y sistema del delito*, Bosch, Barcelona, 2001.
— *La cuadratura del dolo: problemas irresolubles, sorites y Derecho penal*, Univ. Externado de Colombia, Bogotá, 2007.
— «El razonable «regreso» del dolo a la culpabilidad», en AA. VV., *Estudios de derecho penal. Homenaje al profesor Santiago Mir Puig*, BdF, Montevideo, 2017, pp. 735-751.
— «Societas peccare non potest…: nec delinquere», en Bacigalupo Sagesse, S./Feijoo Sánchez, B./Echano Asaldua, J. I. (Coords.), *Estudios de Derecho Penal: homenaje al profesor Miguel Bajo*, Ed. Ramón Areces, Madrid, 2017, pp. 361-416.
— «La responsabilidad penal de las personas jurídicas y el valor de π», Blog de la facultad de derecho de la UAM, 14 de febrero de 2018.
— «La responsabilidad penal de las personas jurídicas y el valor de Pi», *Almacén del Derecho*, 20 de febrero de 2018.
— *Antijuridicidad penal y sistema del delito*, BdF, Montevideo-Buenos Aires, 2021.
— «Acción y tipicidad en los delitos activos», en Molina Fernández, F. (Dir.), *Memento práctico Penal*, Francis Lefebvre, 2025.
— «Delitos contra la seguridad vial», en Molina Fernández, F. (Dir.), *Memento práctico Penal*, Francis Lefebvre, 2025.

Montero, R., *La ridícula idea de no volver a verte*, Seix Barral, Barcelona, 2013.

Mosco, G. D., «AI and the Board Within Italian Corporate Law: Preliminary Notes», *European Company Law Journal*, 3/2020, pp. 87-96.

Mozo Seoane, A., «La revolución tecnológica y sus retos: medios de control, fallos de los sistemas y ciberdelincuencia», en Rogel Vide C. (coord..), *Los robots y el Derecho,* Reus, Madrid, 2020, pp. 79-98.

Mucci, T., «La historia de la IA», IBM blog, 21 de octubre de 2024.

Mulligan, C., «Revenge against robots», *South Carolina Law Review*, 69/3, 2018, pp. 579-596.

Muñoz Aranguren, A., «La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación», *InDret*, 2/2011, pp. 1-39.

Muñoz Conde, F., «Protección de bienes jurídicos como límite constitucional del Derecho penal», en AAVV, *El nuevo Derecho penal español. Estudios penales en memoria del Profesor José Manuel Valle Muniz*, Aranzadi, Cizur Menor, 2001, pp. 561-574.

Muñoz Conde, F., *Derecho Penal. Parte especial*, 25.ª Ed, revisada y puesta al día con la colaboración de Carmen López Peregrín, Tirant Lo Blanch, Valencia, 2023.

Muñoz García, C., *Regulación de la inteligencia artificial en Europa. Incidencia en los regímenes jurídicos de protección de datos y de responsabilidad por productos*, Tirant lo Blanch, Valencia, 2023.

Navarro Cardoso, F., «Anotaciones sobre la responsabilidad penal de las «personas artificiales», *Revista Penal*, 56/2025, pp. 198-207.

Navarro Dolmestch, R., «¿Qué haremos cuando la inteligencia artificial mate?», blog UCM (Universidad Católica de Maule), 3 de enero de 2024.

— «Brechas de responsabilidad penal por la actuación de máquinas dotadas de inteligencia artificial». *Polít. Crim*., vol. 20, 40, 2025, pp. 77-107.

— «Inteligencia artificial como «actuante» en el derecho penal. Una primera aproximación», *IDP. Revista de Internet, Derecho y Política*, (43), 2025, pp. 1-16.

Navas Navarro, S., *Daños ocasionados por sistemas de inteligencia artificial: Especial atención a su futura regulación*, Comares, Granada, 2022.

Nehm, K., «Autonomes Fahren – Bremsen Ethik und Recht den Fortschritt aus», *JZ*, 8, 2018, pp. 398.

Nevejans, N., «Study for Juri Committee, European Civil Law Rules In Robotics, Policy Department, Citizens' Rights and Constitutional Affairs», European Parliament, Bruselas, 2016.

— «Open letter to the European Commission: Artificial Intelligence and Robotics», *Robotics Openletter*, 2018.

Nieto Martín, A., «Introducción al derecho penal económico y de la empresa», en De la Mata Barranco, N./Dopico Gómez-Aller, J./Lascuraín Sánchez, J. A./Nieto Martín, A., *Derecho Penal económico y de la empresa*, Dykinson, Madrid, 2018, pp. 39-60.

— «Responsabilidad penal de la persona jurídica y programas de cumplimiento: de la gestión de la legalidad a la legitimidad», *La Ley compliance penal*, 4, 2021.

— «Introducción al derecho penal económico y de la empresa», en De la Mata Barranco, N./Dopico Gómez-Aller, J./Lascuraín Sánchez, J. A./Nieto Martín, A., *Derecho Penal económico y de la empresa*, Dykinson, Madrid, 2.ª ed., 2024, pp. 45-83.

Nietzsche, F., *Así habló Zaratustra* (trad. Sánchez Pascual), Alianza Editorial, Madrid, 1972.

Nilsson, N. J., *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge, UK, Cambridge University Press, 2010.

Novelli, C./Taddeo, M./Floridi, L., «Accountability in artifcial intelligence: what it is and how it works», *AI & Soc*, 39, 2024, pp. 1871-1882.

Nyholm, S., «Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility», *Science and Engineering Ethics*, 24, 2018, pp. 1201-1219.

O'Neil, C., *Weapons of math destruction: How big data increases inequality and threatens democracy*, Crown Publishing Group, Nueva York, 2016.

Obregón Fernández, A./Lazcoz Moratinos, G., «La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional Humanitario y el Derecho de la Unión Europea», *Revista Electrónica de Estudios Internacionales*, 42, 2021, pp. 1-29.

ORTEGA GIMÉNEZ, A., «El ámbito de aplicación territorial del Reglamento de inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp.123-136.

ORTIZ DE ZÁRATE, L., «Explicabilidad (de la inteligencia artificial)», *Eunomía. Revista en Cultura de la legalidad*, 22, 2022, pp. 328-344.

— «Sesgos de género en la inteligencia artificial», *Revista de occidente*, 502, 2023, pp. 5-20.

OSMANI, N., «The Complexity of Criminal Liability Systems», *Masaryk University Journal of Law and Technology*, 14/2020, pp. 53-82.

OUBIÑA BARBOLLA, S., «Límites a la utilización de algoritmos en el sector público. Reflexiones a propósito del caso SyRI», en BARONA VILAR, S. (Dir.), *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, Tirant lo Blanch, Valencia, 2021, pp. 655-683.

PAGALLO, U., *The Laws of Robots: Crimes, Contracts and Torts*, Springer, Nueva York, 2013.

«When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law», *Minds&Machines*, 27, 2017, pp. 625-638.

PALMA ORTIGOSA, A., «La evaluación de la conformidad en el diseño y producción de sistemas basados en inteligencia artificial en el contexto del «Nuevo Marco Legislativo»«, en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 427-446.

PARASURAMAN, R./MANZEY, D. H., «Complacency and Bias in Human Use of Automation: An Attentional Integration», *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52, 3, 2010, pp. 381-410.

PARASURAMAN, R./SHERIDAN, T. B./WICKENS, C. D., «A Model for Types and Levels of Human Interaction with Automation», *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(3), 2000, pp. 286-297.

PAREDES CASTAÑÓN, J. M., «El límite entre imprudencia y riesgo permitido en Derecho Penal: ¿es posible determinarlo con criterios utilitarios?», *ADPCP*, XLIX, 1996, pp. 909-942.

— «El riesgo como construcción conceptual: sobre el uso y el abuso de las ciencias sociales en el discurso político-criminal del «Derecho Penal del riesgo», en *Revista Catalana de Seguretat Pública*, 13/2003, pp. 11-29.

PAREDES CASTAÑÓN, J. M., «Los delitos de peligro como técnica de incriminación en el derecho penal económico: bases político-criminales», *RDPC*, 11/2003, pp. 138-189.

— «¿Al otro lado del discurso jurídico penal? Sobre las bases político-criminales de la «modernización» del Derecho penal. Reseña crítica de Luis Gracia Martín», *RDPC*, 17, 2006, pp. 437-472.

— «Caso del aceite de colza», en Sánchez-Ostiz Gutiérrez, P. (Coord.), *Casos que hicieron doctrina en derecho penal*, *La Ley*, Madrid, 2011, pp. 425-439.
— «Responsabilidad penal por productos defectuosos», en Camacho Vizcaíno, A. (Dir.), *Tratado de Derecho penal económico*, Tirant Lo Blanch, Valencia, 2019, pp. 601-633.
Parfit, D., *Razones y personas*, trad. de Mariano L. Rodríguez González, A., Machado Libros, Boadilla del Monte, 2004.
Pasquale, F., *The black box society: the secret algorithms that control Moner and information*, Harvard University Press, Cambridge, 2015.
Peñaranda Ramos, E., «Sobre la responsabilidad en comisión por omisión respecto de hechos cometidos en la empresa (y en otras organizaciones)», en AA. VV., *Liber amicorum en homenaje al profesor Antonio González-Cuéllar García*, Colex, 2006, pp. 411-430.
— «Sobre el alcance del art. 65.3 del Código Penal. Al mismo tiempo una contribución a la crítica de los delitos de infracción de deber», en AA. VV., *Estudios penales en homenaje a Enrique Gimbernat*, Edisofer, Madrid, 2008, Tomo II, pp. 1419-1452.
— *La participación en el delito y el principio de accesoriedad*, BdeF, Buenos Aires, 2015.
Pérez Alonso, E., «Expansión, riesgo y principio de precaución en el Derecho penal», en Pérez Alonso, E./Arana García, E./Serrano Moreno, J. L./Mercado Pacheco, P. (Coords.), *Derecho, globalización, riesgo y medio ambiente*, Tirant Lo Blanch, Valencia, pp. 393-414.
Pérez Arias, J., «Algoritmos y big data en la responsabilidad penal», en Peris Riera, J./Massaro, A. (Coords.), *Derecho Penal, Inteligencia Artificial y Neurociencias*, Roma Tre-Press, 2023, pp. 159-190.
Pérez Luño, A. E., «El posthumanismo no es humanismo», *Derechos y libertades*, 44, Época II, 2021, pp. 17-40.
Pérez Manzano, M., «El tiempo de la consciencia y la libertad de decisión: Bases para una reflexión sobre neurociencia y responsabilidad penal», *Doxa, Cuadernos de Filosofía del Derecho*, 35, 2012, pp. 471-498.
— «Justicia penal y neuroderechos», en AA. VV., *Estudios político-criminales, jurídico-penales y criminológico. Llibro homenaje al profesor José Luis Díez Ripollés*, Tirant Lo Blanch, Valencia, 2023, pp. 329-348.
— «¿Un estatuto constitucional singular para el cerebro y las neurotecnologías? Sobre los neuroderechos», *RGDC*, 41, 2024, pp. 136-190.
Pérez Manzano, M., Lascuraín Sánchez, J. A., (Dirs.)., La tutela multinivel del principio de legalidad penal, Marcial Pons, 2016.
Pérez-Sauquillo, C., «Delitos de peligro abstracto y bienes jurídicos colectivos», *Revista Foro FICP*, 3, 2015, pp. 115-145.
— *Legitimidad y técnicas de protección penal de bienes jurídicos supraindividuales*, Tirant Lo Blanch, Valencia, 2019.

PERROW, C., *Normal Accidents: Living with High-Risk Technologies*, Basic Books, Harper Collins Publisher, 1984.

PICOTTI, L./PANATTONI, B. (Coords.), «Traditional criminal law categories and AI: Crisis or palingenesis?», *RIDP*, 94, 2023.

PIERGALLINI, C., *Danno da Prodotto e Responsabilità Penale,* Giuffrè, Milán, 2004

— «Intelligenza Artificiale: ¿De 'mezzo' ad 'autore' del reato?», *Rivista italiana di diritto e procedura penale*, 4, 2020, pp. 1745-1774.

— «Intelligenza Artificiale: De 'mezzo' ad 'autore' del reato?», en ACQUAROLI, R./FRONZA, E./GAMBERRINI, A. (Coords.), *La giustizia penale tra ragione e prevaricazione*, Aracne, Roma, 2021, pp. 219-244.

PINTO NETO, M., «Blur», en SABARIEGO, J./AMARAL, A. J./SALLES, E. B. (Orgs.), *Algoritarismos*, Tirant Lo Blanch, Sao Paulo, 2020, pp. 141-157.

PONGILUPPI, C., «Riflesione sulle funzione della pena nei contesti precauzionali», en FOFFANI, L. (Coord.), *La sicurezza agroalimentare nella prospettiva europea. Precauzione, prevenzione, repressione*, Ed. Giufre, Milán, 2014, pp. 549-558.

POZUELO PÉREZ, L., *Las penas privativas de derechos en el Código penal*, Colex, Madrid, 1998.

— «De nuevo sobre la denominada «expansión» del derecho penal: una relectura de los planteamientos críticos», en AAVV., *El funcionamiento en derecho penal. Libro Homenaje al profesor Günther Jakobs*, Universidad Externado de Colombia, Tomo II, 2003, pp. 107-135.

— «Notas sobre la denominada «expansión del Derecho penal»: un análisis al hilo de los delitos contra la ordenación del territorio», *RDPP*, 9/2003, pp. 13-47.

PRESNO LINERA, M. A., *Derechos fundamentales e inteligencia artificial*, Marcial Pons, Madrid, 2022.

— «La prohibición de sistemas de inteligencia artificial que evalúan y clasifican a las personas a partir de datos que no guardan relación con el contexto donde se generaron y que provocan discriminaciones», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 237-250.

PRITTWITZ, C., *Strafrecht und Risiko: untersuchungen zur Krise von Strafrecht und Kriminalpolitik in der Risikogesellschaft*, Vittorio Klostermann, Frankfurt am Main, 1993.

— «Sociedad de riesgo y derecho penal», en ARROYO ZAPATERO, L./NEUMANN, U./NIETO MARTÍN, A. (Coords.), *Crítica y justificación del derecho penal en el cambio de siglo. El análisis crítico de la Escuela de Frankfurt*, Ediciones de la Universidad de Castilla-La Mancha, Cuenca, 2003, pp. 259-268.

PRITTWITZ, C., «La función del Derecho Penal en la sociedad globalizada del riesgo: defensa de un rol necesariamente modesto», en PÉREZ ALONSO, E./ARANA GARCÍA, E./SERRANO MORENO, J. L./MERCADO PACHECO, P. (Coords.), *Derecho, globalización, riesgo y medio ambiente*, Tirant Lo Blanch, Valencia, 2012, pp. 415-428.

— *Derecho penal y riesgo: Investigaciones sobre la crisis del Derecho penal y la política criminal en la sociedad del riesgo*, Marcial Pons, Madrid, 2021.

Procaccino, A., «Securitizzazione dell'Unione europea e poteri concorrenti. Dall'investigazione, alla prevenzione, all'osservazione», *Diritto Penale Contemporaneo*, 1/2023, pp. 145-170.

Provaznick, J., «Czech Republic national report on criminalisation of AI-related offences», *RIDP,* 95, 2024, pp. 245-274.

Puppe, I., « «Naturgesetze» vor Gericht. Die sogenannte generelle Kausalität und ihr Beweis, dargestellt an Fällen strafrechtlicher Produkthaftung», *JZ*, 1994, pp. 1147-1151.

— «Problemas de imputación del resultado en el ámbito de la responsabilidad penal por el producto», en Mir Puig, S./Luzón Peña, D. M. (Coords.), *Responsabilidad penal de las empresas y sus órganos y responsabilidad por el producto*, Bosch, Barcelona, 1996, pp. 215-230.

Puyol Montero, J., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 273-282.

Quintero Olivares, G., «La robótica ante el derecho penal el vacío de respuesta jurídica a las desviaciones incontroladas», *Revista Electrónica de Estudios penales y de la seguridad*, 1, 2017, pp. 1-23.

Ramírez, H., «Inteligencia Artificial e igualdad de género», Grupo Ático blog.

Ramón Fernández, F., en Barrio Andrés, M., (Dir.), *Comentarios al Reglamento Europeo de Inteligencia artificial*, La Ley, Madrid, 2024, pp. 681-686.

Ramos Vázquez, J. A., «La pena de inhabilitación especial para profesión, oficio, industria o comercio», en Faraldo Cabana, P./Puente Aba, L. M. (Dirs.), *Las penas privativas de derechos y otras alternativas a la privación de libertad*, Tirant Lo Blanch, Valencia, 2013, pp. 115-128.

Rawte, V./Sheth, A./Das, A., «A survey of hallucination in large foundation models», *arXiv preprint arXiv:2309.05922*, 2023.

Reche Tello, N., «Nuevos derechos frente a la neurotecnología: la experiencia Chilena», *Revista de Derecho Político*, 112, 2021, pp. 415-446.

Ricoeur, P., *Reflections on the just*, trad. David Pellauer, Chicago, University of Chicago Press, 2007.

Rivas Vallejo, P., «Discriminación algorítmica: detección, prevención y tutela», XXXI Jornades Catalanes de Dret Social, Barcelona, abril de 2021.

Robles Planas, R., «Participación en el delito e imprudencia», *Revista de Derecho penal y criminología*, 9/2000, pp. 223-251.

— «¿Delitos de personas jurídicas?», *InDret*, 2/2006, pp. 1-25.

— *Pena y persona jurídica. Crítica del artículo 31 bis CP*, *Diario La Ley*, 7705, 2011.

Rodotà, S., «Del ser humano al posthumano», en Quadra-Salcedo, T./Fernández del Castillo, T. (Dirs.), *Sociedad digital y derecho*, BOE, Madrid, 2018, pp. 87-94.

RODRÍGUEZ HORCAJO, D., *Comportamiento humano y pena estatal: disuasión, cooperación y equidad*, Marcial Pons, 2017.

— «La pena de multa como imposible eje del sistema de consecuencias jurídicas del delito de la persona física», en RODRÍGUEZ HORCAJO, D./BASSO, G. (Dirs.), *Repensando las consecuencias jurídicas del delito*, Tirant lo Blanch, Valencia, 2025, pp. 219-253.

RODRÍGUEZ LÓPEZ, S., «Hacia un nuevo delito de difusión no consentida de material íntimo manipulado (deepfakes)», en AAVV., *Un derecho penal moderno para una democracia del siglo XXI, Libro homenaje al Prof. Dr. Carlos Martínez-Buján Pérez*, Tirant Lo Blanch, Valencia, 2025, pp. 628-636.

RODRÍGUEZ MONTAÑÉS, T., *Delitos de peligro, dolo e imprudencia*, Centro de Estudios Judiciales, Madrid, 1994.

— «Incidencia dogmática de la jurisprudencia del caso de la colza y otros casos en materia de productos defectuosos», en BOIX REIG, F. J./CAMPOS CRISTÓBAL, R./BERNARDI, A. (Coords.), *Responsabilidad penal por defectos en productos destinados a los consumidores*, Iustel, Madrid, 2005, pp. 115-132.

RODRÍGUEZ MOURULLO, G., *Derecho penal. Parte general*, Civitas, Madrid, 1978.

— *La omisión de socorro en el código penal*, Tecnos, Madrid, 1996.

RODRÍGUEZ-IGLESIAS, I., «Desespacialización y privilegio en el discurso glotopolítico en la prensa sobre mujeres andaluzas. Implicaciones para la comunicación intercultural y el multilingüismo», en GUERRERO SALAZAR, S. (ed.), *Lenguaje y discurso sobre las mujeres en la prensa española*, Círculo de Lingüística Aplicada a la Comunicación 91, 2022, pp. 97-112.

ROFF, H. M./MOYES, R., «Meaningful Human Control, Artificial Intelligence and Autonomous Weapons», Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons.

ROFF, H. M., «Lethal autonomous weapons and jus ad bellum proportionality», *Case W. Res. J. Int'l L.*, 47, 2015.

ROMEO CASABONA, C. M., «La peligrosidad y el peligro en la estructura del tipo del delito imprudente», en AA. VV., *La ciencia del derecho penal ante el nuevo siglo. Libro homenaje al profesor doctor don José Cerezo Mir*, Tecnos, Madrid, 2002, pp. 950-969.

ROMEO CASABONA, C. M., «Aportaciones del principio de precaución al Derecho penal», en ROMEO CASABONA, C. M. (Ed.), *Principio de Precaución, Biotecnología y Derecho*, Comares, Granada, 2004, pp. 385-422.

— «Criminal responsability of robots and autonomous artificial intelligent systems?», *Comunicaciones en propiedad industrial y derecho de la competencia*, 91, 2020, pp. 167-188.

— «La atribución de responsabilidad penal por los hechos cometidos por sistemas autónomos inteligentes, robots y tecnologías conexas», en SÁNCHEZ, J. M./GARCÍA PÉREZ, O./CEREZO DOMÍNGUEZ, A. I./GARCÍA ESPAÑA, E. (Dirs.), *Es-*

*tudios político-criminales, jurídico-penales y criminológicos. Libro homenaje al profesor José Luis Díez Ripollés*, 2023, pp. 721-733.

Romeo Casabona, C. M./Lazcoz Moratinos, G., «Inteligencia artificial aplicada a la salud: ¿qué marco jurídico?», *Revista de Derecho y Genoma Humano*, 52/2020, pp. 139-167.

Rosa, H., *Alienación y aceleración: hacia una teoría crítica de la temporalidad en la modernidad tardía*, Katz Editores, Madrid, 2016.

— *Resonancia: Una Sociología de la Relación con el Mundo*, Katz Editores, Bogotá, 2019.

Roso Cañadillas, R., «Derecho penal adaptado a la IA: entre el derecho penal analógico y el derecho penal digital», *En Letra*, 17, 2023, pp. 43-94.

— «Los retos de la IA para la teoría del delito: ¿un Derecho penal digital?*», Libertas,* Nº. 15, 2026, pp. 90-157.

Roxin, C., *Derecho penal. Parte General, Tomo I. Fundamentos. La estructura de la teoría del delito*, Civitas, Madrid, 1997.

— «El concepto de bien jurídico como instrumento de crítica legislativa sometido a examen», *RECPC*, 15-01, 2013, pp. 1-27.

Roxin, C./Greco, L., *Strafrecht. Allgemeiner Teil. Band I: Grundlagen. Der Aufbau der Verbrechenslehre*, C. H. Beck, Munich, 5.ª ed., 2020.

Rudin, C., «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead», *Nature Machine Intelligence*, 1, 2019, pp. 206-215.

Rueda Martín, M. Á., «El bien jurídico protegido en los delitos relativos a productos de consumo masivo», *Estudios Penales y Criminológicos*, XXX, 2010, pp. 395-445.

— «La concreción del deber objetivo de cuidado en el desarrollo de la actividad médico-quirúrgica curativa a la luz de los avances de los sistemas de Inteligencia Artificial», *InDret*, 1/2025, pp. 178-231.

Ruga Riva, C., «Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica», en Dolcini, E./Paliero, C. E. (Coords.), *Studi in onore di Giorgio Marinucci*, II, Giufre, Milán, 2006, pp. 1743-1777.

Rusca, B., «Los delitos de peligro abstracto como presunciones refutables. Nuevos argumentos en defensa de una teoría clásica», *Revista Chilena de Derecho*, 49, 1, 2022, pp. 101-126.

Russell, S. J./Norvig, P., *Artificial intelligence: a modern approach*, Prentice Hall, Nueva Jersey, 1995.

Salardi, S., «Robótica e inteligencia artificial: retos para el derecho», *Derechos y libertades*, 42, Época II, enero 2020, pp. 203-232.

Salazar, I./Liébanas, M. A., «Vigilancia poscomercialización en los sistemas de inteligencia artificial de alto riesgo en el Reglamento. Descripción, medidas y casos de uso», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado*

*sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 743-754.

Salvador Coderch, P./Rubí Puig, A., «Causas de exoneración de la responsabilidad: Excepción por riesgos de desarrollo», en Salvador Coderch, P./Gómez Pomar, F. (Coords.), *Tratado de responsabilidad civil del fabricante*, Civitas, Madrid, 2008, pp. 568-584.

— «Riesgos de desarrollo y evaluación judicial del carácter científico de dictámenes periciales. Avance del Tratado de Responsabilidad Civil por Daños Causados por Productos Defectuosos», *InDret*, 1/2008, pp. 1-63.

Salvador Coderch, P./Ruiz García, J. A./Seuba Torreblanca, J. C./Solé i Feliu, J./Luna Yerga, Á./Carrasco Martín, J., «Los riesgos de desarrollo», *InDret*, 1/2001, pp. 1-30.

Salvador Serna, M., «Inteligencia artificial y gobernanza de datos en las administraciones públicas: relaciones y evidencias para su desarrollo», *Gestión y Análisis de Políticas Pública*, 26, julio 2021, pp. 20-32.

Salvadori, I., «Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal», *Cuadernos de política Criminal*, 133, 2011, pp. 137-174.

Samoili, S./Lopez Cobo, M./Gomez Gutierrez, E./De Prato, G./Martinez-Plumed, F./Delipetrev, B., «AI Watch. Defining Artificial Intelligence», Publications Office of the European Union, 2020.

Sánchez Acevedo, M. E., «Ciberseguridad en sistemas de inteligencia artificial de alto riesgo en el artículo 15 del Reglamento», en Cotino Hueso, L./Simón Castellano, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 721-742.

Sánchez Monedero, J./Dencik, L., «The Politics of Deceptive Borders: «Biomarkers of Deceit» and the Case of iBorderCtrl», *Information, Communication & Society*, 3, 2020, pp. 413-430.

Santana Vega, D. M., «Reseña bibliográfica relativa a la obra: «Bienes jurídicos intermedios y delitos de peligro. Aproximación a los presupuestos de la técnica …», *Rev. cienc. Juríd*, 04, 1999, pp. 529-531.

Santisteban Galarza, M., «Garantías frente a la moderación de contenidos en la propuesta de Reglamento único de servicios digitales», *CESCO*, 41, 2022, pp. 159-179.

— «La criminalización de las ultrafalsificaciones (con especial atención a las implicaciones de la normativa europea de servicios digitales e inteligencia artificial)», *RDPC*, 31, 2024, pp. 211-249.

Santoni de Sio, F./Mecacci, G., «Four responsibility gaps with artificial intelligence: Why they matter and how to address them», *Philosophy & Technology*, 34/2021, pp. 1057-1084.

Santoni de Sio, F./Van den Hoven, J., «Meaningful Human Control Over Autonomous Systems: A Philosophical Account», *Frontiers in Robotics and AI*, 2018.

SAUER, F., «Autonomous weapons systems» en ELLIOTT, A. (Ed.), *The Routledge Social Science Handbook of AI*, Routledge, Londres y Nueva York, 2022, pp. 237-250.

SCHARRE, P. D., «The Opportunity and Challenge of Autonomous Systems,» en WILLIAMS, A. P./SCHARRE, P. D., *Autonomous Systems Issues for Defence Policymakers*, HQ SACT, 2015, pp. 3-26.

SCHNEEBERGER, *et al.,* «The Tower of Babel in Explainable Artificial Intelligence (XAI)», en HOLZINGER, A., ET AL. (Eds.), *Machine Learning and Knowledge Extraction*, Springer, 2023, pp. 65-81.

SCHRÖDER, H., «Abstrakt-konkrete Gefährungsdelikte?», *JZ*, 1, 1967, pp. 522-525.

SCHÜNEMANN, B., «Moderne Tendenzen in der Dogmatik der Fahrlässigkeits und Gefährdungsdelikte», *JA*, 1975.

— *Unternehmenskriminalität und Strafrecht: eine Untersuchung der Verantwortlichkeit der Unternehmen und ihrer Führungskräfte nach geltendem und geplantem Straf-und Ordnungswidrigkeitenrecht*, Heymann, Köln, 1979.

— «Cuestiones básicas de dogmática jurídico-penal y de política criminal acerca de la criminalidad de empresa», *ADPCP*, XLI, 1988, pp. 538 y ss.

— «Kritische Anmerkungen zur geistigen Situation der deutschen Strafrechtswissenschaft», *GA* (5), 1995, p. 201-229.

— «Consideraciones críticas sobre la situación espiritual de la ciencia jurídico-penal alemana», *ADPCP*, 49, 1996, pp. 529-558.

— «El principio de protección de bienes jurídicos como punto de fuga de los límites constitucionales de los tipos penales y de su interpretación», en HEFENDEHL, R./VON HIRSCH, A./WOHLERS, W. (Eds.), *La teoría del bien jurídico. ¿Fundamento de legitimación del Derecho penal o juego de abalorios dogmático?*, ed. en español a cargo de Rafael Alcácer Guirao, María Martín Lorenzo e Íñigo Ortiz de Urbina, Marcial Pons, Madrid, 2016, pp. 191-220.

SCOLLO, L., «*,intelligenza artificiale entra nel Codice penale*», *Diritto Penale e processo,* 6/2025.

SCOPINO, G., *Algo bots and the Law. Technology, Automation, and the Regulation of Futures and Other Derivatives*, Cambridge University Press, 2020.

SEARLE, J. R./VILLANUEVA, L. M. V., *Mentes, cerebros y ciencia*, Cátedra, Madrid, 2001.

SEHER, G., «Intelligente Agenten als «Personen» im Strafrecht», en Gless, S./Seelmann, K. (Coords.), *Intellegente Agenten und das Recht*, Nomos, Baden-Baden, 2016, pp. 45-60.

SELBST, A., «Disparate Impact in Big Data Policing», *Georgia Law Review*, 52, 2017, pp. 109-195.

— «Negligence and AI's Human Users», *Boston University Law Review*, 4, 2020, pp. 1331-1340.

SELBST, A. D./BOYD, D./FRIEDLER, S. A./VENKATASUBRAMANIAN, S./VERTESI, J., «Fairness and abstraction in sociotechnical systems», *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 59-68.

SEMPERE, J. F., «El régimen sancionador en el Reglamento de inteligencia artificial», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 873-892.

SENENT VIDAL, M. J./VENTURA FRANCH, A., «Inteligencia artificial e igualdad de mujeres y hombres: una regulación jurídica necesaria. Especial referencia al proyecto de reglamento europeo», *IgualdadES*, 10/2024, pp. 45-71.

SGUBBI, F., *Il diritto penale come rischio sociale. Ricerche sulle scelte di allocazione dell'illegalità penale*, Il Mulino, Bolonia, 1990.

— *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Il Mulino, Bolonia, 2019.

SHINER, R. A., «Corporations and the Presumption of Innocence», *Criminal Law and Philosophy*, 8, 2014, pp. 485-503.

SILVA SÁNCHEZ, J. M., *El delito de omisión: concepto y sistema*, Bosch, Barcelona, 1986.

— *Aproximación al Derecho penal contemporáneo*, Bosch, Barcelona, 1992.

— *La expansión del derecho penal: aspectos de la política criminal en las sociedades postindustriales*, 3.ª ed., BdeF, Madrid-Montevideo, 2011.

— «La responsabilidad penal de las personas jurídicas en Derecho español», en SILVA SÁNCHEZ, J. M. (Dir.) /MONTANER FERNÁNDEZ, R. (Coord.), *Criminalidad de empresa y Compliance*, Atelier, Barcelona, 2013, pp. 15-39.

— «A vueltas con la actio libera in causa», *InDret*, 4/2016.

— *El riesgo permitido en Derecho penal económico*, Atelier, Barcelona, 2022.

— *Derecho Penal. Parte General*, Civitas, Madrid, 2025.

— «Expansión 2.0: los nuevos riesgos», *InDret*, 2/2025.

SIMESTER, A. P., «Is strict liability always wrong?», en SIMESTER, A. (ed.), *Appraising Strict Liability*, Oxford University Press, 2005, pp. 21-50.

SIMESTER, A. P./VON HIRSCH, A., «Remote harms and non-constitutive crimes», *Criminal Justice Ethics*, 28(1), 2009, pp. 89-107.

SIMMLER, M./FRISCHKNECHT, R., «A taxonomy of human–machine collaboration: capturing automation and technical autonomy», *AI & SOCIETY*, 36, 2020, pp. 239-250.

SIMMLER, M./MARKWALDER, N., «Roboter in der Verantwortung? –Zur Neuauflage der Debatte um den funktionalen Schuldbegriff», *Zeitschrift für die gesamte Strafrechtswissenschaft*, 129(1), 2017, pp. 20-47.

— «Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence», *Criminal Law Forum*, 30, 2019, passim.

SIMÓN CASTELLANO, P., «El resto de los sistemas de inteligencia artificial prohibidos o inaceptables en el Reglamento de inteligencia artificial», COTINO HUESO, L./ SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 275-288.

— «Los sistemas de gestión de riesgos como obligación específica para los sistemas de inteligencia artificial de alto riesgo en el artículo 9 del Reglamento», en

COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea,* Aranzadi, Madrid, 2024, pp. 535-564.

SKITKA, L. J./MOSIER, K. L./BURDICK, M., «Does automation bias decision-making?», *Int. J. Human-Computer Studies*, 51, 1999, pp. 91-1006.

SKITKA, L. J./MOSIER, K. L./BURDICK, M./ROSENBLATT, N., «Automation bias and errors: Are crews better than individuals?», *The International Journal of Aviation Psychology*, 10, 1, 2000, pp. 85-97.

SMIT, A. M., «The Netherlands national Report on criminalisation of AI-related offences», *RIDP*, 2024.

SOLA RECHE, E., «Principio de precaución y tipicidad penal», en ROMEO CASABONA, C. M., (Coord.), *Principio de precaución, biotecnología y derecho,* Comares, Bilbao-Granada, 2004, pp. 475-491.

SOLARI MERLO, M. N., «Desarrollo tecnocientífico y bienes jurídicos. Cuando la tecnología revela lo importante», *ADPCP*, VOL. LXXVII, 2024, pp. 255-303.

SOLUM, L. B., «Legal Personhood for Artificial Intelligences», *70 North Carolina Law Review*, 1992, pp. 1231-1287.

SORIANO ARNANZ, A., «La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de Inteligencia Artificial», en BONORINO RAMÍREZ, P. R./FERNÁNDEZ ACEVEDO, R./VALCÁRCEL FERNÁNDEZ, P. (Dirs.), *Nuevas normatividades. Inteligencia artificial, derecho y género*, Thomson Reuters Aranzadi, Cizur Menor, 2021, pp. 63-88.

— «Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules», *Revista d'internet, dret y política*, 38, 2023, pp. 1-12.

SPARROW, R., «Killer Robots», *Journal of Applied Philosophy*, 24(1), 2007, pp. 62-77.

SPENCER, J. R./PEDAIN, A., «Approaches to strict and constructive liability in continental criminal law», en SIMESTER, A. (ed.), *Appraising Strict Liability*, Oxford University Press, 2005, pp. 237-283.

STORTONI, L., «El riesgo procedente del desconocimiento tecnológico», en AA. VV, *Responsa iurisperitorum digesta*, Ediciones Universidad Salamanca, vol. 3, 2000, pp. 83-104.

— «Angoscia tecnologica ed esorcismo penale», *Rivista italiana di diritto e procedura penale*, vol. 47, 2004, pp.71-89.

STRATENWERTH, G., «Zukunftssicherung mit den Mitteln des Strafrechts?», *ZstW*, 105, 1993.

SUNSTEIN, C., «Beyond the Precautionary Principle», *University of Pennsylvania Law Review*, 3, 2003, pp. 1003-1058.

— «*Leyes de miedo: más allá del principio de precaución*, Katz Editores, Buenos Aires-Madrid, 2009.

SURESH, H./GUTTAG, J., «A framework for understanding sources of harm throughout the machine learning life cycle», *Equity and access in algorithms, mechanisms, and optimization* (EAAMO '21), October 5-9, 2021, pp. 1-9.

SUTTON, S. G./ARNOLD, V./HOLT, M., «How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work», *Journal of Emerging Technologies in Accounting*, 15, 2, 2018, pp. 15-25.

— «An extension of the theory of technology dominance: Capturing the underlying causal complexity», *International Journal of Accounting Information Systems*, 50, 2023, pp. 1-18.

SZYMBORSKA, W., BELTRÁN, G./MURCIA, A./PONIATOWSKA, E./BELTRÁN, G./MURCIA, A., *Poesía no completa*, Fondo de Cultura Económica, 2013.

TADDEO, M./FLORIDI, L., «How AI can be a force for good», *Science*, 361(6404), 2018, pp. 751-752.

TEIJÓN ALCALÁ, M., «Los delitos de peligro en el derecho contemporáneo», *RECPC*, 25, 2023, pp. 1-41.

TERRADILLOS BASOCO, J. M., «Peligro abstracto y garantías penales», *Nuevo Foro Penal*, 62, 1999, pp. 67-94.

— «Delitos contra los derechos de los trabajadores: veinticinco años de política legislativa errática», *Estudios Penales y Criminológicos*, 41, 2021, pp. 631-687.

TEUBNER, G., «Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten», *Archiv Für Die Civilistische Praxis*, 218(2/4), 2018, pp. 7-8.

THOMPSON, D. F., «Designing responsibility: The problem of many hands in complex organizations», en VAN DEN HOVEN, J./ MILLER, S./ POGGE, T. (Eds.), *Designing in Ethics*, Cambridge University Press, 2017, pp. 32-56.

TIEDEMANN, K., *Tatbestandsfunktionen im Nebenstrafrecht: untersuchungen zu einem rechtsstaatlichen Tatbestandsbegriff entwickelt am Problem des Wirtschaftsstrafrechts*, Mohr, Tübingen, 1969.

TIEDEMANN, K., *Manual de Derecho penal económico: parte general y especial*, Tirant lo Blanch, Valencia, 2010.

TOMÁS-VALIENTE LANUZA, C., «Deberes positivos del Estado y Derecho penal en la jurisprudencia del TEDH», *InDret*, 3/2016, pp. 1-72.

TORÍO LÓPEZ, A., «Cursos causales no verificables en Derecho penal», en *ADPCP*, 1983, pp. 221-238.

— «Los delitos del peligro hipotético (Contribución al estudio diferencial de los delitos de peligro abstracto)», *ADPCP*, 33, 1991, pp. 825-848.

TRETTER, M./SAMHAMMER, D./DABROCK, P., «Künstliche Intelligenz in der Medizin: Von Entlastungen und neuen Anforderungen im ärztlichen Handeln», *Ethik in der Medizin*, 36, 2024, pp. 7-29.

TULA MOLINA, F., «Hartmut Rosa (2016), Alienación y aceleración: hacia una teoría crítica de la temporalidad en la modernidad tardía, Buenos Aires, Katz, 192 pp.», *Redes: Revista de estudios sociales de la ciencia y la tecnología*, 46/2018, p. 155-160.

TURÉGANO, I., «Los valores detrás de la privacidad», *Doxa,* 43, 2020.
TURING, A., «Computing machinery and intelligence», *Mind*, 59, 236, 1950, pp. 433-460.
TURNER, J., *Robot rules: Regulating artificial intelligence*, Springer, 2019.
VALERIUS, B., «Strafrechtliche Grenzen lernender Künstlicher Intelligenz», *GA*, 2022, 121, 3, 2022, pp.
— «Sorgfaltspflichten beim autonomen Fahren», en Hilgendorf, E. (Coord.), *Autonome Systeme und neue Mobilität*, Nomos, Baden-Baden, 2017, pp. 9-22.
VALLESPÍN PÉREZ, D., «Responsabilidad civil extracontractual en materia de IA: especial referencia a la carga de la prueba y la aplicación de presunciones», *IDP (Revista de Internet y política)*, 42, 2025, pp. 1-11.
VALLOR, S./VIERKANT, T., "Find the Gap: AI, Responsible Agency and Vulnerability", *Minds and Machines*, 34, 2024.
VALLS PRIETO, J., *Inteligencia artificial, Derechos humanos y bienes jurídicos*, Thomson Reuters Aranzadi, Cizur Menor, Navarra, 2021.
— «Sobre la responsabilidad penal por la utilización de sistemas inteligentes», *Revista electrónica de ciencia penal y criminología*, 24-27, 2022, pp. 1-35.
— «La inteligencia artificial frente a la colectividad. Una protección de los bienes jurídicos colectivos frente al auge de los sistemas inteligentes», *Estudios Penales y Criminológicos*, 44, 2023, pp. 1-27.
VAN DE POEL, I./NIHLÉN FAHLQUIST, J./DOORN, N./ZWART, S./ROYAKKERS, L., «The problem of many hands: Climate change as an example», *Science and Engineering Ethics*, 18(1), 2012, pp. 49-67.
VAN DER HOF, S. S./GEORGIEVA, I./SCHERMER, B./KOOPS, B. J. (Eds.), *Sweetie 2.0: Using Artificial Intelligence to Fight Webcam Child Sex Tourism*, Springer, Berlin, 2019.
VAROUFAKIS, Y., *Tecnofeudalismo: El sigiloso sucesor del capitalismo*, Ediciones Deusto, Barcelona, 2024.
VEALE, M./BINNS, R., «Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data», *Big Data & Society*, July–December, 2017, pp. 1-17.
VEALE, M./ZUIDERVEEN BORGESIUS, F., «Demystifying the Draft EU Artificial Intelligence Act», *Computer Law Review International*, 4/2021, pp. 97-112.
VEGA, L. DE, *Fuente Ovejuna*, edición de Rinaldo Froldi, apéndice de Abraham Madroñal, 19.ª ed., Espasa-Calpe, Madrid, Colección Austral, 1998.
VÉLIZ, C., *Privacy is power: Why and how you Should take back control of your data*, Bantam Press, Londres, 2020.
VILLAVERDE MENÉNDEZ, I., «Protección de datos personales, derecho a ser informado y autodeterminación informativa del individuo. A propósito de la STC 254/1993», *Revista Española de Derecho Constitucional*, 41, 1994.
— Principio de taxatividad. Una reflexión jurisprudencial», en PÉREZ MANZANO, M./LASCURAÍN SÁNCHEZ, J. A. (Dirs.), *La tutela multinivel del principio de legalidad penal*, Marcial Pons, 2016, pp. 81-116.

VOGEL, J., «Verbraucherschutz durch strafrechtliche Produkthaftung — Kriminologische und funktionale Aspekte», *GA*, 1990.

VOJTUŠ, F./KORDÍK, M./DRAŽOVÁ, P., «Artificial Intelligence and the criminal responsibility-challenges, obstacles and possible solutions», *2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2022.

VON HIRSCH, A., «Extending the harm principle: 'remote' harms and fair imputation», en A. P. SIMESTER/A. T. H. SMITH (eds.), *Harm and Culpability*, Oxford University Press, 1996, pp. 259-276.

VON WRIGHT, G. H., *Of Human Freedom*, en VON WRIGHT, G. H., *In The Shadow of Descartes. Essays in the Philosophy of Mind*, Kluwer Academic Publishers, Dordrecht/Boston/Londres,1998.

WAGNER, B., «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems», *Policy & Internet*, 11, 1, 2019, pp. 104-122.

WAGNER, G., «Liability rules for the Digital Age –Aiming for the Brussels Effect–», *Journal of European Tort Law*, 13 (3), 2022, p. 196.

WEIZENBAUM, J., *Computer power and human reason: From judgment to calculation*, W. H. Freeman & Co, San Francisco, 1976.

WELLS, C., *Corporations and criminal responsibility*, Oxford University Press, 2001.

WILDE, O., *De Profundis*, Trad. Maritza Izquierdo, Verbum, Madrid, 2020.

WILLIAMS, A. P., «Defining autonomy in systems: challenges and solutions», en WILLIAMS, A. P./SCHARRE, P. D., *Autonomous Systems Issues for Defence Policymakers*, HQ SACT, 2015, pp. 27-62.

WILLIAMS, G., *Criminal Law. The General Part*, Stevens & Sons, London, 1953.

WOLF, S., «The legal an Moral Responsiblity of Organisations», en PERNOCK, J./CHAPMAN, J. (eds.), *Criminal Justice*, New York University Press, Nueva York, 1985.

WOOLLEY, S. C./HOWARD, P. N., *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, Oxford University Press, Nueva York, 2019.

XENIDIS, R., «Tuning EU equality law to algorithmic discrimination: Three pathways to resilience», *Maastricht Journal of European and Comparative Law*, 27, 2020, pp. 736-758.

YUSTE, R./GENSER, J./HERMANN, S., «It's time for Neuro-rights», Horizons, Center for International Relations and Sustainable Development, 18, 2021.

ZARKADAKIS, G., *In Our Image: Will Artificial Intelligence Save or Destroy Us?*, Rider, Londres, 2015.

ZARSKY, T. Z., «Understanding Discrimination in the Scored Society», *Wash. L. Rev.*, 89, 2014, pp. 1375-1385.

ZHAO, S., *Principle of criminal imputation for negligence crime involving artificial intelligence*, Springer, Singapur, 2024.

ZIELINSKY, D., *Handlungs-und Erfolgsunwert im Unrechtsbegriff*, Duncker-Humblot, Berlín, 1973.

ZILLER, J., «El Convenio del Consejo de Europa de inteligencia artificial frente al Reglamento de la Unión Europea: dos instrumentos jurídicos muy diversos», en COTINO HUESO, L./SIMÓN CASTELLANO, P., (Dirs.), *Tratado sobre el Reglamento de Inteligencia artificial de la Unión Europea*, Aranzadi, Madrid, 2024, pp. 67-84.

ZLIOBAITE, I., «Fairness-aware machine learning: a perspective», 2017, disponible en arXiv:1708.00754.

ZUBOFF, S., *The age of surveillance capitalism. The fight for a human future at the new frontier of power*, Profile Books, Londres, 2019.

ZUGALDÍA ESPINAR, J. M., «¿Qué queda en pie del Derecho penal del principio mínima intervención, máximas garantías?», *Cuadernos de Política Criminal*, 79, 2003, pp. 109-124.

# ANEXO III. REFERENCIAS NORMATIVAS

## Normativa internacional


Convenio para la Protección de los Derechos Humanos y las Libertades, Consejo de Europa, hecho en Roma el 4 de noviembre de 1950

Convención sobre Ciertas Armas Convencionales (CCW), Naciones Unidas, hecha en Ginebra el 10 de octubre de 1980

Convenio para la protección del Mar Mediterráneo contra la contaminación (Convenio de Barcelona), Programa de las Naciones Unidas para el Medio Ambiente, hecho en Barcelona el 16 de febrero de 1976. Actualizado como Convenio para la protección del medio marino y de la región costera del Mediterráneo (1995)

Convenio para la protección del medio marino del Atlántico nordeste, OSPAR Commission, hecho en París el 22 de septiembre de 1992

Convenio sobre la protección del medio marino de la zona del Mar Báltico (Helsinki, de 24 de marzo de 1992.

Convenio sobre la ciberdelincuencia (Convenio de Budapest), Consejo de Europa, hecho en Budapest el 23 de noviembre de 2001.

Recomendación CM/Rec(2018)1 del Comité de Ministros a los Estados miembros sobre el pluralismo mediático y la transparencia en la propiedad de los medios, adoptada en Estrasburgo el 7 de marzo de 2018

Convenio Marco sobre inteligencia artificial, Derechos Humanos, Democracia y Estado de Derecho, Consejo de Europa, hecho en Estrasburgo, 2024


## Normativa de la Unión Europea


Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, sobre responsabilidad por productos defectuosos

Comunicación de la Comisión sobre el recurso al principio de precaución, COM(2000) 1 final, de 2 de febrero de 2000

Decisión Marco del Consejo, de 13 de junio de 2002, relativa a la orden de detención europea y a los procedimientos de entrega entre Estados miembros (2002/584/JAI)

Directiva 2009/48/CE del Parlamento Europeo y del Consejo, de 18 de junio de 2009, sobre la seguridad de los juguetes

Tratado de la Unión Europea y Tratado de Funcionamiento de la Unión Europea. Versiones consolidadas. DOUE. núm. 83, de 30 de marzo de 2010

Reglamento (UE) n.° 1025/2012 del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, sobre la normalización europea

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD)

Directiva 2016/681 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa al uso de datos del registro de nombres de pasajeros (PNR)

Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL))

Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales (2016/2225(INI))

Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre derechos de autor en el mercado único digital

Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre ENISA y la ciberseguridad de la UE, y por el que se establece la certificación de ciberseguridad de la Unión

Directiva (UE) 2019/770 del Parlamento Europeo y del Consejo de 20 de mayo de 2019 relativa a determinados aspectos de los contratos de suministro de contenidos y servicios digitales

Libro blanco sobre la inteligencia artificial, COM(2020) 65 final, de 19 de febrero de 2020

Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre el régimen de responsabilidad civil para la inteligencia artificial (2020/2012(INL)).

Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre inteligencia artificial, COM/2021/206 final, de 21 de abril de 2021

Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital (2020/2266(INI)

Resolución del Parlamento de 2021 (Resolución del Parlamento Europeo sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI))

Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales)

Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 sobre mercados disputables y equitativos en el sector digital

Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre ciberseguridad (NIS2)

Propuesta de Directiva del Parlamento Europeo y del Consejo, de 28 de septiembre de 2022, relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial

Reglamento (UE) 2023/988 del Parlamento Europeo y del Consejo, de 5 de junio de 2023, sobre la seguridad general de los productos. Diario Oficial de la Unión Europea L 147, 9 de junio de 2023, pp. 1-39.

Directiva (UE) 2024/1385 del Parlamento Europeo y del Consejo, de 14 de mayo de 2024, sobre la lucha contra la violencia hacia mujeres y la violencia doméstica

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, relativo a la inteligencia artificial

Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos

Decisión (EU) 2024/2218 del Consejo, sobre la firma del Convenio Marco de Inteligencia Artificial

Comisión Europea (19 de noviembre de 2025). *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se modifican los Reglamentos (UE) 2024/1689 y (UE) 2018/1139 en lo que respecta a la simplificación de la aplicación de determinadas normas digitales de la Unión*, COM(2025) 836, Bruselas.

## Normativa española

Ley 25/1964, de 29 de abril, sobre energía nuclear (BOE núm. 107, de 4 de mayo 1964)

Constitución española (BOE núm. 311, de 29 de diciembre de 1978)

Ley Orgánica 5/1992, de 29 de octubre, de regulación del tratamiento automatizado de los datos de carácter personal (BOE núm. 261, de 31 de octubre de 1992)

Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal (BOE núm. 281, de 24 de noviembre de 1995)

RD 1464/1999, de 17 de septiembre, sobre actividades de la primera parte del ciclo de combustible nuclear (BOE núm. 238, de 5 de octubre de 1999)

Real Decreto 783/2001, de 6 de julio, por el que se aprueba el Reglamento sobre protección sanitaria contra radiaciones ionizantes (BOE núm. 178, de 26 de julio de 2001)

Ley Orgánica 15/2003, de 25 de noviembre, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal (BOE núm. 283, de 26 de noviembre de 2003)

Ley 9/2003, de 25 de abril, sobre el régimen jurídico de la utilización confinada, liberación voluntaria y comercialización de organismos modificados genéticamente (BOE núm. 99, de 25 de abril de 2003)

Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias (BOE núm. 287, de 30 de noviembre de 2007)

Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la administración electrónica (BOE núm. 25, de 29 de enero de 2010)

Ley Orgánica 5/2010, de 22 de junio, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal (BOE núm. 152, de 23 de junio de 2010)

Instrumento de Ratificación del Convenio sobre la Ciberdelincuencia, hecho en Budapest el 23 de noviembre de 2001 (BOE núm. 226, de 17 de septiembre de 2010)

Real Decreto 97/2014, de 14 de febrero, por el que se regulan los procedimientos de transporte terrestre de mercancías peligrosas (BOE núm. 41, de 17 de febrero de 2014).

Real Decreto 102/2014, de 21 de febrero, para la gestión responsable y segura del combustible nuclear gastado y los residuos radiactivos (BOE núm. 47, de 24 de febrero de 2014).

Ley Orgánica 1/2015, de 30 de marzo, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal (BOE núm. 77, de 31 de marzo de 2015)

Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (BOE núm. 236, de 2 de octubre de 2015

Ley Orgánica 1/2019, de 20 de febrero, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal (BOE núm. 45, de 21 de febrero de 2019)

Carta de Derechos Digitales en España, publicada en 2021

Ley 29/2022, de 30 de diciembre, de fomento del ecosistema de las empresas emergentes (BOE núm. 314, de 31 de diciembre de 2022)

Ley 18/2022, de 28 de septiembre, de creación y crecimiento de empresas (BOE núm. 233, de 29 de septiembre de 2022)

Proposición de Ley Orgánica (122/000011) de regulación de las simulaciones de imágenes y voces generadas por inteligencia artificial, de 13 de octubre de 2023

Real Decreto 817/2023, de 28 de noviembre, sobre entornos controlados de pruebas (*sandboxes*) (BOE núm. 284, de 29 de noviembre de 2023)

Anteproyecto de Ley para el buen uso y la gobernanza de la inteligencia artificial (aprobado por el Consejo de Ministros el 11 de marzo de 2025)

Proyecto de Ley Orgánica para la protección integral de las personas menores de edad en entornos digitales (aprobado por Consejo de Ministros el 21 de marzo de 2025)

Anteproyecto de Ley Orgánica de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen (aprobado por Consejo de Ministros el 13 de enero de 2026).

## Normativa extranjera

Bundes-Immissionsschutzgesetz (Ley de protección contra los efectos nocivos para el medio ambiente causados por la contaminación atmosférica, el ruido, las vibraciones y procesos similares, BImSchG), hecha en Bonn el 15 de marzo de 1974

*Strafprozeßordnung (StPO*), Código de Procedimiento Penal en la versión publicada el 7 de abril de 1987

Dangerous Dogs Act (Ley de Perros Peligrosos de Reino Unido), hecha en Londres el 12 de agosto de 1991

*StGB,* Código Penal alemán, en la versión publicada el 13 de noviembre de 1998 modificado por última vez por el artículo 2 de la Ley de 22 de noviembre de 2021

Decreto legislativo italiano de 8 de junio de 2001, n. 231 («Reglamentación de la responsabilidad administrativa de las personas jurídicas, de las sociedades y de las asociaciones incluso carentes de personalidad jurídica, según el artículo 11 de la ley n.º 300 de 29 de septiembre de 2000»), hecho en Roma el 8 de junio de 2001. Gazzetta Ufficiale, 19 de junio de 2001.

Dodd-Frank Wall Street Reform and Consumer Protection Act (Estados Unidos), hecha en Washington D. C. el 21 de julio de 2010

DoD Directive 3000.09 (Directiva del Departamento de Defensa de los Estados Unidos, Autonomy in Weapon Systems), hecha en Washington D. C. el 21 de noviembre de 2012

Propuesta de modificación del Código Penal de Singapur (Penal Code Review Committee Report), hecha en Singapur en agosto de 2018. Ministry of Home Affairs

Loi n° 2019-222 du 23 mars 2019 («de programmation 2018-2022 et de réforme pour la justice»), hecha en París el 23 de marzo de 2019. Journal Officiel de la République Française, 24 de marzo de 2019

Legge 23 settembre 2025, n. 132: Disposizioni e deleghe al Governo in materia di intelligenza artificiale. Gazzetta Ufficiale, n. 223.

# ANEXO IV. SENTENCIAS Y RESOLUCIONES

## JURISPRUDENCIA ESPAÑOLA

### Tribunal Constitucional

STC 25/1981, de 14 de julio
STC 62/1983, de 11 de julio
STC 120/1990, de 27 de junio
STC 127/1990, de 5 de julio
STC 254/1993, de 20 de julio
STC 161/1997, de 2 de octubre
STC 111/1999, de 14 de junio
STC 87/2001, de 2 de abril
STC 24/2004, de 24 de febrero
STC 17/2017 de 2 de febrero
STC 76/2019, de 22 de mayo
STC 27/2020, de 24 de febrero

### Sentencias del Tribunal Supremo

STS (Sala 2.ª), de 23 de noviembre de 1970, no disponible en repositorios digitales, citada en Boldova Pasamar, M. A./ Alastuey Dobón, C. (Coords.), *Tratado de las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023 (nota al pie 42)

STS (Sala 2.ª), de 13 de febrero de 1987, no disponible en repositorios digitales, citada en Boldova Pasamar, M. A./ Alastuey Dobón, C. (Coords.), *Tratado de las*

*consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023 (nota al pie 42)
STS, de 12 de noviembre de 1985 (ROJ: 1046/1985), ponente: Martín Jesús Rodríguez López
STS (Sala 2.ª), de 14 de septiembre de 1990, no disponible en repositorios digitales, citada en Boldova Pasamar, M. A./ Alastuey Dobón, C. (Coords.), *Tratado de las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023 (nota al pie 42)
STS (Sala 2.ª), 2521/1992, de 24 de noviembre, no disponible en repositorios digitales, citada en Boldova Pasamar, M. A./ Alastuey Dobón, C. (Coords.), *Tratado de las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023 (nota al pie 42)
STS 3654/1992, de 23 de abril
STS (Sala 2.ª), 2063/1994, de 28 de noviembre, no disponible en repositorios digitales, citada en Boldova Pasamar, M. A./ Alastuey Dobón, C. (Coords.), *Tratado de las consecuencias jurídicas del delito*, Tirant Lo Blanch, Valencia, 2.ª ed., 2023 (nota al pie 42)
STS 895/1997, de 26 de septiembre
STS 492/1997, de 15 de abril
AT 812/1998, de 11 de marzo
STS 1237/1998, de 24 de octubre
STS 538/2000, de 25 de abril
STS 30/2001, de 17 de enero
STS 1210/2001, de 11 de junio
STS 2139/2001, de 15 de noviembre
STS 349/2002, de 22 de febrero
STS 753/2002, de 26 de abril
STS 1442/2002, de 14 de septiembre
STS 1757/2003, de 2 de enero
STS 653/2004, de 24 de mayo
STS 969/2004, de 29 de julio
STS 63/2006, de 31 de enero
STS 136/2006, de 15 de febrero
STS 1007/2006, de 10 de octubre
STS 569/2007, de 29 de junio
STS 828/2009, de 13 de julio
STS 1136/2009, de 4 de noviembre
STS 611/2010, de 21 de junio
STS 1117/2011, de 31 de octubre
STS 626/2012, de 17 de julio
STS 510/2014, de 23 de junio
STS 695/2016, de 28 de julio

STS 419/2017, de 8 de junio (citada en el cuerpo del texto, por cita literal, por el número ROJ: STS 2315/2017)
STS 622/2017, de 19 de septiembre
STS 398/2018, de 12 de septiembre
STS 668/2019, de 14 de enero
STS 539/2019, de 28 de noviembre
STS 481/2020, de 29 de septiembre
STS 679/2020, de 11 de diciembre
ATS 457/2021, de 3 de junio
STS 364/2022, de 8 de abril
STS 623/2022, de 22 de junio
STS 844/2022, de 26 de octubre

## Otras

Sentencia de la Audiencia Provincial de Zaragoza de 10 de noviembre de 1993, no disponible en repositorios digitales, citada en Muñoz Conde, F., *Derecho Penal. Parte especial, 25.ª Ed, revisada y puesta al día con la colaboración de Carmen López Peregrín*, Tirant Lo Blanch, Valencia, 2023, p. 644
SAP Zaragoza (sección 1.ª) 117/1998, de 11 de marzo
SAP Asturias (sección 3.ª) 504/1998, de 10 de noviembre
SAP Baleares (sección 1.ª) 122/2000, de 29 de septiembre
SAP Cuenca (sección 1.ª) 55/2001 de 26 de mayo
Auto de la Audiencia Provincial de Cádiz de 31 de mayo de 2004, núm. de recurso 32/2004
SAP Madrid (sección 5.ª) 43/06, de 16 de mayo
Sentencia de la Audiencia Nacional (Sala de lo Penal) 11/2007, de 28 de marzo
Sentencia de la Audiencia Nacional (Sala de lo Contencioso) 2350/ 2020, de 30 de septiembre de 2020
Sentencia del Juzgado Central de lo Contencioso-Administrativo n.º 8, 143/2021, de 30 de diciembre

## Sentencias de tribunales extranjeros

Sentencia del Tribunal Constitucional Federal Alemán BVerfG 1958, 869 (caso Lüth)
United Kingdom House of Lords, Tesco Supermarkets Ltd v. Nattrass [1972] A. C. 153 (31 de marzo de 1971)
Sentencia del Tribunal Constitucional Federal Alemán BVerfG NJW 1979 (caso Kalkar)

Sentencia del Tribunal Supremo Federal Alemán (BGHSt) 37, 106, de 6 de julio de 1990 (Caso del spray de cuero)
United Kingdom Court of Appeal (Criminal Division), R v P&O European Ferries (Dover) Ltd, 1991, 93 Criminal Appeal Reports 72
Sentencia del Tribunal Supremo Federal Alemán (BGHSt) 1995, 590 (Caso *Holzschutzmittel*)
Sentencia del Tribunal de Justicia de las Comunidades Europeas, STJCE 1998/80, de 5 de mayo de 1998 (Caso de las «vacas locas»)
United States Supreme Court of Wisconsin, State v. Loomis, 881 N. W.2d 749 (Wis. 2016)
United States Court of Appeals, United States v. Michael Coscia, 866 F.3d 782, 786 (7th Cir. 2017)
United States Supreme Court, Packingham v. North Carolina, 582 U. S. 98; 137 S. Ct. 1730 (19 de junio de 2017)
Sentencia del Tribunal de Bolonia, n. 2949/2020, de 31 de diciembre de 2020
Sentencia del Tribunal Europeo de Derechos Humanos**,** Sentencia de 25 de mayo de 2021. Caso Big Brother Watch y otros contra Reino Unido
Tribunal Constitucional Federal Alemán (Bundesverfassungsgericht). Sentencia de 16 de febrero de 2023. Expedientes 1 BvR 1547/19 y 1 BvR 2634/20
Sentencia del Tribunal de Justicia de la Unión Europea, asunto C-817/19, Ligue des droits humains, de 21 de junio de 2022
Sheriff Court de Glasgow (Escocia), Caso HMA v. Margaret Ferrier, de 13 de septiembre de 2022
Sentencia del Tribunal Europeo de Derechos Humanos, caso Glukhin contra Rusia, núm. 11519/20, de 4 de julio de 2023
Sentencia del Tribunal de Justicia de la Unión Europea, caso Schufa, asunto C-634/21, de 7 de diciembre de 2023

## Colección
## Derecho Penal y Procesal Penal

*Director:*
**Luis Rodríguez Ramos**

### Últimos títulos publicados:

25. Cibercrimen: Tendencias y desafíos actuales
    **Coordinadores: Marina Mínguez Rosique, David Gallego Arribas**
26. El delito de trato degradante discriminatorio
    **David Miras Estévez**
27. Verdes y justas: responsabilidad penal y diligencia debida en las organizaciones multinacionales
    **Directores: Jacobo Dopico Gómez-Aller, Marta Muñoz de Morales Romero, Adán Nieto Martín**
    **Coordinador: Luis Miguel Vioque Galiana**
28. Ideología y proceso en los Anteproyectos de Ley de Enjuiciamiento de Falange (1938)
    **Brian Buchhalter Montero**
29. Las agresiones sexuales cualificadas a personas en edad legal de consentimiento sexual
    **María Marta González Tascón**
30. Fraude Carrusel, responsabilidad penal de las personas jurídicas y *Compliance*
    **Miguel Cáceres Casado**
31. La arquitectura del concurso de delitos. Problemas prácticos y fundamentos dogmáticos de su regulación
    **Nicolás Cantard Vanni**
32. La desinformación en la sociedad digital. Aspectos criminológicos, dogmáticos y de política criminal
    **Abraham Devís Matamoros**
33. Libro homenaje al Profesor Enrique Peñaranda Ramos
    **Juan Alberto Díaz López et al.** (coordinadores)
34. El sistema de consecuencias jurídicas del delito a examen
    **Marina Mínguez Rosique y David Gallego Arribas** (coordinadores)

La investigación analiza los problemas de imputación penal que plantean los sistemas de Inteligencia Artificial (IA), entendidos a partir de su autonomía operativa como rasgo distintivo. Examina las lagunas que su imprevisibilidad y falta de transparencia generan en la causalidad y en la atribución de culpa. Se descartan los principales modelos doctrinales: la imputación directa a la máquina, por contradecir principios de control, conocimiento y dignidad; la imputación directa a la persona jurídica, por aproximarse a esquemas de responsabilidad objetiva basados en la asunción de riesgos; y la responsabilidad penal por el producto, por resultar insuficiente en causalidad, imputación subjetiva y objetiva, especialmente ante múltiples concausas desconocidas.

En este libro se estudian los peligros de la sociedad del riesgo algorítmico, que intensifica problemas del moderno derecho penal. Aunque los bienes jurídicos afectados son tradicionales, surgen nuevas formas de ataque colectivas y difusas derivadas de la opacidad y capacidad inferencial de estos sistemas. Se propone un modelo de nuevos delitos de peligro abstracto de idoneidad, que evite interpretaciones formalistas y exija comprobar la aptitud del hecho concreto según las circunstancias. La propuesta se basa en la organización del propio ámbito, considerando la alta especialización y compartimentalización en el diseño de IA. Finalmente, se analiza la influencia del Reglamento de Inteligencia Artificial de la UE como marco del riesgo permitido, identificando obligaciones vinculadas a la protección penal y formulando una propuesta de *lege ferenda*.


## Andrea Bravo Bolado

Andrea Bravo Bolado es graduada en Derecho por la Universidad Autónoma de Madrid. Máster de abogacía y Máster en Derecho penal, ambos por la Universidad Autónoma de Madrid. Doctora en Derecho penal por la UAM (con beca FPU otorgada por el Ministerio de Universidades). Ha impartido clases en el Grado en Derecho y en Derecho y Administración y Dirección de Empresas de la UAM, en el Máster de Derecho penal de la UAM, así como en el Máster de Contratación digital e inteligencia artificial de la UAM. Ha sido profesora invitada en el *Honours level Criminal Responsibility course,* degree of LLB Law, University of Glasgow. Ha realizado estancias doctorales en las universidades de Glasgow, Bolonia, y Würzburg.