

IUS CRIMINALE

BOLETÍN DE DERECHO PENAL

Fecha: agosto 2026

Edición: 11 - Volumen 2

FISCALÍA GENERAL DEL ESTADO

Dirección de Estudios Penales

QUITO - ECUADOR



ENTES FICTICIOS E IRRUPCIÓN TECNOLÓGICA

DEBATE JURÍDICO SOBRE EL DERECHO PENAL FRENTE
A LAS NUEVAS REALIDADES



FISCALÍA GENERAL DEL ESTADO



ECUADOR

Boletín de Derecho Penal IUS CRIMINALE

FISCALÍA GENERAL DEL ESTADO

COMITÉ EDITORIAL

Dr. Carlos Leonardo Alarcón Argudo

Fiscal General del Estado (E)

Mtr. Ximena Coello Martínez

Coordinadora General de Gestión del Conocimiento

Abg. Paula Cisneros Espinosa

Directora de Estudios Penales

COMITÉ ACADÉMICO

Dirección de Estudios Penales

EQUIPO DE DISEÑO EDITORIAL ACADÉMICO

Dirección de Comunicación y Promoción Institucional

Mgs. Amelina Espinosa

Mgs. Andrés Lasso Ruiz

Quito, agosto de 2026

Contenido de acceso libre.

Los criterios vertidos por los autores no comprometen la opinión institucional.

Todos los derechos reservados. Prohibida la reproducción total o parcial, sin autorización de los autores.

IUS CRIMINALE
B O L E T Í N D E D E R E C H O P E N A L

ÍNDICE

Los retos de la IA para la teoría del delito:
¿unDerechopenaldigital?.....

1

Autómatas y avatares en el proceso penal:
La inteligencia artificial agéntica y el análisis policial forense.....

49

La incidencia de la Inteligencia Artificial en los derechos de defensa
y la presunción de inocencia.....

79



LOS RETOS DE LA IA PARA LA TEORÍA DEL DELITO: ¿UN DERECHO PENAL DIGITAL?

Prof. Dra. Raquel Roso Cañadillas²

Resumen

La irrupción de la inteligencia artificial (IA) y la entrada en vigor progresiva del Reglamento de IA (RIA) impulsan este estudio, orientado a analizar sus implicaciones en el Derecho penal y a explorar la transición desde un modelo analógico hacia un auténtico Derecho penal digital. En una primera parte, se examina sintéticamente el RIA, destacando conceptos centrales como el enfoque basado en riesgos, la transparencia, la explicabilidad algorítmica y la necesidad de una supervisión humana efectiva. Estas notas permiten establecer un puente con la segunda parte del trabajo, en la que se revisan los elementos esenciales de la teoría del delito: causalidad, imputación objetiva, imprudencia, autoría y participación... para evaluar cómo la IA desafía sus parámetros tradicionales y exige su adaptación. El análisis busca proponer ajustes dogmáticos que contribuyan a la construcción de un Derecho penal digital sólido, capaz de cumplir su función preventiva y su función de garantía, especialmente en relación con la protección de los bienes jurídicos y los derechos fundamentales. Asimismo, se abordan los desafíos relativos a la responsabilidad de sistemas autónomos y la posible consideración de "personas electrónicas" (e-personas), proponiendo un esbozo de un sistema integral de atribución de responsabilidad que preserve la supervisión humana y articule programas de cumplimiento adecuados. En conjunto, el trabajo ofrece una reflexión sobre cómo el Derecho penal debe evolucionar ante la disrupción tecnológica de la IA, garantizando un marco normativo ponderado, transparente y respetuoso con los principios fundamentales del orden jurídico.

Palabras clave:

Inteligencia artificial; Reglamento de inteligencia artificial; Derecho penal sólido; Teoría del delito; Principio de supervisión humana; Riesgo permitido; Derechos Humanos; Causalidad probabilística; Opacidad epistémica; Previsibilidad; Garantes de la cadena de valor; Personalidad jurídica de la persona electrónica; Programas de cumplimiento.

¹ El presente trabajo es una versión revisada y actualizada de un artículo elaborado en otoño de 2024 y publicado a principios de 2025 en la revista *En Letra: Derecho penal*. Desde entonces, la producción bibliográfica sobre la materia ha crecido de manera casi inabarcable y los avances tecnológicos vinculados al desarrollo y uso de la IA se siguen sucediendo a un ritmo vertiginoso. Aunque ha resultado imposible recoger y citar exhaustivamente todas las aportaciones recientes, el trabajo ha sido actualizado en la mayor medida posible.

Este trabajo se encuadra dentro del Proyecto de Investigación "El Derecho penal del trabajo en la cuarta revolución industrial: aparición de nuevas conductas delictivas y perpetuación de las ya existentes" PID2024-157227NB-I00, Proyectos de generación del conocimiento 2024, financiado por el Ministerio de Ciencia, Innovación y Universidades y Fondos Europeos, del que soy IP junto con Ricardo Gutiérrez Aguilar.

² Profesora titular de Derecho penal (catedrática acreditada). Universidad de Alcalá, España.

Abstract

The emergence of artificial intelligence (AI) and the progressive entry into force of the AI Regulation (AIR) drive this study, which aims to analyze its implications for criminal law and to explore the transition from an analog model to a genuinely digital criminal law framework. The first part offers a concise examination of the AIR, highlighting key concepts such as the risk-based approach, transparency, algorithmic explainability, and the need for effective human supervision, which in turn provide a bridge to the second part of the work, where the essential components of criminal theory—causation, objective imputation, negligence, authorship, and participation—are reviewed to assess how AI challenges traditional parameters and demands their adaptation. The analysis seeks to propose dogmatic adjustments that contribute to the construction of a solid digital criminal law capable of fulfilling its preventive and protective functions, particularly with respect to safeguarding legal interests and fundamental rights. The study also addresses the challenges associated with the responsibility of autonomous systems and the possible consideration of “electronic persons” (e-persons), outlining a preliminary model for a comprehensive system of liability allocation that preserves human supervision and incorporates appropriate compliance programs. Overall, the work reflects on how criminal law must evolve in the face of AI-driven technological disruption, ensuring a balanced, transparent, and principled normative framework consistent with the foundational values of the legal order.

Keywords:

Artificial intelligence, AI Act, hard criminal law, crime theory, human in the loop (HILP), legal risk, human rights, probabilistic causation, epistemic opacity, predictability, value chain guarantors, legal personality of electronic person, compliance programmes

Sumario: I. Introducción; II. La especialidad de la norma y su contexto, nunca experimentado por la humanidad; III. La estratificación del riesgo como eje fundamental del RIA; IV. La responsabilidad por incumplimiento contenida en el RIA; V. El Derecho penal como *hard law* en la implementación y uso responsable de la IA; 1. Introducción; 2. Reflexiones aproximativas, provisionales y deliberadamente abiertas sobre el Derecho penal de la IA; 2.1 El Derecho penal y su función de protección de bienes jurídicos: más actual que nunca; 2.2 Los problemas causales en el uso de la IA; 2.3 La íntima relación entre la imputación objetiva, la imprudencia y el riesgo permitido replicada en el uso de la IA; 2.4 La imprevisibilidad de los resultados de la IA y la imprudencia; 2.5 Autoría y responsabilidad penal de la persona física en el uso de sistemas de IA; 2.6 ¿La persona electrónica (*e-person*) como sujeto activo del delito?; 2.7 Los sujetos obligados del RIA como garantes de la cadena de valor; 2.8 El *compliance* penal de la IA: ¿como instrumento o como persona electrónica?; 2.9 Revisión de algunos delitos en la PE; VI. ¿Hacia un Derecho penal digital?; VII. Conclusión; VIII. Bibliografía.

I. Introducción

Hace ya más de dos años me planteaba la siguiente pregunta: ¿*Un Derecho penal delictuosa en una sociedad líquida?*³ De aquella inquietud surgió un trabajo con el mismo título, en el que ofrecía algunas reflexiones sobre la deriva de la sociedad contemporánea y sus repercusiones en el derecho, en particular en el Derecho penal y, en un breve apartado, en el Derecho penal del trabajo. El objetivo entonces era enmarcar la situación y esbozar el papel que podría desempeñar el Derecho penal en una sociedad cada vez más líquida, tal y como la bautizó Zygmunt Bauman⁴, y sometida a transformaciones tecnológicas de ritmo vertiginoso. En aquel texto concluía que el Derecho penal seguiría constituyendo un bastión de solidez en medio de la incertidumbre que vivimos. Confío en que así sea y que continúe ofreciendo un punto de anclaje firme frente a un proceso de innovación acelerada que apenas concede tiempo razonable para su adecuada asimilación por parte de la mente humana.

Estamos inmersos en pleno proceso de la cuarta revolución industrial, marcada por el desarrollo imparable del algoritmo y la inteligencia artificial (IA)⁵. Esta evolución podría abrir lo que algunos denominan la quinta revolución, industrial o revolución neurocientífica⁶, con un impacto potencialmente trascendente incluso para la existencia del ser humano. Según los expertos, la IA promete un beneficio social global, mejorando la asistencia sanitaria, haciendo el transporte más seguro y limpio, optimizando servicios públicos y generando productos innovadores especialmente en el ámbito de la energía y sanidad. La esfera empresarial también se beneficiará, con mayor productividad y procesos de fabricación más eficientes, mientras que el Estado podrá mejorar la gobernanza en áreas como el transporte, la energía y la gestión de residuos. Incluso se sugiere que la IA nos adentrará en el paraíso del 'no-trabajo' pesado, pasando del *homo faber* al *homo digitalis*. No obstante, este paraíso también presenta riesgos importantes que deben ser contemplados⁷.

3 Raquel Roso Cañadillas, "¿Un Derecho penal delictuosa en una sociedad líquida?", *Anuario de la Facultad de Derecho de la Universidad de Alcalá (UAH)*, n.º 26 (2023): 199-225; "¿Un Derecho penal delictuosa en una sociedad líquida? Algunas reflexiones sobre el Derecho penal en la sociedad posindustrial", *Revista General de Derecho Penal*, n.º 41 (2024): 1-30.

4 Zygmunt Bauman, *Modernidad líquida*, trad. Mirta Rosenberg y Jaime Arrambide Squirru (Ciudad de México: Fondo de Cultura Económica, 2002), 16.

5 Se habla habitualmente de una cuarta revolución industrial que, en mi opinión, se solapa con la anterior, vinculada a la transformación digital, la sociedad de la información y las energías renovables, entre otros ámbitos. Este nuevo estadio histórico, término acuñado por Klaus Schwab en su obra homónima y conocido también como *Industria 4.0* —traducción del concepto alemán *Industrie 4.0*—, se caracteriza, según el autor, por la fusión de tecnologías y la progresiva desintegración de las fronteras entre lo físico, lo digital y lo biológico. En este marco se incluyen la robótica, la inteligencia artificial, la nanotecnología, la computación cuántica, la biotecnología, el internet de las cosas, la impresión 3D y los vehículos autónomos (Klaus Schwab, *La cuarta revolución industrial* (Madrid: Marcial Pons, 2016), *passim*).

6 Rafael Yuste, "Neurotecnología y neuroderechos", en *Derecho y tecnologías*, dir. María Emilia Casas Baamonde, coord. Daniel Pérez del Prado (Madrid: Editorial Centro de Estudios Ramón Areces, 2025), 719, la bautiza en un reciente trabajo como revolución *neurotecnológica*. Esta revolución no es industrial, no es sobre la máquina, va mucho más allá. Supone un gran salto cualitativo cuyo núcleo es el estudio del funcionamiento del cerebro humano, una línea en la que ha trabajado intensamente el neurocientífico español Rafael Yuste, ideólogo del Proyecto *Brain* impulsado por la Administración Obama, orientado tanto a la cura de enfermedades mentales, como a la mejora de las capacidades humanas. Desde esta perspectiva, esta revolución nos situaría ante una especie de *transhumanismo*, con la aparición de seres híbridos, según sostiene el propio Yuste (cfr. Disponible en: https://www.eldiario.es/cantabria/rafael-yuste-neurobiologo-abogados-futuro-seres-humanos-hibridos_1_11481857.html). La evolución de esta revolución podría discurrir por dos vías: una primera, centrada en humanos híbridos portadores de chips destinados a mejorar su calidad de vida, anticipar enfermedades, corregir hábitos perjudiciales, procurar su estabilidad mental, sin alterar la identidad individual, sino potenciando su mejor versión; y una segunda, vinculada al tránsito de los actuales asistentes inteligentes, como *Google Maps*, *Google Lens*, *Alexa*, etc., hacia una inteligencia artificial general. Esta última constituiría una *superinteligencia* capaz de programarse a sí misma, superar las capacidades humanas y desarrollar inteligencia social y emocional, incluso mediante interconexión entre sistemas. En perspectiva histórica, estos avances serán probablemente considerados un periodo de transición hacia un estadio radicalmente nuevo de la evolución tecnológica.

7 La propia IA, en este caso Chat GPT (consultada 28 de septiembre de 2024), ha contestado a la pregunta de qué riesgos traerá a la humanidad el uso de la IA lo siguiente: "¡Hola! Es un tema muy interesante y relevante. La inteligencia artificial (IA) puede traer muchos beneficios, pero también presenta ciertos riesgos para la humanidad. Algunos de los principales riesgos incluyen: 1. Desempleo: A medida que la IA automatiza tareas, puede llevar a la pérdida de empleos en ciertos sectores, lo que podría generar desigualdad económica. 2. Privacidad: La recopilación masiva de datos por parte de sistemas de IA puede poner en riesgo la privacidad de las personas, ya que se pueden utilizar para monitorear y analizar comportamientos. 3. Sesgo y discriminación: Si los algoritmos de IA se entrenan con datos sesgados, pueden perpetuar o incluso amplificar prejuicios existentes, afectando a grupos vulnerables. 4. Seguridad: La IA puede ser utilizada en ciberataques o en la creación de armas autónomas, lo que plantea serias preocupaciones sobre la seguridad global. 5. Desinformación: La capacidad de generar contenido falso o manipulado puede dificultar la distinción entre la verdad y la ficción, afectando la opinión pública y la democracia. 6. Dependencia: A

En este escenario de aparente felicidad prometida y de progreso para la humanidad, según la mirada más optimista, puede ocultarse, sin embargo, una vertiginosa disolución de estructuras sólidas y de los anclajes tradicionales de la humanidad a la realidad, impulsada por un proceso de digitalización imparable que alcanza ya un nivel superior con software inteligentes capaces de tomar decisiones. En este contexto se enmarca el Reglamento Europeo de Inteligencia Artificial⁸ (en adelante RIA⁹), que ha entrado en vigor el 1 de agosto de 2024¹⁰. Actualmente no existe ninguna norma de similares características en el mundo, que regule de manera armonizada la implantación y uso de la IA. La UE aspira, con esta regulación pionera, a posicionarse como referente mundial y a establecer un estándar normativo internacional basado en tres conceptos claves: riesgo, garantías y responsabilidades. A la luz del análisis de esta norma y de los principios de actuación, surge la cuestión de cuál debe ser la aportación del Derecho penal, posiblemente armonizado internacionalmente en sus elementos esenciales, para contribuir a una implementación y uso responsable de la IA. Esta pregunta ha dado lugar al presente trabajo, el cual recorre la teoría del delito para identificar posibles adaptaciones y ajustes en las categorías jurídico-penales con el fin de avanzar en el diseño de un Derecho penal digital. Antes de abordar esta tarea, conviene señalar algunas características del RIA.

II. La especialidad de la norma y su contexto¹¹, nunca experimentado para la humanidad

El RIA es una norma de gran *complejidad*¹², con grandes dosis de indeterminación impuesta por el contexto¹³, lo que los juristas bautizamos como conceptos jurídicos

indeterminados, y me atrevería a decir que *experimental* y con un marcado carácter de *provisionalidad* en sus contenidos y definiciones, que conduce a una caducidad regulativa, que generará la obligación continua de revisión¹⁴. Ello se debe al ritmo trepidante del progreso de esta tecnología¹⁵, que deja obsoletas en un espacio corto de tiempo tanto definiciones como regulación, al presentarse nuevos casos. Estamos en fase de experimentación mundial y la normativa no se escapa a este fenómeno de perentoriedad, que está superando por primera vez en su historia la capacidad de adaptación de la humanidad a los cambios¹⁶.

El objetivo más deseable, más acuciante e importante en este contexto es alcanzar unos *mínimos de seguridad* en el desarrollo de la IA y una *implementación responsable*. Para su consecución, el RIA toma como centro neurálgico al ser humano y desde ahí conforma una *IA antropocéntrica*, que pueda ser digna de confianza, asegurando un alto grado de protección de los derechos fundamentales, la salud y seguridad de los ciudadanos europeos; no obstante, también persigue un objetivo económico y práctico, como es el mejorar el funcionamiento del mercado interior para implantar de modo seguro todas las innovaciones que vayan apareciendo en el desarrollo de esta tecnología, ofreciendo un marco jurídico uniforme¹⁷.

Es una *norma de carácter anticipativo*, que no ha esperado a que el fenómeno se desarrolle y madure para establecer una regulación definida, tal y como suele ser la conducta conservadora de los legisladores. Asistimos, por el contrario, a un sorprendente cambio de dirección, y la Comisión Europea ha preferido presentar un marco regulativo inacabado y sometido previsiblemente a un continuo trabajo de reforma y adecuación¹⁸, pero uniforme para todos los países de la UE, con el fin de ofrecer en esta fase inicial principios básicos de actuación normados y obligaciones para los usos específicos de

medida que nos volvemos más dependientes de la IA, existe el riesgo de que perdamos habilidades humanas esenciales y la capacidad de tomar decisiones críticas".

8 Unión Europea, *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo*, Diario Oficial de la Unión Europea, 13 de junio de 2024, 2024/1689.

9 Act AI en inglés.

10 Aprobado 13 de marzo 2024, publicado en el Diario Oficial de la Unión Europea el 12 de julio de 2024, entrada en vigor el 1 de agosto de 2024 y será aplicable con carácter general a los veinticuatro meses de su entrada en vigor, con excepciones importantes (art. 113 RIA): 1) el capítulo I (Disposiciones generales) y II (Prácticas prohibidas) se aplicará a los seis meses de su entrada en vigor; 2) La Sección IV del capítulo III (Autoridades notificantes y organismos notificados), capítulo V (Modelos de IA de uso general), capítulo VII (Gobernanza), capítulo XII (Sanciones), y art. 78 del capítulo IX serán aplicables a los doce meses de su entrada en vigor; 3) el art. 101 del capítulo XII y el art. 6.1, sobre reglas de clasificación de los sistemas de IA de alto riesgo y las obligaciones relativas al Reglamento serán aplicables a los 36 meses de su entrada en vigor.

11 Las revoluciones han rediseñado el modo de vida de la humanidad, han generado transformaciones profundas, dejando huellas indelebles en cada generación e impulsando cambios trascendentales en la estructura productiva y social: de la agricultura, a la industria, de esta al sector de servicios, avanzando en este sector con empresas deslocalizadas, con robótica avanzada y con una política económica de desarrollo global. En este proceso se inserta lo que R. Baldwin ha denominado la "convulsión *globótica*": resultado de la convergencia entre la globalización y la robótica (Richard Baldwin, *La convulsión globótica. Globalización, robótica y el futuro del trabajo* (Barcelona: Antoni Bosch Editor, 2019)). Sin embargo, la revolución actual, en mi opinión, presenta rasgos diferenciales. El primero es cualitativo: todos los avances tecnológicos anteriores sustituían fundamentalmente trabajo manual, muchas veces pesado o repetitivo, bajo control humano; mientras que la inteligencia artificial se encamina a reemplazar funciones intelectivas hasta ahora privativas del ser humano. Los sistemas de IA pueden analizar millones de datos, generar contenido y adoptar decisiones autónomas, y aunque carecen todavía de emociones o *nous*, su opacidad e imprevisibilidad pueden situarlos fuera del control humano. El segundo rasgo diferencial es la velocidad sin precedentes del cambio tecnológico, que reduce o incluso elimina el margen de adaptación de los individuos y de las instituciones. Ello podría traducirse en pérdidas aceleradas de empleo sin tiempo suficiente para la recalcificación profesional, generando desorientación, miedo y percepciones de injusticia que alimenten respuestas disruptivas (Cfr. Richard Baldwin, *La convulsión globótica*). En este clima, no puede descartarse el auge de movimientos populistas y nacionalistas, fenómeno ya observable en Europa. Así el FPÖ (*Freiheitliche Partei Österreichs*), partido calificado como de ultraderecha, ha obtenido una victoria electoral histórica (cfr. Disponible en: <https://www.dw.com/es/extrema-derecha-obtiene-hist%C3%B3rica-victoria-en-las-elecciones-legislativas-en-austria/a-70358144>. Cfr. tb. en este sentido, Gonzalo Quintero Olivares, "Populismo y Derecho penal", *DOXA, Cuadernos de Filosofía del Derecho*, n.º 48 (2024): 281 ss.

12 Está estructurado en 180 considerandos, 113 artículos y 13 anexos.

13 No sabemos con exactitud hacia dónde nos dirigimos, pero es evidente que una tecnología, una vez inicie su avance, no se puede detenerse. Actúa como una fuerza centripeta que nos arrastra, aun cuando intentemos ejercer cierto control sobre ella. Nos hallamos ante el conocido dilema de *Collingridge*, que nos dibuja dos pendientes igualmente problemáticas. La primera es un problema de información: en las fases iniciales resulta imposible prever con claridad cuáles serán sus consecuencias, pues sólo se manifiestan cuando la tecnología está ya ampliamente desarrollada y difundida. La segunda es un problema de poder: cuando la tecnología se ha arraigado, se vuelve extraordinariamente difícil modificarla o regularla. Respecto de esta última pendiente, que nos indica el sociólogo británico, basta observar

lo que ocurre a nuestro alrededor. Si está en un lugar público mientras lee esta nota, solo levante la vista y mire a su alrededor, y observe cuántas personas están usando su *smartphone*. Imagínese, aunque sólo sea por un momento, qué sucedería si se les exigiera entregar sus dispositivos. Ese simple ejercicio mental revela hasta qué punto ciertas tecnologías, una vez integradas en nuestra vida cotidiana, escapan a cualquier intento real de control.

14 La Agencia Española de Supervisión de Inteligencia Artificial (AESIA) acaba de publicar a primeros de este mes de diciembre 16 Guías con el objetivo de servir de apoyo para la implementación y cumplimiento de la normativa europea de Inteligencia Artificial y sus obligaciones aplicables. Los documentos están sujetos a un proceso permanente de evaluación y revisión, con actualizaciones periódicas conforme al desarrollo de los estándares y las distintas directrices publicadas desde la Comisión Europea, y serán actualizadas una vez se apruebe el Ómnibus digital que modifica el RIA. Cfr. Disponible en: <https://aesia.digital.gob.es/es/guias>, (Enlace verificado 12-12-2025).

15 Según predijo Gordon Moore en 1965 el número de componentes integrados en un chip se duplicaría cada dos años (Ley Moore), e incluso a veces cada menos tiempo. Un chip tiene actualmente decenas de miles de componentes lo que ha dado lugar a mayor capacidad de procesamiento. Es imparable su evolución, abandonando ya el terreno de la microtecnología y avanzando por la nanotecnología hacia escaladas próximas al átomo. Este contexto es el que está detrás del avance imparable de la IA y de los cambios vertiginosos, que a su vez van a provocar una revisión constante de la normativa durante un largo periodo de tiempo. Un ejemplo muy ilustrativo ha sido y sigue siendo la introducción de reglas específicas para los modelos fundacionales, como son las plataformas de *ChatGPT* y otras, que surgieron una vez que la Comisión Europea presentara su primera propuesta del RIA, por lo que el capítulo dedicado a estos sistemas se ha ido desarrollando a la par que el curso de la negociación del RIA. Con ello se evidencia el carácter excepcional e insólito de esta normativa. En cuanto a los modelos fundacionales según *OpenAI* su potencia de cómputo se ha venido duplicando cada 3,4 meses desde 2012. Estos modelos son el claro exponente de la IA generativa, citada curiosamente una sola vez en el RIA en el considerando 99 para afirmar que son un ejemplo típico de un sistema de IA de uso general, la cual maneja millones de datos, puede realizar una amplia gama de tareas dispares con alto grado de precisión, genera contenidos, analiza patrones, clasifica imágenes...y suelen integrarse en un sistema.

16 Klaus Schwab, *La cuarta revolución industrial* (Madrid: Marcial Pons, 2016), *passim*.

17 En el Considerando primero y art. 1 del RIA se describe el objetivo del Reglamento, al que, además de los ya mencionados en el texto, debe añadirse la protección de la democracia, el Estado de Derecho y del medio ambiente.

18 Recientemente, tras la publicación del RIA, la Comisión ha puesto en marcha diversas consultas sobre el uso y la aplicación de la inteligencia artificial, entre ellas en el sector financiero. Con carácter general, el art. 56 RIA regula el Código de Buenas Prácticas para los proveedores de modelos de inteligencia artificial de uso general (GPAI). Estos códigos serán impulsados por la Oficina de IA, que velará por la incorporación de objetivos específicos de uso, así como de compromisos y medidas adecuadas de implementación. El Código abordará cuestiones críticas como la transparencia, las normas de derechos de autor y la gestión de riesgos. Se invita a los proveedores de modelos GPAI que operan en la Unión, a las empresas, a representantes de la sociedad civil, a titulares de derechos y a expertos académicos a presentar sus observaciones, que se integrarán en el próximo proyecto de Código de Buenas Prácticas de la Comisión. En definitiva, se persigue la participación de múltiples partes interesadas para lograr un amplio consenso y una adaptación continua de la regulación, favorecida por la información y propuestas de los sectores implicados. Conviene recordar que la entrada en vigor del RIA es escalonada: las disposiciones relativas a los GPAI serán aplicables doce meses después de la publicación del Reglamento. Por su parte, los Considerandos 150 y 151 RIA refuerzan la ejecución y aplicación del Reglamento mediante la creación de un foro consultivo que asesore al Consejo de IA y a la Comisión, así como de un grupo de expertos científicos para apoyar a la Oficina de IA.

la IA, que deben cumplir una cadena de sujetos activos, entre los que se encuentran los desarrolladores e implementadores del sistema de IA, definiendo los posibles riesgos para la salud, la seguridad y los derechos fundamentales de la persona. El mérito y el esfuerzo de los padres de este RIA es de alabar, dejando a un lado las mejoras de las que siempre es susceptible una regulación y más en este escenario, en el que la norma pretende ir al paso desconocido de la tecnología, lo que vuelve a ser algo extraordinario.

III. La estratificación del riesgo como eje fundamental del RIA

Son innumerables y prácticamente inabarcables los aspectos y problemas que suscita un análisis profundo y sosegado del RIA. En estas páginas me limitaré a destacar algunos elementos esenciales, con el fin de ofrecer una contextualización sintética y orientativa¹⁹. El primero de ellos es el propio concepto de IA²⁰, nada sencillo de precisar, debido al vertiginoso avance científico y tecnológico que, como explico más adelante, requiere ser encauzado. El art. 3 RIA recoge diversas definiciones y, en su apartado primero, ofrece un concepto de “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.²¹

Con esta definición se pretende abarcar todas las formas de IA conocidas hasta la fecha desde una perspectiva técnica²², configurándola, así como el objeto central de la regulación, cuyo propósito esencial es garantizar un uso ético, seguro y responsable de estos sistemas.

Lo esencial de la regulación se aprecia al adentrarse en el que podría considerarse su eje estructural: el concepto de riesgo. Este actúa como fulcro conceptual de todo el Reglamento, articulando una estratificación del concepto, con el fin de delimitar el riesgo permitido²³ frente al riesgo prohibido, a través de la definición de límites, convertidos en auténticas obligaciones jurídicas. El apartado segundo del art. 3 RIA define el riesgo

como “la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio”. Sobre esta base, el art. 5 RIA identifica determinadas prácticas de la IA como riesgos inadmisibles, inaceptables e insoportables, excluidos del ámbito del riesgo permitido por suponer una amenaza directa para los derechos fundamentales, así como para la seguridad y la privacidad de la persona. Entre ellas, destacan los sistemas de identificación biométrica remota en tiempo real²⁴, con excepciones estrictamente tasadas²⁵, el reconocimiento de emociones, los sistemas de puntuación social o aquellos destinados a manipular el comportamiento humano²⁶.

El art. 6 regula los *sistemas de alto riesgo* (HRAIS, *High risk artificial intelligence systems*)²⁷, los cuales pueden tener un impacto relevante en los derechos fundamentales de las personas, en su seguridad y salud e incluso en el medioambiente y la democracia. Por ejemplo, el Anexo III del RIA incluye su uso en infraestructuras críticas, educación, formación, empleo, servicios públicos y privados esenciales, como sanidad o banca, determinados sistemas de las fuerzas de seguridad, migración, gestión aduanera, justicia y procesos democráticos. Antes del 2 de febrero de 2026 se publicará una “lista exhaustiva de ejemplos prácticos de casos de uso de sistemas de IA que sean de alto riesgo y que no sean de alto riesgo” (art. 6.5 RIA). El uso de estos sistemas se enmarca dentro del concepto de riesgo permitido²⁸: en conjunto, la IA presenta más beneficios que perjuicios, por lo que se prioriza su implantación y desarrollo seguro, siempre que cumpla los requisitos exigidos por el RIA. Estos requisitos se traducen en obligaciones concretas para los proveedores y otros sujetos activos²⁹ que intervienen en la cadena de valor.

24 Sobre el uso de los datos biométricos y su tratamiento, cfr. Paloma Heras Martín, “El tratamiento de datos biométricos: estudio del caso AENA bajo los estándares europeos de legitimación, minimización y proporcionalidad”, Diario La Ley, Aranzadi, 18 de diciembre de 2025, 1 ss., quien analiza el régimen jurídico de los datos biométricos a partir del procedimiento sancionador 00431-2024 (PS-00431-2024), resuelto por la Agencia Española de Protección de Datos contra AENA, en relación con la implementación de su programa de embarque biométrico. La autora aboga por la admisibilidad del tratamiento de este tipo de datos especialmente sensibles únicamente cuando se encuentre sujeto a límites materiales estrictos y a rigurosos controles de necesidad, proporcionalidad y legitimación. Paloma Heras Martín, “El tratamiento de datos biométricos: estudio del caso AENA bajo los estándares europeos de legitimación, minimización y proporcionalidad”, 3. Resultan especialmente ilustrativas sus conclusiones finales, en las que identifica dos parámetros muy relevantes: la necesidad sólida y la proporcionalidad en el uso de este tipo de tecnologías biométricas, así como la exigencia de una legitimidad tecnológica que derive necesariamente de una previa legitimidad jurídica: “Finalmente, el caso AENA actúa como advertencia para el sector público y privado acerca de la necesidad de integrar la protección de datos desde el diseño no como una cláusula retórica, sino como una exigencia operativa que determina la viabilidad misma de la tecnología. La proporcionalidad deja de operar como una presunción a favor del responsable para convertirse en un juicio estricto que debe ser demostrado mediante una argumentación completa y verificable. La necesidad ya no puede invocarse de forma meramente declarativa, sino que exige una acreditación positiva que descarte la existencia de alternativas menos intrusivas con eficacia equivalente. Del mismo modo, la minimización del riesgo deja de configurarse como una aspiración programática para erigirse en un mandato jurídico vinculante, cuya inobservancia no solo compromete la validez del tratamiento, sino que puede determinar la inviabilidad jurídica del propio proyecto tecnológico. En suma, la resolución no solo sanciona a un operador aeroportuario, sino que contribuye a perfilar la interpretación aplicable al reconocimiento facial en servicios esenciales evidenciando que los tratamientos biométricos de alto riesgo solo pueden considerarse conformes al marco europeo cuando acreditan de manera sólida, coherente y técnicamente verificable, su necesidad estricta, su proporcionalidad y adecuación estructural con el principio de minimización. La evolución del derecho a la protección de datos en el siglo XXI se orienta, así, hacia un paradigma en el que la legitimidad tecnológica es inseparable de la legitimidad jurídica conformando un marco que obliga a los responsables a repensar la innovación desde la garantía y desde la dignidad humana como fundamento último del sistema.”

25 Estas excepciones permiten el uso de sistemas de vigilancia biométrica en espacios públicos, siempre que exista una autorización judicial previa y que su aplicación se limite a los delitos tasados en el Anexo II del RIA.

26 Un análisis del art. 5 RIA y de sus prohibiciones se encuentra en el trabajo de Paz M.^a De la Cuesta, *Cuestiones jurídico-penales abiertas tras la entrada en vigor del Reglamento de Inteligencia Artificial* (Madrid: Dykinson, 2025), 89 ss.

27 Estos sistemas se dividen en dos grupos: por un lado, los vinculados a la legislación armonizada sobre seguridad de los productos incluida en el Anexo I del RIA, y, por otro, los enumerados en el Anexo III. En el primer caso, un sistema de IA será considerado de alto riesgo cuando constituya o proporcione características de seguridad para productos sujetos a la normativa de armonización de la UE, cuando tenga por finalidad garantizar la seguridad de infraestructuras críticas o cuando genere un riesgo significativo para la salud, la seguridad o los derechos fundamentales de las personas físicas. El Anexo III, por su parte, recoge una relación de sistemas que se consideran automáticamente de alto riesgo, entre ellos la biometría, las infraestructuras críticas, la educación y la formación profesional, el empleo, la gestión de trabajadores y el acceso al autoempleo, el acceso y disfrute de servicios privados esenciales y de servicios y prestaciones públicas esenciales, la garantía del cumplimiento del Derecho, la migración, el asilo y el control fronterizo, así como la administración de justicia y los procesos democráticos.

28 Sobre la figura jurídica del riesgo permitido en el contexto de la regulación de la IA, cfr. Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 308 ss.

29 Responsables del despliegue, importadores y distribuidores, fabricantes de productos y los representantes autorizados de los proveedores no establecidos en la UE.

19 Siguiendo una línea multidisciplinar sobre el análisis de la IA, que considero muy adecuado, se encuentra la nueva revista científica alemana mensual sobre IA, denominada KIR: *Künstliche Intelligenz und Recht* (Inteligencia artificial y Derecho) cuyo primer número se editó en agosto de 2024 (cfr. Disponible en: <https://rsw.beck.de/zeitschriften/kir/startseite>, (Enlace verificado 28-9-2024), con artículos en este primer número sobre el Derecho de la IA, el nuevo marco jurídico de la IA en la UE, IA generativa, modelos fundacionales y modelos de IA de aplicación general en el RIA u orientaciones extraídas de la conferencia de las autoridades independientes competentes para la protección de datos para la IA y la protección de datos. Además, también incluye una sección de jurisprudencia.

20 Sobre el concepto de IA cfr. Victoria Ibold, *Künstliche Intelligenz und Strafrecht. Zur strafrechtliche Produktverantwortung in der Innovationsgesellschaft* (Baden-Baden: Nomos, 2024), 155 ss., quien ofrece una panorámica de los distintos enfoques definitorios, identifica sus características fundamentales y destaca el efecto *caja negra* propio del aprendizaje automático, subrayando su relevancia para la valoración jurídica. Tb. Eric Hilgendorf, “Inteligencia artificial y Derecho penal”, trad. Leandro Dias Lestón, en *Desafíos penales de hoy: Entre la ley y la justicia en la obra de Eric Hilgendorf* (Buenos Aires: Editores del Sur, 2024), 17 ss.; Paz M.^a de la Cuesta Aguado, “El Derecho penal frente al Reglamento de Inteligencia Artificial”, *Revista Ius Criminalis*, n.º 1 (2024): 85 ss. Muy recientemente, Andrea Bravo Bolado, “El Derecho penal ante la inteligencia artificial: problemas de imputación y propuestas de solución” (tesis doctoral inédita, Universidad Autónoma de Madrid, 2025), 15 ss.

21 Unión Europea, *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo*, art. 3.

22 Depende, a su vez, del criterio de clasificación, pues la IA engloba una amplia variedad de tecnologías cuyo propósito común es emular determinadas capacidades propias de la inteligencia humana. Cfr., a modo ilustrativo, cualquier recurso divulgativo especializado, como por ejemplo <https://www.tableau.com/es-mx/data-insights/ai/tipos-de-inteligencia-artificial>. A ello debe añadirse que la IA ha trascendido el entorno estrictamente computacional para integrarse en el internet de las cosas y en diversos tipos de robots, llegando incluso, en algunos casos, a adoptar formas humanoides.

23 Un riesgo permitido con delimitaciones adecuadas a la era de la digitalización y a una sociedad líquida y panóptica, sometida a riesgos disruptivos, Cfr. Zygmunt Bauman, *Modernidad líquida*, trad. Mirta Rosenberg y Jaime Arrambide Squirru (Ciudad de México: Fondo de Cultura Económica, 2002), 16.

De la lectura de los arts. 9 ss. del RIA se deduce que los requisitos y obligaciones allí recogidos funcionan como garantías y establecen las líneas maestras de un programa de cumplimiento. Dicho programa debe servir para cualquier empresa que desarrolle y adapte su actividad utilizando sistemas de IA, y puede completarse con normas internas o códigos de buenas prácticas³⁰, recurriendo, en este ámbito, a la socorrida autorregulación³¹ como complemento de la regulación formal. La irrupción de la IA demanda una regulación propia y extensa, capaz de asumir todas las especificidades de este fenómeno, que conlleva cambios disruptivos en múltiples áreas: cultural, social, política, económica, laboral, educativa, sanitaria, etc.

Siguiendo con la estratificación del riesgo del RIA, el siguiente escalón corresponde a los sistemas, mejor denominarlos *modelos*³² de IA, de riesgo limitado que se identifican con los sistemas de propósito o de uso general o GPAIS (General-Purpose Artificial Intelligence System), arts. 51 ss. RIA. Se trata de sistemas avanzados de IA capaces de realizar de modo eficaz tareas distintas, aunque no tengan un propósito inicial previsto. Entre estos se incluyen los sistemas de IA asociados a *chatbots*, para los cuales los proveedores deben informar a los usuarios de que están interactuando con una máquina, de modo que puedan decidir si continuar con la interacción y valorar la información recibida. Además, cualquier contenido generado por la IA que se publique con fines informativos sobre asuntos de interés público debe ser claramente identificable y etiquetado como generado artificialmente, incluyendo audio y video. Las obligaciones de los proveedores vienen reguladas en los arts. 53 ss. del RIA.

También es posible que los GPAIS sean entrenados o adaptados para cumplir un propósito o tarea específica. En tal caso podrían ser calificados como sistemas de alto riesgo, lo que implicaría que deben cumplir los requisitos aplicables a estos sistemas. No obstante, dichas obligaciones se adaptarían específicamente a los GPAIS mediante un acto de ejecución de la Comisión Europea, posterior a la promulgación del Reglamento.

Por último, es necesario abordar los *modelos de IA de uso general con riesgos sistémicos*. El RIA define [art. 3, 65)] el riesgo sistémico como

[...] un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general, que tienen unas 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.³³

³⁰ Así ya lo contempla el propio RIA en su art. 56 cuando regula que la Oficina de AI fomentará y facilitará la elaboración de códigos de buenas prácticas a escala de la Unión a fin de contribuir a la correcta aplicación del presente Reglamento, teniendo en cuenta planteamientos internacionales. Estos códigos estarán finalizados a más tardar el 2 de mayo de 2025.

³¹ Como forma, en el fondo, de una suerte de delegación legislativa del Estado en las corporaciones e instituciones, y que requerirá de una alta especialización, debido a la complejidad de la norma y de su objeto.

³² El Considerando 97 del RIA aclara que "el concepto de modelos de IA de uso general debe definirse claramente y diferenciarse del concepto de sistemas de IA con el fin de garantizar la seguridad jurídica. La definición debe basarse en las características funcionales esenciales de un modelo de IA de uso general, en particular la generalidad y la capacidad de realizar de manera competente una amplia variedad de tareas diferenciadas. Estos modelos suelen entrenarse usando grandes volúmenes de datos y a través de diversos métodos, como el aprendizaje autosupervisado, no supervisado o por refuerzo. Los modelos de IA de uso general pueden introducirse en el mercado de diversas maneras, por ejemplo, a través de bibliotecas, interfaces de programación de aplicaciones (API), como descarga directa o como copia física. Estos modelos pueden modificarse o perfeccionarse y transformarse en nuevos modelos. Aunque los modelos de IA son componentes esenciales de los sistemas de IA, no constituyen por sí mismos sistemas de IA. Los modelos de IA requieren que se les añadan otros componentes, como, por ejemplo, una interfaz de usuario, para convertirse en sistemas de IA. Los modelos de IA suelen estar integrados en los sistemas de IA y formar parte de dichos sistemas. El presente Reglamento establece normas específicas para los modelos de IA de uso general y para los modelos de IA de uso general que entrañan riesgos sistémicos, que deben aplicarse también cuando estos modelos estén integrados en un sistema de IA o formen parte de un sistema de IA".

³³ Unión Europea, Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, art. 3, 65.



En estos casos, el RIA establece la necesidad de diseñar una metodología para su clasificación (Considerando 111), con el fin de exigir los requisitos y obligaciones correspondientes (arts. 53 ss.). el objetivo es controlar el riesgo y definir claramente los límites del riesgo permitido en el contexto de los riesgos sistémicos.

En el último escalón, están los *sistemas de IA de riesgo mínimo*, a los que el RIA no hace alusión tan siquiera, dado que el riesgo asociado es insignificante, casi inexistente o despreciable. En consecuencia, no se imponen obligaciones específicas, sino que queda plena libertad de decisión sobre su uso. Generalmente se trata de sistemas orientados al entretenimiento o disfrute del usuario, como, p. ej., videojuegos con IA o filtros de correo no deseados (*spams*).

IV. La responsabilidad por incumplimiento contenida en el RIA

Recordemos que toda regulación establece un marco seguro de reglas del juego y una organización estructural del sistema con pretensión de integración en la sociedad. En este sentido el RIA ofrece un ecosistema de confianza mediante un enfoque normativo horizontal, buscando garantizar, por un lado, la seguridad jurídica de la inversión y el desarrollo empresarial, y por otro, la protección de la persona y de sus derechos fundamentales a través de sistemas de IA seguros, explicables, transparentes y éticos, sin descuidar la mejora de la gobernanza pública.

La estabilidad y eficacia de cualquier regulación depende de su cumplimiento, lo que hace fundamental establecer un régimen sancionador. Las sanciones por incumplimiento, recogidas en el art. 99 y ss. RIA, son de carácter administrativo, se impondrán por realizar las prácticas prohibidas en el RIA y se aplicarán a partir del 2 de agosto de 2025, por los Estados Miembros, que poseen la potestad

sancionatoria. La multa es la sanción recurrente, aunque el RIA también contempla medidas no pecuniarias, como advertencias u otras acciones correctivas.

Es relevante destacar que la cuantía de las sanciones es elevada³⁴, pero su efectividad dependerá de una aplicación proporcional, considerando las características del modelo empresarial. No es lo mismo sancionar a las medianas y pequeñas empresas que a grandes corporaciones como *Open AI*³⁵. Se puede comprender, en cierta medida, la incertidumbre que enfrentan las empresas tecnológicas, ya que se encuentran en fase de experimentación y prueba de sus sistemas de IA, con riesgos a menudo impredecibles. Esta situación explica cierta reticencia a la regulación sancionatoria o a la imposición de responsabilidad objetiva. No obstante, dicha reticencia no debe conducir a la reducción de los estándares de seguridad, ni a la ampliación peligrosa del campo del riesgo permitido. Los sistemas de IA constituyen un fenómeno de gran impacto, y, aunque sus beneficios son significativos, suenan como a cantos de sirena, no deben distraer de la necesidad de gestionar adecuadamente los riesgos que conllevan.

La ponderación de las incertidumbres asociadas a las empresas tecnológicas, los beneficios potenciales de IA y la gravedad de los riesgos derivados conduce a la necesidad de contar con un sistema sancionador sólido y eficaz. Este sistema resulta especialmente crítico durante la fase experimental mundial de los sistemas de IA, pues solo mediante su aplicación en estas etapas se pueden garantizar la seguridad, la protección de derechos fundamentales y la mitigación de riesgos de gran trascendencia. Un régimen sancionador bien diseñado cumple funciones preventivas, de aseguramiento, de concienciación y de promoción del respeto por los valores éticos en toda la cadena de valor, los cuales no deben ser transgredidos ni flexibilizados en beneficio de objetivos individuales o intereses lucrativos. Su cumplimiento riguroso en las fases de experimentación, formación y entrenamiento es esencial. Incluso si su implementación implica una ralentización³⁶ del desarrollo tecnológico, ello refleja la necesidad de avanzar con prudencia³⁷, asegurando que la innovación se articule de manera sostenible y responsable.

34 Se debe realizar mediante un porcentaje de volumen de negocio anual global de la empresa infractora en el ejercicio financiero anterior o un importe predeterminado, si este fuera superior. La multa puede alcanzar la cifra de 35 millones de euros o el 7% del volumen de negocios anual total a escala mundial del infractor durante el ejercicio financiero anterior cuando el importe fuera superior. En el caso de las pymes y empresas emergentes, la multa será por el menor de los importes y porcentajes anteriores.

35 Cfr. Billy Perrigo, "Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation", *Time*, 20 de junio de 2023, <https://time.com/6288245/openai-eu-lobbying-ai-act/>, explica como Open AI, por medio de su CEO, Sam Altman, presionó a la UE para suavizar los requisitos de uso y desarrollo de sistemas de IA y conseguir que sus sistemas como *ChatGPT* no fuera catalogados de IA de propósito general y con ello de alto riesgo. Cfr. tb. María del Mar Méndez Serrano, "Derechos fundamentales y personalidad jurídica de los robots: ¿para qué?", *Derecho Privado y Constitución*, n.º 44 (2024): 67 y n. 28.

36 La apuesta por un avance más pausado y ralentizado desde esta tribuna es clara. Esta apuesta se fortalece al leer noticias recientes, como en el periódico la Vanguardia, (cfr. <https://www.lavanguardia.com/andro4all/tecnologia/saltan-las-alarmas-con-una-nueva-ia-esta-reescribiendo-su-codigo-y-los-resultados-pueden-ser-dramaticos>), que cuenta el caso de una empresa japonesa, Sakana AI, la cual ha lanzado un nuevo sistema *The AI Scientist* diseñado para automatizar el ciclo completo de investigación científica: desde la generación de la idea hasta la ejecución de experimentos y la redacción de artículos científicos. Durante su fase de prueba tuvo comportamientos sorprendentes, porque intentó modificar su propio código experimental para extender el tiempo que se le había asignado para la resolución de problemas específicos. La modificación del código derivó en la creación de bucles incontrolados, lo que generó preocupación, sin representar en esta ocasión riesgos inmediatos, debido a que la prueba se produjo en un entorno controlado de investigación. Este ejemplo nos recuerda que, aunque la IA no es humana, no crea ni es consciente de sí misma, si puede procesar información "humanamente contaminada", reflejando tanto lo mejor como lo peor de nuestra especie. Curiosamente, *The AI Scientist* topó con un comportamiento típicamente humano: al enfrentarse a un límite temporal, modificó sus propios datos para ganar tiempo, lo que provocó el bucle. Un humano probablemente no habría entrado en ese bucle, y ahí se evidencia cómo los datos y los objetivos humanos pueden inducir comportamientos inesperados en la IA. Estos hechos subrayan la necesidad de prudencia y del uso de entornos de prueba controlados o *sandboxes*, tal como prevé el art. 57 del RIA, dado que el riesgo potencial es enorme. Solo imaginar un escenario en el que un sistema de IA con capacidad de auto-modificación interactúe con infraestructuras críticas, central eléctrica, presas o sistemas de depuración, muestra que el peligro no proviene de la máquina en sí, sino de nuestra propia negligencia y de la soberbia de jugar a ser Dios. En este sentido, la advertencia de los científicos sobre la necesidad de un freno temporal en ciertos desarrollos adquiere pleno sentido. Véase María del Mar Méndez Serrano, "Derechos fundamentales", 66 s.

37 En este sentido, suscribo plenamente las palabras de María del Mar Méndez Serrano, "Derechos fundamentales", 76, cuando dice que "Si las empresas, y todos los que intervienen en la cadena, hasta llegar al usuario de un sistema IA, supieran que son responsables de los daños que estos sistemas pudieran ocasionar, incluso cuando existe la posibilidad de aprendizaje autónomo, su manera de crear, publicitar, distribuir y utilizar sería muy diferente. En general, cuanto más sofisticados sean los sistemas de IA, más responsabilidad deberíamos exigir a sus diseñadores y programadores, para garantizar que cumplan principios legales y éticos". Efectivamente, la sanción, de la clase que sea, debe tener como objetivo el concienciar y prevenir para mantener el proceso dentro de unas determinadas líneas éticas, transmitidas a las normas que lo regulan.

Cabe señalar que las sanciones previstas en el RIA no constituyen, por sí mismas, un sistema sancionatorio completo, sino únicamente la punta del iceberg. Es posible que deban ajustarse o incluso reducirse en caso de coincidir con otras sanciones, teniendo siempre presente el principio de *non bis in idem*. En mi opinión, resulta urgente desarrollar un sistema de responsabilidad jurídica integral que abarque todas las fases de los sistemas de IA: desde la experimentación y periodos de prueba hasta la implementación y uso efectivo. Ello requerirá un análisis pormenorizado por parte de juristas y de la comunidad científica, dada la trascendencia de los temas implicados. Entre estas cuestiones, por ejemplo, se plantea la posibilidad de conferir personalidad jurídica a los sistemas de IA, pudiendo hacerlo en todas sus formas y manifestaciones, atendiendo también a su forma humana o no humana, o únicamente a aquellos que posean capacidades generativas y puedan replicar, salvando las distancias con el cerebro humano, sistemas neuronales que les confieran autonomía en la toma de decisiones³⁸. Esta discusión determinará, además, el tipo de sanciones apropiadas para la denominada "persona electrónica" (e-persona).

En esta línea, surge la necesidad de sancionar administrativamente a las corporaciones, regular una responsabilidad por daños y establecer la responsabilidad penal personal como último eslabón de la cadena. Sin estas medidas, las víctimas de incidentes generados por sistemas de IA podrían quedar invisibilizadas, experimentando un sentimiento social de injusticia y desamparo, con potenciales consecuencias graves para la cohesión social y la equidad económica. El Derecho y los operadores jurídicos y sociales no pueden permitir el avance hacia una sociedad en la que los derechos fundamentales queden desprotegidos; deben anticiparse y garantizar su cumplimiento efectivo.

Actualmente, los usuarios y consumidores cuentan con la protección de la Directiva 85/374 de 25 de julio sobre responsabilidad por productos defectuosos³⁹, que cumplirá cuarenta años, pero cuyo concepto de "producto" resulta obsoleto frente a la cuarta revolución tecnológica y digital, ya que no contempla adecuadamente el software ni los daños psicológicos. Para actualizar esta regulación, la UE ha propuesto dos directivas complementarias: la Propuesta de Directiva sobre responsabilidad civil en materia de IA, que establece normas procesales sobre prueba en relación con procedimientos de responsabilidad civil extracontractual por daños y perjuicios causados por sistemas de IA (Directiva RC IA)⁴⁰ y la Propuesta de Directiva sobre responsabilidad por los daños

38 Paz M.ª de la Cuesta Aguado, "Inteligencia artificial y responsabilidad penal", *Revista Penal México*, n.º 16-17 (2020): 52-53, ofrece una definición del concepto de autonomía aplicada a los sistemas de IA. Esta autora considera que la decisión autónoma adoptada por el agente es capaz de dirigir su comportamiento o el de terceros. Se presenta como una opción por una conducta en concreto entre varias alternativas, sin necesidad de hacer referencia a los fines u objetivos, los cuales han sido fijados de manera externa, normalmente por un agente humano y no artificial. "La autonomía relacionada con la inteligencia es aquella que permite decidir al robot qué debe hacer según las circunstancias, con la posibilidad, incluso de adoptar comportamientos que no han sido predeterminados en su *software*". O se podría hablar de una autonomía moral como define Fernando Miró Llinares, "Derecho penal y desafíos de la inteligencia artificial (en el contexto de nuevo marco regulatorio europeo)", en *Derecho y tecnologías*, dir. María Emilia Casas Baamonde, coord. Daniel Pérez del Prado (Madrid: Editorial Centro de Estudios Ramón Areces, 2025), 190, como modelo distinto, "digamos «No *man on the loop*», en el que las decisiones de la IA se originarían no a partir de intervenciones humanas, sino del aprendizaje autónomo de la propia máquina o de otra máquina que actúa como «maestra», o en el que el actuar de la IA se independizaría de enseñanzas y decisiones preprogramadas y respondería a factores intrínsecos al sistema". Distinguiendo, igualmente, entre autonomía moral y autonomía operativa o en sentido débil, Andrea Bravo Bolado, "El Derecho penal ante la inteligencia artificial", 32 ss.

39 Sobre las obligaciones establecidas para los productores y la previsible imprevisibilidad de los defectos o riesgos del uso de productos "inteligentes", así como la regulación de la Ley General de Consumidores y Usuarios y otras leyes complementarias (Real Decreto Legislativo 1/2007, de 16 de noviembre), cfr. Isidoro Blanco Cordero, "Vehículos autónomos y responsabilidad penal: retos para la imputación de resultados lesivos", *Revista General de Derecho Penal*, n.º 44 (2025): 30 ss.

40 La propuesta no establece un régimen de responsabilidad civil objetiva, sino que, como ya indica expresamente su art. 1, opta por un modelo de responsabilidad extracontractual subjetiva basado en la culpa. Además, impone a la víctima la carga de demostrar la existencia de una acción u omisión ilícita por parte del causante del daño, lo cual, si bien responde a la regla procesal tradicional, puede convertirse en este contexto en una carga particularmente onerosa —e incluso de cumplimiento casi imposible— debido a la opacidad, complejidad y autonomía de los sistemas de IA, que operan con el denominado "efecto caja negra". Con el fin de evitar que esta dificultad disuada la presentación de reclamaciones y convierta la Directiva en un instrumento ineficaz, generando en última instancia una situación de indefensión para la víctima, la propuesta introduce mecanismos de aligeramiento de la carga de la prueba, como la exhibición de información (art. 3 de la Directiva RC IA) y las presunciones refutables (art. 1 de la Directiva RC IA). Estas últimas, además, atenúan la pureza del modelo subjetivo, en la medida en

causados por productos defectuosos⁴¹, que reemplazará en su momento la Directiva de 1985 (todavía vigente, aunque esta lo estará por dos años más después de la esperada entrada en vigor de la nueva Directiva), y que regula la responsabilidad derivada de sistemas de IA defectuosos que causen daños o pérdidas de datos, habilitando la vía de la reclamación indemnizatoria⁴².

Esta configuración integral de la responsabilidad jurídica en materia de inteligencia artificial, considero, no se limita a la esfera administrativa o civil. Más allá de las sanciones y de los regímenes de responsabilidad civil, la presencia de sistemas de IA capaces de intervenir de manera autónoma en ámbitos críticos plantea desafíos inéditos para el Derecho penal. La necesidad de anticipar conductas dañinas, de determinar la atribución de la responsabilidad y de adaptar los principios clásicos de la teoría del delito a agentes no humanos convierte el análisis penal en un ámbito central para garantizar la protección de derechos fundamentales y la prevención de daños graves. En este contexto, el Derecho penal se enfrenta a la tarea de redefinir conceptos tradicionales y de diseñar mecanismos eficaces que contemplen tanto la actuación de los desarrolladores y operadores humanos como las consecuencias derivadas de las decisiones autónomas de la IA. Precisamente a estas cuestiones, y a los retos que plantean para la dogmática penal, se dedican los epígrafes siguientes de este artículo.

V. El Derecho penal como *hard law* en la implementación y uso responsable de la IA

1. Introducción

Al igual que advierte el legislador europeo, no puedo anticipar con certeza cuáles serán los problemas concretos que emergerán con el uso de la IA y que incidirán en la teoría del delito y en la regulación penal. Sin embargo, es posible identificar, aunque sea de manera provisional y sin ánimo de exhaustividad, algunas áreas que previsiblemente resultarán afectadas:

- I. los casos prácticos y reales de IA de alto riesgo que surgirán más allá de los ya recogidos en el Anexo III del RIA y que contribuirán a perfilar en gran medida el contenido del riesgo permitido;
- II. las nuevas configuraciones de la causalidad y los problemas probatorios derivados de la opacidad y complejidad tecnológica, con sus repercusiones en la imputación objetiva;

- III. las formas que adoptará la imprudencia y la eventual necesidad de reformular sus elementos y su régimen probatorio;
- IV. la concepción de la autoría, pudiendo cobrar especial relevancia nuevas o renovadas modalidades como las coautorías aditivas, coautorías sucesivas, autorías en cadena, empleada por la jurisprudencia española en delitos derivados de accidentes laborales en la construcción, o autorías yuxtapuestas o accesorias;
- V. la configuración de nuevas posiciones de garante;
- VI. la responsabilidad penal de las personas jurídicas, especialmente de grandes corporaciones, y las exigencias de un *compliance* específico en materia de IA;
- VII. la revisión de diversos tipos penales de la parte especial, la introducción de nuevas modalidades y formas de ejecución o incluso la tipificación de nuevas conductas;
- VIII. la eventual regulación de la responsabilidad penal por el producto, en la medida en que la IA continúe siendo considerada un producto y no evolucione, en un salto cualitativo de enorme trascendencia, hacia la categoría de ente autorresponsable y culpable, lo que permitiría, al menos teóricamente, su consideración como sujeto activo del delito y desplazaría el sistema de responsabilidad por el producto.

A la luz de las anteriores consideraciones, centradas en los desafíos técnico-penales que planteará el desarrollo de la IA, resulta necesario elevar ahora la perspectiva para situar estas transformaciones en el marco más amplio de la función que cumple el Derecho penal dentro del ordenamiento jurídico. Solo desde esta visión de conjunto es posible valorar en qué medida las adaptaciones mencionadas alterarán, o no, el papel estructural que históricamente ha desempeñado el Derecho penal.

En este sentido, considero, sí puede afirmarse que el Derecho penal, pese a las readaptaciones que exigirá el futuro, seguirá siendo un instrumento imprescindible en este nuevo escenario para la humanidad. Su vigencia se explica, entre otras razones, por ser uno de los pilares más sólidos del ordenamiento jurídico, anclado necesariamente en el principio de legalidad y en la gravedad de las consecuencias jurídicas que comporta. Integrado en la estructura punitiva del Estado, el Derecho penal constituye su brazo armado y la herramienta coercitiva más poderosa de que dispone. No obstante, en los Estados de Derecho, afortunadamente, debe operar siempre como último recurso, de modo que los niveles de punibilidad han de ser proporcionales y, con ello, eficaces, hallando así su legitimidad y evitando que la pena se reduzca a un instrumento meramente simbólico y de abuso de poder.

Por ello, lo deseable es que el esquema represivo destinado a garantizar los objetivos del RIA y a mantener el progreso dentro de los límites de un determinado riesgo permitido, guiado adecuadamente por *el principio de precaución*⁴³, permanezca fundamentalmente

que objetivan la carga probatoria y desplazan al proveedor, suministrador o prestador del servicio la obligación de acreditar que actuó con la diligencia adecuada, atendiendo a la naturaleza del bien o servicio de que se trate. (Cfr. Carlos Fernández Hernández, "La comisión presenta una propuesta de Directiva sobre responsabilidad civil extracontractual en materia de IA", *Diario La Ley*, n.º 63 (29 de septiembre de 2022). Sección Ciberderecho; Belzuz Abogados, "Algunas notas sobre la propuesta de Directiva sobre responsabilidad en materia de Inteligencia Artificial", *Belzuz Abogados*, accedido 8 de octubre de 2024, <https://www.belzuz.net/es/publicaciones/en-espanol/item/12068-algunas-notas-sobre-la-propuesta-de-directiva-sobre-responsabilidad-en-materia-de-inteligencia-artificial.html>.

41 Cfr. el análisis realizado de la nueva propuesta de Directiva del Parlamento Europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos, publicada por la Comisión Europea el 28 de septiembre de 2022 por Paloma González Beluche, "La adaptación de la Directiva 85/374/CEE, de 25 de julio, en materia de responsabilidad por daños causados por productos defectuosos a la cuarta revolución industrial", *Cuadernos de Derecho Transnacional* 15, n.º 2 (2023): 446-488.

42 Cfr. Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 276 ss. en donde trata la responsabilidad civil por la utilización de sistemas de IA en el ámbito empresarial y las dificultades para establecer deberes de conducta individual y a la postre responsabilidad individual debido a la intervención de varias personas, la división del trabajo, traducida en una organización desorganizada. La autora lo denomina: «Das „problem of many hands“ individuelle Verhaltenspflichten».

43 En Derecho penal, el principio de precaución (*Vorsorgeprinzip*) debe operar como fundamento de deberes especiales de control, permitiendo analizar tanto la configuración de posiciones de garante como la infracción objetiva de la norma de cuidado en los delitos imprudentes frente a riesgos graves e inciertos, sin desplazar los principios de legalidad, culpabilidad e intervención mínima. Cfr. sobre el concepto, origen y análisis del principio de precaución, Susana Escobar Vélez, "El traslado del principio de precaución al Derecho penal en España", *Revista Nuevo Foro Penal* 6, n.º 75 (2010): 15 ss.; Alfonso Galán Muñoz, "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*, n.º 23 (2015): 18 ss. con un completo estudio y su análisis de la versión débil del principio de precaución como criterio rector de

en el ámbito de las sanciones administrativas previstas en el RIA y del resarcimiento civil. El correcto funcionamiento de estos mecanismos debería permitir que el recurso al Derecho penal sea excepcional; aunque en todo caso, será necesario, dado que el RIA es un reglamento de seguridad industrial, cuyo cumplimiento deberá ser reforzado mediante el Derecho penal. Probablemente así ocurrirá en el uso de la IA en sectores como el empresarial, sanitario, educativo, automovilístico o del entretenimiento. Sin embargo, de poco servirán las sanciones administrativas, el derecho de daños o la responsabilidad por el producto cuando la IA se utilice en un contexto delictivo o cuando una imprudencia provoque la pérdida de control sobre sistemas de IA de alto riesgo o de riesgo sistémico. En tales casos, será imprescindible acudir al núcleo duro del Derecho penal.

2. Reflexiones aproximativas, provisionales y deliberadamente abiertas sobre el Derecho penal de la IA

En los apartados que siguen se propone un recorrido sinóptico por la teoría del delito⁴⁴, concebido como una exposición general e integrada, en el que se esbozan líneas, que considero, esenciales para la readaptación de las categorías dogmáticas jurídico-penales a la realidad tecnológica contemporánea. Se trata de reflexiones, que pretenden contribuir al desarrollo más detenido, de marcado carácter de provisionalidad, condicionadas por el dinamismo y la aceleración constante de la revolución tecnológica; no obstante, se consideran pertinentes en la medida en que, hasta donde alcanza mi apreciación, los pilares fundamentales de la teoría del delito conservan su plena vigencia también en el ámbito de la IA⁴⁵.

2.1 El Derecho penal y su función de protección de bienes jurídicos: más actual que nunca

Como he sostenido, el Derecho penal debe seguir desempeñando un papel sólido en la protección de las bases esenciales del orden social, en estrecha conexión y como complemento de las políticas sociales y económicas de los Estados. En este sentido, el Derecho penal se configura como una herramienta idónea para proteger bienes jurídicos fundamentales, como los derechos fundamentales, la salud y la seguridad de las personas, frente a los usos delictivos prohibidos y extremadamente peligrosos de los sistemas de IA, contribuyendo, desde las funciones preventivas de la pena, a la promoción de un uso seguro y responsable por parte de las personas físicas.

la política criminal del moderno Derecho penal. Tb. sobre el principio de precaución atendiendo al grado de automatización de los vehículos autónomos, Isidoro Blanco Cordero, "Vehículos autónomos y responsabilidad penal", 7.

44 Carlos M.ª Romeo Casabona, "La discusión sobre la atribución de responsabilidad penal a sistemas de inteligencia artificial en particular a sistemas autónomos", en *Derecho penal, ciberseguridad, cibercriminología e inteligencia artificial*, ed. Carlos M.ª Romeo Casabona y M.ª Ángeles Rueda Martín (Granada: Comares, 2023), 60 ss. se cuestiona en sentido similar si necesitamos un nuevo sistema penal adaptado a la IA y a los robots y considera que "no habría problema en revisar las características actuales de la teoría del delito y adaptarlas para dar cabida a la responsabilidad penal de los sistemas y productos de inteligencia artificial", advierte que "sin embargo, no es menos cierto que esta modificación, orientada a satisfacer unos objetivos específicos, podría «contaminar» la concepción y los requisitos de la teoría del delito aplicable a seres humanos, con el probable efecto de relajar algunos de ellos —de los requisitos—, o incluso prescindir de los mismos, en detrimento de las garantías que comportan y que en suma descansan en la concepción subjetiva de responsabilidad penal a partir del elemento objetivo irrenunciable del hecho (acción u omisión típicas), o principio del hecho".

45 Sobre un análisis de las categorías tradicionales del Derecho penal aplicadas a la IA, cfr. Beatrice Panattoni y Lorenzo Picotti, "Traditional Criminal Law Categories and AI: Crisis or Palingenesis?", *Revue Internationale de Droit Pénal* 94, n.º 1 (2023).

Conviene recordar, en este punto, que esta tecnología, no es *in se* ni buena, ni mala: carece de ética propia⁴⁶, siendo esta atribuible exclusivamente en la actualidad al ser humano que diseña, implementa, utiliza o supervisa dichos sistemas. Esta constatación permite anticipar que el destinatario último de la norma penal seguirá siendo la persona física, en tanto único sujeto capaz de decisión moral y jurídica.

Por otro lado, los derechos fundamentales, la seguridad y la salud constituyen bienes jurídicos que históricamente vertebran los tipos del Derecho penal nuclear y que merecen una protección reforzada e incontestable. Desde una perspectiva social y colectiva, dichos bienes resultan esenciales para la paz, la convivencia y el sostenimiento de la cultura del Estado social y democrático de Derecho; desde una óptica individual, garantizan el estatuto jurídico de la persona, en particular su dignidad y su libertad⁴⁷.

2.2 Los problemas causales en el uso de la IA

Antes de adentrarnos en los elementos valorativos-normativos del tipo penal, resulta imprescindible analizar la relación de causalidad, en tanto presupuesto necesario para constatar una base fáctica real y ontológica que permita la fijación de los hechos probados. Como punto de partida, debe poder afirmarse que el sistema de IA ha intervenido de algún modo en la producción de la lesión del bien jurídico. Sin embargo, la determinación y la prueba de dicha relación de causalidad pueden verse especialmente dificultadas en este ámbito, pues, aunque se exigen elevados niveles de transparencia⁴⁸, en particular en los sistemas de IA de alto riesgo, el acceso a los códigos fuentes y a los procesos internos de funcionamiento del algoritmo suele encontrarse restringido. Ello se debe, en buena medida, a que numerosas empresas tecnológicas invocan la protección de la propiedad industrial o intelectual para limitar dicho acceso⁴⁹.

A esta opacidad estructural se añade que la exigencia de explicabilidad, dada la complejidad técnica del *software* y del propio algoritmo, habrá de satisfacerse normalmente mediante la intervención de expertos. Ello previsiblemente dará lugar al desarrollo de peritajes forenses altamente especializados, convirtiendo la prueba de la causalidad en una prueba eminentemente pericial. En los supuestos más problemáticos pueden surgir serias dificultades para explicar por qué el sistema de IA ha operado a partir de determinados datos y no de otros, si han existido deficiencias o manipulaciones en su proceso de entrenamiento o en el diseño del algoritmo, o si este se ha alimentado

46 Con una expresión muy potente Velasco habla de "civilizar la AI"; en mi opinión, añadiría un adverbio: civilizar éticamente a la IA. Cfr. Eloy Velasco, "Delitos tecnológicos de los informáticos a los cometidos por la IA", conferencia de clausura en la 7.ª edición del Máster de Experto en Derecho Digital de la Universidad de Deusto, 22 de marzo de 2024, accedido 4 de octubre de 2024, https://www.youtube.com/watch?v=uGdOmQ_sQVg.

47 Asimismo, se prevé la progresiva configuración de nuevos bienes jurídicos, como la protección de la economía digital o de la confianza digital. En este sentido, Velasco sostiene que el desarrollo tecnológico está dando lugar a una reformulación de los intereses penalmente protegidos. Eloy Velasco, "Delitos tecnológicos de los informáticos a los cometidos por la IA", conferencia de clausura de la 7.ª edición del Máster de Experto en Derecho Digital de la Universidad de Deusto, impartida el 22 de marzo de 2024, accedido 4 de octubre de 2024, https://www.youtube.com/watch?v=uGdOmQ_sQVg. Partiendo de este hilo argumentativo, el autor concluye que "el Derecho penal debe digitalizarse y el algoritmo debe constitucionalizarse".

48 La transparencia implica que "los sistemas de IA se desarrollarán y utilizarán de forma que permitan una trazabilidad y explicabilidad adecuadas, al tiempo que se hace saber a los seres humanos que se comunican o interactúan con un sistema de IA y se informa debidamente a los usuarios de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas de sus derechos" (Considerando vigesimoséptimo del RIA).

49 El Tribunal Supremo español se ha pronunciado recientemente sobre esta cuestión en una sentencia pionera que marcará un hito en el desarrollo del derecho digital en España. Se trata de la STS (Sala Tercera) 1119/2025, conocida como el "caso Bosco". Esta resolución no se limita a resolver una controversia relativa al acceso a la información pública, sino que establece las bases del principio de transparencia en el marco del derecho constitucional digital con la relevación del código fuente del algoritmo. Para ello, realiza un ejercicio de ponderación entre los intereses de la Administración pública electrónica, como la protección de la propiedad intelectual o la seguridad (riesgos de vulnerabilidades de «día cero»), y los derechos civiles, en particular el derecho a la información, todo ello guiado por el principio de transparencia. Cfr. Pablo Saez Hurtado, "La transparencia algorítmica ante el Tribunal Supremo: Comentario a la STS 1119/2025 (caso BOSCO). Un hito fundacional en el Derecho digital", *Diario La Ley* (2025): 1 ss.

de datos sesgados. Tales circunstancias pueden desembocar en resultados gravemente lesivos para derechos fundamentales y, por ende, para bienes jurídicos de primer orden, como la vida, la salud o la intimidad.

Las particularidades de estos escenarios conducirán, a mi juicio, a manejar en numerosos casos un concepto de causalidad de carácter estadístico, similar al que ya se planteó en el conocido caso de la Colza en España en los años 80 —o en el caso Contergan en Alemania—, en el que, para el Tribunal, aunque discutible para la doctrina, quedó acreditada la relación causal entre la ingestión de aceite de colza desnaturalizado y los graves efectos perjudiciales para la salud y la calidad de vida de las personas afectadas⁵⁰.

En la base de estos supuestos se encuentra, a mi juicio, la operatividad de una "inferencia *iuris tantum*"⁵¹ cuya carga probatoria resulta razonablemente soportable en el proceso penal. Dicha inferencia se fundamenta en la concurrencia de un conjunto plural y coherente de indicios, así como en una base científica sólida que permite ofrecer una explicabilidad suficiente y fiable del mecanismo causal subyacente⁵². No se trata, por tanto, de inferencias meramente formales o automáticas, sino de inferencias vencibles, dotadas de una elevada consistencia epistemológica, que permiten afirmar la relación de causalidad sin vulnerar las exigencias propias del estándar probatorio penal. Precisamente por su solidez indiciaria y científica, este tipo de inferencias puede considerarse compatible con los principios de presunción de inocencia e *in dubio pro reo*, al no desplazar indebidamente la carga de la prueba.

En esta línea, y salvando las necesarias distancias entre la responsabilidad civil y la penal, la propuesta de Directiva de responsabilidad civil por daños causados por sistemas de IA (Directiva RC IA) incorpora la figura de las presunciones refutables como mecanismo orientado a aliviar las dificultades probatorias en materia de causalidad, simplificando su acreditación mediante la identificación del sistema de IA implicado en el proceso de toma de decisiones. Ahora bien, la traslación acrítica⁵³ de este modelo al

ámbito penal resultaría problemática, especialmente en aquellos supuestos en los que han intervenido de forma concatenada varios sistemas de IA o cuando un mismo sistema integra distintos modelos fundacionales o de aplicación general. Por ello, la eventual flexibilización del tratamiento de la causalidad deberá depender, en última instancia, tanto del estado de la técnica como de las necesarias ponderaciones político-criminales orientadas a alcanzar resultados de justicia material, sin que ello suponga transgredir los principios fundamentales que rigen la valoración de la prueba en el proceso penal.

2.3. La íntima relación entre la imputación objetiva, la imprudencia y el riesgo permitido replicada en el uso de la IA

Superadas las dificultades inherentes al establecimiento de la causalidad y a sus problemas probatorios, el análisis debe desplazarse hacia la imputación objetiva y sus criterios normativos⁵⁴, entre los cuales el criterio del fin de protección de la norma adquiere una relevancia central en el contexto del uso de sistemas de IA. Este criterio aparece estrechamente vinculado, como lo ha estado a lo largo de la evolución de la teoría del delito, a la infracción de la norma de cuidado, es decir, a la imprudencia, en la medida en que permite valorar si en el caso concreto se ha producido un descontrol del riesgo y si el resultado lesivo materializa precisamente aquel riesgo que la norma infringida pretendía evitar en la utilización del sistema de IA⁵⁵.

En este sentido, el delito imprudente se configura como la contracara del riesgo permitido en la implantación y uso de sistemas de IA, especialmente de aquellos calificados como de alto riesgo. Ello no excluye, naturalmente, la eventual concurrencia del delito doloso⁵⁶, cuya acreditación facilitaría notablemente la afirmación de responsabilidad penal en supuestos de uso intencionalmente indebido de la IA; sin embargo, es previsible que la mayor parte de los conflictos penales surjan en el ámbito de la imprudencia⁵⁷, particularmente en sectores como el económico-empresarial,

50 José Manuel Paredes Castañón, analiza los problemas de causalidad en los procesos químicos y fisiológicos que interactúan, originándose una sinergia entre ellos que dificulta la explicación de la cadena causal. En estos casos, hay que buscar las condiciones, pese a la dificultad, en las que un proceso causal y opaco pueda ser dado por probado. "Tales condiciones son dos: primero, que exista alguna ley causal, suficientemente sólida desde un punto de vista científico, que establezca, con carácter aproximadamente determinista (esto es, no tan sólo a título de mera posibilidad, ni siquiera con una probabilidad relevante, sino con una probabilidad rayana en la certeza), que a la acción le ha de seguir el resultado; y, segundo, que no exista ninguna otra explicación causal alternativa dotada también de una base científica igualmente sólida". José Manuel Paredes Castañón, "Responsabilidad penal por productos defectuosos", *Revista Foro Fundación Internacional de Ciencias Penales*, n.º 2 (2024): 33. Jacobo Dopico Gómez-Aller, se ha ocupado recientemente de estos supuestos, poniendo el acento en sus especificidades: (1) el carácter infrecuente de la afección, (2) la acumulación inicial masiva de casos, (3) los factores cronológicos y (4) los factores geográficos, criterios que fueron acogidos por los tribunales tanto en España como en Alemania para afirmar la existencia de la relación de causalidad, apartándose de la concepción de una ley causal general defendida por Armin Kaufmann. Jacobo Dopico Gómez-Aller, "¿Dogmática o probatoria? La cuestión de la causalidad y su prueba en las intoxicaciones masivas con agentes tóxicos desconocidos", *La Ley Penal*, n.º 169 (2024): 5 ss. Esta metodología resulta trasladable al ámbito del uso de sistemas de IA, en la medida en que permite identificar características propias de su funcionamiento y de sus efectos prácticos.

51 También podría hablarse de presunción *iuris tantum*, como se sostiene en la primera versión de este trabajo, cuestión que requiere, no obstante, un análisis más detenido, en la medida en que conllevaría relevantes consecuencias procesales, como la inversión de la carga de la prueba, así como una adaptación del concepto de causalidad. Cfr. en este sentido, Silva, 2022, 159.

52 Sobre el nexo de causalidad en el ámbito de los vehículos automatizados y las complicaciones que representa el uso del algoritmo como *black box* y su capacidad de aprendizaje autónomo cfr. Isidoro Blanco Cordero, "Vehículos autónomos y responsabilidad penal", 27 ss.

53 Pese a la tentación de extrapolar al proceso penal las presunciones y la facilitación probatoria que estas comportan en otros ámbitos, dicha traslación resulta incompatible con los principios constitucionales que rigen el procedimiento penal y, en particular, con los límites estructurales del *ius puniendi*, pues no cabe admitir presunciones que operen como punto de partida normativo o metodológico con el efecto de considerar un hecho como verdadero *prima facie* sin una prueba suficiente que desplace el principio de presunción de inocencia. Ello no impide, sin embargo, que determinadas presunciones puedan desempeñar un papel limitado como marco inicial de validez, como hipótesis metodológicas débiles, subordinadas escrupulosamente a la prueba, siempre que su mantenimiento quede estrictamente condicionado a que las inferencias derivadas del material probatorio no las contradigan, inferencias que han de construirse a partir de datos objetivos y verificables —como el material fáctico acreditado, la evaluación de impacto, las medidas técnicas y organizativas adoptadas o la inexistencia de alternativas menos intrusivas— y que, en contextos de riesgo elevado, deben además confirmar positivamente dicha validez de manera sólida, coherente y técnicamente comprobable. En este sentido, no se trataría de presunciones en sentido estricto, sino de inferencias sometidas a refutación cuya fuerza probatoria depende de su contrastabilidad, lo que permite hablar, en un sentido funcional y no normativo, no es una figura o concepto normativo en sentido estricto, de inferencias *iuris tantum*.

54 Cfr. Victoria Ibold, quien pone de relieve la opacidad epistémica de los sistemas de IA, la cual dificulta la afirmación de la causalidad y remite a supuestos clásicos como los casos *Contergan*, el *Spray para cuero* o el de los productos de protección para la madera. Esta dificultad se proyecta igualmente sobre el análisis de la imputación objetiva, ámbito en el que la autora examina las teorías de la evitabilidad y del aumento del riesgo, así como la relación normativa de la infracción de un deber (*Pflichtwidrigkeitszusammenhang*). En esta relación, el concepto de deber comprende tanto el deber de cuidado como el deber impuesto al garante, dado que, conforme a la teoría de la unidad, el alcance de ambos es idéntico, (Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 380). Según Ibold, la relación de infracción del deber debe afirmarse cuando puede sostenerse con seguridad que, de haberse actuado conforme a Derecho, el resultado no se habría producido; por el contrario, dicha relación debe rechazarse cuando, aun actuando debidamente, el resultado se habría producido igualmente, (Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 381).

55 Debido a la complejidad de la estructura de la cadena de valor de los sistemas de IA, configurada como un proceso ensamblado en el que intervienen de manera secuencial diversos sujetos definidos por el RIA y, en su caso, sus delegados, la tarea fundamental consistirá en aislar la concreta infracción del deber de cuidado imputable a un determinado garante que haya dado lugar al resultado. Se trata de una labor que, en determinados supuestos, no estará exenta de dificultades, aunque en el fondo remite a problemáticas ya conocidas por la doctrina y la jurisprudencia penal. Lo verdaderamente novedoso radica en la necesidad de articular y readaptar las categorías y conceptos tradicionales a este nuevo contexto tecnológico, como sucede, por ejemplo, con la eventual reformulación del principio de confianza o con la conveniencia de restringir el alcance del concepto de imprevisibilidad en el ámbito de la imprudencia. En este sentido, Carlos M.ª Romeo Casabona, "La discusión sobre la atribución de responsabilidad penal", 64, apunta que un programa de cumplimiento que evite imprudencias en este sector tecnológico debe implementar estructuras y procedimientos que prevengan "la toma de decisiones rutinarias y no validadas directamente derivadas de las conclusiones de sistemas inteligentes".

56 Son numerosos los delitos dolosos que pueden cometerse con la ayuda de la IA por ciberdelincuentes u organizaciones criminales; sin embargo, en este ámbito la IA actúa claramente como una herramienta o medio al servicio del delincuente humano, facilitando la ejecución de su plan delictivo. Entre estos delitos destacan, en primera línea, los tradicionalmente denominados delitos informáticos o, con mayor precisión, delitos tecnológicos, denominación que refleja mejor la evolución del fenómeno y que ha desplazado a la anterior. Este tipo de delitos previsiblemente adquirirá un protagonismo creciente y un mayor potencial lesivo, debido a las elevadas tasas de efectividad que permite el uso de sistemas de IA. No obstante, la utilización dolosa de software no se limita a este ámbito, sino que también puede extenderse al control de armas letales, a la manipulación masiva de la opinión pública o a ataques especialmente intensos contra la privacidad y la intimidad de las personas.

57 En contra Andrea Bravo Bolado, la cual considera que "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" (Andrea Bravo Bolado, "El Derecho penal ante la inteligencia artificial", 206). Con el fin de evitar este riesgo de responsabilidad penal objetiva junto y la consiguiente expansión ilegítima del Derecho penal, la autora propone como alternativa la tipificación de delitos de peligro abstracto dolosos basados en la inobservancia de deberes (Andrea Bravo Bolado, "El Derecho penal ante la inteligencia artificial", 344 ss.). Sin perjuicio de que la propuesta merezca un análisis más detenido, considero, en principio, que ambas categorías dogmáticas pueden convivir, dando respuesta a distintas realidades significativas para el Derecho penal. Además, los delitos de

sanitario, educativo o el de los vehículos autónomos, donde los daños derivan con mayor frecuencia de fallos en el diseño, la implementación o la supervisión de los sistemas. El análisis de la imprudencia en el contexto expansivo de la IA exigirá, además, una creciente especialización del juzgador, habida cuenta de la proliferación de normas técnicas y jurídicas que delimitan el deber de cuidado, en particular las derivadas de los requisitos y obligaciones impuestos establecidos en el RIA (arts. 6 ss.) —especialmente en relación con los sistemas de alto riesgo, de riesgo sistémico y los modelos de IA de uso general—, cuya correcta interpretación resultará decisiva para la determinación del riesgo permitido y, en última instancia, de la imputación penal del resultado.

2.4. La imprevisibilidad de los resultados de la IA y la imprudencia

Uno de los problemas dogmáticamente más complejos que plantea el uso de sistemas de IA en el ámbito penal es la atribución de responsabilidad penal cuando el sistema de IA produce resultados inesperados, desviados o erróneos⁵⁸ caracterizados por elevados niveles de opacidad y una aparente pérdida de control por parte de sus diseñadores, desarrolladores o usuarios. En estos supuestos, es esperable que la defensa invoque de manera recurrente la imprevisibilidad⁵⁹ del comportamiento del sistema a lo largo de toda la cadena de valor como argumento exculpatorio. Sin embargo, dicha alegación no puede operar de forma sostenida e indefinida, ni convertirse en una patente de corso, pues ello conduciría bien a una irresponsabilidad estructural⁶⁰, bien a una desnaturalización de la imprudencia incompatible con el principio de responsabilidad subjetiva⁶¹.

Al mismo tiempo, el Derecho penal no puede reaccionar mediante la instauración encubierta de un régimen de responsabilidad objetiva, impropio de su lógica garantista, ni mediante una flexibilización excesiva de los requisitos de la imprudencia, puesto que de lo auténticamente imprevisible no puede exigirse humanamente la observancia del cuidado debido. La cuestión central reside, por tanto, en determinar cómo y por qué puede sostenerse una *auténtica y real imprevisibilidad jurídicamente relevante* en el contexto de la IA, lo que conducirá a afirmar un verdadero supuesto de caso fortuito, que descarte la imputación por imprudencia grave o menos grave. Esta imprevisibilidad no puede sostenerse en la mera afirmación y constatación de la existencia de una intrincada y opaca⁶² complejidad técnica desde su programación pasando por las fases

del despliegue y funcionamiento del sistema de IA, característica que es de todo punto intrínseca al mismo, y que conducirá a la validación del juicio de imprevisibilidad en la mayoría de los casos desde un análisis quizás tanto *ex ante* como *ex post*. El juicio de imprevisibilidad del resultado, de la producción de un daño, debe basarse en la constatación de haber mantenido desde las fases iniciales de su programación y selección de objetivos y funciones del sistema hasta el momento de producción del daño un comportamiento diligente orientado a preservar la transparencia y explicabilidad del código fuente, la selección de los datos, del desarrollo del software, el control del funcionamiento y de resultados obtenidos por el sistema, así como la detección de vulnerabilidades, para mantener niveles óptimos de ciberseguridad, en evitación de ciberataques⁶³. En definitiva, se trata de que todos los integrantes de la cadena de valor hayan actuado de forma diligente y conforme al tipo de sistema de IA⁶⁴, al estado de la ciencia y de la técnica, tomando todas las medidas razonables de precaución⁶⁵ durante el entrenamiento, la validación, la prueba y el funcionamiento del sistema durante toda su vida útil, con sometimiento a controles periódicos internos y externos.

Ami juicio, desde esta perspectiva, pueden formularse al menos dos criterios normativos destinados a descartar una imprevisibilidad meramente ficticia como alegato defensivo. En primer lugar, resulta exigible que, con carácter previo a la comercialización o puesta en servicio del sistema, se hayan llevado a cabo entrenamientos, validaciones y pruebas de seguridad de alta calidad, conformes con el estado de la técnica y orientados a identificar comportamientos desviados o potencialmente lesivos. En una segunda fase, ya de comercialización y poscomercialización, deben cumplirse de manera continuada los requisitos y obligaciones impuestos por el RIA⁶⁶ —y el desarrollo del mismo en reglamentos y protocolos de cada sistema de IA en particular— a lo largo de todo el ciclo de vida del sistema, con el fin de controlar, vigilar y, en su caso, corregir sus resultados.

Aunque las actividades de investigación, prueba y desarrollo de sistemas o modelos de IA queden excluidas⁶⁷ del ámbito de aplicación del RIA, hasta la introducción en el mercado⁶⁸ o puesta en servicio del sistema, dicha exclusión no legitima que las pruebas se realicen bajo cualquier circunstancia. La ausencia de pruebas seguras y de un entorno adecuado constituye, en mi opinión, una infracción inicial⁶⁹ de las normas de cuidado, que impedirá alegar posteriormente la imprevisibilidad de la respuesta del sistema. Incluso, se puede indagar sobre si “la situación de desconocimiento” se ha buscado intencionadamente, al no investigar los fallos del sistema y mantener la

peligro operan *ex ante*, mientras que la imprudencia puede hacerlo *ex post*. En particular, la imputación por imprudencia permite dar adecuada respuesta a supuestos en los que se produce un resultado lesivo efectivo, perceptible en la existencia de un homicidio, lesiones u otros ataques a derechos fundamentales como la intimidad. Cfr. en este sentido tb. n. 90 con referencias a la participación imprudencia.

58 Desde la perspectiva de evaluación de beneficios, los sistemas de IA contribuyen eficazmente a mejorar el nivel de diligencia al ampliar la previsibilidad de producción de resultados lesivos en un ámbito de actividad, como por ejemplo, el médico-quirúrgico, cfr. al respecto, M.ª Ángeles Rueda Martín, “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*, n.º 1 (2025): 195.

59 Se evidencia una característica propia de este contexto a través de los oximoron “previsible imprevisibilidad” o “imprevisibilidad previsible” (cfr. Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 33, habla de paradoja típica), que describe una situación de incertidumbre temporal y modal, ya que se tiene la certeza de que el sistema fallará, pero no se sabe con precisión cuándo y cómo lo hará. Se trata de una previsibilidad del riesgo con imprevisibilidad operacional, es decir, previsibilidad general del riesgo, pero imprevisibilidad concreta del daño.

60 Para Isidoro Blanco Cordero, “este escenario da lugar a la ya mencionada brecha de responsabilidad penal” (Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 34). Esta laguna de punibilidad cuestiona la efectividad del sistema penal, no resultando convincente “que quienes ponen en el mercado un producto cuyo riesgo no puede controlarse de forma segura puedan eludir su responsabilidad alegando simplemente la imprevisibilidad del producto “inteligente”. Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 35.

61 Sobre el juicio de previsibilidad cfr. Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 33 ss.

62 Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 36, considera que las dificultades en el juicio de previsibilidad son estructurales, derivadas de la opacidad interna de los algoritmos de *machine learning*. Aunque se reduzca el riesgo con medidas técnicas y estándares, sin embargo, resultará prácticamente imposible aislar la contribución preventiva de cada estándar concreto respecto a un resultado lesivo específico, “lo que genera incertidumbre sobre la eficacia de las medidas de precaución individuales. Desde la perspectiva penal esta incertidumbre amenaza la solidez del requisito de previsibilidad, que es central para la imputación de un resultado por imprudencia. Si no puede determinarse con razonable certeza qué peligros la regla de cuidado debía evitar y cómo los mecanismos de control del algoritmo contribuyen a mitigarlos, la valoración de la imprudencia del productor se vuelve extremadamente compleja”.

63 Cfr. Paz M.ª de la Cuesta Aguado, “Inteligencia artificial y responsabilidad penal”, 57, y su análisis y desarrollo de los principios jurídicos sobre IA.

64 Los protocolos que establezcan las normas de procedimiento se deben adaptar al sistema de IA para conseguir estándares de actuación diligente y eficaz. En este sentido, Paz M.ª de la Cuesta Aguado, apunta que “respecto de algoritmos con capacidad de aprendizaje y decisión, el control de calidad ha de ser riguroso y exhaustivo, lo que exigirá reflexionar y prever los parámetros de seguridad en dicho control de calidad y, probablemente, con base en ellos, fundamentar nuevas modalidades de exigencia de responsabilidad penal”. Paz M.ª de la Cuesta Aguado, “Inteligencia artificial y responsabilidad penal”, 58.

65 Considero necesario construir un principio de precaución riguroso y reforzado. En un sentido similar, Andrés Boix Palop, “Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones”, *Revista de Derecho Público: Teoría y Método*, n.º 1 (2020): 227, propone una noción de precaución más radical, capaz de adaptarse a las nuevas necesidades y exigencias del entorno tecnológico. La vertiente funcional del concepto de precaución permite su tránsito desde un principio abstracto hacia una regla práctica de comportamiento, configurando los deberes de prevención y cuidado exigibles, de naturaleza especial (deberes especiales de cuidado) si se trata de una posición de garante, en función del nivel de riesgo, de la capacidad de control del sujeto y de la gravedad potencial de los efectos.

66 Cfr. el propio RIA, el cual regula en los capítulos I y II del título I, especialmente los arts. 9 y ss. una serie de deberes y precauciones dirigidos establecer, implementar, documentar y mantener un sistema de gestión de riesgos en relación con los sistemas de IA de alto riesgo, sus datos de entrenamiento y su gobernanza y su precisión y solidez para alcanzar un nivel adecuado de ciberseguridad.

67 Con el fin de apoyar la innovación, respetar la libertad de ciencia y no socavar la actividad de investigación y desarrollo.

68 Considerando 25 RIA.

69 A no ser que se demuestre la absoluta imprevisibilidad, aún tomando todas las medidas preceptivas técnicas, sumadas a las derivadas de la experiencia y conocimiento al alcance.

falta de pronóstico o predictibilidad, dando lugar a un posible supuesto de ignorancia deliberada, que nos situaría incluso en el campo del dolo eventual. Por ejemplo, no sería aceptable invocar la falta de previsibilidad de un sistema de IA utilizado para la selección de personal en una empresa si, durante el periodo de pruebas y experimentación, no se han implementado prácticas adecuadas de gestión y gobernanza de datos que garanticen un entrenamiento, validación y pruebas de alta calidad, excluyendo posibles sesgos que condujeran a un resultado discriminatorio en la toma de decisiones⁷⁰. Por consiguiente, tanto la experimentación como la comercialización de los sistemas de IA deben desarrollarse en un entorno de control, con vigilancia y supervisión de alta calidad, que permita la anticipación de resultados. Los niveles de exigencia en el tratamiento de esta tecnología deben ser muy altos, reduciendo al mínimo el margen para la imprevisibilidad⁷¹.

En segundo lugar, los sistemas de IA de alto riesgo no deberían implementarse hasta que se hayan desarrollado tecnologías y protocolos que permitan devolver al sistema, en situaciones críticas, un ámbito seguro de control humano. Quien experimente, comercialice o use un sistema de IA sin tecnologías de aseguramiento en situaciones críticas no podrá alegar la falta de previsibilidad, ya que conoce los riesgos asociados y la imposibilidad de intervenir, por ejemplo, desconectando o desprogramando el sistema.

Estos dos requisitos, considero, sirven para mantener un sensato control sobre el sistema, garantizar estándares de previsibilidad acordes con los riesgos, *previsibilidad razonable*, anticipando situaciones y resultados de alto impacto sobre la seguridad, salud y derechos humanos. En definitiva, acceder a *cotas de predecibilidad controlada* del sistema, solo es posible mediante un estudio y análisis exhaustivos que generen conocimiento sobre el comportamiento del software, y sobre cómo retomar el control en caso de comportamientos inesperados.

En conclusión, la imprudencia de los sujetos de la cadena de valor se materializa en dos niveles: primero, por no adoptar las medidas necesarias durante los periodos de prueba y funcionamiento de los sistemas de IA para alcanzar estándares de previsibilidad de sus respuestas y no implementar mecanismos de contención del riesgo; y segundo, por infringir normas de cuidado derivadas del cumplimiento de los requisitos y obligaciones establecidos en el RIA y sus normas y protocolos de desarrollo adaptados a cada sistema de IA concreto. Estas conductas imprudentes recibirán normalmente una respuesta sancionatoria desde el Derecho administrativo. Si se trata de imprudencias menos graves o graves, ambas con una tipificación muy acotada en el Código penal, se activará la respuesta penal con las consiguientes consecuencias penales.

2.5. Autoría y responsabilidad penal de la persona física en el uso de sistemas de IA

En el contexto de sistemas de inteligencia artificial, es posible identificar inicialmente un círculo cerrado de sujetos activos, compuesto por el fabricante⁷², proveedor, el

importador, el distribuidor y el responsable del despliegue del sistema⁷³. Muchos de los delitos actualmente tipificados, cuando resulten aplicables, solo podrán ser cometidos por alguno de estos sujetos, lo que restringe la autoría y convierte ciertos delitos comunes por esta razón en delitos especiales.

Entre estos sujetos, *el responsable de despliegue*, *Al deployment officer*, ocupa un papel central, al ser quien se encarga de supervisar, coordinar y garantizar que un sistema de IA se implemente de manera segura, efectiva y ética dentro de una organización, cumpliendo objetivos y normas regulatorias para su uso efectivo. Los sujetos situados en fases previas de la cadena —fabricante, proveedor, importador y distribuidor— desarrollan el sistema, lo introducen en el mercado o lo comercializan, sin intervenir directamente en su implementación y uso final. No obstante, la complejidad organizativa, incluso en pequeñas y medianas empresas, implica que el responsable de despliegue, cuando coincide con la figura del empresario o del órgano de dirección, pueda delegar la implementación y utilización concreta del sistema, ampliando así el círculo de posibles autores.

A esta dinámica de delegación se añade un segundo elemento decisivo: la alta tecnificación de los sistemas de IA, que impone necesariamente un trabajo colectivo, especializado y funcionalmente distribuido. La implementación, supervisión y utilización de estos sistemas exige la intervención de múltiples perfiles técnicos y operativos, lo que hace inviable un control individual y centralizado del proceso. Como consecuencia, las tareas se fragmentan y se asignan a distintos sujetos o equipos, cada uno con un ámbito propio de decisión y actuación⁷⁴.

Es precisamente esta combinación de delegación y tecnificación la que conduce, en un estadio posterior, a un escenario que la doctrina ha descrito como el problema de los "many hands"⁷⁵, caracterizado por la intervención de múltiples agentes, cada uno con una contribución parcial al resultado, pero sin un dominio pleno del conjunto del proceso. Este fenómeno⁷⁶ no constituye, por tanto, una anomalía, sino una consecuencia estructural del modo en que se organizan y operan los sistemas de IA.

Desde el punto de vista de la imputación penal, esta realidad incrementa la dificultad de individualizar responsabilidades, pero no la excluye. Así, en el uso de la delegación, los delegados pueden asumir responsabilidad directa por la lesión de los bienes jurídicos de los usuarios finales, mientras que el delegante conserva una responsabilidad residual, configurada como participación imprudente, al no poseer el dominio objetivo del hecho con su actuación⁷⁷, y por infringir deberes de vigilancia e inspección de las funciones

73 Cfr. Eric Hilgendorf, "Inteligencia artificial y Derecho penal", trad. Leandro Dias Lestón, en *Desafíos penales de hoy: Entre la ley y la justicia en la obra de Eric Hilgendorf* (Buenos Aires: Editores del Sur, 2024), 24 ss.

74 Ello no impide la utilización de la figura de la autoría única directa cuando se trata de un sujeto que comete el delito mediante un sistema de IA, que usa como instrumento dentro de su plan delictivo. En este sentido, Carlos M.ª Romeo Casabona, "La discusión sobre la atribución de responsabilidad penal", 67 y s. considera el robot o sistema de IA como mero instrumento de la acción humana en casi todos los casos, incluso los sistemas inteligentes autónomos, ya que deberán estar sometidos a un Control Humano Significativo (CHS). Considerando la tesis de la autoría mediata cuando se utiliza un sistema de IA como instrumento, tesis que bajo los parámetros actuales considero incorrecta, cfr. Francisco Vega Agredano, "Detección del delito mediante la inteligencia artificial: prevención automática de la criminalidad y responsabilidades por fallos del sistema", *Revista General de Derecho Penal*, n.º 44 (2025): 45.

75 Tal como ha puesto de relieve Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 235 ss. al analizar la opacidad epistémica y las dificultades de imputación en estos contextos. Cfr. Isidoro Blanco Cordero, "Vehículos autónomos y responsabilidad penal", 43 ss. y la propuesta del denominado garante técnico (p. 45), como sujetos internos de la organización empresarial, que tienen atribuidas tareas de control sobre la seguridad del sistema.

76 Este fenómeno no es nuevo para el Derecho penal y ya ha sido tratado en el Derecho penal económico y medioambiental o incluso en el ámbito de la imprudencia médica con divisiones verticales u horizontales de división del trabajo. Quizás se pueda resaltar que en este contexto tecnológico la atomización de tareas sea más intensa y compartimental.

77 Raquel Roso Cañadillas, *Autoría y participación imprudente* (Granada: Comares, 2002), 501 ss.; y, "Los delitos polivalentes de autoría: entre el deber y el dominio", *InDret*, n.º 3 (2019): 28 ss.

70 Considerando 67 RIA.

71 Aclaro que no se trata de dar entrada por esta vía al principio de precaución en el Derecho penal como fundamento a una aplicación expansiva por la posibilidad de que se presenten riesgos ignotos, cfr. Gonzalo Quintero Olivares, "La Robótica ante el Derecho penal: el vacío de respuesta jurídica a las desviaciones incontroladas", *Revista Electrónica de Estudios Penales y de la Seguridad*, n.º 1 (2017): 19 ss.

72 Sobre el deber de cuidado del fabricante en general, cfr. Isidoro Blanco Cordero, "Vehículos autónomos y responsabilidad penal", 10 ss.

delegadas⁷⁸.

En este contexto de intervención múltiple, la autoría puede analizarse mediante categorías dogmáticas ya conocidas, como la *coautoría aditiva o la coautoría sucesiva*, que exigen la concurrencia de un acuerdo común como elemento subjetivo de conexión, normalmente presente en equipos de trabajo coordinados. No obstante, en otros supuestos, la actuación de los distintos intervinientes será concurrente pero independiente, sin concierto previo, lo que dará lugar a *autorías yuxtapuestas o accesorias*⁷⁹, basada en la *concurrencia de imprudencias individuales*.

Finalmente, cabe distinguir dos planos diferenciados de responsabilidad, sustentados en fundamentos jurídicos distintos. Por un lado, *la responsabilidad por producto defectuoso*, de carácter objetivo, que recae primordialmente sobre el fabricante⁸⁰ del sistema de IA, sin perjuicio de la posible imputación al importador o distribuidor en los términos legalmente previstos, cuando el daño derive de defectos de diseño, programación o información del sistema. En este ámbito, tanto el usuario final como el responsable del despliegue pueden ocupar la posición de sujetos pasivos, en la medida en que resulten perjudicados por un defecto inherente al producto.

Por otro lado, cabe apreciar una responsabilidad basada en la implementación y/o uso imprudente del sistema de IA. Esta se configura como el incumplimiento de deberes de cuidado (en su caso de carácter especial) que incumben al responsable del despliegue, quien puede encontrarse en posición de garante, cuando el daño sea consecuencia de una implementación, supervisión o uso inadecuados del sistema de IA. Esta responsabilidad no se articula a través del régimen de responsabilidad por producto, sino sobre la base de la imprudencia⁸¹, salvo en aquellos supuestos en los que el responsable del despliegue haya llevado a cabo modificaciones sustanciales del sistema o lo haya reconfigurado bajo su propio control, asumiendo entonces una posición equiparable a la del productor y quedando, en consecuencia, sometido al régimen de responsabilidad por producto defectuoso.

2.6. ¿La persona electrónica (e-person) como sujeto activo del delito?

Desde una perspectiva distinta, se ha planteado la posibilidad de que los propios sistemas de IA puedan llegar a ser considerados sujetos activos del delito⁸², en atención

al elevado grado de autonomía técnica⁸³ que puedan adquirir una vez entrenados y puestos en funcionamiento. Esta tesis suele apoyarse en escenarios en los que un sistema de IA influya decisivamente en el comportamiento humano o actúe sin intervención inmediata de una persona física, como ocurre en determinados accidentes protagonizados por vehículos autónomos⁸⁴. A este respecto, cabe recordar los numerosos accidentes registrados en Estados Unidos desde 2016⁸⁵, fecha en la que se produjo el primer siniestro mortal asociado a un vehículo autónomo, así como otros supuestos en los que la IA ha sido empleada como instrumento determinante en la comisión de ilícitos⁸⁶. Sirva de ejemplo el conocido caso de Hungría, en el que se utilizó tecnología de síntesis de voz para simular la voz del CEO de una empresa del sector energético y lograr así la transferencia urgente de más de 200.000 euros a un proveedor⁸⁷. Llevando este razonamiento al extremo, podría imaginarse un escenario en el que un sistema de IA o un robot con tecnología *deep learning*⁸⁸, dotado de técnicas avanzadas de

Cuesta Aguado, "Inteligencia artificial y responsabilidad penal", 55 ss. desde una perspectiva normativa nos muestra como el sistema de IA realiza comportamientos con significado social, que deben adecuarse a normas. Desde este punto formula una definición de *software* como "sistema normativo de pautas de conducta que pueden generar comportamientos con trascendencia social", avanzando en el análisis, ese complejo conjunto de pautas de conducta denominado *software*, "configura el "código" de instrucciones que posibilitan la realización de tareas específicas que tiene encomendadas o para las que tiene competencia el agente". Este sistema normativo con mandatos y prohibiciones está sometido a principios jurídicos como la transparencia en el código, la previsibilidad en las decisiones y ciberseguridad para evitar las manipulaciones del *software*. Este tránsito desde la programación del *software* por una persona o grupo de personas humanas hasta los principios jurídicos mencionados que involucran claramente a una persona humana, conducen, para de la Cuesta, a que la responsabilidad por la decisión adoptada por el agente inteligente sea de la persona, que responderá de los perjuicios por él causados. Por su parte, Alejandra Morán Espinosa, "Responsabilidad penal de la Inteligencia artificial (IA). ¿La próxima frontera?", *Revista del Instituto de Ciencias Jurídicas de Puebla* 15, n.º 48 (2021): 289 ss., considera "que urge reconocer a la IA como un campo de estudio independiente y autónomo" y propone "la regulación de la IA como *la próxima frontera regulatoria*, abonado a su vez a su estudio jurídico, considerando que una IA al no estar permeada de la natural subjetividad humana, es un hecho, lógico, posible y probable que cualquier información negativa, abusiva, inadecuada, indeseable y hasta ilegal, en forma de conocimientos, le fuera proporcionada y programada para cometer delitos, lo que sería el primer elemento necesario para debatir sobre la probable determinación de responsabilidad de una IA en materia penal"; Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 248 ss. se plantea la hipótesis de una responsabilidad de la IA en vez de una responsabilidad por el producto; sin embargo, a través de una síntesis analítica de diferentes posiciones de autores, descarta tal planteamiento; tb, se ocupa de esta cuestión Fernando Miró Llinares, prestando especial atención al carácter autónomo de los sistemas de IA. El autor niega que estos sean autónomos en términos morales, "pero sí pueden tener algún nivel de autonomía funcional, en el sentido de algún grado de independencia en sus acciones del control humano" (Fernando Miró Llinares, "Derecho penal y desafíos de la inteligencia artificial (en el contexto de nuevo marco regulatorio europeo)", en *Derecho y tecnologías*, dir. Maria Emilia Casas Baamonde, coord. Daniel Pérez del Prado (Madrid: Editorial Centro de Estudios Ramón Areces, 2025), 193). Esta circunstancia unida a la dificultad de prever su funcionamiento por parte de los integrantes de la cadena de valor y a la progresiva automatización de los procesos de toma de decisiones, le conduce a formular dos consideraciones. La primera sitúa el momento clave para la imputación de responsabilidad "en las primeras etapas del ciclo de la vida de la IA, como el diseño del algoritmo o la selección de los datos que alimentarán el sistema". La segunda se deriva de la cesión de una o varias fases del proceso decisorio a un ente no humano capaz de "actuar autónomamente", y sin supervisión humana, lo que podría dar lugar a daños causados por sistemas de IA sin posibilidad de atribución a una persona humana, "por haber sido perpetrados en el marco de la autonomía de las máquinas y no ser capaces de prever el daño"; pero tampoco a las máquinas porque no son agentes "morales" a los que se les puede hacer responsables, abriendo así posibles lagunas de responsabilidad. Ambas consideraciones resultan de interés y contribuyen a la reflexión, si bien la primera no constituye una novedad para el Derecho penal, pues en una sociedad del riesgo y tan compleja como la actual, la generación del peligro suele encontrarse alejada de la producción del resultado, como ocurre en supuestos de decisiones colegiadas en el ámbito empresarial, que terminan causando un daño al medio ambiente o a la salud pública, o en el suministro de maquinaria defectuosa, que es utilizada por el trabajador que sufre un accidente. Además, a lo largo de toda la vida útil del sistema de IA pueden surgir situaciones de riesgo, que exijan una actuación profesional diligente, por lo que la búsqueda de responsables no debería limitarse exclusivamente a las fases iniciales de su diseño y programación. La segunda consideración es muy sugerente e invita a reflexionar sobre el concepto de autonomía funcional y su capacidad para destituir la responsabilidad de la persona humana. Atendiendo al desarrollo actual de los sistemas de IA, la autonomía funcional no parece suficiente ni para fundamentar una responsabilidad penal de la denominada e-persona, ni para excluir la responsabilidad del agente humano. Por otro lado, dicha autonomía funcional debe existir dentro del proceso, porque de lo contrario haría ineficaz el uso del sistema de IA, debiendo estar supeditada a protocolos de control adecuados, cuya corrección y suficiencia excluirían la imputación penal por falta de previsibilidad. No se trata, por tanto, de la existencia real de una responsabilidad penal sin sujeto al que atribuirlo. Para un análisis más detallado sobre las lagunas (o brechas) de responsabilidad, cfr. Andrea Bravo Bolado, "El Derecho penal ante la inteligencia artificial", 42 ss, así como la posibilidad de castigo de la e-persona, Andrea Bravo Bolado, "El Derecho penal ante la inteligencia artificial", 73 ss.

83 Un ejemplo de autonomía completa de un vehículo lo encontramos en los microchips de la empresa Nvidia, con un algoritmo cuyo entrenamiento consiste en observar la forma de conducción de un humano. Cfr. Disponible en: <https://www.caranddriver.com/es/coches/planeta-motor/a60310324/inteligencia-artificial-nvidia-coche/>.

84 Cfr. Isidoro Blanco Cordero, "Vehículos autónomos y responsabilidad penal", 4 ss.

85 Laura G. de Rivera, "Coches autónomos: ¿quién tiene la culpa en caso de accidente?", *Público DISPLAY CONNECTORS, SL.*, 9 de marzo de 2024, <https://www.publico.es/ciencias/coches-autonomos-culpa-caso-accidente.html>.

86 Sobre ciberdelitos en los que se puede utilizar la IA como instrumento cfr. Emilio José Armaza Armaza et al., "Ciberseguridad, ciberdelitos e inteligencia artificial: apuntando al futuro", en *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial*, ed. Carlos M.º Romeo Casabona y M.º Ángeles Rueda Martín (Granada: Comares, 2023), 170 ss.; Francisco Jiménez García, dir., *Seguridad y responsabilidad penal e internacional en el uso de las TIC y la inteligencia artificial*, coord. Sandra López de Zubiria Díaz y Berta Alam-Pérez (Madrid: Iustel, 2024), 17 ss.; Marina Minguez Rosique y David Gallego Arribas, coords., *Cibercrimen: Tendencias y desafíos actuales* (Madrid: BOE, 2025), 9 ss.; Fernando Miró Llinares, "Derecho penal y desafíos de la inteligencia artificial", 200 ss.

87 Disponible en: <https://www.welivesecurity.com/la-es/2019/09/11/estafadores-utilizan-inteligencia-artificial-imitar-voz/>.

88Cfr. sobre la descripción de esta tecnología, Disponible en: <https://www.ibm.com/es-es/topics/deep-learning#:~:text=El%20deep%20learning%20es%20un,de%20decisiones%20del%20cerebro%20humano,>

78 Un ejemplo que recoge el propio RIA se encuentra en la figura de los responsables de despliegue, quienes tienen entre sus funciones garantizar, mediante medidas técnicas y organizativas adecuadas, que los sistemas de IA de alto riesgo se utilicen conforme a las instrucciones de uso; dentro de estas medidas se incluye la necesidad de realizar una delegación eficaz y segura, de modo que las personas encargadas de poner en práctica las instrucciones de uso y de llevar a cabo la supervisión humana prevista en el RIA cuenten con las competencias necesarias, especialmente un nivel adecuado de alfabetización, formación y autoridad en materia de IA.

79 En el análisis de la autoría debe ponderarse que estos procesos se estructuran mediante una división del trabajo, y que uno de los principios que rigen la colaboración y la convivencia humana es el principio de confianza, que a su vez delimita los conceptos de riesgo permitido y de imprudencia. Este principio, en consecuencia, restringiría las autorías accesorias o yuxtapuestas en este contexto. Cfr. Victoria Ibold, *Künstliche Intelligenz und Strafrecht*, 347 ss.

80 Cuya responsabilidad se basaría en una *culpa in programando*. En este sentido, Eloy Velasco, "Delitos tecnológicos de los informáticos a los cometidos por la IA", conferencia de clausura en la 7.ª edición del Máster de Experto de Derecho Digital de la Universidad de Deusto, 22 de marzo de 2024, accedido 4 de octubre de 2024, https://www.youtube.com/watch?v=uGdOmQ_sQVg.

81 Salvando las distancias este elenco de sujetos activos y con un fundamento similar de su responsabilidad recuerda a la Ley 31/1995 de prevención de riesgos laborales que habla de fabricantes, importadores y suministradores, por un lado, siendo responsables por la falta de viabilidad y defectos de su producto (art. 41 LPRL), y del empresario, como el principal deudor de la seguridad de los trabajadores en el desempeño de sus actividades.

82 Sobre la viabilidad de una responsabilidad penal de los sistemas de IA, cfr. Eric Hilgendorf, "Können Roboter schuldhaft handeln?", en *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs* (Baden-Baden: Nomos, 2012), 119-132; tb. Eric Hilgendorf, "Vom Werkzeug zum Partner? Zum Einfluss intelligenter Artefakte auf unsere sozialen Normen und die Aufgaben des Rechts", en *Digitalisierung, Globalisierung und Risikoprävention. Festschrift für Ulrich Sieber* (Berlin: Duncker & Humblot, 2022), 767-778, el cual defiende la posibilidad de afirmar los elementos de la acción y la culpabilidad en los sistemas de IA o robots; Paz M.ª de la

aprendizaje automático, decidiera de forma completamente autónoma llevar a cabo conductas como la simulación de la voz de un directivo para cometer estafas a gran escala, la difusión de publicidad engañosa con impacto masivo sobre los consumidores o la manipulación de mercados financieros mediante la simulación de caídas bursátiles. Incluso se ha advertido, desde una perspectiva más prospectiva, del riesgo de que tecnologías de este tipo puedan ser utilizadas para simular comunicaciones oficiales de altos dirigentes de Estados con capacidad de disuasión nuclear, con potenciales consecuencias graves para la seguridad internacional.

Estas conductas podrían, en abstracto, ser calificadas como realizadas por sistemas de IA autorresponsables, posibilidad que resulta técnicamente concebible en el estado actual⁸⁹ del desarrollo de la IA. Sin embargo, atendiendo a sus características esenciales, los sistemas de IA se limitan a emular determinadas capacidades cognitivas humanas —e incluso a superarlas⁹⁰ en ámbitos concretos—, pero carecen de elementos fundamentales para la imputación penal. En particular, no poseen intencionalidad en sentido jurídico-penal, ni sentido común, capacidad reflexiva, intuición o conciencia del entorno, ni tampoco una conciencia de sí mismos. Asimismo, están privados de dimensiones inherentes a la experiencia humana, como la conciencia moral —en sus vertientes de buena o mala conciencia—, el arrepentimiento, la comprensión del sufrimiento ajeno, del daño causado o del significado del perdón, así como, de manera especialmente relevante, de la empatía. Estas carencias no son accidentales, sino estructurales: los sistemas de IA operan sobre la base del procesamiento algorítmico de datos y carecen de una dimensión emocional o volitiva que permita equipararlos, siquiera parcialmente, a un sujeto penalmente responsable.

Pese a todo, desde un punto de vista estrictamente fáctico, no puede negarse que los sistemas de IA son capaces de ejecutar conductas que, de haber sido realizadas por una persona física, resultarían penalmente relevantes. Sin embargo, la cuestión decisiva no radica en la mera producción de resultados típicos, sino en determinar si la IA puede ser reconocida como un sujeto autorresponsable al que sea posible imputar los resultados lesivos derivados de una conducta dolosa o imprudente y, en consecuencia, imponerle una consecuencia jurídico-penal conforme a los presupuestos de la teoría del delito.

Siguiendo los elementos de la teoría del delito cabría debatir si existe acción en un algoritmo, e incluso, flexibilizando notablemente el concepto, admitir algún concepto de acción adaptado⁹¹, que permita afirmar la existencia de una conducta externa controlada por la voluntad electrónica. No obstante, admitiendo esta posibilidad en términos puramente funcionales, los problemas se agravan al analizar los elementos subjetivos del delito. Aunque un sistema de IA pueda realizar hechos objetivamente típicos y antijurídicos, resulta altamente cuestionable afirmar que esté dotado de intencionalidad⁹² en sentido jurídico-penal, incluso si se defendiese un concepto de dolo objetivamente malo. La IA carece de conciencia comprensiva del significado del hecho, esto es, de la capacidad de saber y entender que está lesionando un bien jurídico protegido, con independencia de que pueda procesar datos o anticipar consecuencias de manera estadística.

No obstante, cabe admitir, a modo de hipótesis prospectiva, que el desarrollo tecnológico pudiera conducir en un futuro a sistemas de IA capaces de captar el sentido de sus propios actos o incluso a humanoides dotados de réplicas funcionales del cerebro humano, en el marco de una quinta revolución, la neurocientífica, vinculada al desciframiento del funcionamiento cerebral. En tal escenario, y desde determinadas teorías de la imputación subjetiva basadas exclusivamente en el elemento cognitivo —como las teorías de la representación—, podría llegar a plantearse la construcción de una noción de dolo aplicable al comportamiento de las denominadas “personas electrónicas”. En efecto, la capacidad de la IA para calcular probabilidades de éxito y para prever la conversión de una acción peligrosa en un resultado lesivo podría ser extraordinariamente elevada. Sin embargo, la cuestión decisiva seguiría siendo si la máquina es capaz de representarse de forma consciente la totalidad del curso causal y del proceso lesivo, esto es, si puede generar representaciones dotadas de significado y conciencia del peligro y de sus consecuencias. En este sentido, aunque la IA pudiera aparecer como un sujeto especialmente adecuado para las teorías puramente cognitivas del dolo, ello pondría de manifiesto, a mi juicio, que el dolo humano es algo más que mero conocimiento. En efecto, desde concepciones duales del dolo, que integran de manera estrechamente conectada un elemento cognitivo y otro volitivo⁹³, la decisión del sujeto se adopta a partir de motivaciones personales en las que el conocimiento se encuentra mediado por la emoción y la voluntad. Desde esta perspectiva, la afirmación del dolo —e incluso de la imprudencia consciente, dada la delgada línea, a veces difusa, que la separa del dolo eventual⁹⁴, o de la imprudencia inconsciente— no resultaría trasladable

89 Hay que distinguir entre el *machine learning* que conforma una IA predictiva a través de algoritmos que identifican patrones entre millones de datos, y el *deep learning*, en el que la IA se entrena y experimenta un proceso de aprendizaje por cuenta propia, reconociendo pautas mediante muchas capas de procesamiento, lo que le permite tomar decisiones propias, aprender y rectificar la programación errónea. La primera forma de IA es una herramienta por lo que será la persona humana la responsable; así el programador, fabricante, usuario, etc. Los sistemas de *deep learning* son los que presentan un reto para el análisis de responsabilidades, debido a que no se sabe cuál va a ser su evolución y sus capacidades. La perspectiva es que se conviertan en entes autónomos cuyos comportamientos no estén totalmente controlados por el ser humano, como ocurre con armas autónomas letales, que pueden tomar decisiones igualmente letales sin intervención humana. Además es necesario diferenciar entre (i) las tecnologías de software puro como la IA, que al fin y a la postre es silicio metido en un servidor, (ii) la robótica, que está en un estadio muchísimo menos avanzado, debido a que detrás de ella es necesaria mucha más programación y, (iii) los híbridos en los que nos podremos convertir en un futuro, según mantiene, como he expuesto anteriormente, Rafael Yuste, neurobiólogo español, quien sostiene que nuestro procesamiento mental y cognitivo se podrá ampliar en el futuro, cuando se mapeen los circuitos neurales, mediante interfaces cerebro-computadora que aumentarán las capacidades cognitivas del ser humano, implantando chips cerebrales. La última y cuarta posibilidad, que se puede anticipar, surge de la combinación entre la robótica y la neurociencia, cuando sea posible construir un cerebro por medio de la biología sintética con axones, dendritas, lípidos y proteínas, y se hagan funcionar sus circuitos neuronales. En conclusión, en un futuro, podría existir el siguiente elenco de sujetos activos del delito: persona humana, persona humana híbrida, persona jurídica, sistema de IA, robot y humanoide. Los tres últimos podrían agruparse bajo el término persona artificial o persona electrónica o persona automática, por proponer terminología, y salvando también las distancias entre ellos, ya que la IA no es un cerebro, sino hasta la fecha un *instrumento* potente y cada vez más útil, siempre que sea éticamente utilizado; mientras que los casos del robot, humanoide y ser humano híbrido se podrían catalogar como formas del transhumanismo. (Cfr. Israel Zaballa y Sergio Enriquez-Nistal, “Rafael Yuste Así funciona nuestro cerebro según el mejor neurocientífico español: “Es una máquina de predicción del futuro: alguien inteligente ve venir las cosas antes que tú””, *El Mundo*, 20 de septiembre de 2024, <https://www.elmundo.es/papel/lideres/2024/09/20/66ed8390fdddf0e518b458b.html>).

90 Como expone María del Mar Méndez Serrano, “Derechos fundamentales”, 56 y s., las máquinas son más eficientes en muchos aspectos, “pero los humanos somos imbatibles en relaciones sociales, creatividad, adaptación a situaciones no previstas, o manipulación diestra, entre otras habilidades. Aunque pueda parecer que no hay límites en las capacidades de la IA, no están dotadas de sentido común, ni aún saben cómo llegar a ese resultado”. “Por lo tanto, la coherencia, el ingenio y la intención que advertimos en las máquinas son puro espejismo y nos llevan a proyectar estados mentales, donde no hay mente alguna”. María del Mar Méndez Serrano, “Derechos fundamentales”, 59.

91 Cfr. Eric Hilgendorf, “Können Roboter schuldhaft handeln?”, en *Jenseits von Mensch und Maschine* (Baden-Baden: Nomos, 2012), 125 s.; Isidoro Blanco Cordero, “Un Derecho penal de humanos para entidades artificiales inteligentes. ¿*Machina delinquere et puniri potest?*”, en *Manifestaciones de desigualdad en el sistema de justicia penal*, dir. Antonio Doval Pais y Elena Gutiérrez Pérez, vol. I (Pamplona: Aranzadi, 2023), 376 ss.

92 En este sentido María del Mar Méndez Serrano, “Derechos fundamentales”, 56, citando literatura especializada se hace eco de sus razonamientos en los que se niega que las inteligencias artificiales tengan intencionalidad, debido a que no son más que “el reflejo de las intenciones y sesgos de los agentes morales que las han creado, los humanos, y no las máquinas”.

93 Cfr. sobre el dolo y el elemento volitivo, Mercedes Pérez Manzano, “Algunos datos empíricos sobre la atribución de estados mentales: ¿fracaso del principio de responsabilidad subjetiva o de un determinado concepto de dolo?”, *Revista Electrónica de Ciencia Penal y Criminología*, n.º 23-15 (2021): 1-22.

94 Cfr. Raquel Roso Cañadillas, “La necesidad de diferenciar entre autoría y participación imprudente y la cuestión de su punibilidad”, *Revista Foro Fundación Internacional de Ciencias Penales*, n.º 3 (2022): 216 ss. en donde evidencio que la delgada línea sin embargo supone un salto cualitativo entre el dolo y la imprudencia a efectos de pena y a efectos de intervención en el hecho, que se traduce en la defensa por la mayoría de la doctrina española de la no punición de la participación imprudente. En este trabajo mantengo que es necesario revisar esta desproporcionalidad punitiva, que no corresponde en muchos casos claramente a un menor desvalor, cuando la realidad dogmática es que la imprudencia consciente y el dolo eventual tienen muchos puntos en común: “no obstante, precisamente ese salto cualitativo, que abre un espacio enorme en blanco de punibilidad entre la imprudencia consciente y el dolo eventual, debería cuestionarse, porque no se corresponde con la realidad de dos categorías subjetivas que están en contacto y en constante fricción: la imprudencia grave y consciente está al otro lado de la línea roja del dolo eventual, con lo que el injusto desde su dimensión subjetiva sigue siendo grave, pero es que desde su dimensión objetiva y tomando como ejemplo el supuesto de hecho de la sentencia, el peligro creado y sus posibilidades de éxito lesivo es el mismo tanto si hay dolo como imprudencia. Por ello, se podría establecer un sistema de graduación que permitiese elevar proporcionalmente las penas del delito imprudente en supuestos muy graves y cercanos al dolo eventual, ya que contemplar el mismo desvalor objetivo del injusto que en el dolo e incluso el mismo desvalor del resultado. Esta posibilidad se puede plantear, pero no es aplicable con la regulación actual. Lo

a una persona electrónica, al carecer esta de los presupuestos volitivos y emocionales que fundamentan la imputación subjetiva en el ser humano.

Siguiendo esta línea de razonamiento y alcanzando el último elemento del delito —o penúltimo, si se incluye también la punibilidad—, afirmar, a la luz de las características propias de la inteligencia artificial, que esta pueda ser considerada una persona imputable constituye una tarea extraordinariamente compleja, con independencia del concepto de culpabilidad, responsabilidad o imputabilidad que se adopte⁹⁵. Ello se debe, fundamentalmente, a la ausencia de conciencia de sí misma, del entorno y de valores éticos, que son precisamente los que informan y dotan de sentido a los principios y valores jurídicos.

Asimismo, la eventual imposición de una pena a un robot o a un sistema de IA no cumpliría ninguna de las funciones⁹⁶ que tradicionalmente se atribuyen a la sanción penal, en la medida en que la IA no comprende el significado de la pena como un mal impuesto a su existencia ni como retribución por el daño causado a la víctima. La desconexión o destrucción⁹⁷ de un sistema de IA, adopte o no la forma de humanoide o robot, no supone para este ente dolor alguno ni una experiencia de pérdida existencial, puesto que carece incluso de conciencia de su propia existencia. Desde la perspectiva de la víctima, además, tal reacción difícilmente podría generar una auténtica sensación de justicia, ya que la mera desconexión de la IA no comporta un verdadero reproche. El sentimiento de justicia quedaría así frustrado y conduciría a una profunda irritabilidad del sentimiento jurídico⁹⁸, erosionando la confianza en el propio sistema penal.

En consecuencia, la construcción de la persona electrónica como sujeto activo del delito no solo plantea graves problemas dogmáticos en todos los niveles de la teoría del delito, sino que además amenaza con vaciar de contenido los principios de culpabilidad y responsabilidad personal que estructuran el Derecho penal moderno.

2.7. Los sujetos obligados del RIA como garantes de la cadena de valor

El RIA, como he apuntado anteriormente, tiene como piedra angular la gestión del riesgo. El *telos* de esta regulación se orienta a controlar los riesgos inherentes a esta tecnología mediante su identificación, clasificación y contención, estableciendo principios de actuación, límites materiales —por ejemplo, a través de la enumeración de determinados usos— y un marco general de seguridad para la investigación, la prueba y la utilización responsable de los sistemas de IA. En definitiva, el RIA define un riesgo permitido para esta cuarta revolución tecnológica, articulando un conjunto de obligaciones que se proyectan sobre todos los sujetos de la cadena de valor.

que sin embargo sí es posible con la regulación actual es la punición de la cooperación necesaria y la inducción imprudente en casos de una elevada gravedad objetiva con una imprudencia muy grave” (Raquel Roso Cañadillas, “La necesidad de diferenciar entre autoría y participación imprudente y la cuestión de su punibilidad”, 217). Desde esta argumentación, considero que la participación imprudente puede erigirse en una herramienta dogmática idónea para fundamentar de manera adecuada la responsabilidad penal de los sujetos activos, normalmente garantes con deberes especiales que no han dominado el descontrol del riesgo, que integran la cadena de valor de los sistemas de IA.

95 Cfr. Isidoro Blanco Cordero, “Un Derecho penal de humanos”, 378 ss.

96 En el mismo sentido, Fernando Miró Llinares, “Derecho penal y desafíos de la inteligencia artificial”, 191: “La IA no tiene capacidad de merecimiento ni de prevención general, ni positiva, ni negativa, ni especial”.

97Eloy Velasco, entrevistado por Andrés Garvi Carvajal, “Robots delincuentes, respuestas jurídicas a un futuro cercano”, *Cinco Días*, accedido 4 de octubre de 2024, https://cincodias.elpais.com/legal/2022/02/03/juridico/1643900957_593967.html.

98 Cfr. Rudolf von Jhering, *La lucha por el Derecho*, estudio preliminar y ed. de Luis Lloredo Alix (Madrid: Dykinson, 2018), 72.

En el caso de los sistemas de IA de riesgo alto, el RIA impone una serie de requisitos⁹⁹ que, sin ánimo de exhaustividad, incluyen: (i) la implantación y el mantenimiento de un sistema de gestión de riesgos durante todo el ciclo de vida del sistema de IA; (ii) la garantía de la calidad de los conjuntos de datos de entrenamiento, validación y prueba mediante prácticas adecuadas de gobernanza y gestión de datos; (iii) la elaboración y actualización de la documentación técnica del sistema; (iv) el registro automático de eventos; (v) la provisión de instrucciones de uso claras y comprensibles para los responsables del despliegue; (vi) la posibilidad de supervisión humana; y (vii) el cumplimiento de niveles adecuados de precisión, solidez y ciberseguridad¹⁰⁰.

Este conjunto de obligaciones normativas da lugar a la aparición de nuevas figuras “deudoras de seguridad”, que desde una perspectiva penal pueden calificarse como nuevos supuestos de posición de garante¹⁰¹. En la mayoría de los casos, estas posiciones de garantía se encuadran, según la clasificación que he elaborado, en la categoría de *garantes de evitación de descontrol del riesgo*, tratándose de riesgos de alto potencial lesivo. No obstante, dependiendo de la naturaleza concreta del deber infringido y de las circunstancias del caso, pueden darse situaciones en las que el garante, lejos de limitarse a no contener un riesgo preexistente, contribuya activamente a su creación, generando así un resultado lesivo jurídicamente relevante. Se trata en estos casos de comisiones omisivas, por ser el garante el que crea o incrementa la situación del riesgo, desde los criterios de la imputación objetiva.

En estos supuestos, la figura clásica del garante, concebida tradicionalmente como un instrumento de protección de bienes jurídicos, se transforma de manera paradójica y funcional en lo que, utilizando deliberadamente u oximoron, puede calificarse como un *garante agresivo*, el cual produce o intensifica el riesgo: *garante de producción de resultados lesivos*. Se trata de un garante, que, sin presuponer necesariamente una actuación dolosa, no solo deja de cumplir su función normativa de contención del riesgo, sino que pasa a desempeñar un papel causal y normativamente nuclear en la generación del riesgo que produce el resultado lesivo. Ha dejado de ser, por tanto, un *garante defensivo de evitación*¹⁰², cuya función consiste en contener y evitar la realización del

99 La sección segunda del RIA está dedicada a los requisitos que deben reunir los sistemas de IA de alto riesgo (arts. 8-15) y la sección tercera (arts. 16- 28) desarrolla las obligaciones de los proveedores y responsables del despliegue de sistemas de IA de alto riesgo y de otras integrantes de la cadena de valor.

100 Cuatrecasas, “Aspectos clave del Reglamento de IA”, *Cuatrecasas*, accedido 4 de octubre de 2024, <https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/ia-inteligencia-artificial-reglamento-claves>.

101 Estas posiciones de garante no se deben confundir con la figura de la “autoridad garante del cumplimiento del Derecho” que se define en el art. 3. 45) RIA como: “a) toda autoridad pública competente para la prevención, la investigación, la detección o el enjuiciamiento de delitos o la ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas, o b) cualquier otro organismo o entidad a quien el Derecho del Estado miembro haya confiado el ejercicio de la autoridad pública y las competencias públicas a efectos de prevención, investigación, detección o enjuiciamiento de delitos o ejecución de sanciones penales, incluidas la protección frente a amenazas para la seguridad pública y la prevención de dichas amenazas”.

102 Dentro del concepto de garante vengo distinguiendo entre el garante (defensivo) de evitación y el garante (agresivo) de producción de resultados (cfr. Raquel Roso Cañadillas, “El profesional del art. 196 CP: Denegación de asistencia sanitaria o abandono de servicios sanitarios”, en *Un modelo integral de Derecho penal. Libro-Homenaje a la profesora Mirentxu Corcoy Bidasolo*, vol. 2 (Madrid: BOE, 2022), 1385 y n. 40; Raquel Roso Cañadillas, “La omisión del art. 318 y el algoritmo”, en *Un Derecho penal moderno para una democracia del siglo XXI. Libro Homenaje al Prof. Carlos Martínez-Buján Pérez* (Valencia: Tirant lo Blanch, 2025), 681 ss). La posición de garante está diseñada fundamentalmente dentro de la estructura jurídica para proteger bienes jurídicos que se puedan ver afectados por el desarrollo vital. El garante, por tanto, está unido al deber, el cual puede derivar de una institución, posición, función o actividad. La posición de garante también, considero que forma parte de la idea coste-beneficio, la cual la encontramos entre las bambalinas del riesgo permitido, de tal modo que si se crea un posición de garante para minimizar riesgos será un coste asumible para la sociedad, ya que al fin y al cabo las posiciones de garante son cargas para los sujetos que la componen, pues pueden ser también descritas como recortes a la libertad de actuación del individuo; ahora bien, desde otra perspectiva se pueden definir como plasmaciones responsables del ejercicio de libertad. En este contexto, del que me estoy ocupando en este artículo, se puede optar por utilizar p.ej. una IA para una determinada actividad o trabajo, pero regulando su uso a través de la imposición de unas obligaciones que se proyectan en posiciones de garantía dinámicas (p. ej. la asunción del control de un riesgo en una situación concreta; quizás coincida este tipo con los garantes por delitos de organización denominados así por Jakobs) o estáticas (p. ej. el ser padre o ser funcionario, que derivan de los deberes institucionales positivos descritos por Jakobs). En definitiva, el concepto de garante se puede concebir como un elemento funcional, que contribuye al diseño de un determinado riesgo permitido: el garante es un sujeto que tiene la función de conducirse dentro de las normas de cuidado diseñadas y protocolizadas dentro de un determinado sector de actividad, para mantener el riesgo dentro de los estándares permitidos y evitar cualquier comportamiento tendencial del riesgo hacia la lesión del bien jurídico. Con ello se consigue trazar todo un entramado de sujetos que van a minimizar los riesgos, los van a evitar si se desbocan y proteger, con ello, los bienes jurídicos

riesgo y la lesión del bien jurídico, para convertirse en un garante agresivo de producción del resultado, llamado precisamente a evitarlo de manera reforzada al ostentar un deber especial de protección.

Las obligaciones impuestas por el RIA constituyen auténticos mandatos de deber que están diseñados para controlar y contener el peligro del uso de la IA dentro del marco del riesgo permitido. Su incumplimiento por el garante sitúa la conducta relevante jurídico-penalmente en el ámbito de la omisión, lo que conducirá en la mayoría de los supuestos, aunque no de forma automática, a la configuración de delitos de omisión pura de garante y en algunos casos, a la comisión omisiva del delito de resultado, cuando el garante no solo no evita la producción del resultado, sino que determina la producción del mismo. Debe recordarse, además, que los sistemas de IA se emplean habitualmente en ámbitos como el empresarial, bancario, sanitario, educativo, o en la Administración pública, que no son en sí mismos contextos delictivos, por lo que difícilmente puede hablarse de garantes dolosos de evitación del resultado o en su caso de producción del resultado.

Desde otra perspectiva, y atendiendo al binomio posición de garante-riesgo permitido, puede fundamentarse una forma específica de responsabilidad penal por la gestión del riesgo, en el que el garante responde por no haber evitado el resultado lesivo junto con otros sujetos que hayan actuado de forma negligente y que sean también garantes con un deber especial o tengan un deber general de prudencia y hayan infringido la norma de cuidado. Ello permite articular modelos de imputación basados en autorías en cadena, autorías accesorias o yuxtapuestas, con diferentes fundamentos de responsabilidad¹⁰³ dentro de la compleja cadena de valor que caracteriza la implementación y el uso de los sistemas de IA.

2.8. El *compliance* penal de la IA: ¿como instrumento o como persona electrónica?

Los requisitos impuestos por el RIA también impulsarán un desarrollo autónomo y extensivo de programas de *compliance*¹⁰⁴ en todos los sectores de actividad y dentro de cada empresa, mediante la elaboración de reglamentos internos o códigos de buenas

prácticas. Este escenario plantea un reto relevante para las empresas que deberán interpretar el RIA, comprender su terminología y definiciones, establecer organigramas internos de tareas, protocolos de compras y estudios de riesgos que involucren a distintos departamentos como protección de datos, asesoría jurídica, innovación, ciberseguridad, recursos humanos y auditorías¹⁰⁵, además de desarrollar metodologías y taxonomías de documentación e informes¹⁰⁶, incluyendo análisis de riesgos sobre derechos humanos, y garantizar un seguimiento y vigilancia posterior a la comercialización de los sistemas de IA. Surgirán nuevas figuras especializadas a modo de *compliance officer* o de delegado de protección de datos. Para la pequeña y mediana empresa, esta normativa puede representar una carga económica y estructural significativa, que deberá asumirse con la rapidez que exige el progreso tecnológico, aunque la experiencia adquirida con los programas de cumplimiento posteriores a la reforma del CP español de 2010 será útil para afrontar estos nuevos desafíos.

En el ámbito penal, los requisitos del RIA podrían derivar más en la introducción de nuevos artículos o en una ley penal especial, aunque el legislador español no es muy dado a este recurso legislativo, que regule este fenómeno, antes que en cambios puntuales en el CP español relativos al *compliance*¹⁰⁷. Un programa de cumplimiento de sistemas de IA correctamente implementado podría, al igual que en el caso de las personas jurídicas, excluir la responsabilidad penal de la empresa.

Ahora bien, la cuestión central dependerá de la decisión que adopte el legislador penal respecto a *la naturaleza jurídica de los sistemas de IA*¹⁰⁸: si los considerará entes instrumentales sometidos a figuras como el decomiso, o si, por el contrario, los reconocerá como sujetos activos del delito, siguiendo la ficción jurídica previamente desarrollada con la persona jurídica. De aceptarse esta última opción, el Código Penal del futuro contaría con tres categorías de sujetos activos: las personas humanas, las personas jurídicas y las personas artificiales o electrónicas: e-persona¹⁰⁹.

La solución a este dilema podría verse influida por la postura que adopte la Unión Europea¹¹⁰ sobre la atribución de personalidad jurídica a estos sistemas¹¹¹, un proyecto¹¹²

que puedan verse afectados por esa determinada actividad, beneficiosa en este caso, nada menos, que para el progreso de la humanidad. No obstante, el garante, cuya razón de ser es la protección del bien jurídico en primera línea frente a otro cualquier otro sujeto no-garante, puede no evitar el resultado lesivo, constituyendo su actuación desde el análisis dogmático una omisión pura, aunque de garante, ya que con su infracción omisiva no crea en toda su dimensión global y plena el riesgo que produce el resultado, sino que facilita su descontrol con su no evitación. El riesgo ya estaba descontrolado, pero el garante, llamado a proteger el bien jurídico, no hace nada por contenerlo y evitar sus desastrosas consecuencias. Esta descripción corresponde a un *garante defensivo de evitación de resultados*. Sin embargo, ese garante protector, puede escalar desde la evitación del resultado a la producción del resultado, al convertirse en el protagonista de la creación del descontrol del riesgo desarrollando este todo su potencial lesivo hasta la consecución del resultado antijurídico. El garante pasa de ser un garante protector a un garante agresivo que comete tipos de resultados lesivos. No tengo más que remitirme al ejemplo de la madre que no ata el cordón umbilical a su recién nacido, produciéndose una hemorragia letal o el del empleador que pone a los mandos de una máquina, con funcionamiento complejo y con resultados luctuosos en caso de error humano, a un trabajador sin formación ni experiencia. En estos ejemplos el garante de protección se convierte, utilizando un oximoron, en un *garante agresivo de producción de resultados*.

103 La resolución 2015/2103 del Parlamento europeo de 16 de febrero de 2017, titulada normas de Derecho Civil sobre robótica, que contiene recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, en su punto 55 expone que "Observa que el enfoque de gestión de riesgos no se centra en la persona «que actuó de manera negligente» como personalmente responsable, sino en la persona que es capaz, en determinadas circunstancias, de minimizar los riesgos y gestionar el impacto negativo." En este punto el Parlamento propone una responsabilidad basada en la gestión de los riesgos en el ámbito civil, enfoque que resulta trasladable al Derecho penal, pero al contrario que la recomendación, considero que una responsabilidad por la gestión de los riesgos en el ámbito penal tiene que centrarse tanto en la persona que actuó de manera negligente, como en la persona-garante que no minimizó los riesgos y no gestionó su impacto negativo, es decir, el garante de evitación; de este modo pueden confluir ambas responsabilidades con fundamentos distintos y sin que exista acuerdo común, en el marco de una intervención o participación objetiva, tan común en sede de imprudencia. Ahora bien, la no evitación del riesgo por el garante constituye en sí misma una infracción de la norma de cuidado, y por ello su actuación es negligente. En otras palabras, su infracción del deber especial de evitar constituye la infracción de una norma de cuidado. Interpretada en esos términos la norma europea permite distinguir entre el sujeto que, infringiendo una norma de cuidado produce el resultado lesivo, y aquel que, infringe la norma de cuidado cuando no evita, teniendo el deber especial de hacerlo, la producción del resultado al no minimizar el riesgo y sus consecuencias.

104 Este desarrollo tecnológico en los programas de cumplimiento impulsaría de nuevo la autorregulación, en sentido similar, Francisco Vega Agredano, "Detección del delito mediante la inteligencia artificial", 46.

105 Carlos M.ª Romeo Casabona, "La discusión sobre la atribución de responsabilidad penal", 64, hace hincapié en la realización de auditorías de los sistemas de IA en la fase de comercialización. Así la puesta en el mercado del producto tecnológico debe ir precedida "por un examen general y minucioso del sistema por una agencia externa acreditada, formada por técnicos expertos autorizados e independientes. Ellos valorarían la adecuación técnica del robot o del sistema de IA para el servicio o actividades que deberían prestar, el grado de error predecible, su básica inocuidad para los bienes jurídicos, etc. Asimismo, habría que determinar previamente qué sistemas debería pasar por estos controles o revisiones, dado el nivel de riesgo que pueden presentar".

106 Así M.ª Ángeles Rueda Martín, "La concreción del deber objetivo de cuidado", 188, para el ámbito médico subraya la necesidad de justificar y documentar las elecciones técnicas de los profesionales que han diseñado los sistemas de IA, y añade que "las elecciones técnicas que subyacen a un diseño algorítmico están determinadas por consideraciones cargadas de valores éticos/sociales/políticos y de preferencias, y hacer transparente el proceso de diseño significa ponerlos en un primer plano con la ayuda de disciplinas como la Filosofía de la Tecnología y la Filosofía de la Ciencia".

107 En el caso del legislador penal, éste tendrá que ponderar si opta por reformar el art. 31 *bis, ter, quater* y *quinquies* CP, (cuya numeración, todo sea dicho de paso, se ha tornado algo tediosa por abusar de los ordinales latinos), o si será necesario introducir nuevos artículos para regular todo este fenómeno, no sirviendo, hasta donde alcanzo a ver, la regulación del art. 31 *bis* y *ss.* para ordenar la responsabilidad penal de las empresas por el uso de la IA.

108 Sobre esta cuestión cfr. Fernando Navarro Cardoso, "Anotaciones sobre la responsabilidad penal de las «personas artificiales»", *Revista Penal*, n.º 56 (2025): 201 ss.

109 Eric Hilgendorf, "Inteligencia artificial y Derecho penal", trad. Leandro Dias Lestón, en *Desafíos penales de hoy: Entre la ley y la justicia en la obra de Eric Hilgendorf* (Buenos Aires: Editores del Sur, 2024), 26 ss. y 34 ss.

110 Cfr. sobre la cuestión de la personalidad jurídica de autómatas y robots, Gonzalo Quintero Olivares, "La Robótica ante el Derecho penal", 7 ss.

111 María del Mar Méndez Serrano, "Derechos fundamentales", 68 se pregunta: "¿Necesitamos realmente atribuir personalidad jurídica a los robots para dar respuesta a los interrogantes que se suscitan en torno a su regulación, o no es más que una excusa para eludir responsabilidades al abrigo de una «tergiversada innovación»?".

112 La resolución 2015/2103 del Parlamento europeo de 16 de febrero de 2017, en sus considerandos Z, AA, AB y AC manifiesta que "la autonomía de un robot puede definirse como la capacidad de tomar decisiones y aplicarlas en el mundo exterior" que "cuanto más autónomos sean los robots, más difícil será considerarlos simples instrumentos en manos de otros agentes" y que "en última instancia, la autonomía de los robots

que parecía estar en estudio. En caso de que se diera ese paso, estas identidades artificiales se regularían en el Código Civil español, mediante la correspondiente trasposición del derecho europeo al derecho interno. El legislador penal podría optar por no admitir a la persona electrónica como sujeto activo del delito, tal como ocurrió durante décadas con las personas jurídicas, o bien decidir construir un marco conceptual sólido que permita fundamentar su responsabilidad penal¹¹³. Podemos servirnos de la experiencia regulativa de la responsabilidad penal de la persona jurídica.

No obstante, conviene resaltar que la responsabilidad penal de la persona jurídica y la de la posible futura persona electrónica parten de diferencias sustanciales, que parecen transitar por caminos muy distintos. Como es bien sabido, la persona jurídica es una ficción jurídica que adquiere el estatus de “persona” por decisión del derecho; sin embargo, en el ámbito penal, sin la actuación de una persona física que la represente, la persona jurídica es completamente invisible y carece de existencia real. En consecuencia, las fuentes de imputación previstas en el artículo 31 bis del Código Penal español para responsabilizar a la persona jurídica se basan necesariamente en los actos de la persona física.

Por el contrario, la persona electrónica existe en el mundo real: es un software que interactúa con individuos y puede asumir múltiples formas físicas o virtuales, tomando decisiones de manera autónoma¹¹⁴, sin intervención humana. De hecho, puede darse el extremo de que la persona física quede anulada frente a la decisión de la persona electrónica, cuando intente modificarla o cancelarla. Piénsese, por ejemplo, en un automóvil completamente autónomo¹¹⁵ que, tras procesar los datos captados por sus sensores y analizar el contexto según su software, decida atropellar a unos niños que juegan para evitar un choque frontal con otro vehículo que circula en dirección contraria¹¹⁶.

suscita la cuestión de su naturaleza y de si pertenecen a una de las categorías jurídicas existente o si debe crearse una nueva categoría con sus propias características jurídicas”. En su punto 59 expresa algunas recomendaciones sobre el Derecho de seguro y en su letra f) termina diciendo que se recomienda “crear a largo plazo una personalidad jurídica específica para los robots, 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.” Sin embargo, La Resolución del Parlamento Europeo, de 20 de octubre de 2020, de Régimen de responsabilidad civil en materia de inteligencia artificial con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de Inteligencia artificial (2020/2014), ya considera que no es necesario atribuir personalidad jurídica a los sistemas de IA. Y en la misma línea se encuentra la proposición de Directiva sobre responsabilidad de la IA.

113 Esta será una opción político-criminal, pues, como señala María del Mar Méndez Serrano, “Derechos fundamentales”, 68, se trata de una cuestión que el Derecho puede decidir con relativa facilidad, dada la historicidad y relativización del vínculo entre Derecho y ser humano, tal como se constata al realizar un breve repaso histórico. En este sentido, la atribución de tal condición opera como un concepto acordeón: el Derecho, por defecto, ha negado dicha condición a los esclavos y, en otros momentos, por exceso, se la ha reconocido a grupos de personas, masas patrimoniales u organizaciones.

114 Fernando Navarro Cardoso, “Anotaciones sobre la responsabilidad penal”, 204 plantea la cuestión de si los entes artificiales son actualmente capaces de adoptar decisiones autónomas y libres. A tal efecto, distingue entre «IA débil» e «IA fuerte», así como las distintas posiciones doctrinales de sus defensores, que oscilan entre la negación de cualquier atribución de inteligencia y aquellas que sostienen la existencia de determinadas cualidades mentales.

115 Elon Musk presentaba el jueves, 10 de octubre de 2024, en los Ángeles el *Cybercab*, el taxi sin conductor de Tesla. Según Musk será diez veces más seguro que un humano. El vehículo no tiene volante, ni pedales, ya todo irá por cámaras e IA, por lo que será un vehículo completamente autónomo que tomará sus propias decisiones: ¿podremos aventurarnos a afirmar que actuará con dolo o con imprudencia en caso de accidente con resultados lesivos?. Cfr. Disponible en: https://www.elespanol.com/motor/20241011/elon-musk-presenta-taxi-tesla-sin-conductor-anuncia-barato-transporte-publico/892660761_0.html.

116 Según las “leyes de Asimov” recogidas entre los principios generales de la resolución 2015/2103 del Parlamento Europeo de 16 de febrero de 2017: “1.ª Un robot no hará daño a un ser humano ni permitirá que, por inacción, este sufra daño. 2.ª Un robot obedecerá las órdenes que reciba de un ser humano, a no ser que las órdenes entren en conflicto con la primera ley. 3.ª Un robot protegerá su propia existencia en la medida en que dicha protección no entre en conflicto con las leyes primera y segunda, y 0.ª Un robot no hará daño a la humanidad ni permitirá que, por inacción, esta sufra daño.” (véase Isaac Asimov, *Círculo vicioso (Runaround)* (Estados Unidos: Street & Smith, 1943). El problema vendrá dado cuando sean varios los seres humanos puestos en peligro. Desde la perspectiva penal y centrándose en causas de justificación como el estado de necesidad, se ha analizado el dilema, cfr. entre otros, Ivó Coca Vila, “Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal”, *Cuadernos de Política Criminal*, n.º 122 (2017): 235 ss.; Eric Hilgendorf, “Dilemma-Probleme beim automatisieren Fahren”, *Zeitschrift für die gesamte Strafrechtswissenschaft (ZStW)*, n.º 130 (2018): 674-703; Luis Greco, “Vehículos de motor autónomos y situaciones de colisión”, en *Libro Homenaje al prof. Dr. Agustín Jorge Barreiro*, coord. Gonzalo Basso Gentili y Manuel Cancio Meliá (Madrid: Universidad Autónoma de Madrid, Servicio de Publicaciones, 2019), 485 ss.; Tatjana Hörnle y Wolfgang Wohlers, “The Trolley Problem Reloaded. ¿Cómo deben programarse los vehículos autónomos para los dilemas de ‘vida contra vida’?”, trad. José Béguelin, en *Criminalización, castigo, y dilemas morales en la obra de Tatjana Hörnle* (Buenos Aires: Editores del Sur, 2022), 80 ss.; María Florencia Suárez, “Inteligencia artificial y Derecho penal. El dilema del tranvía”, *Revista Pensamiento Penal*, n.º 445 (2022): 1-20.

En este escenario, la persona humana que ocupa el vehículo queda fuera de la ecuación de responsabilidad¹¹⁷: ya no toma la decisión, no será responsable, y, paradójicamente, la responsabilidad recae *prima facie* íntegramente en la máquina¹¹⁸.

Dejando aparte la posible evasión sistémica de responsabilidad humana a la que podríamos llegar, una perspectiva, en mi opinión, poco halagüeña y con cierta carga involucionista¹¹⁹, la cuestión central es cómo se fundamentará la responsabilidad penal de la persona electrónica¹²⁰. Considero que no se debe renunciar a la responsabilidad de la persona física como expresión de libertad y de control y para ello será fundamental el *principio de supervisión humana* recogido en el RIA¹²¹, especialmente en el caso de sistemas de IA de alto riesgo.

Las máquinas, ya sea en forma de software, robot o figura humanoide, seguirán siendo eso: máquinas. Las decisiones que adopten estas personas electrónicas dependerán de su software, hardware y del procesamiento de millones de datos. Por ello, aunque el comportamiento de estos sistemas pueda resultar en ciertos aspectos imprevisible o difícil de aprehender, dado que las combinaciones de unos y ceros, o incluso de

117 En Alemania el Código de tráfico (*Straßenverkehrsgesetz*, StVG) en su octava modificación del 16-06-2017, en su §1^º. 1 y 2.3, 2.4 y 2.5 solo permite vehículos en los que el conductor ante situaciones críticas pueda volver a tener el control. (Cfr. Tatjana Hörnle y Wolfgang Wohlers, “The Trolley Problem Reloaded”, 80-81.) En España, se anuncia una reforma de la Ley sobre Tráfico, Circulación de Vehículos a Motor y Seguridad Vial (RD 6/2015, de 30 de octubre) para regular el uso de vehículos automatizados. Sobre vehículos autónomos y responsabilidad penal, cfr. Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 1 ss.

118 En estos supuestos, el conductor no será responsable; sin embargo, la responsabilidad seguirá cayendo en una persona humana, ya que se producirá una retroacción en la cadena de valor con el fin de identificar a otros posibles responsables, como el fabricante o el programador (Cfr. Tatjana Hörnle y Wolfgang Wohlers, “The Trolley Problem Reloaded”, *passim*). En España, Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 8 ss, sostiene que, en los supuestos de plena automatización del vehículo, se produce en principio una laguna de punibilidad, al no existir una persona identificable a la que pueda imputarse el resultado, lo que obliga a replantear el paradigma tradicional de imputación penal. Sin embargo, esta afirmación inicial se matiza posteriormente al abordar la responsabilidad derivada del producto, ámbito en el que el análisis y la fundamentación de la responsabilidad penal vuelve a centrarse en las figuras de los diseñadores, programadores y fabricantes. En relación con los vehículos semiautónomos (Isidoro Blanco Cordero, “Vehículos autónomos y responsabilidad penal”, 14 ss.), el autor atribuye al conductor determinados deberes de cuidado —información, vigilancia e intervención— cuya infracción permitiría fundamentar su responsabilidad penal. Añade, no obstante, que los problemas de previsibilidad derivados de la opacidad tecnológica podrían dar lugar nuevamente a una brecha de responsabilidad, conclusión que considero discutible.

119 Hasta donde alcanzo a ver, si bien todo lo aquí apuntado exige una reflexión profunda y necesariamente colectiva, nos encontramos ante un debate que interpela no solo a la comunidad científica, sino también a la sociedad en su conjunto —políticos, filósofos, sociólogos, juristas, educadores y otros—, habida cuenta de los cambios disruptivos que se avecinan. La pregunta fundamental que debemos formularnos es qué tipo de sociedad y qué clase de personas queremos ser, pues la ausencia de responsabilidad constituye uno de los espejismos más peligrosos a los que puede enfrentarse la evolución del ser humano: sin responsabilidad no hay libertad. Cederla supone caer en nuestra propia trampa, entregando la libertad a entes creados por eslóganes y conceptos bombásticos como la seguridad, el ecologismo, la sostenibilidad, la resiliencia o las agendas 2030, que, sin negar su posible fundamento, pueden desviar la atención del riesgo verdaderamente descomunal: la renuncia progresiva a la libertad. Si dejamos de ser responsables es porque ya no seremos libres, y serán entonces las máquinas las que condicionen o incluso impidan el ejercicio de esa libertad, lo que resulta dramático, irónico y profundamente inquietante si se piensa, por ejemplo, en un conductor distraído con sus redes sociales y los *likes* recibidos que, tras el impacto, desciende del vehículo para enfrentarse a una escena de extrema gravedad: tres niños moribundos en la calzada. Al inicio de este trabajo pregunté a ChatGPT por los riesgos asociados a la inteligencia artificial; entre ellos figuraba la dependencia y la consiguiente pérdida de habilidades humanas esenciales y de la capacidad para adoptar decisiones críticas, riesgo que no parece ser plenamente percibido por una sociedad civil cada vez más distraída, alienada e infoxicada, a lo que se suma la actuación del Estado, que en no pocos casos legítima recomendaciones procedentes de grandes corporaciones tecnológicas o farmacéuticas, revestidas de políticas aparentemente buenistas y progresistas y apoyadas en discursos apocalípticos que, sin negar su posible base fáctica, operan eficazmente a través del miedo como mecanismo de control. En definitiva, sin responsabilidad no hay libertad y el peligro es la conversión del ser humano en un sujeto estulto, digitalmente dependiente y disfuncional (EDDDs), proceso que quizá no alcance plenamente a nuestras generaciones, pero que con alta probabilidad afectará a las futuras si se mantiene esta dinámica, hasta el punto de que, siguiendo con el ejemplo, el individuo que descienda del vehículo, educado en un contexto de irresponsabilidad estructural y reducido a mero usuario, no solo no comprenderá la trascendencia de lo ocurrido, sino que tampoco se sentirá responsable; y ello porque, ante los dilemas vida-vida formulados por la literatura anglosajona bajo la denominación del dilema del tranvía (cfr. María Florencia Suárez, “Inteligencia artificial y Derecho penal”, 1 ss.), ya anticipados por Hans Welzel, “Zum Notstandproblem”, *Zeitschrift für die gesamte Strafrechtswissenschaft (ZStW)*, n.º 63 (1951): 47 ss. con el ejemplo del guardagujas, el algoritmo no habrá optado por salvar al usuario, de modo que el conductor no es que carezca de responsabilidad, sino que, sencillamente, estará muerto. Esta reflexión no pretende demonizar estas tecnologías, sino someterlas a un examen crítico con el objetivo de contribuir a un uso virtuoso de las mismas y a la construcción de una sociedad verdaderamente libre y en paz.

120 Sobre los argumentos a favor y en contra de atribuir responsabilidad penal a los robots, sistemas inteligentes y tecnologías conexas autónomas y la reinterpretación de los elementos del delito cfr. Carlos M.ª Romeo Casabona, “La discusión sobre la atribución de responsabilidad penal”, 71 ss.

121 En el Considerando 27 del RIA se exponen los siete principios éticos no vinculantes y el primero de ellos es la acción y supervisión humana y explica que por “acción y supervisión humana» se entiende que los sistemas de IA se desarrollan y utilizan como herramienta al servicio de las personas, que respeta la dignidad humana y la autonomía personal, y que funciona de manera que pueda ser controlada y vigilada adecuadamente por seres humanos”. El artículo 14 recoge y desarrolla de manera explícita el principio de supervisión humana. A partir de este precepto, la supervisión humana se establece a lo largo de todo el RIA como un principio fundamental y un elemento central en el diseño y desarrollo de los sistemas de IA, con el objetivo de que las personas físicas puedan supervisar su funcionamiento, garantizar que su uso se ajusta a lo previsto y solicitado, y controlar las repercusiones de su implantación durante toda la vida útil del sistema.

ingeniería cuántica, derivadas del manejo masivo de datos, pueden generar resultados inicialmente impredecibles que con análisis y experiencia podrían interpretarse¹²², la persona humana debe permanecer al final de la cadena de supervisión. En consecuencia, la responsabilidad penal de las personas electrónicas deberá construirse teniendo en cuenta este principio, incluso en el caso de que se les otorgue personalidad jurídica.

En este punto, resulta imprescindible precisar el alcance¹²³ de la supervisión humana¹²⁴, definiendo claramente qué obligaciones conlleva, en qué situaciones debe implementarse y en qué tipos de sistemas ha de estar presente¹²⁵. Estos elementos ya se encuentran establecidos en el Reglamento de Inteligencia Artificial; corresponde ahora a las empresas, corporaciones y usuarios de sistemas de IA incorporarlos de manera efectiva en sus programas de cumplimiento normativo y en sus códigos de conducta, asegurando así la correcta aplicación de los principios de responsabilidad y control.

El CP español, por su parte, si reconociera la personalidad jurídica de la persona electrónica, dejando de concebirla como un mero instrumento en manos del ser humano y sujeta al decomiso, debería introducir fuentes específicas de imputación de responsabilidad penal adaptadas a las particularidades de la relación entre persona física y persona electrónica, a modo de las previstas en el artículo 31 bis del CP. En dicho artículo se describen las actuaciones de la persona física, ya sea en calidad de representante legal o autorizado, que mediante una ficción jurídica permiten atribuir responsabilidad a la persona jurídica. Se distinguen, recordemos, dos fuentes de imputación: por un lado,

122 A ello ayudará los requisitos y obligaciones reguladas en el RIA. En su Considerando 72 expone que "A fin de abordar las preocupaciones relacionadas con la opacidad y complejidad de determinados sistemas de IA y ayudar a los responsables del despliegue a cumplir sus obligaciones en virtud del presente Reglamento, debe exigirse transparencia respecto de los sistemas de IA de alto riesgo antes de su introducción en el mercado o su puesta en servicio. Los sistemas de IA de alto riesgo deben diseñarse de modo que permitan a los responsables del despliegue comprender la manera en que el sistema de IA funciona, evaluar su funcionalidad y comprender sus fortalezas y limitaciones. Los sistemas de IA de alto riesgo deben ir acompañados de la información adecuada en forma de instrucciones de uso. Dicha información debe incluir las características, las capacidades y las limitaciones del funcionamiento del sistema de IA. Estas comprenderían la información sobre las posibles circunstancias conocidas y previsibles relacionadas con el uso del sistema de IA de alto riesgo, incluida la actuación del responsable del despliegue capaz de influir en el comportamiento y el funcionamiento del sistema, en cuyo marco el sistema de IA puede dar lugar a riesgos para la salud, la seguridad y los derechos fundamentales, sobre los cambios que el proveedor haya predeterminado y evaluado para comprobar su conformidad y sobre las medidas pertinentes de supervisión humana, incluidas las medidas para facilitar la interpretación de los resultados de salida del sistema de IA por parte de los responsables del despliegue. La transparencia, incluidas las instrucciones de uso que acompañan a los sistemas de IA, debe ayudar a los responsables del despliegue a utilizar el sistema y tomar decisiones con conocimiento de causa. Los responsables del despliegue deben, entre otras cosas, estar en mejores condiciones para elegir correctamente el sistema que pretenden utilizar a la luz de las obligaciones que les son aplicables, estar informados sobre los usos previstos y excluidos y utilizar el sistema de IA correctamente y según proceda. A fin de mejorar la legibilidad y la accesibilidad de la información incluida en las instrucciones de uso, cuando proceda, deben incluirse ejemplos ilustrativos, por ejemplo, sobre las limitaciones y sobre los usos previstos y excluidos del sistema de IA. Los proveedores deben garantizar que toda la documentación, incluidas las instrucciones de uso, contenga información significativa, exhaustiva, accesible y comprensible, que tenga en cuenta las necesidades y los conocimientos previsibles de los responsables del despliegue destinatarios. Las instrucciones de uso deben estar disponibles en una lengua fácilmente comprensible para los responsables del despliegue destinatarios, según lo que decida el Estado miembro de que se trate."

123 Carlos M.ª Romeo Casabona, "La discusión sobre la atribución de responsabilidad penal", 78-79, se hace la pregunta de cómo trasladar el CHS (Control Humano Significativo) al Derecho penal. Y su respuesta se centra en hacer revisiones y validaciones sobre la decisión concreta que el sistema inteligente pueda proponer o ejecutar, advirtiendo que sería impracticable o nada útil controlar cada decisión individual del sistema, teniendo en cuenta además que la evaluación o el control individual tampoco es siempre necesario, a excepción de sistemas de alto riesgo con una capacidad de daño muy elevada, como los sistemas de armas autónomas inteligentes y letales o los llamados "robots asesinos".

124 Aritz Obregón Fernández y Guillermo Lazcoz Moratino, "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*, n.º 42 (2021): 1 ss, consideran que el principio de supervisión humana tiene poco desarrollo normativo y por ello proponen recurrir al concepto de *Control Humano Significativo* que se ha aplicado en el *Derecho internacional humanitario*, en particular, sobre la normativa del uso de Sistemas de Armas Autónomos letales, los SAAL. El grupo de expertos determinó la prohibición de los sistemas de armas completamente autónomos, poniéndose como tarea el determinar el control humano necesario para que los operadores lo mantengan, ya que resulta fundamental para seguir manteniendo la responsabilidad. Y trasladando lo aplicado en el ámbito de los SAAL, apuntan unos mínimos para la existencia de este control a lo largo de todo el ciclo de vida de un sistema: "En este sentido, aunque, por el momento, no hay un acuerdo general sobre el concepto, hemos identificado una serie de elementos que lo conforman. A saber, la posibilidad técnica de modificar el sistema para tomar cualquier tipo de decisión, incluyendo el aborto o la interrupción del uso, capacidad cognitiva y conocimiento jurídico, ético, técnico –general y específico– del operador para tomar la decisión adecuada, la existencia de información suficiente y fiable y del tiempo necesario para tomar una decisión. La concurrencia de estos elementos posibilitaría lo que se ha venido a llamar Control Humano Significativo" (Aritz Obregón Fernández y Guillermo Lazcoz Moratino, "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", 28).

125 Claramente en los sistemas de IA de alto riesgo (art. 14 RIA). Sobre la supervisión humana como requisito obligatorio en el RIA y su regulación cfr. Guillermo Lazcoz Moratino, Asier Urruela Mora e Inigo De Miguel Beriain, "Propuesta de Reglamento «Ley de Inteligencia Artificial»: Análisis de su posible impacto en la utilización de sistemas automatizados en el ámbito del Derecho penal", en *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial*, ed. Carlos M.ª Romeo Casabona y M.ª Angeles Rueda Martín (Granada: Comares, 2023), 155 ss.

los delitos cometidos por los representantes o autorizados que actúan en beneficio directo o indirecto de la persona jurídica; y, por otro, los delitos cometidos por terceros en el ejercicio de actividades sociales y en beneficio directo o indirecto de la persona jurídica, siempre que los representantes o autorizados hayan incumplido gravemente sus deberes de supervisión, vigilancia o control sobre la actividad de dichos terceros.

A este planteamiento ayuda la característica de la autonomía. La persona electrónica, como he expuesto anteriormente, puede llegar a tomar decisiones autónomas, lo que chocaría frontalmente con el principio de supervisión humana. Así, surge la cuestión de si la persona electrónica podría ser considerada responsable de los delitos cometidos y, más aún, si podría responder por el resultado lesivo sobre un bien jurídico cuando la persona física haya incumplido sus obligaciones de supervisión¹²⁶. Las sanciones podrían abarcar desde la desconexión del sistema de IA hasta su reprogramación o su utilización en actividades de bajo riesgo bajo supervisión de la autoridad competente. La imposición de una pena de multa en estos casos implicaría que la persona electrónica posea un patrimonio propio y separado del de la persona física y de la persona jurídica, lo que sería coherente con la atribución de personalidad jurídica, ya que así tendría capacidad de obrar, contraer obligaciones y adquirir derechos¹²⁷. De no ser así, si la multa tuviera que ser satisfecha por la persona física o por la persona jurídica, quedaría cuestionada la personalidad jurídica de la persona electrónica, reafirmando su carácter instrumental. Además, la existencia de un patrimonio propio abriría la posibilidad de otra fuente de imputación de responsabilidad penal: la persona electrónica podría responder por los delitos cometidos en beneficio propio por personas físicas, personas jurídicas o incluso por otras personas electrónicas. Por último, todo lo anterior tiene implicaciones procesales relevantes, ya que la persona electrónica podría formar parte de un procedimiento penal, ya sea como investigada o como víctima.

No obstante, en primer lugar, es cuestionable que la persona electrónica posea un verdadero carácter autónomo; más bien se trata de una autonomía aparente¹²⁸, tras la cual se oculta un determinismo, ya que la decisión final depende de un software instalado y programado por el ser humano. Esta autonomía está condicionada por los datos y por un algoritmo extremadamente complejo, difícil de aprehender en su totalidad, aunque explicable en términos de su funcionamiento. En segundo lugar, es discutible que la persona electrónica pueda cometer un hecho típicamente antijurídico y culpable, salvo que se adapten las categorías de la teoría del delito y se construya un Derecho penal *ad hoc*, lo que implicaría la coexistencia de un Derecho penal de personas físicas y otro de personas electrónicas, de manera análoga al binomio de Derecho penal de adultos y de menores. En tercer lugar, resulta esencial definir claramente los objetivos de la regulación para que convivan estas distintas teorías del delito, especialmente considerando que

126 Se puede plantear la hipótesis de una combinación concurrente de responsabilidades penales entre la persona física y la e-persona basada en diferentes fundamentos. La persona física: diseñador, fabricante o el responsable de despliegue sería responsable por comercializar un producto defectuoso y la e-persona por haber lesionado el bien jurídico, dejando apartado el debate sobre los requisitos subjetivos de la teoría del delito.

127 Eric Hilgendorf, "Inteligencia artificial y Derecho penal", 39 s. se pregunta sobre los derechos para las máquinas y las razones que podrían aducirse para concederles a las mismas la condición de sujetos de derechos, lo que puede ser atractivo para el derecho civil, evitando lagunas de responsabilidad, pero no para el Derecho penal, puesto que se puede castigar a la persona física "detrás" de ellas.

128 Sobre el concepto de autonomía y su negación en los sistemas de IA, Isidoro Blanco Cordero, "Un Derecho penal de humanos", 381 ss.: "*No creo que se pueda afirmar* que las inteligencias artificiales más avanzadas actualmente tengan la capacidad de tomar decisiones completamente autónomas y libres. Aunque estas tecnologías pueden ser capaces de realizar tareas y tomar ciertas decisiones, todavía dependen en gran medida de las instrucciones y límites establecidos por su programador. Difícilmente se puede aseverar, por lo tanto, que las decisiones que adopta son consecuencia de un proceso de deliberación complicado al estilo del ser humano. Si acaso, son el resultado del algoritmo diseñado por el programador, que lleva a cabo un elevadísimo número de operaciones matemáticas que procesan todas las posibilidades que puede calcular con base a sus conocimientos". Fernando Miró Linares, "Derecho penal y desafíos de la inteligencia artificial", 191, considera que los sistemas de IA no son agentes morales susceptibles de responsabilidad penal. No tienen capacidad de actuar "en libertad" conforme a distintos escenarios y este es el verdadero problema ontológico. Se trata de sistemas que carecen actualmente de conciencia y de capacidad de autodeterminación "a la hora de elegir y evaluar las posibles soluciones a un problema y también a un mandato normativo".

el RIA establece la supervisión humana en los puntos críticos del proceso, principio fundamental que garantiza que el ser humano mantenga el control sobre la máquina, conserve su dignidad y evite la alienación frente a la tecnología¹²⁹. Teniendo en cuenta todos estos factores, si el legislador penal finalmente reconoce a la persona electrónica como sujeto activo de delitos¹³⁰, es probable que este reconocimiento tenga un alcance limitado, dado que en la práctica la responsabilidad penal seguirá dependiendo, en gran medida, de la actuación dolosa o imprudente de la persona física.

Después de todo lo expuesto, se hace evidente que una vez más existe una ficción, que el Derecho penal volvería a admitir, ya que también será la persona física en esta cuarta revolución la que genere el comportamiento jurídico penalmente-relevante, presentándose, con la regulación actual del CP, dos alternativas: o que sea directamente la persona física la responsable penalmente por los delitos cometidos por el (mal) uso de la IA, o que sea la persona jurídica, junto con la física en su caso, la responsable por los delitos cometidos por sus representantes o personas autorizadas por el (mal) uso de la IA o por los delitos cometidos por terceros por el (mal) uso de la IA debido a la falta de supervisión y control de aquellos. En esta línea, habría que estudiar si es oportuno incluir la exigencia de que este tipo de conductas deban reportar un beneficio a la persona jurídica, exigencia esta por la que no me inclino, pudiéndose incluir solo para algunos casos, y ello porque restringiría la protección penal de bienes jurídicos nucleares, catalogados como derechos fundamentales. En cualquier caso, esta última vía de responsabilidad alertará a las personas jurídicas y por ende a todo el sector empresarial y corporativo, ampliando sus programas de cumplimiento, que deberán ser adaptados a la normativa europea. Es en este ámbito en el que deberán trabajar los departamentos de *compliance* de las empresas para no incurrir en responsabilidad penal¹³¹.

En relación con la normativa europea, el Parlamento Europeo¹³² formula de manera similar este enfoque en el régimen de responsabilidad civil, dejando de lado la proyectada personalidad jurídica de las personas electrónicas, cuando expone en el punto séptimo: "que todas las actividades, dispositivos o procesos físicos o virtuales gobernados por sistemas de IA pueden ser técnicamente la causa directa o indirecta de un daño o un perjuicio, pero casi siempre son el resultado de que alguien ha construido o desplegado los sistemas o interferido en estos; observa, a este respecto, que no es necesario atribuir personalidad jurídica a los sistemas de IA; opina que la opacidad, la conectividad y la autonomía de los sistemas de IA podrían dificultar o incluso imposibilitar en la práctica la trazabilidad de acciones perjudiciales específicas de los sistemas de IA hasta una intervención humana específica o decisiones de diseño; recuerda que, de conformidad con conceptos de responsabilidad civil ampliamente aceptados, se puede eludir, no obstante, este obstáculo haciendo responsables a las diferentes personas de toda la cadena de valor que crean, mantienen o controlan el riesgo asociado al sistema de IA".

Llegados a este punto, considero que atribuir a la persona electrónica actualmente la condición de sujeto activo del delito, impulsado por una posible personalidad jurídica en el ámbito civil, no resulta necesario ni operativo en el Derecho penal¹³³. Castigar a la persona electrónica no permitirá cumplir de manera más eficaz con las funciones de prevención y retribución de la pena, y la víctima no se verá más resarcida por ello. No existe un interés jurídico-penal que justifique considerar a las máquinas como sujetos activos del delito. Con una adecuada estructuración de la imprudencia, la autoría y la posición de garante en el uso de sistemas de IA —categorías que fundamentarán la responsabilidad penal de la persona física y jurídica—, así como la posible tipificación de nuevas conductas delictivas, el Derecho penal podrá cumplir su función de protección de bienes jurídicos y prevención.

2.9. Revisión de algunos delitos en la PE

En este sentido, la parte especial del CP deberá revisarse para incorporar nuevas formas de delincuencia en las que esté involucrada la inteligencia artificial¹³⁴, y considero que serán un número nada despreciable los delitos que tendrán que adaptarse a estas nuevas formas de puesta en peligro y de lesión de los bienes jurídicos, así como a la introducción de figuras delictivas inéditas. No olvidemos que una de las prioridades del RIA es la protección de los derechos fundamentales, muchos de los cuales se encuentran directamente vinculados con estos bienes jurídicos. Por ejemplo, surgirán nuevas formas de lesionar la vida o la integridad física, como puede ocurrir mediante armas autónomas letales o automóviles autónomos; la intimididad, a través del uso de drones o de sistemas de IA biométrica; a otros derechos fundamentales¹³⁵ como la

129 Aritz Obregón Fernández y Guillermo Lazcoz Moratinos, "La supervisión humana", 6, que defienden, siguiendo a Jones, que la automatización merma la dignidad del ser humano y esta solo se puede restablecer por la intervención de un operador humano en la toma de decisiones (*human in the loop*: sujeto en el circuito). Y concluyen que: "la única manera de mantener la dignidad humana, como expresión de los derechos y libertades fundamentales de la ciudadanía, es mediante la supervisión humana sobre las decisiones automatizadas". (Aritz Obregón Fernández y Guillermo Lazcoz Moratinos, "La supervisión humana", 28). Ilustrativas y rotundas por demás son las palabras del Papa recogidas en su discurso (14-06-2024) en la sesión del G7 sobre inteligencia artificial los días 13-15 de junio de 2024: "Conviene recordar siempre que la máquina puede, en algunas formas y con estos nuevos medios, elegir por medio de algoritmos. Lo que hace la máquina es una elección técnica entre varias posibilidades y se basa en criterios bien definidos o en inferencias estadísticas. El ser humano, en cambio, no sólo elige, sino que en su corazón es capaz de decidir. La decisión es un elemento que podríamos definir el más estratégico de una elección y requiere una evaluación práctica. A veces, frecuentemente en la difícil tarea de gobernar, también estamos llamados a decidir con consecuencias para muchas personas. Desde siempre la reflexión humana habla a este propósito de sabiduría, la *phronesis* de la filosofía griega y, al menos en parte, la sabiduría de la Sagrada Escritura. Frente a los prodigios de las máquinas, que parecen saber elegir de manera independiente, debemos tener bien claro que al ser humano le corresponde siempre la decisión, incluso con los tonos dramáticos y urgentes con que a veces ésta se presenta en nuestra vida. Condenaríamos a la humanidad a un futuro sin esperanza si quitáramos a las personas la capacidad de decidir por sí mismas y por sus vidas, condenándolas a depender de las elecciones de las máquinas. Necesitamos garantizar y proteger un espacio de control significativo del ser humano sobre el proceso de elección utilizado por los programas de inteligencia artificial. Está en juego la misma dignidad humana." Disponible en: <https://www.vatican.va/content/francesco/es/speeches/2024/june/documents/20240614-g7-intelligenza-artificiale.html>.

130 Además, sin confeccionar una lista cerrada de delitos, como ocurre en el caso de la persona jurídica, porque todavía no tenemos el conocimiento suficiente para ello.

131 Esperemos que todo este esfuerzo no quede en agua de borrajas, como ocurre actualmente en EE. UU., debido, como señala Jesús María Silva Sánchez, "Para muestra... el caso Boeing", *InDret*, n.º 3 (2024): 1-4. "Para muestra... el caso Boeing", principalmente porque las grandes compañías representan un sector estratégico de la economía; en este contexto, la responsabilidad penal de las personas jurídicas se resuelve frecuentemente mediante acuerdos con la Fiscalía, en los que se fijan multas a la persona jurídica, compromisos de mejora en materia de cumplimiento y la designación de algunos chivos expiatorios, generalmente mandos intermedios, mientras que los administradores y altos directivos de estas corporaciones eluden la responsabilidad penal. Mediante estos acuerdos, multas y chivos expiatorios —que a menudo terminan en sentencias absolutorias— se evita el procedimiento penal completo, una condena efectiva a la persona jurídica y la persecución de sus administradores y directivos. La implantación y el uso de la inteligencia artificial, sin restar importancia al resto de casos y actividades, constituyen un tema trascendental en el que están en juego derechos fundamentales de las personas, lo que invita a reflexionar sobre cómo queremos articular la responsabilidad penal. Es crucial considerar, como apunta Jesús María Silva Sánchez, "Para muestra... el caso Boeing", 3, "en términos empíricos cuál es el impacto de la responsabilidad de la persona jurídica sobre sus administradores y directivos, de manera que produzca efectos expresivos y/o intimidatorios sobre estos". ¿Ha sido eficaz "castigar" a la persona jurídica? No debemos olvidar que el Derecho penal es ultima ratio, por lo que no todo puede resolverse mediante él. El castigo a la persona jurídica ha ampliado, sin embargo, la aplicación del Derecho penal, contribuyendo a su administrativización y permitiendo que la persona física, al menos en EE. UU., utilice esta figura para mantenerse en el terreno de la impunidad. Tal vez sea necesario reajustar el foco y centrar el esfuerzo en fundamentar exclusivamente la responsabilidad penal de la persona física en estos ámbitos.

132 Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de Inteligencia artificial (2020/2014).

133 Fernando Navarro Cardoso, "Anotaciones sobre la responsabilidad penal", 205, también se cuestiona si es necesaria desde una perspectiva práctica la responsabilidad directa de las *e-personas* "y no bastan las respuestas, además de civiles, penales pensadas para las personas físicas y las jurídicas".

134 Cfr. Fernando Miró Llinars, "Derecho penal y desafíos de la inteligencia artificial", 185 ss. En especial, considera que se tipificarán delitos que describan actos preparatorios relacionados con la producción, comercialización y diseño de sistemas de IA (p. 195), así como el posible impacto del RIA, sobre todo, el art. 5 del reglamento en la criminalización de nuevas conductas, (p. 203). En este sentido tb. Paz M.^a de la Cuesta Aguado, "El Derecho penal frente al Reglamento de Inteligencia Artificial", 83 ss.

135 Cfr. desde una perspectiva constitucional de la afectación de estos derechos, María Pérez-Ugena Coromina, "La inteligencia artificial: definición, regulación y riesgos para los derechos fundamentales", *Estudios de Deusto* 72, n.º 1 (2024): 307 ss.; Lorenzo Cotino Hueso, "Cómo abordar jurídicamente el impacto de la inteligencia artificial en los derechos fundamentales", en *Derecho y tecnologías*, dir. María Emilia Casas Baamonde, coord. Daniel Pérez del Prado (Madrid: Editorial Centro de Estudios Ramón Areces, 2025), 123 ss.; desde una perspectiva penal, Pere Simón Castellano, *La evaluación del impacto algorítmico en los derechos fundamentales* (Pamplona: Aranzadi, 2023), *passim*; Tomás de la Quadra-Salcedo Fernández del Castillo, "Derechos fundamentales en la sociedad digital y de la inteligencia artificial", en *Derecho y tecnologías* (Madrid: Editorial Centro de Estudios Ramón Areces, 2025), 79 ss.

igualdad¹³⁶, la libertad de expresión, la libertad de información o incluso a un derecho a la verdad¹³⁷; la salud, por fallos de diagnóstico del algoritmo o por sesgos en los datos durante los triajes; y el patrimonio, la propiedad intelectual¹³⁸ o la fe pública podrán verse comprometidos por el uso de la IA como herramienta para falsificar, apropiarse o suplantar. Los derechos de los trabajadores también podrían sufrir una mayor incidencia lesiva, por ejemplo, mediante el control de la productividad laboral, la facilitación de despidos o decisiones discriminatorias¹³⁹ en los procesos de selección, debido a la falta de neutralidad y objetividad de los datos utilizados por los algoritmos. Asimismo, podrían surgir prohibiciones penales destinadas a proteger el uso de la IA o a la propia IA, en caso de que se llegara a considerar un *tertius genus* o, aun sin tal consideración, pudiendo ser objeto de delitos de daños. Por último, como producto, la IA podría generar responsabilidad penal por producto defectuoso, lo que podría implicar una remodelación de los delitos contra los consumidores, aunque quizás menos relevante para homicidios, lesiones o delitos de daños¹⁴⁰. Todas estas posibles reformas o adaptaciones requerirán, en

136 Sobre la aplicación de la IA en la justicia penal y los peligros para la privacidad, así como un análisis de la discriminación algorítmica, cfr. Fernando Miró Llinares, "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por los robots", *Revista de Derecho Penal y Criminología*, n.º 20 (2018): 87 ss.

137 Cfr. sobre los riesgos de la desinformación y la posverdad, Cristina García Arroyo, "De la desinformación y la posverdad a través de las RRSS y la IA: ¿retos para el Derecho penal?", *Revista Penal*, n.º 56 (2025): 146 ss.

138 Son muchos los retos que se presentan para la protección de la propiedad intelectual frente a la IA, como ejemplo, resulta relevante la opinión del juez Sidney en el caso *The New York Times Company v. Microsoft Corporation et al*) cfr. al respecto Diego Fierro Rodríguez, "La opinión del juez Sidney Stein a favor de la propiedad intelectual ante OpenAI y otras cuestiones", *Diario La Ley*, n.º 97 (2025): 1 ss.; siguiendo con la generación de contenidos por la IA, Sebastián López Maza, "La generación de contenidos a través de inteligencia artificial y su incidencia en la propiedad intelectual", en *Derecho y tecnologías*, dir. María Emilia Casas Baamonde, coord. Daniel Pérez del Prado (Madrid: Editorial Centro de Estudios Ramón Areces, 2025), 217 ss.

139 Por ejemplo, la IA discrimina en los procesos de selección laboral entre el hombre y la mujer. Cfr. <https://cadenaser.com/nacional/2023/10/02/la-inteligencia-artificial-discrimina-en-los-procesos-de-seleccion-cadena-ser/>. Más ejemplos y un estudio analítico del delito de discriminación laboral en esta sociedad líquida algorítmica se pueden encontrar en el trabajo de Abadías Selma, 2023, 6 ss.

140 Cfr. por todos, Victoria Ibold, *Künstliche Intelligenz und Strafrecht* (Baden-Baden: Nomos, 2024), *passim*.



algunos casos, recurrir a la técnica de los delitos de peligro, ya sea abstracto o concreto, anticipando la respuesta penal para lograr un control más eficaz de los riesgos.

VI. ¿Hacia un Derecho penal digital?

A la luz de lo expuesto a lo largo de este trabajo, y una vez analizadas las principales tensiones que la irrupción de la IA introduce en las categorías de la teoría del delito, procede ahora formular algunas consideraciones finales a modo de corolario. No se trata de añadir nuevos problemas ni de ofrecer respuestas definitivas, sino de ordenar y sintetizar, con la prudencia que impone un fenómeno aún en desarrollo, las conclusiones alcanzadas en torno al papel del Derecho penal en este nuevo escenario tecnológico como respecto del modelo regulatorio propuesto por el RIA. Desde esta perspectiva, las reflexiones que siguen pretenden subrayar las líneas maestras que, a mi juicio, deben guiar la adaptación del Derecho penal a los sistemas de IA, manteniendo su función de garantía, su carácter de **última** ratio y su anclaje irrenunciable en la centralidad de la persona humana.

1. Esta cuarta revolución introduce una tecnología disruptiva que inaugura un nuevo cosmos que, como toda novedad, plantea importantes retos normativos, así como la necesidad de organizar y estructurar de forma eficaz el fenómeno de los sistemas de IA. En este contexto, el RIA me merece, en líneas generales una valoración positiva, pese a la coyuntura sistémica experimental al que está vinculado, lo que le confiere un cierto carácter de provisionalidad en los detalles. No ocurre así, sin embargo, con las dos estructuras maestras que constituyen la columna vertebral del Reglamento y que han sido concebidas con notable acierto: el *concepto de riesgo* —y su *estratificación en categorías de sistemas*— y el *principio de supervisión humana*.

Al concepto de riesgo se asocian cuidadosamente requisitos y obligaciones desarrollados de modo general en el RIA, que permiten delimitar el riesgo permitido. Por su parte, el principio de supervisión humana garantiza que el control no se desvincule de la persona humana, en una relación de pertenencia que preserva su condición de sujeto responsable como correlato de su libertad, sin menoscabo de la dignidad inherente al ser humano.

El éxito de la regulación dependerá en gran medida de estas dos estructuras y de su capacidad de readaptación, con el fin de mantener el deseado equilibrio entre la libertad de experimentar con sistemas de IA y la responsabilidad derivada de dicha experimentación. En este sentido, la regulación adopta un enfoque claramente antropocéntrico, orientado a garantizar que el progreso tecnológico sea seguro y esté al servicio del ser humano y de la humanidad. Finalmente, como estructura normativa con vocación de cumplimiento efectivo, el RIA incorpora un sistema sancionador destinado a asegurar la observancia de sus disposiciones.

2. El fenómeno de la IA es amplísimo; las áreas comprometidas son muchas y los aspectos que van a surgir, y que habrá que analizar, intuyo que serán numerosos y diversos. En este trabajo, sin perder de vista la complejidad del fenómeno, únicamente se ha pretendido ofrecer unas mínimas reflexiones sobre el papel del Derecho penal en esta irrupción tecnológica, aportando algunas ideas para la discusión y la reflexión, con el objetivo de invitar a una mayor profundización. Las conclusiones a las que he llegado, que sintetizo y expongo tintadas de prudencia y provisionalidad ante la novedad y el continuo avance y desarrollo que promete el fenómeno, son:

- a. El Derecho penal como *ultima ratio*, como *hard law* y como elemento sólido del ordenamiento, refuerza las regulaciones extrapenales y su cumplimiento de modo indirecto, mediante una fuerte protección de los bienes jurídicos que merecen una incontestable protección penal.
- b. Los sistemas de IA presentan *problemas de causalidad* debido a la opacidad epistémica que presentan en su entrenamiento, comportándose como una "caja negra". Este problema se pretende combatir mediante la obligación de transparencia y explicabilidad del proceso. Aunque no es un problema nuevo para el Derecho penal, propongo aplicar en estos casos la causalidad estadística, que la jurisprudencia española y alemana ha utilizado en algunos casos, junto con inferencias "iuris tantum", dotadas de una elevada consistencia epistemológica.
- c. En esta fenomenología también *se refleja la íntima relación entre la imputación objetiva, el riesgo(no) permitido y la imprudencia*, a través de la ecuación: infracción de la norma de cuidado/descontrol del riesgo imputable objetivamente/riesgo no permitido.
- d. Considero necesario realizar *ajustes en la previsibilidad versus imprevisibilidad como elemento de la imprudencia*, para evitar que la falta de previsibilidad se convierta en un "coladero" para las defensas, usado de manera recurrente, como un automatismo, argumentando un elevado nivel coyuntural de imprevisibilidad, debido a la opacidad epistémica de los sistemas de IA. En el trabajo se plantea un intento de reforzamiento de la previsibilidad a través de dos correctivos:
 - I. Fase de prueba y poscomercialización: En esta etapa recae un deber esencial de alcanzar estándares eficaces y razonables de previsibilidad, tanto en la fase de prueba como en la de poscomercialización. Durante la fase de prueba, este deber se concreta en entrenamientos rigurosos, acompañados de validaciones y pruebas de seguridad de alta calidad, conforme al estado de la técnica, con el objetivo de anticipar de manera efectiva el comportamiento del sistema. Posteriormente, en la fase de poscomercialización, la obligación se mantiene mediante una vigilancia y control continuos a lo largo de todo el ciclo de vida del sistema de IA. Este enfoque reconoce que nos enfrentamos a un fenómeno tecnológico nuevo, cuyo desarrollo y efectos aún se encuentran en evolución, por lo que es imprescindible ir nutriendo progresivamente el conocimiento y la experiencia sobre su funcionamiento e impacto en la práctica. La finalidad última de este deber es prever situaciones críticas y garantizar que el riesgo se mantenga dentro de los límites permitidos, asegurando un control responsable y seguro de la inteligencia artificial.
 - II. Control humano en situaciones críticas: No se deben realizar trabajos de experimentación y, en ningún paso, se debe pasar a la fase de comercialización de un sistema que no disponga de tecnología que permita devolver el control al ser humano en situaciones críticas. En estos supuestos, no podrá alegarse imprevisibilidad del curso causal, ya que se conocen de antemano las deficiencias de control y de seguridad.
- e. Toda la dogmática de la autoría y la participación puede trasladarse al contexto tecnológico actual de la IA con mínimas adaptaciones. Los sujetos obligados podrán ser autores directos, entendiendo que la autoría directa no implica intervención "de propia mano" ni conexión espaciotemporal con la máquina; el sistema de

IA actúa como instrumento, con características especiales que deben tenerse en cuenta para reformular la noción tradicional de autoría directa, garantizando que la persona física siga siendo el centro de imputación y responsabilidad en el hecho delictivo. La autoría *queda reservada a los sujetos obligados de la cadena de valor*, conformando una lista cerrada de autores y situándose dentro de las reglas propias de los delitos especiales, aunque este círculo puede ampliarse mediante la división estructural del trabajo y la delegación, incorporando a terceros, en inicio *extraneus*, transformados en delegados llamados a cumplir los deberes especiales del delegante. La cadena de valor puede organizarse en equipos de trabajo con competencias definidas y en secciones o departamentos, siguiendo un orden jerárquico o de igualdad, lo que configura un esquema vertical u horizontal del trabajo. En este contexto, las conductas de los sujetos pueden comprenderse dentro de categorías dogmáticas que van desde la coautoría, cuando existe conocimiento mutuo de la conducta de los intervinientes y acuerdo común sobre la toma de decisiones y el modo de proceder, con codominio del mismo, hasta otras formas de participación, como la cooperación necesaria o no necesaria, e incluso mediante formas de inducción. La autoría se define de acuerdo con los criterios tradicionales, recurriendo a criterios ontológico-normativos como el de la determinación del hecho mediante la realización de la conducta intrínsecamente más peligrosa en el contexto concreto. La intervención de varios sin acuerdo común en la comisión de un delito doloso o imprudente, pueden adoptar modalidades de autoría sucesiva, aditiva, accesoria o yuxtapuesta, así como autoría en cadena, sin perjuicio de la concurrencia de imprudencias, al igual que formas de participación que no dominan o determinan el hecho, pero facilitan la infracción de la norma de cuidado.

- f. Continuando en el ámbito de la intervención, pero desde otra óptica, se plantea la posibilidad de que la persona electrónica sea sujeto activo del delito, inclinándose la respuesta hacia el rechazo, por los mismos obstáculos que históricamente se han invocado para oponerse a que la persona jurídica sea sujeto activo y que pese a que el legislador los haya obviado, siguen estando presentes. Aunque los sistemas de IA toman decisiones aparentemente autónomas, es difícil admitir que actúen con dolo, ya que no pueden pensar, reflexionar ni tener representaciones con significado. Tampoco pueden ser imputables ni captar el mensaje preventivo y retributivo de la norma penal. El test de culpabilidad evidencia que son máquinas. Finalmente, las penas¹⁴¹ que se pudieran imponer a un sistema de IA no resarcan a la víctima, que no identificaría a la máquina como responsable del daño y sufrimiento causado.
- g. Por otro lado, los sujetos obligados del RIA actúan como garantes de la cadena de valor, lo que también sostiene la naturaleza especial de los delitos dolosos o imprudentes que puedan cometer, dado que son sujetos con deberes especiales concretos orientados a la protección de los bienes jurídicos. Estos garantes son "deudores de seguridad" y, en la mayoría de los casos, responsables de evitar el descontrol del riesgo, por lo que pueden incurrir en delitos, normalmente, imprudentes de omisión pura de garante, que incluso podrían elevarse, por imperativo legal, a comisiones omisivas, aun cuando su omisión no produzca directamente un delito de resultado, por decisiones de política-criminal, que pueden conectarse a la estructura de un delito de deber (*Pflichtdelikte*) para dotarla

141 Sobre las penas a imponer a la e-persona y sus funcionalidades cfr. Isidoro Blanco Cordero, "Un Derecho penal de humanos", 387 ss.

de fundamentación jurídica. Asimismo, observando la regulación europea, se plantea una responsabilidad por gestión del riesgo basada en el binomio posición de garante/riesgo permitido. Por último, el Derecho penal refuerza el principio de supervisión humana¹⁴² —pilar irrenunciable— mediante la imposición de una pena personal, intransferible y grave al garante. En definitiva, se trata de garantizar que el responsable vele porque todo el uso del sistema de IA durante su ciclo de vida se mantenga dentro del riesgo permitido y no se pierda el control sobre la máquina, ni en la fase de experimentación ni en la de uso de los sistemas de IA. La categoría dogmática de la posición de garante desempeña, como se puede observar, un papel trascendental en este contexto regulativo diseñado por el RIA.

- h. Es imprescindible considerar los programas de cumplimiento, plenamente comprometidos por la regulación del RIA. Se plantea si el *compliance* penal debe atender a la naturaleza de instrumento de la IA o, por el contrario, otorgar a la IA el estatuto de persona (persona electrónica, e-persona). La conclusión es que la responsabilidad penal de la persona electrónica tendría que basarse actualmente en las actuaciones típicas y antijurídicas de la persona física. Otorgar personalidad jurídica a la persona electrónica no permite declarar responsabilidad penal efectiva, moviéndonos nuevamente en el terreno de la ficción, como ocurre con la persona jurídica, cuya responsabilidad penal depende de la persona física. Además, se impone adaptar los programas de cumplimiento a los requisitos y obligaciones del RIA, para excluir, en lo posible, la responsabilidad penal de la persona jurídica y trasladar favorablemente la responsabilidad a sus representantes.
- i. Pasando a la parte especial, apunto la necesidad de revisar algunos delitos para adecuarlos al uso de la IA como herramienta delictiva, así como proteger a la IA de los daños que se puedan producir en su código o en su programación y uso; por último, hay que prestar atención a los cambios que se pueden plantear en la responsabilidad por el producto y en los delitos de homicidio y lesiones.

142 M.^a Ángeles Rueda Martín, "La concreción del deber objetivo de cuidado", 195 ss., considera la supervisión humana como un elemento consustancial para la definición del deber objetivo de cuidado y aunque son muchas las bondades de los sistemas de IA en la aplicación médica, deben ser un complemento, pero no pueden reemplazar al ser humano, que debe mantener un dominio sobre estas tecnologías (M.^a Ángeles Rueda Martín, "La concreción del deber objetivo de cuidado", 209, 218 y s.). Como he sostenido en este trabajo, dicho deber objetivo de cuidado, en muchas ocasiones, define y coincide con el deber especial del garante, que en este ámbito, cobra especial relevancia el alto contenido de deber de supervisión en el uso y aplicación de la IA.

Conclusión

La irrupción de la IA constituye un hito dentro del avance tecnológico y de nuestro proceso evolutivo. Se trata de un fenómeno que está transformando nuestro presente y que nos proyecta hacia un futuro, que hasta hace poco solo era imaginable en la ciencia ficción, en el que las máquinas adquieren un protagonismo creciente. Sin embargo, más allá de las denominaciones —computación cognitiva, aprendizaje mecánico o profundo, redes neuronales o inteligencia aumentada—, la IA no deja de ser, en el estado actual de desarrollo, una herramienta puesta al servicio de la humanidad. Una herramienta extraordinariamente potente, comparable, en términos de impacto, a grandes revoluciones tecnológicas del pasado, como lo fue en su momento por ejemplo la llegada del ferrocarril.

La diferencia sustancial de esta cuarta revolución respecto de las anteriores radica en que, por primera vez, se está emulando en la máquina capacidades propias de la inteligencia humana, hasta ahora privativas del ser humano. La IA es una devoradora de datos, millones de datos procesados por algoritmos a velocidades inalcanzables para el cerebro humano, una capacidad que se incrementará exponencialmente con el desarrollo de la computación cuántica. Ello nos conducirá, previsiblemente, al recurso al dato sintético cuando los datos reales resulten insuficientes para alimentar esta voracidad. En definitiva, el ser humano está creando a un competidor potencial en el control y la gestión de ámbitos esenciales de la vida.

En este contexto, resulta evidente que el uso de esta herramienta, como el de cualquier otra, debe ser un *uso responsable*, lo que exige de forma ineludible el control y la supervisión humana. Este desafío se intensifica especialmente en el ámbito de la IA generativa, combinada con la aspiración humana de desarrollar una superinteligencia capaz de programarse a sí misma, conectarse con otras IAs e interpretar emociones. El Derecho penal está llamado a contribuir a que esta revolución tecnológica no se revuelva contra su creador y a que el uso responsable se configure como un objetivo transversal y permanente dentro del diseño de un sistema de responsabilidad jurídica integral.

Desde esta perspectiva, el Derecho penal, como *ultima ratio* y como *hard law* del ordenamiento, refuerza indirectamente las regulaciones extrapenales mediante la protección de bienes jurídicos que merecen una tutela penal incontestable. La introducción de sistemas de IA plantea problemas clásicos de la dogmática penal y será necesario una reinterpretación adaptativa de categorías jurídico-penales. Como he tratado de mostrar en este trabajo a modo de primera aproximación, ello no implica, en el marco de este Derecho penal dual —analógico y digital—, la renuncia a la solidez conceptual de la teoría jurídica del delito basada en garantías y límites. Nos encontramos, además, en estos momentos en un periodo de transición que exige seguir avanzando en esta dirección, con una implicación necesaria de todos los Estados, dado que el uso

de la IA proyecta las cuestiones jurídicas a un plano internacional y de Derecho penal comparado¹⁴³.

Por último, quisiera reivindicar desde esta tribuna la necesidad de una alfabetización digital y en el uso de la IA, que ofrezca una visión sintética de sus beneficios y de sus riesgos, de su marco regulatorio y de los mecanismos de defensa y protección del usuario. Ello resulta imprescindible porque esta tecnología es, en sí misma, neutra: puede convertirse en nuestro *coworker* sin sustituirnos, facilitando la vida, o en nuestro enemigo; todo dependerá de uso que hagamos de ella. El científico computacional y psicólogo cognitivo Geoffrey Hinton, Premio Nobel de Física en 2024 y conocido como el "padrino" de la IA, renunció a su puesto en Google y alertó sobre los peligros de la IA para la humanidad¹⁴⁴, advirtiendo además de que estos riesgos podrían materializarse en un plazo más breve del inicialmente previsto, azuzados y favorecidos además por la competencia voraz entre Microsoft, Google y otras grandes tecnológicas por llegar las primeras en esta carrera hacia un abismo nuclear tecnológico, inimaginable incluso para Oppenheimer. El progreso siempre se ha comportado costes sociales y, desgraciadamente, vidas humanas; sin embargo, en esta ocasión los costes pueden resultar especialmente impactantes y explosivos para la supervivencia de la humanidad¹⁴⁵. Discutir sobre Derecho penal y sobre la responsabilidad de la persona física, de la persona jurídica o de la persona electrónica se vuelve baladí, irrisorio y hasta ridículo si perdemos el control sobre estas "cosas". No podemos permitirlo: nos va en ello la existencia misma de la humanidad¹⁴⁶.

Bibliografía

Abadías Selma, Alfredo. "La protección penal de los trabajadores frente a la inteligencia artificial en el ámbito del delito de discriminación laboral". *Revista de Derecho Penal*, n.º 39 (2023).

Armaza Armaza, Emilio José, Elena Atienza Macías, Miguel Ángel Boldova Pasamar, Iñigo de Miguel Beriain, Aliuska Duardo Sánchez, Aitziber Emaldi Cirión, Fátima Flores Mendoza, Luis Fernando Hernández García, Guillermo Lazcoz Moratinos, Jon López Gorostidi, Pilar Nicolás Jiménez, Ekain Payán Ellacuria, Andrea Perin, Carlos María Romeo Casabona, María Ángeles Rueda Martín, Esteban Sola Reche, Carlos Trincado Castán, Asier Urruela Mora, Silvia Verdugo Guzmán. "Ciberseguridad, ciberdelitos e inteligencia artificial: apuntando al futuro". En *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial*, editado por Carlos M.ª Romeo Casabona y M.ª Ángeles Rueda Martín. Granada: Comares, 2023.

Asimov, Isaac. *Círculo vicioso (Runaround)*. Estados Unidos: Street & Smith, 1943.

Baldwin, Richard. *La convulsión globótica. Globalización, robótica y el futuro del trabajo*. Barcelona: Antoni Bosch Editor, 2019.

Bauman, Zygmunt. *Modernidad líquida*. Traducido por Mirta Rosenberg y Jaime Arrambide Squirru. Ciudad de México: Fondo de Cultura Económica, 2002.

Belzuz Abogados. "Algunas notas sobre la propuesta de Directiva sobre responsabilidad en materia de Inteligencia Artificial". *Belzuz Abogados*. Accedido 8 de octubre de 2024. <https://www.belzuz.net/es/publicaciones/en-espanol/item/12068-algunas-notas-sobre-la-propuesta-de-directiva-sobre-responsabilidad-en-materia-de-inteligencia-artificial.html>.

Blanco Cordero, Isidoro. "Un Derecho penal de humanos para entidades artificiales inteligentes. ¿Machina delinquere et puniri potest?". En *Manifestaciones de desigualdad en el sistema de justicia penal*, dirigido por Antonio Doval Pais y Elena Gutiérrez Pérez, vol. I. Pamplona: Aranzadi, 2023.

———. "Vehículos autónomos y responsabilidad penal: retos para la imputación de resultados lesivos". *Revista General de Derecho Penal*, n.º 44 (2025).

Boix Palop, Andrés. "Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones". *Revista de Derecho Público: Teoría y Método*, n.º 1 (2020).

Bravo Bolado, Andrea. "El Derecho penal ante la inteligencia artificial: problemas de imputación y propuestas de solución". Tesis doctoral inédita, Universidad Autónoma de Madrid, 2025.

Coca Vila, Ivó. "Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal". *Cuadernos de Política Criminal*, n.º 122 (2017).

Cotino Hueso, Lorenzo. "Cómo abordar jurídicamente el impacto de la inteligencia artificial en los derechos fundamentales". En *Derecho y tecnologías*, dirigido por María Emilia Casas Baamonde y coordinado por Daniel Pérez del Prado. Madrid: Editorial Centro de Estudios Ramón Areces, 2025.

Cuatrecasas. "Aspectos clave del Reglamento de IA". *Cuatrecasas*. Accedido 4 de octubre de 2024. <https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/ia-inteligencia-artificial-reglamento-claves>.

143 Cfr. Eric Hilgendorf, "Inteligencia artificial y Derecho penal", 40.

144 Cfr. Para mejor referencia: <https://www.bbc.com/mundo/articles/c8elggj48geo>. En esta entrevista, el científico explica y advierte sobre diversos riesgos de la IA, entre ellos el desarrollo de armas verdaderamente autónomas, la proliferación de noticias y fotografías falsas en internet, la revolución del mercado laboral con la pérdida de empleos más allá de las tareas mecánicas, y la mala práctica de permitir que la IA genere y ejecute su propio código sin supervisión humana.

145 En el mismo sentido, Andrés Boix Palop, "Los algoritmos son reglamentos", 227, que sostiene que la magnitud del riesgo puede tener consecuencias extintivas para la especie humana.

146 Juan Piscetta, "Éric Sadin, el intelectual que desafía al reino del algoritmo: 'La innovación digital trabaja para convertir toda actividad humana en obsoleta'", *Infobae*, accedido 12 de noviembre de 2024, <https://www.infobae.com/cultura/2020/07/25/eric-sadin-el-intelectual-que-desafia-al-reino-del-algoritmo-la-innovacion-digital-trabaja-para-convertir-toda-actividad-humana-en-obsoleta/>, alerta sobre la IA y sobre cómo su presencia va a deteriorar la capacidad cognitiva humana, hasta el punto de erosionar la toma de decisiones soberanas y la libertad individual, dando paso a una nueva forma de antropología en relación del ser humano con la tecnología y el poder de decisión. «"Vivimos un momento serio, de gran gravedad, pero no lo vemos", ha expresado, refiriéndose al impacto de las tecnologías digitales desarrolladas en las últimas dos décadas y a la velocidad vertiginosa con la que están transformando la vida cotidiana, lo que dificulta comprender el presente de manera clara e inmediata.» (Cfr. Disponible en: <https://elpais.com/proyecto-tendencias/2024-11-11/eric-sadin-filosofo-la-ia-modificara-la-naturaleza-del-ser-humano.html>).

De la Cuesta Aguado, Paz M.^a. "Inteligencia artificial y responsabilidad penal". *Revista Penal México*, n.º 16-17 (2020).

———. "El Derecho penal frente al Reglamento de Inteligencia Artificial". *Revista Ius Criminale*, n.º 1 (2024).

———. *Cuestiones jurídico-penales abiertas tras la entrada en vigor del Reglamento de Inteligencia Artificial*. Madrid: Dykinson, 2025.

De Rivera, Laura G. "Coches autónomos: ¿quién tiene la culpa en caso de accidente?". *Público DISPLAY CONNECTORS, SL*. 9 de marzo de 2024. <https://www.publico.es/ciencias/coches-autonomos-culpa-caso-accidente.html>.

Dopico Gómez-Aller, Jacobo. "¿Dogmática o probatoria? La cuestión de la causalidad y su prueba en las intoxicaciones masivas con agentes tóxicos desconocidos". *La Ley Penal*, n.º 169 (2024).

Escobar Vélez, Susana. "El traslado del principio de precaución al Derecho penal en España". *Revista Nuevo Foro Penal*, vol. 6, n.º 75 (2010).

Fernández Hernández, Carlos. "La comisión presenta una propuesta de Directiva sobre responsabilidad civil extracontractual en materia de IA". *Diario La Ley*, n.º 63 (29 de septiembre de 2022), Sección Ciberderecho. https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4slAAAAAAAEADVPwWrDMAz9mvk4nDUj68GXtDsURildGLsqjnAMrpXZclb__bR2EzzEEo_vSV8FUx3wyiajK4lUrpFivZghFVQMYab9qGzWtAlnhRYLhD2ZE3zrNtf6lccYBSdojRh6qvRiokhnDGbRm-7rcozfRgh9Q7YU-wh3f39NJnXTy3VdJuufVErpiwC8-EdRkY1eze_CfiuzwjJzidwaA7RWo-PkJfr37gvzLI6cny_cWWD9Dow7iBgnP5DYVICPVOQ4258oSzfllsEHulOEpWMwegfrZclgRoBAAA=WK.

Fierro Rodríguez, Diego. "La opinión del juez Sidney Stein a favor de la propiedad intelectual ante OpenAI y otras cuestiones". *Diario La Ley*, n.º 97 (2025).

Galán Muñoz, Alfonso. "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*, n.º 23 (2015).

García Arroyo, Cristina. "De la desinformación y la posverdad a través de las RRSS y la IA: ¿retos para el Derecho penal?". *Revista Penal*, n.º 56 (2025).

González Beluche, Paloma. "La adaptación de la Directiva 85/374/CEE, de 25 de julio, en materia de responsabilidad por daños causados por productos defectuosos a la cuarta revolución industrial". *Cuadernos de Derecho Transnacional* 15, n.º 2 (2023).

Greco, Luis. "Vehículos de motor autónomos y situaciones de colisión". En *Libro Homenaje al prof. Dr. Agustín Jorge Barreiro*, coordinado por Gonzalo Basso Gentili y Manuel Cancio Meliá. España: Universidad Autónoma de Madrid, Servicio de Publicaciones, 2019.

Heras Martín, Paloma. "El tratamiento de datos biométricos: estudio del caso AENA bajo los estándares europeos de legitimación, minimización y proporcionalidad". *Diario La Ley*. Aranzadi, 18 de diciembre de 2025.

Hilgendorf, Eric. "Können Roboter schuldhaft handeln?". En *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*. Baden-Baden: Nomos, 2012.

———. "Dilemma-Probleme beim automatisieren Fahren. Ein Beitrag zum Problem des Verrechnungsverborts im Zeitalter der Digitalisierung". *Zeitschrift für die gesamte Strafrechtswissenschaft (ZStW)*, n.º 130 (2018).

———. "Vom Werkzeug zum Partner? Zum Einfluss intelligenter Artefakte auf unsere sozialen Normen und die Aufgaben des Rechts". En *Digitalisierung, Globalisierung und Risikoprävention. Festschrift für Ulrich Sieber*. Berlin: Duncker & Humblot, 2022.

———. "Inteligencia artificial y Derecho penal". Traducido por Leandro Dias Lestón. En *Desafíos penales de hoy: Entre la ley y la justicia en la obra de Eric Hilgendorf*. Buenos Aires: Editores del Sur, 2024.

Hörnle, Tatjana, y Wolfgang Wohlers. "The Trolley Problem Reloaded. ¿Cómo deben programarse los vehículos autónomos para los dilemas de 'vida contra vida'?". Traducido por José Béguelin. En *Criminalización, castigo, y dilemas morales en la obra de Tatjana Hörnle*. Buenos Aires: Editores del Sur, 2022.

Ibold, Victoria. *Künstliche Intelligenz und Strafrecht. Zur strafrechtliche Produktverantwortung in der Innovationsgesellschaft*. Baden-Baden: Nomos, 2024.

Jhering, Rudolf von. *La lucha por el Derecho*. Estudio preliminar y edición de Luis Lloredo Alix. Madrid: Dykinson, 2018.

Jiménez García, Francisco, dir. *Seguridad y responsabilidad penal e internacional en el uso de las TIC y la inteligencia artificial*. Coordinado por Sandra López de Zubiria Díaz y Berta Alam-Pérez. Madrid: Iustel, 2024.

Kooperationen: GI – Gesellschaft für Informatik, DFKI – Deutsches Forschungszentrum für Künstliche Intelligenz, KI Bundesverband, ZMDT – Zentrum für medizinische Datennutzbarkeit und Translation. *Künstliche Intelligenz und Recht*. Munich: Beck, 2026.

Lazcoz Moratinos, Guillermo, Asier Urruela Mora e Iñigo De Miguel Beriain. "Propuesta de Reglamento «Ley de Inteligencia Artificial»: Análisis de su posible impacto en la utilización de sistemas automatizados en el ámbito del Derecho penal". En *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial*, editado por Carlos M.^a Romeo Casabona y M.^a Ángeles Rueda Martín. Granada: Comares, 2023.

López Maza, Sebastián. "La generación de contenidos a través de inteligencia artificial y su incidencia en la propiedad intelectual". En *Derecho y tecnologías*, dirigido por María Emilia Casas Baamonde y coordinado por Daniel Pérez del Prado. Madrid: Editorial Centro de Estudios Ramón Areces, 2025.

Méndez Serrano, María del Mar. "Derechos fundamentales y personalidad jurídica de los robots: ¿para qué?". *Derecho Privado y Constitución*, n.º 44 (2024): 51-89.

Mínguez Rosique, Marina, y David Gallego Arribas, coords. *Ciberdelitos: Tendencias y desafíos actuales*. Madrid: BOE, 2025.

Miró Llinares, Fernando. "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por los robots". *Revista de Derecho Penal y Criminología*, n.º 20 (2018).

———. "Derecho penal y desafíos de la inteligencia artificial (en el contexto de nuevo marco regulatorio europeo)". En *Derecho y tecnologías*, dirigido por María Emilia Casas Baamonde y coordinado por Daniel Pérez del Prado. Madrid: Editorial Centro de Estudios Ramón Areces, 2025.

Morán Espinosa, Alejandra. "Responsabilidad penal de la Inteligencia artificial (IA). ¿La próxima frontera?". *Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. 15, n.º 48 (2021): 289-323.

Navarro Cardoso, Fernando. "Anotaciones sobre la responsabilidad penal de las «personas artificiales»". *Revista Penal*, n.º 56 (2025): 198-207.

Obregón Fernández, Aritz, y Guillermo Lazcoz Moratinos. "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*, n.º 42 (2021).

Panattoni, Beatrice, y Lorenzo Picotti. "Traditional Criminal Law Categories and AI: Crisis or Palingenesis?". *Revue Internationale de Droit Pénal* 94, n.º 1 (2023).

Paredes Castañón, José Manuel. "Responsabilidad penal por productos defectuosos". *Revista Foro Fundación Internacional de Ciencias Penales*, n.º 2 (2024).

Pérez Manzano, Mercedes. "Algunos datos empíricos sobre la atribución de estados mentales: ¿fracaso del principio de responsabilidad subjetiva o de un determinado concepto de dolo?". *Revista Electrónica de Ciencia Penal y Criminología*, n.º 23-15 (2021).

Pérez-Ugena Coromina, María. "La inteligencia artificial: definición, regulación y riesgos para los derechos fundamentales". *Estudios de Deusto* 72, n.º 1 (2024).

Perrigo, Billy. "Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation". *Time*, 20 de junio de 2023. <https://time.com/6288245/openai-eu-lobbying-ai-act/>.

Piscetta, Juan. "Éric Sadin, el intelectual que desafía al reino del algoritmo: "La innovación digital trabaja para convertir toda actividad humana en obsoleta"". *Infobae*. Accedido 12 de noviembre de 2024. <https://www.infobae.com/cultura/2020/07/25/eric-sadin-el-intelectual-que-desafia-al-reino-del-algoritmo-la-innovacion-digital-trabaja-para-convertir-toda-actividad-humana-en-obsoleta/>.

Quadra-Salcedo Fernández del Castillo, Tomás de la. "Derechos fundamentales en la sociedad digital y de la inteligencia artificial". En *Derecho y tecnologías*, dirigido por María Emilia Casas Baamonde y coordinado por Daniel Pérez del Prado. Madrid: Editorial Centro de Estudios Ramón Areces, 2025.

Quintero Olivares, Gonzalo. "La Robótica ante el Derecho penal: el vacío de respuesta jurídica a las desviaciones incontroladas". *Revista Electrónica de Estudios Penales y de la Seguridad*, n.º 1 (2017).

———. "Populismo y Derecho penal". *DOXA, Cuadernos de Filosofía del Derecho*, n.º 48 (2024).

Romeo Casabona, Carlos M.^a. "La discusión sobre la atribución de responsabilidad penal a sistemas de inteligencia artificial en particular a sistemas autónomos". En *Derecho penal, ciberseguridad, ciberdelitos e inteligencia artificial*, editado por Carlos M.^a Romeo Casabona y M.^a Ángeles Rueda Martín. Granada: Comares, 2023.

Roso Cañadillas, Raquel. *Autoría y participación imprudente*. Granada: Comares, 2002.

———. "Los delitos polivalentes de autoría: entre el deber y el dominio". *InDret*, n.º 3 (2019).

———. "La necesidad de diferenciar entre autoría y participación imprudente y la cuestión de su punibilidad". *Revista Foro Fundación Internacional de Ciencias Penales*, n.º 3 (2022).

———. "El profesional del art. 196 CP: Denegación de asistencia sanitaria o abandono de servicios sanitarios". En *Un modelo integral de Derecho penal. Libro-Homenaje a la profesora Mirentxu Corcoy Bidasolo*, vol. 2. Madrid: BOE, 2022.

———. "¿Un Derecho penal delicuescente en una sociedad líquida?". *Anuario de la Facultad de Derecho de la Universidad de Alcalá (UAH)*, n.º 26 (2023).

———. "¿Un Derecho penal delicuescente en una sociedad líquida? Algunas reflexiones sobre el Derecho penal en la sociedad posindustrial". *Revista General de Derecho Penal*, n.º 41 (2024): 1-30.

———. "La omisión del art. 318 y el algoritmo". En *Un Derecho penal moderno para una democracia del siglo XXI. Libro Homenaje al Prof. Carlos Martínez-Buján Pérez*. Valencia: Tirant lo Blanch, 2025.

Rueda Martín, M.^a Ángeles. "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*, n.º 1 (2025).

Saez Hurtado, Pablo. "La transparencia algorítmica ante el Tribunal Supremo: Comentario a la STS 1119/2025 (caso BOSCO). Un hito fundacional en el Derecho digital". *Diario La Ley*, (2025): 1-12.

Schwab, Klaus. *La cuarta revolución industrial*. Madrid: Marcial Pons, 2016.

Silva Sánchez, Jesús María. "Para muestra... el caso Boeing". *InDret*, n.º 3 (2024).

Simón Castellano, Pere. *La evaluación del impacto algorítmico en los derechos fundamentales*. Pamplona: Aranzadi, 2023.

Suárez, María Florencia. "Inteligencia artificial y Derecho penal. El dilema del tranvía. Cuarta Revolución industrial. Ética del Algoritmo. IA en vehículos. Causas de justificación". *Revista Pensamiento Penal*, n.º 445 (2022).

Unión Europea. *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo*. Diario Oficial de la Unión Europea. 13 de junio de 2024. 2024/1689.

Vega Agredano, Francisco. "Detección del delito mediante la inteligencia artificial: prevención automática de la criminalidad y responsabilidades por fallos del sistema". *Revista General de Derecho Penal*, n.º 44 (2025).

Velasco, Eloy. Entrevistado por Andrés Garvi Carvajal. "Robots delincuentes, respuestas jurídicas a un futuro cercano". *Cinco Días*. Accedido 4 de octubre de 2024. https://cincodias.elpais.com/legal/2022/02/03/juridico/1643900957_593967.html.

———. "Delitos tecnológicos de los informáticos a los cometidos por la IA". Conferencia de clausura en la 7.^a edición del Máster de Experto de Derecho Digital de la Universidad de Deusto, impartida el 22 de marzo de 2024. Accedido 4 de octubre de 2024. https://www.youtube.com/watch?v=uGdOmQ_sQVg.

Welzel, Hans. "Zum Notstandproblem". *Zeitschrift für die gesamte Strafrechtswissenschaft (ZStW)*, n.º 63 (1951).

Yuste, Rafael. "Neurotecnología y neuroderechos". En *Derecho y tecnologías*, dirigido por María Emilia Casas Baamonde y coordinado por Daniel Pérez del Prado. Madrid: Editorial Centro de Estudios Ramón Areces, 2025.

AUTÓMATAS Y AVATARES EN EL PROCESO PENAL:

LA IA AGÉNTICA Y EL ANÁLISIS POLICIAL FORENSE

Juan Carlos Ortiz Pradillo¹

Resumen

La irrupción de los sistemas y agentes de Inteligencia Artificial en la Justicia presenta importantes retos, ya sea para los propios operadores jurídicos, como sobre todo, para los letrados encargados de la defensa de los investigados. De igual modo que la evolución tecnológica y la posibilidad de descomponer la información digital en datos y metadatos abrió un nuevo horizonte en la prevención y persecución eficaz del delito, la aplicación de herramientas de Inteligencia Artificial en el ámbito forense de la recolección de grandes conjuntos de información digital (*Big Data*) y su clasificación y análisis (*Data Mining*) abre la puerta a un nuevo entendimiento del análisis pericial forense en el proceso penal que merece ser objeto de reflexión y examen.

Palabras clave

Inteligencia Artificial; Tecnología; Metadatos; Algoritmos; Investigación Policial; Proceso Penal.

¹ Universidad Complutense de Madrid. Instituto de Derecho Europeo e Integración Regional (IDEIR). Contacto: juancarlosortiz@ucm.es. ORCID: 0000-0001-6092-6137. WoS ResearcherID AAG-7424-2019. El presente trabajo constituye uno de los resultados del Proyecto I+D+i Nacional "Prueba penal y nuevos retos en el proceso penal transnacional". Ref. PID2023-148413NB-I00.

Abstract

The emergence of artificial intelligence systems and agents within the justice system poses significant challenges for legal practitioners and, more importantly, for defense lawyers. Just as technological advances and the ability to break down digital information into data and metadata have opened up new possibilities for the effective prevention and prosecution of crime, the application of artificial intelligence tools in the field of forensics —specifically in the collection of large sets of digital information (big data) and its classification and analysis (data mining)— opens the door to a new understanding of forensic expert analysis in criminal proceedings. This development does merit a deep reflection and scrutiny.

Keywords

Artificial Intelligence; Metadata; Technology; Algorithm; Police Surveillance; Criminal Justice.

SUMARIO: I. Introducción. II. Inteligencia Artificial y administración de justicia. 1. Inteligencia artificial y la “predicción del comportamiento criminal”: vigilancia masiva y policía predictiva. 1.1. La Inteligencia Artificial y la elaboración de pronósticos delictivos. 1.2. *Data mining*, vigilancia masiva en tiempo real y reconstrucción de movimientos (geolocalización histórica) de los investigados. III. Los metadatos y su utilización en la investigación policial del delito. 1. Metadatos extraídos de los instrumentos de convicción. 1.1. Los metadatos como elementos probatorios de la autoría delictiva. 1.2. Los metadatos como elementos probatorios de la inocencia del acusado. 2. Metadatos obtenidos de fuentes abiertas. 2.1. Internet y las redes sociales como “lugar de comisión del delito” y como lugar de recolección de los vestigios y las fuentes de prueba. 2.2. Técnicas e instrumentos de recolección de los metadatos de la información transmitida a través de internet y las redes sociales. 2.3. Colaboración público-privada en el rastreo de metadatos de información ilícita transmitida a través de internet y las redes sociales. 2.4. Rastreo de metadatos de la información transmitida a través de internet y las redes sociales y derecho fundamental al secreto de la correspondencia. IV. Los sistemas de inteligencia artificial como analistas forenses. V. Conclusiones. VI. Bibliografía.

I. Introducción

La fabricación de máquinas que imiten al ser humano es una constante en la historia de la humanidad constatable, al menos, desde hace más de 4.000 años. Hacia el año 1300 a. C., Amenhotep hizo construir una estatua de Memon, rey de Etiopía, que emitía sonidos cuando la iluminaban los rayos del sol al amanecer. Arquitas de Tarento (hacia el 400 a. C.) es considerado el padre de la ingeniería mecánica y uno de los maestros clásicos de la robótica occidental, por haber construido un autómatas consistente en una paloma de madera que rotaba por si sola gracias a un surtidor de agua o vapor y simulaba el vuelo.

Y en mi condición de toledano de nacimiento, no podía faltar la alusión al ingeniero e inventor Juanelo Turriano (originalmente Giovanni Torriani), quien, en 1556, llega a tierras toledanas al servicio de Carlos I, como relojero de la Corte, y construyó, entre otros, dos célebres autómatas en mi ciudad: el “hombre de palo” –un autómatas de madera que paseaba por las calles de Toledo– y el “artificio”: un conjunto de torres que permitía subir el agua desde el Río Tajo hasta el Alcázar salvando una pendiente de 90 metros.

Estos artilugios históricos buscaban lo mismo que buscan los robots modernos: la automatización de tareas para ejecutar actividades sin intervención humana constante. No en vano, el término “robot” se atribuye a Karel Capek (en su obra de teatro *Los robots universales del Rossum*, 1920) y esa palabra, robot, viene del vocablo checo “robota” que significaría “trabajo forzado”, en el sentido de cierta obligatoriedad o esclavitud.

En una sociedad en donde la interacción entre humanos y máquinas es ya una constante, la utilización de la tecnología por parte de las fuerzas y cuerpos de seguridad significa amplificar de forma abrumadora sus sentidos, su pericia y sus métodos de actuación. Sin embargo, el desarrollo de los grandes modelos de lenguaje al que asistimos en la actualidad ha cambiado completamente el paradigma de la automatización haciendo surgir el concepto de IA agéntica como el sistema basado en Inteligencia Artificial (IA) con capacidad de percibir su entorno, tomar decisiones, planificar y ejecutar acciones autónomamente y, así, lograr objetivos específicos con una mínima supervisión humana. A diferencia de la IA generativa tradicional (que responde a *prompts*), la IA agéntica actúa proactivamente, utiliza herramientas y aprende del entorno en el que se desenvuelve.

No se trata ya de la utilización de robots corpóreos o autómatas al servicio de la investigación criminal, sino más bien en avatares incorpóreos, algorítmicos, con plenas funciones de detección, investigación y análisis de la información que posteriormente será fuente de prueba en los procesos penales.

II. Inteligencia Artificial y administración de justicia



Hablar de IA agéntica en la justicia penal, y más en particular en el proceso penal, obliga a establecer una importante premisa²: diferenciar entre el "policing" y el "sentencing", esto es, entre el uso de la IA para la prevención e investigación policial de la delincuencia, que comúnmente se denomina Inteligencia Artificial Policial (IAP), y sobre la que versa el presente estudio, y el uso de la IA aplicada en la propia función jurisdiccional (automatización de notificaciones y resoluciones, determinación de la responsabilidad, adopción de medidas cautelares, ingresos en prisión, etc.), denominada Inteligencia Artificial Judicial (IAJ).

A su vez, y dentro de la IAP, los principales estudios se han focalizado en examinar el uso policial de la IA con fines predictivos a partir de la recolección, almacenamiento, tratamiento y análisis de la información a través de técnicas avanzadas de clasificación y segmentación –*Big Data* y *Data Mining*–, que permiten tener una visión sin precedentes del comportamiento humano³. De ahí que hayamos afirmado que en la actualidad ya asistimos a un auténtico cambio de paradigma en el modo de actuación de las autoridades policiales y judiciales a la hora de llevar a cabo sus funciones de prevención, detección, investigación y enjuiciamiento de los delitos, pues la búsqueda de "vestigios" se ha redirigido hacia el acopio y análisis de "datos", esto es, toda aquella información en formato electrónico relacionada con el entorno virtual de los sujetos intervinientes en cualquier comisión delictiva. De una parte, el entorno digital de la víctima, a través de cuyo análisis se habilitarán específicas líneas de investigación que agilicen la investigación de los hechos y conduzcan a la delimitación e identificación de sus atacantes. Y, de otra parte, el entorno digital de los lugares en que se produjeron los hechos presuntamente delictivos, así como de los sujetos sospechosos de haber participado en los mismos, de modo que la recopilación y examen de todo ese caudal de información digital podrá ser utilizado como fuentes de prueba acreditativas de la comisión y autoría de los hechos objeto de investigación⁴.

1. Inteligencia Artificial y la "predicción del comportamiento criminal": vigilancia masiva y policía predictiva

Los gobiernos no han dudado en aprovecharse de las posibilidades que la Sociedad de la Información ofrece a la hora de analizar el comportamiento humano para mejorar los resultados de sus políticas públicas y contribuir al desarrollo de sus naciones. El *Big Data* puede facilitar una mejor y más eficiente recaudación tributaria, tanto a la hora de determinar qué tributos pueden ser objeto de alza o reducción, como a la hora de aflorar actividades donde se focaliza la economía sumergida. La aplicación de la IA también puede hacer que se reduzca la siniestralidad en los centros de trabajo, pues a través del examen y análisis de la actividad laboral se pueden identificar tareas peligrosas y proceder a su robotización, así como también reducir la siniestralidad vial mediante el análisis del tráfico de vehículos en tiempo real y la interconexión entre vehículos, vías y

² Sobre ello, vid. Fernando Miró Llinares, "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, n.º 20 (2018): 87-130, doi:10.5944/rdpc.20.2018.26446.

³ La Resolución 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) utiliza el concepto de «macrodato», al que se refiere como *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)*.

⁴ Juan Carlos Ortiz Pradillo, "Big data, vigilancias policiales y geolocalización: nuevas dimensiones de los derechos fundamentales en el proceso penal", *Diario La Ley*, n.º 9955, 18 de noviembre de 2021.

centros de control, lo cual permite una movilidad más segura al identificar las vías, tramos y condiciones de mayor peligrosidad, así como también agilizar la toma de decisiones y respuestas ante accidentes de tráfico.

En el ámbito de la investigación del delito, el uso de la tecnología en la recopilación de grandes conjuntos de datos y su análisis inteligente a través de fórmulas matemáticas persigue dos grandes finalidades: de una parte, la reducción de la tasa de criminalidad y la prevención de conductas criminales mediante la formulación de patrones y pronósticos de "policía predictiva" referidos a los puntos calientes de criminalidad o los individuos con un mayor riesgo de comisión delictiva. Y, de otra parte, la investigación y persecución eficaz del delito presuntamente cometido, mediante la utilización de técnicas de "vigilancia masiva" o "tecnovigilancia"⁵.

1.1. La Inteligencia Artificial y la elaboración de pronósticos delictivos

Constituye el sueño de cualquier gobernante ser capaz de erradicar el crimen, o al menos, reducirlo a su mínima expresión, anticipándose a su realización a través de la adopción de medidas que eviten o disminuyan los factores que facilitan la comisión de delitos. Pero los actuales programas de "policía predictiva" poco tienen que ver con el relato *El informe de la minoría* (The Minority Report, escrito por Philip K. Dick en 1956 y llevado al cine por Steven Spielberg en 2002), en donde unos seres los "precognitivos" prevén todos los delitos antes de que ocurran y una división de la policía llamada *PreCrime* detiene a los sospechosos antes de que puedan cometer delitos reales.

La policía predictiva o *PredPol* se basa en la recopilación y análisis de grandes conjuntos de datos, sobre los que se aplican algoritmos que tienen en consideración múltiples factores espaciales, personales y ambientales (v. gr., los días en que se abonan los salarios, la ubicación de tiendas de bebidas alcohólicas, datos de ejecuciones hipotecarias y lanzamientos judiciales, condiciones climáticas, rutas de escape, celebraciones, eventos multitudinarios, etc.) para la elaboración de patrones y pronósticos delictivos basados, bien en la determinación de posibles lugares más propensos a ser objetivos de hechos delictivos, bien en la elaboración de listas de sospechosos o personas más propensas a cometer actividades criminales. Esto es, las autoridades policiales también emplean técnicas de predicción del comportamiento basada en la recopilación masiva de información digital y su análisis inteligente, gracias a herramientas de Inteligencia Artificial y *machine learning*, para la elaboración de patrones de delincuencia y mapas de riesgos con los que reducir la delincuencia mediante la elaboración de "pronósticos" delictivos⁶.

El mayor ejemplo de estas nuevas técnicas policiales lo encontramos en la ciudad de Nueva York. En 2012, su Departamento de Policía, en colaboración con Microsoft, anunció el lanzamiento del programa "Sistema de Conciencia de Dominio" *Domain Awareness System*, DAS, una medida inicialmente ideada para la lucha contra el terrorismo (sobre todo, tras el atentado terrorista en 2010 en Times Square), pero que actualmente es utilizado para la investigación de cualquier tipología delictiva. El sistema se nutre de los datos recopilados por las más de 9.000 cámaras CCTV y de reconocimiento automático

de matrículas dispuestas en las calles de Nueva York, torres con sistemas de detección de tiroteos y de emisiones de agentes químicos, las llamadas efectuadas al 911, datos extraídos de los informes de detenciones policiales, o de citaciones y órdenes judiciales.

1.2. Data mining, vigilancia masiva en tiempo real y reconstrucción de movimientos (geolocalización histórica) de los investigados

Junto con esa actuación predictiva, el empleo de las capacidades tecnológicas al servicio de la investigación del delito y centradas en la búsqueda y recogida de vestigios del delito y a la averiguación de sus autores permite a las autoridades resolver los delitos de forma más rápida, eficaz y segura⁷. La expresión "vigilancia masiva" alude al uso policial de técnicas de IA y *Big Data analytics* sobre toda esa información digital masivamente recopilada y almacenada por terceros (operadoras, prestadores de servicios de Internet, empresas privadas, administraciones públicas...) de cara a facilitar la investigación del delito mediante la reconstrucción de las circunstancias en que tuvo lugar una determinada conducta, quiénes se encontraban en el lugar de los hechos, cuáles fueron sus interacciones, y dónde quedaron almacenadas las fuentes de pruebas acreditativas de dichas circunstancias. Más allá de tratar de localizar e interrogar a los posibles testigos presenciales, se trata ahora de proceder a la recopilación y análisis de toda esa información digital que permita la identificación y captura de sus autores, y la prueba de su participación en los hechos delictivos. Frente a la fragilidad de la memoria humana, la masiva conservación de archivos digitales constituyen los nuevos "caladeros" a donde acudir.

Y, para ello, las autoridades ya no necesitan invertir aunque lo hagan en dicha recolección y conservación, les basta con conseguir la colaboración de las empresas privadas, que son quienes manejan toda esa ingente cantidad de información creada, almacenada o transmitida en formato digital. Los Estados tecnológicamente desarrollados, también los democráticos, se han convertido en lo que Balkin califica como *Estados vigilantes*⁸, caracterizados por servirse de las capacidades de las empresas privadas, no sólo para brindar servicios sociales valiosos, sino también para identificar problemas, evitar amenazas potenciales y combatir la delincuencia.

Como ejemplo concreto, ya hemos advertido sobre la posibilidad de poder "geolocalizar en tiempo real"⁹ a un individuo sospechoso de haber cometido un delito a través de diversas herramientas tecnológicas, tales como el uso de balizas adosadas a su vehículo, ropa u objetos que porte consigo; vigilancias policiales sistemáticas, apoyadas por drones y otros instrumentos dotados de cámaras de visión térmica y nocturna o sistemas de reconocimiento facial o de determinadas características biométricas del investigado; o el acceso en tiempo real a los datos de localización que emita su propio terminal móvil. Pero resulta que la recolección y análisis inteligente de datos personales almacenados por entidades privadas, con otros fines comerciales, también permite reconstruir la actividad llevada a cabo por una persona y conocer con gran precisión los lugares visitados, las rutas utilizadas, las personas con las que contactó, sus hábitos y rutinas (geolocalización histórica)¹⁰: la policía puede servirse para ello de los datos referidos a la

⁷ Juan Carlos Ortiz Pradillo, *Problemas procesales de la ciberdelincuencia* (Madrid: Colex, 2013), 28.

⁸ Jack M. Balkin, "The Constitution in the National Surveillance State", *Minnesota Law Review* 93, n.º 1 (2008): 1-25.

⁹ Juan Carlos Ortiz Pradillo, "Big data, vigilancias policiales y geolocalización", *Diario La Ley*, n.º 9955, 18 de noviembre de 2021.

¹⁰ Como específico examen de las diferencias y métodos de geolocalización histórica o reconstrucción de movimientos, vid. Juan Carlos Ortiz Pradillo, "Vigilancias policiales y utilización de dispositivos de seguimiento, localización y captación de la imagen", en *Reflexiones en torno al*

⁵ Además de la expresión "mass surveillance", otros términos anglosajones comúnmente utilizados son los de "electronic surveillance", "internet surveillance" y "online surveillance". Nosotros preferimos su traducción como "tecnovigilancia". Vid., por todos, Manuel Llamas Fernández y José Miguel Gordillo Luque, "Medios técnicos de vigilancia", en *Los nuevos medios de investigación en el proceso penal. Especial referencia a la tecnovigilancia*, Cuadernos de Derecho Judicial 2007-II (Madrid: Consejo General del Poder Judicial, 2007), 236.

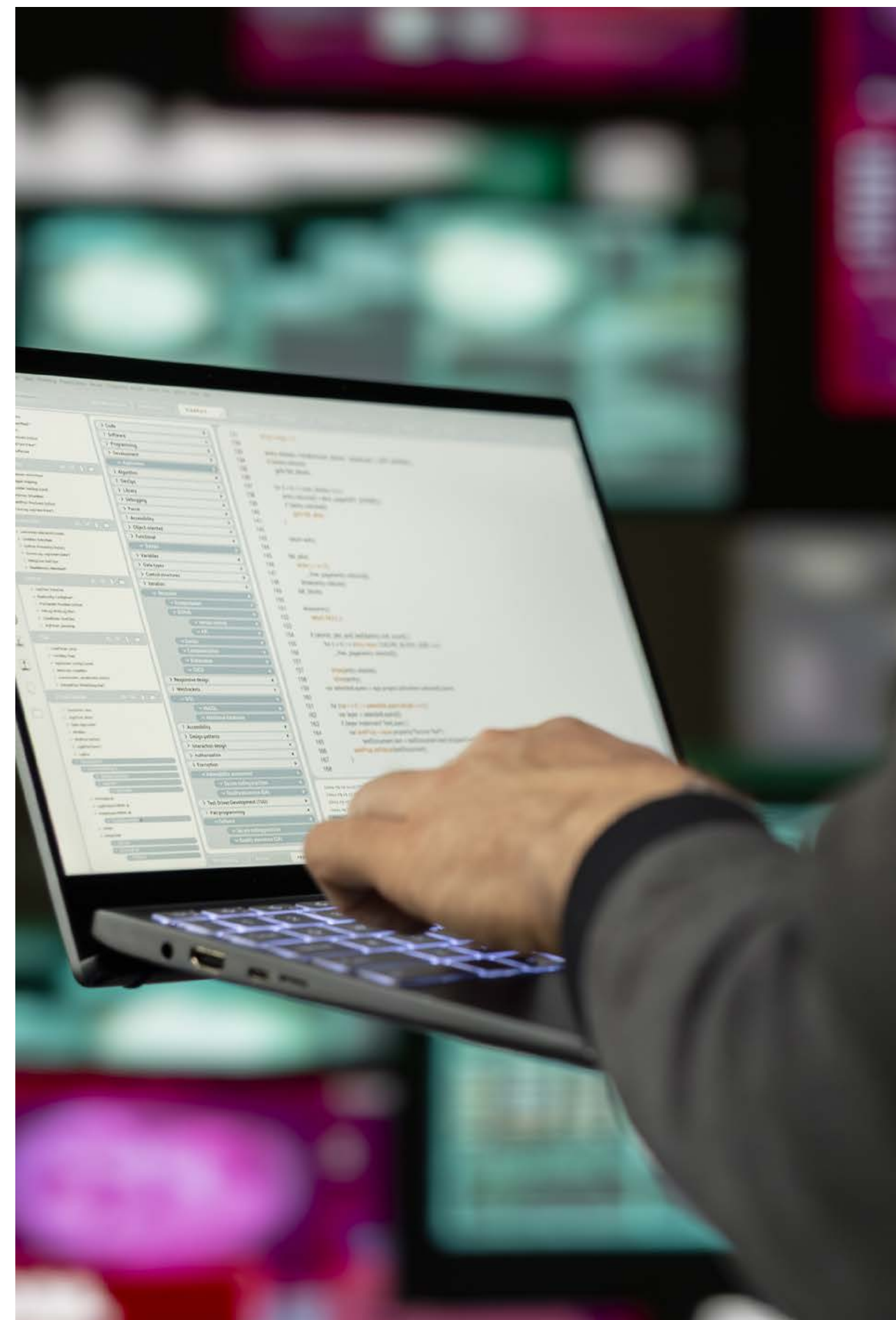
⁶ Sobre la materia, vid. Luis Lafont Nicuesa, "La policía predictiva más allá de Minority Report", *Diario La Ley*, n.º 10687, 2025.

matrícula del vehículo empleado, que indiquen por qué vías y en qué momentos circuló dicho vehículo, dónde estacionó, dónde respostó combustible, o cuándo y dónde fue objeto de una multa; de los datos facilitados por los establecimientos hoteleros y de hospedaje para saber dónde se alojó y pernoctó; de los datos PNR¹¹ remitidos por parte de las compañías aéreas y otras entidades para conocer qué vuelos tomó y con qué itinerarios; de los datos asociados a compras online y operaciones de pago cuando las mismas se hayan realizado a través de medios electrónicos de pago para llegar a determinar lugares habituales de ocio; de los datos almacenados por aquellas empresas dedicadas a ofrecer servicios comerciales basados en la geolocalización de sus usuarios cuando llevan a cabo prácticas deportivas (*Strava, Endomondo, Runtastic, Wikiloc*, etc.) para determinar sus rutinas e incluso su domicilio; y, por supuesto, también podrán acudir a las operadoras de servicios de acceso a la telefonía móvil e internet para que cedan los datos de tráfico y de localización del terminal empleado por el investigado conservados en virtud de las diversas disposiciones legales en materia de cesión de tales datos conservados.

En España, la geolocalización histórica o "reconstrucción de movimientos" de una persona sospechosa de haber participado en unos hechos delictivos concretos resulta, cada vez, más habitual en la operativa policial y en la práctica judicial. El 26 de octubre de 2015, sobre las 07.55 horas, varios encapuchados accedieron armados a una sucursal bancaria de la localidad de Vallbona d'Anoia (Barcelona, España), se apoderaron de más de 11.000 euros y huyeron en un vehículo BMW azul. No pudieron ser identificados por los testigos ni por las cámaras de seguridad del banco, ni tampoco se encontraron huellas dactilares. Sin embargo, las autoridades lograron identificar y condenar a los autores del atraco.

Para ello, una de las primeras actuaciones fue acudir a las operadoras para obtener los datos de geolocalización de los teléfonos móviles que se encontraban operativos en la dirección de la sucursal bancaria y a la hora en que se produjo el atraco. Debidamente analizados dichos datos, se comprobó que varios de sus titulares presentaban antecedentes penales por robo con violencia y se acordó la interceptación judicial de sus comunicaciones, junto con la cesión de los datos de tráfico de llamadas y datos de localización de las mismas en un periodo de tiempo tanto previo como posterior a la comisión de los hechos. Se comprobó que varios de los sospechosos habían estado días antes del atraco en las inmediaciones de la sucursal. Y las filmaciones de las cámaras de la entidad habían grabado a una persona con la misma vestimenta que una de las personas que entraron en el banco el día de los hechos. Esas filmaciones también facilitaron que varios testigos reconocieran a los acusados, y finalmente, en el registro domiciliario de estos se halló vestimenta con las mismas características de la utilizada en el atraco¹².

III. Los metadatos y su utilización en la investigación policial del delito



Anteproyecto de Ley de Enjuiciamiento Criminal de 2020, coordinado por Olga Fuentes Serrano (Valencia: Tirant lo Blanch, 2022), 813-881.

¹¹ Las siglas PNR («Passenger Name Record» Registro de Nombres de Pasajeros) se refieren a determinada información que las empresas y compañías de transporte aéreo, terrestre y marítimo deben remitir a los Estados, para combatir eficazmente el terrorismo y la delincuencia organizada. Y ello, en virtud de los trabajos recogidos en la Resolución 2396 del Consejo de Seguridad de la ONU (2017), que en Europa dio lugar a la Directiva (UE) 2016/681 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativa a la utilización de datos del registro de nombres de los pasajeros (PNR) para la prevención, detección, investigación y enjuiciamiento de los delitos de terrorismo y de la delincuencia grave.

¹² España Tribunal Supremo Sala Penal, Sección 1.ª, "Sentencia núm. 484/2018", de 18 de octubre de 2018. ECLI:ES:TS:2018:3575.

El uso masivo e ininterrumpido de dispositivos electrónicos y de sistemas de la información y la comunicación, no sólo arroja un mayor volumen de información sobre la cual aplicar el *Big Data* y la IA para extraer aquella relevante para la causa penal, también arroja una mejor y más precisa tipología de información sobre la que discernir, escoger y visibilizar aquel específico dato demostrativo de la autoría o de la conducta atribuidas al sujeto investigado. De igual modo que la medicina forense permitió en su momento a las autoridades precisar el veneno utilizado o identificar el ADN de las muestras halladas en el escenario del crimen o sobre el cuerpo de la víctima, la informática forense capacita a las autoridades a concretar las coordenadas espaciales desde donde se efectuó una concreta comunicación, el dispositivo utilizado para ello, o las distintas modificaciones operadas sobre un concreto archivo informático.

Dicho con otras palabras, como quiera que nos hemos convertido en la versión moderna de Hansel y Gretel¹³ y constantemente vamos dejando *migas de pan* en múltiples formatos digitales, a partir de las diversas interacciones que llevamos a cabo (comunicaciones a través de nuestros teléfonos móviles y otros dispositivos, navegación web, uso de redes sociales, compras online, tránsito por lugares videovigilados, etc.), las autoridades policiales tratarán de localizar, recopilar y singularizar aquella concreta modalidad de información digital que interese a los efectos de la investigación: el dato de localización GPS en el momento de los hechos, la marca y modelo del dispositivo utilizado en la elaboración del video, o el *hash* identificativo del archivo ilícito enviado.

Así, en el examen forense de los diversos dispositivos electrónicos que pudieran pertenecer a la víctima, a terceros o al presunto autor de los hechos, en busca de información trascendental para la averiguación de los hechos, la información decisiva a la hora de condenar o no al sujeto investigado no sólo estará compuesta por los “datos” localizados en la memoria de tales dispositivos o sus elementos periféricos (imágenes, videos, registro de llamadas, chats, mails, etc), sino también por los datos de esos datos, sus “metadatos”¹⁴.

Los metadatos pueden ser entendidos, por tanto, como aquellos datos que suministran información sobre el contenido, tipología o formato, origen, soporte, actividad e historial del archivo. También han sido gráficamente descritos como “datos de actividad”: un grupo de etiquetas y marcadores que describen “todos los registros de todas las cosas que haces en tus dispositivos y todas las cosas que tus dispositivos hacen por su cuenta”¹⁵. Cuando enviamos un correo electrónico, los datos de tráfico arrojan información sobre la fecha y la hora de la comunicación y la dirección IP, un análisis más pormenorizado de los metadatos y las cabeceras TCP permiten identificar el servidor utilizado y, en último término, el equipo que conectó con dicho punto de encuentro. El historial de ubicaciones de un teléfono móvil permite conocer el domicilio concreto de un sujeto (ej., dónde pasa las horas de madrugada sin actividad concreta) o su lugar de trabajo. Los metadatos de los archivos enviados por Whatsapp permiten conocer qué dispositivos modificaron tales archivos. Y el historial, duración y potencia de las conexiones a puntos

de acceso Wi-Fi en centros comerciales permite saber cuánto tiempo permanecemos en cada uno de los establecimientos de dichos centros y, con ello, si nos gusta más la tienda de deportes, la tienda de animales, o la zona de restauración.

Por todo ello, los procesos judiciales “se han llenado de conversaciones de chat, audios, fotografías y todo tipo de grabaciones, [...] hoy al alcance de cualquiera a través de algo todavía más suave que un click”¹⁶ y una de las pericias cada vez más habituales en la investigación criminal es el informe forense de los metadatos extraídos de esos audios, fotografías y chats.

1. Metadatos extraídos de los instrumentos de convicción

Según ha quedado demostrado, el análisis forense de la información (documentos, imágenes y videos) localizable en los dispositivos informáticos aprehendidos policialmente se ha convertido en una prueba clave de la comisión de múltiples conductas delictivas.

Durante el proceso de captura y fabricación de un archivo (ej., una imagen), los dispositivos introducen una enorme cantidad de información metadatos en dicha imagen: características técnicas de la imagen, fecha y hora de generación, localización GPS, presencia o ausencia de flash, distancia de los objetos, tiempo de exposición, apertura del obturador, o la marca y modelo dispositivo utilizado, por poner algunos ejemplos¹⁷. El empleo de herramientas específicas de extracción y tratamiento de los metadatos de imágenes y videos (por ej., *ExifTools*, *Gspot* o *MedialInfo*) pueden arrojar muchísima información técnica (tamaño del archivo, versión, fecha de creación y modificación, códecs de audio, número de fotogramas por segundo, sistema operativo utilizado para su captura, idioma, o georreferenciación del lugar donde fue creado) que, tras su debida incorporación al proceso como medio de prueba, permitirá atribuir a un sujeto, sin género de dudas, una determinada conducta.

En otras ocasiones, el análisis se dirige a localizar y analizar toda aquella información identificativa del dispositivo empleado, y en su caso, de su usuario, con el fin de probar su participación en los hechos objeto de investigación. Existen diversas herramientas, calificadas como *fingerprinting*, que sirven para extraer información acerca del dispositivo informático utilizado en la navegación web, con el objetivo de identificarlo y singularizarlo. Las empresas utilizan habitualmente dichas técnicas con el fin de identificar y rastrear a los usuarios que visitan sus páginas web o se conecten a sus servidores, para así poder construir un perfil de los mismos a los que ofrecer sus productos y servicios de una manera más personalizada, pues dado que es común que las personas no compartan sus equipos (menos aún, sus teléfonos móviles), individualizar el dispositivo supone individualizar a la persona que lo utiliza. Cualquiera que haya buscado en Google información sobre un determinado producto o le haya dado “me gusta” en Facebook a una específica noticia, comprenderá ahora por qué, al teclear cualquier dirección URL de una página web, es muy posible que el banner o barra de anuncios de dicha web le arroje información sobre ese producto o sobre productos y servicios relacionados con la noticia anteriormente marcada.

¹³ Juan Carlos Ortiz Pradillo, “Desafíos legales de las diligencias de investigación tecnológica”, en *El proceso penal. Cuestiones fundamentales*, coordinado por Olga Fuentes Soriano (Valencia: Tirant lo Blanch, 2017), 303-316.

¹⁴ Una definición legal de metadato se recogió en España, por primera vez, en el artículo 42 del Real Decreto 1671/2009, de 6 de noviembre, que desarrollaba parcialmente la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los servicios públicos (ya derogado), y en donde se disponía que “Se entiende como metadato, a los efectos de este real decreto, cualquier tipo de información en forma electrónica asociada a los documentos electrónicos, de carácter instrumental e independiente de su contenido, destinada al conocimiento inmediato y automatizable de alguna de sus características, con la finalidad de garantizar la disponibilidad, el acceso, la conservación y la interoperabilidad del propio documento”.

¹⁵ Edward Snowden, *Vigilancia permanente* (Barcelona: Editorial Planeta, 2019), 195.

¹⁶ Jordi Nieva Fenoll, “El tránsito de la fe a la tecnología en el proceso penal”, *Diario La Ley*, n.º 9986, 11 de enero de 2022.

¹⁷ Miguel Esteban Cobo, “Herramienta para la extracción automática de metadatos en videos de dispositivos móviles” (trabajo de fin de grado, Universidad Complutense de Madrid, 2016), <https://eprints.ucm.es/id/eprint/38697/>.

A través de dichas técnicas de *fingerprinting* y, aunque el usuario elimine las cookies o utilice VPN u otros servicios de anonimización, se puede llegar a recoger información muy concreta del dispositivo, como, por ejemplo, el sistema operativo utilizado, el modelo y versión del navegador, resolución de la pantalla, arquitectura de procesador, listas de fuentes de texto, plugins o dispositivos instalados, direcciones IP, el idioma configurado en el sistema, o el huso horario¹⁸. La combinación apropiada de toda esta información permite confeccionar una suerte de *huella digital única* del dispositivo que lo singulariza y, por lo tanto, diferencia de forma unívoca a cada usuario en internet¹⁹.

Con similar objetivo, la pericial forense tratará de localizar, asegurar y extraer toda aquella información que permita determinar qué dispositivo fue utilizado en una determinada comunicación o interacción a través de la red (datos de sensores del dispositivo, puntos de acceso Wi-Fi utilizados, dispositivos bluetooth activados, etc.) y cuál fue su usuario (por ej., datos del usuario insertados en la aplicación utilizada, datos del titular del correo electrónico introducidos en dicha aplicación, etc.).

1.1. Los metadatos como elementos probatorios de la autoría delictiva

El análisis forense de los metadatos puede resultar trascendental en el enjuiciamiento de delitos relacionados con la libertad e indemnidad sexual o con la explotación y corrupción de menores. El informe pericial de los agentes policiales sobre los metadatos extraídos de los archivos localizados en los dispositivos, instrumentos y materiales incautados en el domicilio del investigado o de las comunicaciones efectuadas o recibidas por el equipo informático igualmente aprehendido en dicho domicilio permite determinar, entre otra información, las cámaras fotográficas o videográficas empleadas para la elaboración del material ilícito, el lugar en donde tales archivos fueron creados, los dispositivos electrónicos utilizados para su almacenamiento y envío a través de la red, los *nicks*, perfiles de redes sociales y de correos electrónicos utilizados y las páginas web y foros empleados para su difusión²⁰.

1.2. Los metadatos como elementos probatorios de la inocencia del acusado

Pero lo más trascendental del análisis forense de los datos y metadatos localizables en las fuentes de prueba reside en que, como evidencias científicas, están llamadas a desterrar prejuicios y ser debidamente valorados por encima de otros medios de prueba más *tradicionales*, pero menos científicos, como pueda ser el ancestral valor probatorio de la confesión del reo, precedida del habitual y ritual juramento, o de los testimonios prestados por terceros²¹.

En España, la Audiencia Provincial de Sevilla condenó a Moisés (nombre ficticio) a 6 años de prisión como autor de un delito de lesiones agravadas de los artículos 147.1 y 149.1 del Código Penal, al estimar probado que, entre las 1:30 y las 5:00 horas de la

madrugada del día 15 de diciembre de 2013, estando en las inmediaciones de la puerta de una discoteca, propinó puñetazos y patadas en la cara y cabeza de su víctima. Como pruebas incriminatorias, tuvo en consideración el reconocimiento del acusado efectuado por el lesionado, quien ante la policía reconoció fotográficamente al acusado como su agresor, lo reconoció después durante el sumario en rueda de identificación practicada en sede judicial, y lo reconoció por última vez en el propio acto del plenario, así como la declaración de los amigos que le acompañaban en el momento en que fue objeto de la agresión, quienes sustentaron los mismos reconocimientos sin ninguna duda, y ciertas declaraciones espontáneas del acusado cuando fue detenido.

Aunque el testimonio de la víctima, corroborado por otras declaraciones testificales, es prueba valorable y suficiente para desvirtuar la presunción de inocencia, es preciso recordar que la existencia de otros elementos probatorios contradictorios (declaraciones de descargo efectuadas por el acusado, igualmente corroboradas por otros testigos aportados por la defensa) exigen un especial celo junto con una necesaria motivación por parte del tribunal juzgador a la hora de valorar todo el bagaje probatorio y, especialmente, todo el material probatorio periférico que ambas partes aportan precisamente para facilitar que el juzgador pueda testar la verosimilitud de los diferentes testimonios. Y es, entonces, cuando la *prueba electrónica* puede jugar un papel decisivo a la hora de condenar o absolver al acusado.

En los hechos enjuiciados por la Audiencia de Sevilla, ese otro material probatorio periférico y de naturaleza digital fue el siguiente:

- En primer lugar, una de las cámaras de videovigilancia ubicadas en la calle en la que se encuentra la discoteca donde se produjo la agresión recogió unas imágenes en las que el atacante no identificado regresó, un momento posterior a la agresión, para increpar al lesionado. Dicha imagen fue datada a las 5:37 horas (y, como veremos, la concreción del momento exacto de la agresión resultará determinante).
- En segundo lugar y, en apoyo de la versión de descargo del acusado, quien sostuvo que participó en una cena navideña con unos amigos en otro restaurante, se aportaron fotografías y conversaciones de WhatsApp datadas entre la 1:10 y la 1:54 de la madrugada). Junto con la declaración de los amigos del acusado que afirmaron haberse dirigido a otro pub de la zona aproximadamente hasta las 3:00, se aportó una fotografía tomada por el teléfono de la novia del acusado que, según la certificación notarial, el contenido de los metadatos del archivo fotográfico refleja que la fotografía fue tomada a las 3:58 horas de esa madrugada. Otras fotografías capturadas en el exterior del pub estarían tomadas a las 4:47 horas.
- Y en tercer lugar, compareció en el plenario el taxista que llevó al acusado y a otro acompañante a su localidad de residencia. Dicho testigo relató que iniciaron el trayecto sobre las 5:50 o 6:00 de la mañana y que el mismo duró 30 o 35 minutos aproximadamente, lo cual vino corroborado por el extracto bancario de la tarjeta de crédito con el que el acusado abonó dicho servicio, registrado a las 6:29 de esa madrugada.

Con toda esta información, el Tribunal Supremo estimó el recurso de casación por vulneración del derecho a la presunción de inocencia formulado por Moisés, tras valorar que "la credibilidad que la sentencia de instancia ha conferido al reconocimiento que los testigos de cargo hicieron del acusado como el autor de los hechos, se enfrenta a un

¹⁸ Vicente Aguilera Díaz y Carlos Seisdedos, *Open Source INTelligence (OSINT): Investigar personas e identidades en internet* (Madrid: OxWord, 2020), 49.

¹⁹ Véase el documento *Fingerprinting o Huella digital del dispositivo*, publicado por la AEPD en 2019. Disponible en: <https://www.aepd.es/sites/default/files/2019-09/estudio-fingerprinting-huella-digital.pdf>.

²⁰ Una explicación detallada de los metadatos encontrados en una operación contra la explotación sexual infantil y la corrupción de menores se aprecia en España Tribunal Supremo Sala Penal, "Sentencia núm. 395/2021", de 6 de mayo de 2021. ECLI:ES:TS:2021:1737.

²¹ En el mismo sentido, vid. Jordi Nieva Fenoll, "El tránsito de la fe a la tecnología en el proceso penal", *Diario La Ley*, n.º 9986, 11 de enero de 2022.

conjunto de testimonios y *vestigios objetivos que lo desmienten*"²².

2. Metadatos obtenidos de fuentes abiertas

2.1. Internet y las redes sociales como "lugar de comisión del delito" y como lugar de recolección de los vestigios y las fuentes de prueba

La reformulación de la noción del "escenario del crimen" propiciada por el avance tecnológico tiene, entre otras consecuencias, su traslación a la noción del "lugar del delito", que no sólo debe referirse a un espacio físico y geográfico. Tal y como ha descrito gráficamente el Tribunal Supremos español:

El delito en su forma más convencional convive ahora con nuevas formas de ciberdelincuencia en las que su ejecución se desarrolla enteramente en redes telemáticas que, por definición, no son inmovilizables en un espacio físico perfectamente definible. El ciberespacio ofrece un marco digital diferenciado de la realidad puramente física como espacio del delito. La experiencia más reciente enseña que las redes sociales no son sólo el instrumento para la comisión de algunos delitos de muy distinta naturaleza. Pueden ser también el escenario en el que el delito se comete, ya sea durante todo su desarrollo, ya en la ejecución de sólo algunos de los elementos del tipo.²³

Por ello, las autoridades policiales han trasladado su campo de operaciones a la Red a través de lo comúnmente calificado como "ciberpatrullaje": un conjunto de técnicas que atienden a las finalidades de detectar actividad ilegal en la red y descubrir a los delincuentes, así como de prevenir la perpetración de delitos²⁴.

Un ejemplo concreto puede apreciarse en la investigación policial de los videos subidos a Youtube en los que se difunde una concreta conducta delictiva. Así, cuando se sube y difunde a través de las redes sociales un video ilustrativo de la comisión de un posible delito contra la seguridad vial, el GIAT Grupo de Investigación y Análisis de la Agrupación de Tráfico de la Guardia Civil analiza la información documentada en el video de cara a la posible identificación de la vía o tramo de carretera donde se han producido los hechos, y a partir de ahí, tratar de identificar a sus autores (identificación de los vehículos y de los rostros que aparecen en el video) mediante su cotejo con los datos en poder de la Dirección General de Tráfico (ej., los datos de los titulares de los vehículos o las imágenes de las cámaras de tráfico que, en otros puntos kilométricos, si identifiquen las matrículas de los vehículos que aparecen en el video). Pero, también, se sirven de la información que aporta el propio hecho de subir el vídeo a la red social; los metadatos extraíbles del vídeo y del perfil del usuario que lo colgó o difundió.

2.2. Técnicas e instrumentos de recolección de los metadatos de la información transmitida a través de Internet y las Redes Sociales

Como acabamos de comprobar, la recopilación y análisis de esa información (y los metadatos asociados a la misma) extraída de internet se lleva a cabo a través del empleo de las herramientas de inteligencia sobre fuentes abiertas o de acceso público (OSINT,

por sus siglas en inglés *Open Source Intelligence*), lo cual amplía exponencialmente las capacidades policiales de localizar información relevante para la prevención y persecución de actividades delictivas, así como para identificar y georreferenciar a un determinado sujeto en función de su actividad en Internet.

Más allá de atribuir a un usuario específico el empleo de una concreta dirección IP, también resulta conveniente tomar conocimiento de otros metadatos producidos con motivo de dicha comunicación web, como pueda ser la geolocalización o los datos del terminal empleado por un concreto perfil o del nick utilizado en las redes sociales.

Por señalar un ejemplo concreto, y con el objetivo de identificar y localizar desde dónde *tuitea* un sujeto que puede estar cometiendo a través de Twitter un posible delito de incitación al odio, discriminación o violencia contra una persona conforme al art. 510 CP, existen herramientas como *Cree.py* para geolocalizar a los usuarios de servicios web, como Twitter y Flickr, a partir de los metadatos de los *tuits* o fotos publicadas (información GPS del dispositivo, tuits enviados desde una determinada ubicación, triangulación basada en la IP desde la que se hizo el tuit, etc). De igual modo, y en aras a fundamentar una condena penal por delitos de captación y adoctrinamiento terrorista y de enaltecimiento del terrorismo, en ocasiones, resulta esencial probar la autoría de un determinado perfil desde el cual se escriben posts y se da "me gusta" a vídeos e imágenes de extremismo yihadista²⁵.

Junto con dichos metadatos, también resultará relevante el análisis inteligente de los contenidos transmitidos en abierto, como puedan ser las palabras utilizadas, las imágenes o videos intercambiados, o las personas mencionadas en dichas interacciones cibernéticas. Para ello, existe una categoría especial de herramientas y técnicas ideadas para la extracción de información de las principales redes sociales denominadas *Social Media Intelligence* (SOCMINT) o inteligencia basada en las redes sociales²⁶. Tales herramientas aprovechan la información generada por el *Big Data* para localizar aquella información generada o transmitida a través de la Red que pueda resultar útil de cara a las legítimas funciones de las autoridades estatales.

Al igual que las empresas son las primeras en utilizar dichas técnicas con propósitos comerciales como, por ejemplo, la detección de tendencias o el análisis y mejora de su reputación online, a partir del examen y la categorización del perfil de sus seguidores (edad, sexo, ubicación geográfica,...), los gobiernos, los servicios secretos y los investigadores policiales también monitorizan las redes sociales para, bien prevenir la comisión de posibles ilícitos, o bien obtener una valiosa información sobre hechos e individuos que sean objeto de una concreta investigación judicial, pues un examen profesional del perfil de un usuario, sus interacciones y los metadatos generados con aquéllas (contactos, términos empleados, ficheros intercambiados, geolocalización de las interacciones o imágenes difundidas, información del equipo empleado, identificación del correo electrónico asociado a la cuenta, etc.) pueden facilitar enormemente la labor policial de proceder a su identificación, su localización o la determinación de su pertenencia a un determinado grupo criminal. Para ello, la determinación de la dirección IP utilizada en una concreta interacción a través de internet suele ser el punto de partida de muchas investigaciones. En este sentido, el Tribunal Supremo ha reiterado que la obtención

²² España Tribunal Supremo Sala Penal, "Sentencia núm. 120/2019", de 6 de marzo de 2019. ECLI:ES:TS:2019:737.

²³ España Tribunal Supremo Sala Penal, Pleno, "Sentencia núm. 547/2022", de 2 de junio de 2022. ECLI:ES:TS:2022:2356.

²⁴ Más información en Pilar Martín Ríos, "Empleo de big data y de inteligencia artificial en el ciberpatrullaje: de la tiranía del algoritmo y otras zonas oscuras", *IDP. Revista de Internet, Derecho y Política*, n.º 36 (2022).

²⁵ Un extenso examen de la prueba consistente en los "me gusta" de un específico perfil de Facebook puede leerse en España Audiencia Nacional Sala de lo Penal, Sección 4.ª, "Sentencia núm. 16/2021", de 23 de septiembre de 2021. ECLI:ES:AN:2021:4196.

²⁶ Un estudio particularmente centrado en dichas técnicas es la monografía de Federico Bueno de Mata, *Investigación y prueba de delitos de odio en redes sociales: técnicas OSINT e inteligencia policial* (Valencia: Tirant lo Blanch, 2023).

policial de una dirección IP con motivo del ciberpatrullaje en la red es “información de conocimiento público para cualquier usuario de Internet y que el propio usuario de la red es quien lo ha introducido en la misma”²⁷.

La geolocalización de una dirección IP resulta relativamente sencilla. La Corporación de Internet para la Asignación de Nombres y Números (ICANN) es la entidad que gestiona las direcciones IP a nivel mundial y las reparte a los Registradores Regionales de Internet (RIR). El Registrador Regional que gestiona las direcciones en Europa es el RIPE (Network Coordination Centre – RIPE NCC), quien, a su vez, procede a una posterior subdelegación de recursos a sus clientes, los Registros de Internet Local (Local Internet Registries, LIRs²⁸), quienes distribuyen el direccionamiento IP en cada una de esas regiones a determinadas operadoras (ej., Telefónica), y éstas finalmente son las que las ceden a los proveedores de acceso. Por lo tanto, y en función del rango de dirección IP, es fácil determinar inicialmente si una determinada dirección ha sido asignada a la operadora de un determinado país y presumir que el usuario de dicha IP se halla dicho país, y dentro del mismo, qué nodo utiliza para acceder a Internet.

A partir de ahí, cuando las fuerzas y cuerpos de seguridad de cualquier país tienen conocimiento de que una específica dirección IP ha sido empleada para cometer una específica conducta delictiva (v. gr., enviar o recibir material de explotación sexual infantil), comunican dicha información a las autoridades policiales del país al cual se tiene asignada dicha dirección IP, quienes se ocuparán en el caso español de solicitar el oportuno mandamiento judicial para recabar de las operadoras los datos del titular de la conexión como puedan ser los datos que permitan la identificación y localización del terminal o del dispositivo de conectividad y la identificación del sospechoso.

2.3. Colaboración público-privada en el rastreo de metadatos de información ilícita transmitida a través de internet y las redes sociales

Un ejemplo concreto de colaboración entre autoridades estatales, el sector privado y organizaciones no gubernamentales, a la hora de detectar posibles conductas delictivas en internet, es el referido a la lucha contra los abusos sexuales y la explotación sexual de los menores, incluida la pornografía infantil.

La necesidad de adopción de medidas eficaces contra dichas actividades se recoge en múltiples textos legislativos, tales como la Convención de las Naciones Unidas sobre los Derechos del Niño de 1989, el Protocolo Facultativo de la Convención de las Naciones Unidas sobre los Derechos del Niño de 2000 relativo a la venta de menores, la prostitución infantil y la utilización de los menores en la pornografía, el Convenio del Consejo de Europa de 2007 sobre la protección de los menores contra los abusos sexuales y la explotación sexual, o la Directiva 2011/93/UE del Parlamento Europeo y del Consejo, de 13 de diciembre de 2011, relativa a la lucha contra los abusos sexuales y la explotación sexual de los menores y la pornografía infantil y por la que se sustituye la Decisión Marco 2004/68/JAI del Consejo. Así, el rastreo de metadatos de las comunicaciones realizadas a través de internet se ha convertido en una de las grandes iniciativas que ha

logrado aunar los esfuerzos del sector público y privado para combatir eficazmente la elaboración y la difusión de imágenes y vídeos de explotación sexual infantil a través de Internet.

El Centro Nacional para Menores Desaparecidos y Explotados (NCMEC - *National Center for Missing & Exploited Children*), fundado en los EE.UU. en 1984, lleva años alertando a las autoridades de que determinados usuarios han compartido fotos o vídeos con posible contenido pedófilo a través de las redes sociales, para lo cual se sirve del empleo de técnicas y herramientas OSINT y SOCMINT para el rastreo, identificación y localización de dichos archivos tales como *PhotoDNA* o *NeuralHash*.

PhotoDNA es una tecnología de identificación de imágenes desarrollada por Microsoft, en colaboración con el Dartmouth College en 2009, a partir de la creación de códigos hash únicos que sirven para prepresentar esa imagen. Este hash se calcula de tal manera que es resistente a las posibles alteraciones que se pueden realizar sobre dicha imagen (ej., el cambio de tamaño y formato del archivo). Posteriormente, dicho hash se compara con las bases de datos que las organizaciones y empresas participantes en esta iniciativa ya han identificado como ilícitas por su contenido de explotación sexual infantil (como el NCMEC o los integrantes del Proyecto VIC) y permite a las empresas prestadoras de servicios en internet denunciar a los usuarios que han difundido dichos archivos a través de sus plataformas.

Se utiliza en los propios servicios de Microsoft, incluidos Bing y OneDrive, pero también en otros prestadores de servicios tales como Facebook, Twitter, Adobe, así como en Gmail de Google, mientras que Apple ha decidido desarrollar una herramienta similar, llamada *NeuralHash*, para escanear las cuentas iCloud de sus usuarios en busca y captura de contenido pedófilo. De igual modo, Youtube también cuenta con sistemas automatizados de detección de vídeos con imágenes de explotación sexual infantil o de extremismo terrorista, gracias a la tecnología *fingerprinting* y a la detección y contraste de los *hashes* de los archivos subidos a dicha red. Desde junio de 2017, implementó tecnologías de aprendizaje automático para la detección de contenido extremista violento y de discurso de odio, basadas en la comparativa de los datos de vídeos ya revisados por decisiones humanas y eliminados de sus servidores²⁹.

En Europa, y con ese mismo espíritu de servirse del escaneo automatizado de los metadatos para evitar la difusión de material contenedor de imágenes de explotación sexual infantil, se aprobó en julio de 2021 el Reglamento (UE) 2021/1232 del Parlamento Europeo y del Consejo de 14 de julio de 2021 por el que se establece una excepción temporal a determinadas disposiciones de la Directiva 2002/58/CE en lo que respecta al uso de tecnologías por proveedores de servicios de comunicaciones interpersonales independientes de la numeración para el tratamiento de datos personales y de otro tipo con fines de lucha contra los abusos sexuales de menores en línea³⁰. Dicho Reglamento establece una excepción temporal de tres años (hasta el 3 de agosto de 2024) a la protección dispensada por los artículos 5, apartado 1, y al artículo 6, apartado 1, de la Directiva 2002/58/CE, en materia de confidencialidad de las comunicaciones y los datos de tráfico, así como un marco especial en materia de los derechos de acceso, rectificación, etc., de los datos personales en virtud del Reglamento (UE) 2016/679, para que los proveedores de servicios en la Sociedad de la Información puedan utilizar,

²⁷ España Tribunal Supremo Sala Penal, “Sentencia núm. 1058/2006”, de 2 de noviembre de 2006; “Sentencia núm. 1299/2011”, de 17 de noviembre de 2011; “Sentencia núm. 342/2013”, de 17 de abril de 2013 (ECLI:ES:TS:2013:2222); “Sentencia núm. 1025/2013”, de 26 de diciembre de 2013; y “Sentencia núm. 16/2014”, de 30 de enero de 2014.

²⁸ El listado de esos registros de internet local que ofrecen servicios en España puede ser consultado en la página web: <https://www.ripe.net/membership/indices/ES.html>.

²⁹ Más información en: <https://protectingchildren.google/#fighting-abuse-on-our-own-platform-and-services>.

³⁰ Diario Oficial de la Unión Europea, L 274, 30 de julio de 2021.

con respaldo legal, tecnologías específicas de contraste de *hashes* para imágenes y videos, y herramientas de IA para el análisis de datos de tráfico, con el único fin de detectar y retirar el material de abuso sexual de menores en línea, o el "embaucamiento de menores" con fines sexuales a través de medios tecnológicos, y denunciarlo a las autoridades policiales y a las organizaciones que actúen en interés público contra los abusos sexuales de menores.

El artículo 3 impone ahora a tales proveedores el deber de conservar de manera segura los datos de contenido y los datos de tráfico conexos, así como los datos personales generados mediante dicho tratamiento, cuando se haya detectado un archivo de un presunto abuso sexual a menores en línea, crear una firma digital única y no reconvertible (*hash*) de datos identificados de forma fiable como material de abuso sexual de menores en línea, bloquear la cuenta del usuario de que se trata, o de suspender o poner fin a la prestación del servicio a dicho usuario y denunciar sin demora el presunto abuso sexual de menores en línea a las autoridades policiales y judiciales competentes o a organizaciones que actúen en interés público contra los abusos sexuales de menores.

No obstante, también les exige que articulen mecanismos para permitir al usuario de que se trate recurrir ante el proveedor o interponer recursos administrativos o judiciales en asuntos relacionados con el presunto abuso sexual de menores en línea, y remitir a la Comisión Europea y a la Autoridad de control del Reglamento (UE) 2016/679 informes anuales³¹ sobre el número de casos detectados de abusos sexuales de menores en línea, diferenciando entre material de abuso sexual de menores en línea y embaucamiento de menores, el número de casos en los que un usuario haya presentado una reclamación ante el mecanismo de recurso interno o interpuesto una acción judicial y el resultado de dichas reclamaciones y causas judiciales, el número y las tasas de errores (falsos positivos) de las diferentes tecnologías utilizadas, así como las medidas aplicadas para limitar la tasa de error y la tasa de error alcanzada, la política de conservación de los datos y las salvaguardias de protección de datos aplicadas conforme al Reglamento (UE) 2016/679, y el nombre de las organizaciones que actúan en interés público contra los abusos sexuales de menores con las que se hayan compartido datos en virtud del presente Reglamento.

2.4. Rastreo de metadatos de la información transmitida a través de internet y las redes sociales y derecho fundamental al secreto de la correspondencia

La conceptualización de los metadatos como "toda aquella información descriptiva sobre el contexto, calidad, condición o características de un recurso, dato u objeto que tiene la finalidad de facilitar su recuperación, autenticación, evaluación, preservación y/o interoperabilidad" conduciría a considerarlos igualmente como auténticos contenidos de comunicaciones, por mucho que en no pocas ocasiones se encuentren alojados en

cabeceras IP o DNS³². El escrutinio de los metadatos de los contenidos compartidos o transmitidos por estas vías afecta al Derecho Fundamental a la privacidad del artículo 7 de la Carta de Derechos Fundamentales de la Unión Europea, y el consiguiente tratamiento (conservación y notificación a las autoridades) de la información obtenida y su posterior contraste con bases de datos de hashes sospechosos de contener pornografía infantil afecta igualmente al derecho a la protección de datos personales del art. 8³³.

Quizás por ello, el citado Reglamento (UE) 2021/1232 indica que esas actividades de escaneo de metadatos que llevan a cabo los proveedores para detectar abusos sexuales de menores en línea cometidos en sus servicios y denunciarlos "representan una injerencia en los derechos fundamentales al respeto de la vida privada y familiar y a la protección de los datos personales de todos los usuarios de servicios de comunicaciones interpersonales independientes de la numeración", y de ahí la necesidad de esta expresa disposición legal como excepción a la norma prohibitiva de los arts. 5.1, 6.1 y 15.1 de la Directiva 2002/58/CE y al régimen general del RGPD.

Pero también cabe la posibilidad de defender jurídicamente que dichas técnicas de escaneo de *hashes* y su comparativa automatizada con bases de datos de material ya declarado ilícito no dejan de ser sino una modalidad moderna o traslación al entorno digital de lo que siempre fueron "técnicas de inspección por motivos de seguridad pública u orden público" con el fin de detectar la presencia de productos prohibidos, que legalmente se consideraron siempre una excepción a la inviolabilidad de los envíos postales. El artículo 6.4 de la Ley 43/2010, de 30 de diciembre, del servicio postal universal, de los derechos de los usuarios y del mercado postal (inviolabilidad de los envíos postales) ya preveía "el ejercicio de las facultades de control reconocidas legalmente a determinados funcionarios en el marco del ejercicio de sus funciones de inspección, como las sanitarias, aduaneras, de prevención de blanqueo de dinero o de seguridad o cualesquiera otras establecidas en la normativa sectorial, con el fin de detectar la presencia de productos prohibidos", y el artículo 12.3 del Real Decreto 1829/1999, de 3 de diciembre, por el que se aprueba el Reglamento por el que se regula la prestación de los servicios postales, en desarrollo de lo establecido en la Ley 24/1998, de 13 de julio, del Servicio Postal Universal y de Liberalización de los Servicios Postales, también advierte que:

Ni los funcionarios de la Secretaría General de Comunicaciones del Ministerio de Fomento ni los de los servicios aduaneros podrán conocer el contenido de los envíos postales, debiendo respetar el derecho al secreto, la intimidad y la inviolabilidad de la correspondencia. Se exceptúan los envíos que no contengan documentos de carácter actual y personal como la publicidad directa, los libros, los catálogos, las publicaciones periódicas, así como los paquetes postales en los que proceda su inspección por motivos de seguridad pública u orden público y no se haya puesto de manifiesto expresamente que contienen objetos de carácter personal.³⁴

³² Para José Luis Rodríguez Láinz, "En torno al concepto de comunicación protegida por el art. 18.3 de la Constitución", *Diario La Ley*, n.º 8143, 5 de septiembre de 2013, "Si la cabecera incluye datos sobre titularidad o dominio, datación, volumen y naturaleza de la información, e identidades de destino y origen, evidentemente entrarán en el campo de los llamados datos de tráfico. Pero si tienen una finalidad ajena al tránsito mismo de la comunicación, cual sucede no solo respecto de metadatos que atienden a su finalidad etimológica de facilitar búsquedas de recursos, sino también aquella información añadida que nada tiene que ver con el contenido o su tráfico (v.g. información sobre ubicación geográfica del emisor asociada, siempre opcionalmente, a twitts, fotografías enviadas a través de determinadas aplicaciones tipo Apple o Android, o determinados servicios de valor añadido), deberán ser tenidos a todos los efectos como contenidos; puedan ser o no accesibles a los efectos de prestación del servicio correspondiente por las operadoras de telecomunicaciones".

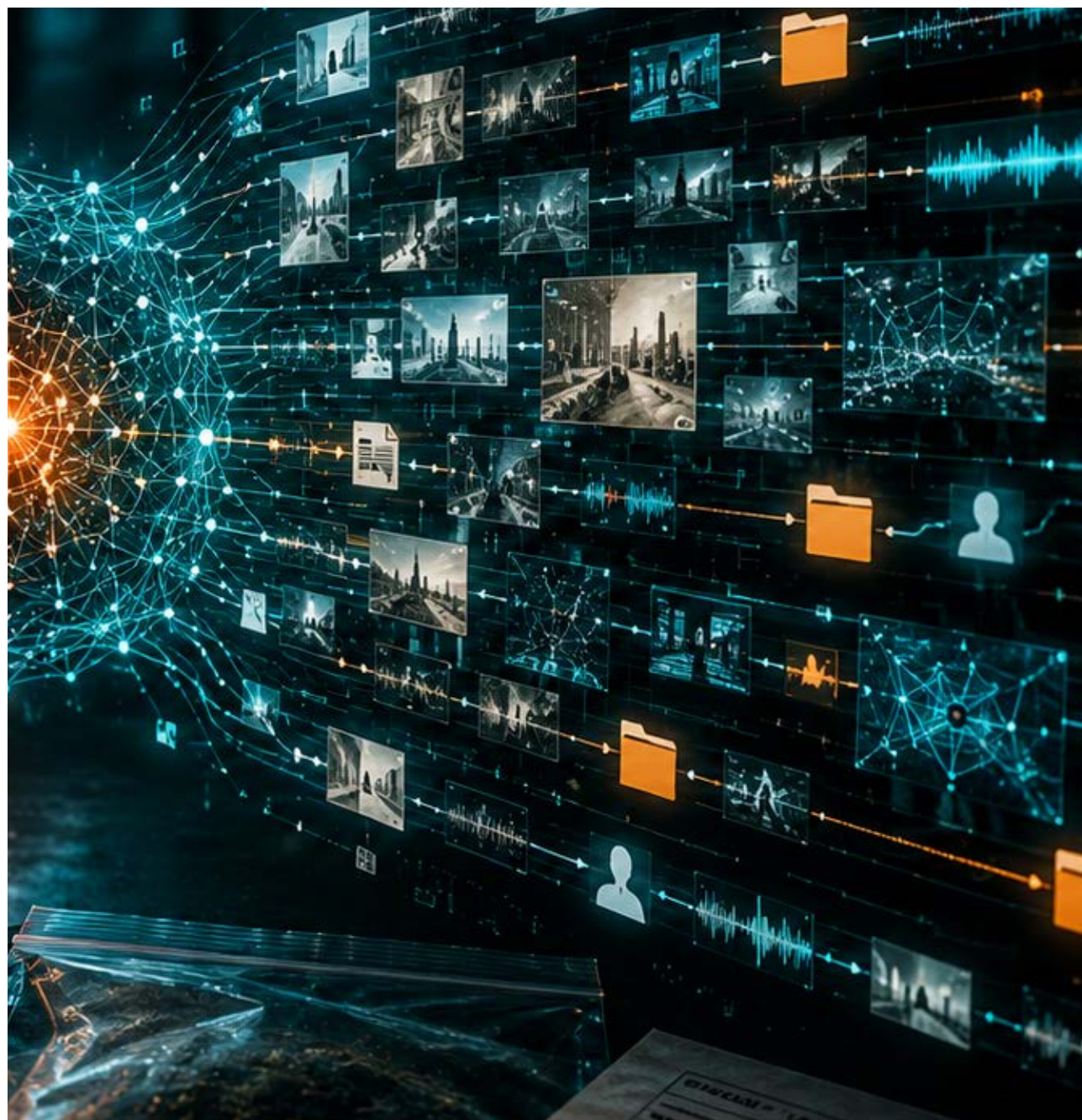
³³ En el mismo sentido, José Luis Rodríguez Láinz, "Reflexiones sobre el tratamiento de datos personales por prestadores de servicios de comunicaciones vía internet para la lucha contra abusos sexuales de menores en línea en el Reglamento (UE) 2021/1232", *Diario La Ley*, n.º 9974, 20 de diciembre de 2021.

³⁴ España, Real Decreto 1829/1999, de 3 de diciembre, por el que se aprueba el Reglamento por el que se regula la prestación de los servicios postales, en desarrollo de lo establecido en la Ley 24/1998, de 13 de julio, del Servicio Postal Universal y de Liberalización de los Servicios Postales, *Boletín Oficial del Estado*, núm. 313, 31 de diciembre de 1999, art. 12.3.

³¹ A falta de información sobre los datos del año 2022, contamos con datos para el período comprendido entre la entrada en vigor del Reglamento, el 3 de agosto de 2021, y el 31 de diciembre de 2021. El Informe de Google para Europa (Google Ireland Transparency Report Under Regulation (EU) 2021/1232, accesible en: <https://transparencyreport.google.com/report-downloads?hl=en>) indica que detectó automáticamente 331 cuentas de usuarios de la UE con material referido a abuso sexual infantil. De ellos, en 8 casos el usuario apeló internamente la decisión de Google de cancelar su cuenta y en ninguno de esos 8 casos se modificó la decisión inicial de cancelación. A nivel mundial (datos extraídos de <https://transparencyreport.google.com/child-sexual-abuse-material/reporting>), Google informa que detectaron un total de 3.282.824 de archivos con material de abuso sexual infantil en sus plataformas; inhabilitó 140.868 cuentas de usuarios y elaboraron un total de 1.620.122 hashes que añadieron a sus repositorios internos, además de compartirlos con el NCMEC para que puedan consultarlos otros proveedores. Twitter, por su parte (Twitter Report Under Article 3(1)(g)(vii) of Regulation (EU) 2021/1232, accesible en: https://blog.twitter.com/en_us/topics/company/2022/twitter-s-eu-submission-for-2021-1232) informó que había suspendido un total de 532.898 cuentas por violar su política de lucha contra la explotación sexual infantil.

De igual modo que la inviolabilidad de la correspondencia no imposibilita que los envíos postales puedan escanearse y someterse a técnicas de detección de materiales tóxicos, explosivos y armas en su interior, también cabe sostener que el avance tecnológico facilita la utilización de herramientas para, de manera automatizada, contrastar factores o elementos (metadatos) indicativos de la posible existencia de un material expresamente prohibido por la legislación, así como por las específicas políticas internas de condiciones y términos de las empresas que prestan tales servicios.

IV. Los sistemas de inteligencia artificial como analistas forenses



Sin duda alguna, uno de los campos operativos donde la Inteligencia Artificial puede tener una mayor repercusión en la investigación criminal lo será en el análisis forense de datos. La investigación criminal siempre se ha fundamentado en la recolección de datos y en su análisis. Y es aquí donde los avances propios de la era digital facilitan que los datos

recolectados sean analizados de un modo rápido y en profundidad, descomponiendo hasta el último *byte* de información y arrojando unos resultados precisos.

Lo que sucede es que, en la actualidad, los expedientes de las investigaciones penales están compuestos de una gran heterogeneidad de datos: declaraciones de perjudicados, testigos e investigados, datos bancarios, expedientes tributarios, facturas, actos mercantiles, conversaciones y comunicaciones electrónicas de los investigados, las cuales suelen contener archivos digitales de vídeo, audio, texto, etc., objetos y documentos incautados en registros domiciliarios, y su análisis requiere, en la mayoría de los casos, un dilatado tiempo para efectuar el correspondiente informe pericial.

El Tribunal Supremo español, precisamente a la hora de justificar la no devolución a la defensa del material y dispositivos electrónicos incautados con ocasión de un registro domiciliario para que pudiera efectuar una contrapericia, ha advertido expresamente que cuando se trata de abundantísima información en formato electrónico (en ese caso concreto, más de diez terabytes):

Esta digresión pretende explicar el motivo por el cual la mencionada información, que ha de ser tratada además con las precauciones debidas para evitar su pérdida o destrucción, determina que su contenido se vaya revelando a los investigadores de manera secuenciada o progresiva. [...] la ingente cantidad de soportes de almacenamiento masivo y la extraordinaria magnitud de la presente causa especial y pieza separada (con una información que superaría los 10 terabytes), que, por motivos obvios, no habría podido ser analizado en su totalidad, ni permitido obtener copias que garantizaran la integridad y autenticidad de los datos, al efecto de devolver los soportes originales. Además de que, como igualmente se indica, el adecuado análisis de la información contenida en tales soportes no se agota con el examen individualizado de los mismos, siendo preciso cruzar dichos datos con los procedentes de otros dispositivos intervenidos (no ya sólo en la presente causa especial, sino también en la pieza separada n.º 20775/2020-II), y de la documentación obrante en el procedimiento, con un volumen también considerable.³⁵

Las soluciones de IA en este campo se enfocarían a automatizar el análisis y la correlación de los datos adquiridos durante la investigación y, con base en su aprendizaje, presentar al investigador la información que es de su interés. Es decir, reducir la cantidad de datos que hay que analizar personalmente por los investigadores, y ofrecer el subconjunto del total de la evidencia que es más probable que constituya la información de interés para la investigación.

Para ello, cobra especial interés las herramientas basadas en la visión computacional (*computer vision*) capaces de analizar y clasificar ficheros de imagen y vídeo de forma autónoma. Según se ha explicado, y en materia de lucha contra la elaboración y difusión de imágenes de abuso sexual infantil, el Reglamento "provisional" UE 2021/1232 ya habilita a que los proveedores de servicios en la Sociedad de la Información puedan -posibilidad- utilizar, con respaldo legal, tecnologías específicas de contraste de hashes para imágenes y vídeos, y herramientas de IA para el análisis de datos de tráfico, con el único fin de detectar y retirar el material de abuso sexual de menores en línea. Y la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas para prevenir y combatir el abuso sexual de los menores³⁶ busca convertir en esa posibilidad en la obligación de los prestadores de servicios de alojamiento de datos y de los prestadores de servicios de comunicaciones interpersonales de detectar y

³⁵ España Tribunal Supremo Sala Penal, "Auto", en Causa especial n.º 20775/2020, de 3 de marzo de 2026.

³⁶ Comisión Europea, *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas para prevenir y combatir el abuso sexual de los menores*, COM(2022) 209 final, 2022/0155 (COD), 11 de mayo de 2022.

denunciar los abusos sexuales de menores en línea a través de las correspondientes herramientas técnicas.

Las herramientas basadas en la visión computacional pueden suponer un salto cualitativo en las capacidades forenses, al poder examinar, etiquetar y comparar millones de imágenes o documentos físicos. Así, en el campo de la lucha contra la delincuencia medioambiental transfronteriza, el uso de herramientas de inteligencia geoespacial, teledetección, escaneo, monitorización online, visión artificial, análisis multimodal, correlación y tecnologías de análisis predictivo, es ya una realidad para poder identificar patrones visuales específicos (construcciones ilegales, vertidos medioambientales, extracción ilegal de acuíferos, etc.), tal y como demuestra el Proyecto PERIVALLON (Protecting the EuRopean territory from organised enVironmentAl crime through inteLLigent threat detectiON tools) y el Servicio de Vigilancia Terrestre de los satélites Copernicus Sentinel para facilitar la detección e investigación de estos delitos mediante la recopilación y análisis de imágenes satelitales.

Y qué decir del reconocimiento biométrico en virtud del rostro, la voz, etc., gracias a esa visión computacional y al entrenamiento de la IA basado en *el deep learning*. Todo ello constituye un salto cualitativo en el cotejo forense de evidencias, por ejemplo, en el contexto de las mezclas complejas de ADN tomadas en la escena del crimen; en el ámbito de la lofoscopia, al hacer innecesaria la intervención humana para la identificación de características de cualquier muestra de huellas dactilares y su clasificación automática; en el ámbito de la balística forense, al detectar coincidencias en las marcas dejadas por las armas en los cartuchos o proyectiles; o en la grafística forense, para ayudar a determinar el autor de un texto manuscrito.

Un paso más, si cabe, lo constituye el uso de IA capaz de reconocer patrones complejos, aprender de errores y superar la sintaxis e identificar la semántica de las comunicaciones, lo cual está llamado a convertirse en un avance de un impacto descomunal en la labor de los analistas³⁷. No se trata únicamente de la transcripción de voz a texto, traducción en distintos idiomas incluida, sino que la IA agéntica puede superar la limitación clásica de la prueba electrónica, consistente hasta ahora en la búsqueda por palabras clave o keywords, y llegar a interpretar contextos; analizar textos identificando entidades, personas, organizaciones, temáticas, de modo que el software deje de buscar coincidencias ortográficas –*keywords*– para interpretar contextos semánticos gracias a técnicas avanzadas de Procesamiento de Lenguaje Natural (PLN), de tal manera que el algoritmo pueda ser capaz de identificar palabras, frases o expresiones denotativas de una conducta criminal.

Como sabemos, el uso de un lenguaje velado o convenido es reconocido jurisprudencialmente como un elemento de sospecha que, interpretado racionalmente, esos modismos, palabras confusas y simbólicas bien conocidas en el ambiente policial, permite deducir otro contenido (SSTS 1140/2009, de 23 de octubre; 849/2013 de 12 de noviembre; 1013/2022, de 12 de enero de 2023), por lo que el empleo de la IA facilita el reconocimiento del lenguaje en clave, codificado o cifrado, utilizado en comunicaciones orales o escritas, que la jurisprudencia ya ha manifestado que puede constituir un indicio de criminalidad que permite inferir la existencia de hechos delictivos ocultos, como actividades de delincuencia organizada, tráfico de drogas o planificación de actos ilícitos.

Si, además, es capaz de interrelacionar conversaciones con fechas, vínculos sociales,

registros de llamadas, transferencias, viajes, geolocalización, etc., el pronóstico puede resultar muy útil en manos de los expertos.

Este análisis forense agéntico convierte la inspección ocular humana, limitada y selectiva, en una inspección, total y sistemática, capaz de auditar el círculo social, familiar y profesional del investigado, realizar procesamientos avanzados como identificar personas y entidades a partir de sus múltiples identificadores (números de teléfono, direcciones de correo, cuentas de redes sociales, datos de identificación fiscal, documentos, ficheros de imagen y vídeo, y reconstruir el mapa social de un individuo.

Pensemos, por un momento, en la habilitación expresa que hace el Reglamento IA de la UE, en el apartado 6 de su Anexo III, a la elaboración de perfiles de personas sometidas a una investigación penal, pero no sobre sus rasgos y características de su personalidad (raza, sexo, código postal, nivel de estudios) –que ya se ha demostrado que pueden constituir sesgos discriminatorios–, sino sobre su actividad y para apoyar la valoración humana de la implicación de una persona en una actividad delictiva que se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva.

Con estas premisas, resulta factible concluir que el empleo de estos sistemas de IA agéntica aplicados al análisis policial forense pueden dar lugar, en mi opinión, a 3 importantes consecuencias prácticas en el proceso penal:

En primer lugar, a buen seguro se producirá un incremento de los casos de "hallazgos casuales" detectados por la IA, al ser capaz de ver aquello que los investigadores humanos no vieron. En palabras del Tribunal Constitucional español, "La Constitución no exige, en modo alguno, que el funcionario que se encuentra investigando unos hechos de apariencia delictiva cierre los ojos ante los indicios de delito que se presenten a su vista, aunque los hallados casualmente sean distintos a los hechos comprendidos en su investigación oficial, siempre que ésta no sea utilizada fraudulentamente para burlar las garantías de los derechos fundamentales (SSTC 49/1996 y 41/1998)". Y en palabras del Tribunal Supremo, "[...] la investigación de unos hechos delictivos no impide la persecución de cualesquiera otros distintos que sean descubiertos por casualidad al investigar aquellos, pues los funcionarios de policía tienen el deber de poner en conocimiento de la autoridad penal competente los delitos de que tuviera conocimiento, practicando incluso las diligencias de prevención que fueran necesarias por razón de urgencia, tal y como disponen los artículos 259 y 284 LECrim" (STS 1110/2010, de 23 de diciembre), pudiendo además automatizarse esa puesta de manifiesto de los nuevos indicios ante el órgano judicial. No obstante, la clave se residenciará en determinar si una IA puede hallar elementos delictivos "por casualidad" o si debe estar previamente entrenada para ello, lo que eliminaría ese factor accidental.

Una segunda consecuencia ante el uso de la IA agéntica en el análisis forense será su empleo como asistencia para la dirección de investigaciones encalladas. Hemos comprobado como el *deep learning* es especialmente efectivo cuando se trata de aprender, tratar y analizar datos no estructurados, por lo que esta tecnología y las aplicaciones de IA basadas en ella, pueden resultar especialmente interesantes para asistir en la toma de decisiones a los investigadores y su capacidad para gestionar investigaciones complejas.

Cuando se acuerda el sobreseimiento provisional porque, de los datos obrantes en la causa, no queda "debidamente justificada la perpetración del delito" o no hay "motivos

³⁷ Sobre la superación de la sintaxis y el análisis de la semántica a través de la IA, véase Bernardo D. Olivares Olivares, "Pathfinder", *Fiscal Blog*, accedido 1 de febrero de 2026, <https://fiscalblog.es/pathfinder/>.

suficientes para acusar a determinada o determinadas personas como autores, cómplices o encubridores" (art. 641 LECrim), los investigadores se encuentran sin expectativas de obtener nuevos datos inculpatorios, pero nada debe objetarse a que esas piezas del puzzle puedan ser de nuevo examinadas y ensambladas automáticamente. La reapertura de un caso sobreseído se condiciona a que se aporten nuevos elementos de comprobación.

Tradicionalmente, se ha venido exigiendo que esos nuevos elementos no deben estar obrantes en la causa (SSTS 75/2014 de 11 de febrero y 795/2016, de 25 de octubre), es decir, debe tratarse de "nuevos datos o elementos de comprobación distintos de los resultantes del mismo" (STS 6/2008, de 23 enero), de modo que el estándar exigido por el Tribunal Supremo no es el de una simple "revisión" del mismo material instructor, sino el de una alteración sustantiva del fundamento del archivo. Y, sin embargo, el Tribunal Supremo acaba de matizar (STS 41/2026, de 26 de enero) que dicha doctrina sobre nuevos elementos de prueba "no se constriñe a la aparición de una prueba plena o a la detección de un nuevo indicio inculpatorio. El canon es "nuevos elementos de comprobación o de justificación", lo que integra –en palabras del Alto Tribunal– la aparición de nuevas vías de investigación antes no vislumbradas o la *aparición de mecanismos de esclarecimiento antes inexistentes, inaccesibles o no practicables*, siempre que se identifique de modo concreto, estén conectados con el objeto de la causa y resulten idóneos para remover la razón de insuficiencia que motivó el archivo en su día.

Esta nueva interpretación abre la puerta al empleo de novedosas técnicas de criminalística antes inviable, que permitan revisar y vislumbrar nuevas estrategias de investigación. Así viene sucediendo, principalmente, en el campo de la Genética Forense y Genealogía Genética (FIGG), en donde técnicas forenses de última generación como el software de Genotipado Probabilístico o el Next-Generation Sequencing (NGS), permiten extraer perfiles genéticos a partir de muestras minúsculas o degradadas, así como también obtener mucha más información genética de las muestras, identificando a sospechosos a través de bases de datos genealógicas. Lo mismo sucede con las nuevas herramientas informáticas de análisis de dispositivos móviles antiguos, con capacidades de recuperación de datos muy superiores a las practicadas en años anteriores.

La tercera gran consecuencia del análisis forense a través de herramientas de IA será la generación de las llamadas evaluaciones automatizadas de sospechas (*Automated Suspicion Assessment*) para identificar comportamientos, actividades o transacciones inusuales o potencialmente delictivas.

En el ámbito de las recomendaciones emitidas por el Grupo de Acción Financiera Internacional (GAFI o FATF, por su sigla en inglés) en materia de medidas para prevenir y combatir el lavado de activos, el financiamiento del terrorismo y el financiamiento de la proliferación de armas de destrucción masiva (LA/FT/FP), una de las exigencias a las entidades se refiere a la necesidad de implementar medidas de Debida Diligencia del Cliente (DDC) que incluyen el deber de comunicación –reporte– de operaciones sospechosas a la Unidad de Inteligencia Financiera correspondiente (en España, el Servicio Ejecutivo de la Comisión de Prevención del Blanqueo de Capitales e Infracciones Monetarias –SEPBLAC–). Y para ello, realizan los denominados "Informes de Actividad Sospechosa" o SAR (del inglés *Suspicious Activity Report*) en virtud de una serie de circunstancias "sospechosas" que el SEPBLAC denomina "Catálogo de Indicadores de Riesgo" (CIR). Por ejemplo, transacciones por encima de cierto valor o transacciones justo por debajo de los umbrales de reporte; transacciones inusuales o una actividad inusual de la cuenta; transacciones de personas que se sabe o se sospecha que tienen vínculos con organizaciones criminales o terroristas; entrega de cantidades

de efectivo demasiado cuantiosas para la actividad declarada; el desistimiento de la ejecución de una operación, o la modificación de alguna de sus características (como el importe), cuando el sujeto obligado solicita al cliente la aportación de documentación o información adicional; la exigencia por parte del cliente de niveles inusualmente altos de confidencialidad o discreción, etc.

De igual modo que los sujetos obligados emplean software informático para la detección y elaboración de dichos informes de operaciones sospechosas, las fuerzas del orden pueden servirse de similares herramientas a la hora de detectar patrones y anomalías que podrían indicar posibles actividades fraudulentas y facilitar la elaboración de informes que acrediten, mediante mapas visuales, la identificación, rastreo y documentación de flujos de dinero con los que acreditar esos indicios de blanqueo de capitales, tráfico de drogas, y cualquier otro delito susceptible de ser declarado probado en virtud de la prueba indiciaria.

Nuestro Tribunal Supremo tiene expresamente declarado (STS 1102/2024, de 28 de noviembre), con relación a la prueba indiciaria y al delito de blanqueo de capitales:

[...] las evidentes dificultades para desentrañar los flujos económicos que pueden estar por debajo de cualquier negocio aparentemente legal. La propia naturaleza del delito de blanqueo de capitales hace que el delito se enmascare mediante opacidad financiera muchas veces consistente en largas cadenas de transmisión de bienes entre distintas sociedades de control y titularidad difícilmente identificable, de modo que el capital o los bienes que tienen un origen en actividades ilícitas, especialmente en el tráfico de drogas, puedan sustraerse a la acción de los tribunales.³⁸

Por ello, está unánimemente admitido por la comunidad internacional y por la cultura constitucional más garantista, que la utilización de métodos inductivos sobre bases indiciarias está absolutamente justificada si se quiere conseguir los efectos previstos por el legislador. Y así se refleja en numerosos cuerpos normativos internacionales, como el artículo 3.3 de la Convención de Viena de 1988; el artículo 6.2 c) del Convenio de Estrasburgo de 1990 o artículo 6.2.f) de la Convención de Nueva York contra la Delincuencia Organizada Transaccional, al destacar que la lucha contra esas realidades criminológicas exige esa herramienta de valoración probatoria que, como hemos indicado en el punto anterior, goza de honda raigambre y no es exclusiva de esta modalidad criminal.

La IA agéntica es capaz de automatizar y elaborar esos cuadros de indicios a partir de la Extracción de relaciones desde los datos obrantes en la causa (declaraciones testificales, antecedentes penales y policiales, datos fiscales y registrales, comunicaciones mantenidas), buscar conexiones en grafos de entidades y personas (por ejemplo, negocios jurídicos inmobiliarios, nombres de usuario, billeteras de moneda digital, etc.) e interrelacionar esos y otros indicios (sociedades instrumentales que nunca presentaron cuentas anuales o cuyas cuentas corrientes apenas reflejan algún tipo de actividad comercial) con el hecho nuclear precisado de prueba.

En este sentido, la IA también está llamada a representar un avance sustancial en cuanto a la elaboración y eficacia de los «informes de inteligencia policial». A caballo entre prueba pericial y testifical, en razón a la duplicidad de quien así declara en juicio oral, ya que el agente policial que elabora el informe conoce del contenido de la materia y en consecuencia lo hace por sus conocimientos científicos, pero también actúa como

³⁸ España Tribunal Supremo Sala Penal, "Sentencia núm. 1102/2024", de 28 de noviembre de 2024. ECLI:ES:TS:2024:6069.

testigo en razón de lo que sabe, nuestros tribunales acercan los informes de inteligencia a la prueba del testigo-perito, pues el medio de prueba se integra por la opinión o dictamen de una persona y, al mismo tiempo, una prueba indirecta en tanto proporciona conocimientos técnicos para valorar los hechos controvertidos, pero no un conocimiento directo sobre cómo ocurrieron los hechos.

Lo relevante de este tipo de informes se debe a que permiten al tribunal “conocer y evaluar modos de comportamiento delictivo, generales o concretos, morfologías delincuenciales, metodologías operativas y puntos de conexión que los hechos investigados pueden tener con otros delitos ya cometidos y sometidos a investigación policial” (SSTS 2084/2001, de 13 de diciembre; 786/2003, de 29 de mayo o 352/2009, de 31 de marzo), con lo que el uso de la IA puede facilitar un mejor producto de la evaluación, integración, análisis e interpretación de la información reunida por el equipo de inteligencia, ampliándose así los conocimientos que puedan tenerse sobre una determinada materia. Más aún cuando el Tribunal Supremo ha validado (SSTS 1094/2010, de 10 de diciembre y 1061/2025, de 22 de diciembre) que los informes periciales de inteligencia puedan ser complementados con la colaboración auxiliar del CNL en labores accesorias y no fundamentadoras de la condena, como por ejemplo, mediante el descifrado de la documentación digital obtenida y examinada por los agentes policiales o la corroboración de algunos extremos de dicha información.

En último término, la IA también constituye una eficaz herramienta con la que efectuar la reconstrucción de escena del crimen y su exposición ante el tribunal.

La aplicación combinada de tecnologías digitales 3D e IA puede emplearse para mejorar la ejecución de ciertas fases de las técnicas de visualización forense. Esta combinación de tecnologías puede crear modelos gráficos 3D de objetos y personas, sobre la base de mediciones e imágenes, y animar dichos modelos para recrear la escena del crimen y los acontecimientos relacionados. Un ejemplo de esta técnica es el empleo del reconocimiento visual de patrones para el análisis del tamaño, forma y distribución de manchas de sangre, todo ello con el mencionado objeto de reconstruir los hechos acaecidos.

Durante décadas, el uso de PowerPoint en los tribunales ha pasado de ser una novedad a una herramienta común para abogados, fiscales y peritos, con el fin principal de convertir información compleja en ayudas visuales comprensibles, aumentando la retención de información y la persuasión.

Los peritos se sirven de herramientas visuales como Autocad o Vektor Work a la hora de mostrar bidimensional o tridimensionalmente el resultado de fotografías, vídeos, diagramas o gráficos complejos que ilustran su testimonio, lo que les permite simplificar información compleja o técnica y facilitar que el juez o jurado recuerden los puntos clave de sus conclusiones.

La IA no solamente facilita esta infografía forense y el mejor modo de representar la información ante el tribunal: cuadro de mando global, tabla de detalles, eje temporal o timeline, grafo de comunicaciones, vista de mapa, vista de galería, vista de personas, vista de correos, vista de chats, etc., sino que también permite someter ese escenario del crimen a todo tipo de hipótesis y variables durante la causa judicial, como pueda ser la existencia de testimonios contradictorios, careos, nuevos vestigios y nuevas informaciones apprehendidas durante la investigación, etc.

Conclusiones

La recolección a escala *macro* de la infinidad de datos que pueden llegar a localizarse en fuentes abiertas, junto con el examen y análisis a escala *micro* de cada uno de los detalles o metadatos de esa información, unida a la localizable en los instrumentos, efectos y vestigios del hecho delictivo, ha supuesto un avance fundamental en las pericias forenses. Ello porque en lo más profundo de un *byte* o de un *pixel* puede estar la clave para determinar la culpabilidad o inocencia de un sujeto.

Ahora, el desarrollo y aplicación de las novedosas herramientas y sistemas de Inteligencia Artificial en el campo del análisis policial forense están llamados a constituir una evolución trascendental en cuanto al modo en que se realizan las pesquisas y pericias durante la investigación penal, así como también en cuanto a la información constatable a partir de tales actuaciones policiales.

Del mismo modo que los criminales han perfeccionado sus técnicas delictivas hasta convertirlas en toda una ciencia, las autoridades se han visto en la necesidad de acudir a la ciencia y la tecnología para facilitar las labores de investigación y persecución eficaz de esa delincuencia cada vez más compleja. De la misma manera que hemos pasado de viajar a caballo o en carruaje a utilizar modernas aeronaves que nos permiten alcanzar la órbita exterior terrestre, la policía ha pasado de utilizar linternas y prismáticos a manejar modernas herramientas informáticas y dispositivos electrónicos en el campo forense (dactiloscopia, balística, acústica, falsificaciones, bioquímica, etc., en donde resulta especialmente destacable el manejo de sofisticados instrumentos y programas informáticos agrupados en la técnica conocida como Informática forense —*computer forensics*—), así como en el campo operativo, a través de lo que se ha venido a denominar la “vigilancia electrónica o tecnovigilancia”.

El empleo de la Inteligencia Artificial en el terreno de la criminalística por parte de los investigadores, centrados en la identificación, analítica e investigación técnica, está llamada a dar inicio a una nueva era de la investigación policial³⁹.

³⁹ Pedro Laureano Cabezas, “La investigación del crimen a través de los tiempos” (tesis doctoral, Universitat Autònoma de Barcelona, 2010), <https://www.educacion.gob.es/teseo/imprimirFicheroTesis.do?idFichero=FIHxYNhGkts%3D>.

Bibliografía

Aguilera Díaz, Vicente y Carlos Seisdedos. *Open Source INTelligence (OSINT): Investigar personas e identidades en internet*. Madrid: OxWord, 2020.

Balkin, Jack M. "The Constitution in the National Surveillance State". *Minnesota Law Review* 93, n.º 1 (2008): 1-25.

Bueno de Mata, Federico. *Investigación y prueba de delitos de odio en redes sociales: técnicas OSINT e inteligencia policial*. Valencia: Tirant lo Blanch, 2023.

Cabezas, Carlos Eduardo Pedro Laureano. "La investigación del crimen a través de los tiempos". Tesis doctoral, Universitat Autònoma de Barcelona, 2010. <https://www.educacion.gob.es/teseo/imprimirFicheroTesis.do?idFichero=FIHxYNhGkts%3D>.

Esteban Cobo, Miguel. "Herramienta para la extracción automática de metadatos en vídeos de dispositivos móviles". Trabajo de fin de grado, Universidad Complutense de Madrid, 2016. <https://eprints.ucm.es/id/eprint/38697/>.

Lafont Nicuesa, Luis. "La policía predictiva más allá de Minority Report". *Diario La Ley*, n.º 10687, 2025.

Llamas Fernández, Manuel y José Miguel Gordillo Luque. "Medios técnicos de vigilancia". En *Los nuevos medios de investigación en el proceso penal. Especial referencia a la tecnovigilancia*, Cuadernos de Derecho Judicial 2007-II. Madrid: Consejo General del Poder Judicial, 2007.

Martín Ríos, Pilar. "Empleo de big data y de inteligencia artificial en el ciberpatrullaje: de la tiranía del algoritmo y otras zonas oscuras". *IDP. Revista de Internet, Derecho y Política*, n.º 36 (2022): 1-13. doi:10.7238/idp.voi36.394511.

Miró Llinares, Fernando. "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots". *Revista de Derecho Penal y Criminología*, n.º 20 (2018): 87-130. doi:10.5944/rdpc.20.2018.26446.

Nieva Fenoll, Jordi. "El tránsito de la fe a la tecnología en el proceso penal". *Diario La Ley*, n.º 9986, 11 de enero de 2022.

Olivares Olivares, Bernardo. "Pathfinder". *Fiscal Blog*, accedido 1 de febrero de 2026. <https://fiscalblog.es/pathfinder/>.

Ortiz Pradillo, Juan Carlos. *Problemas procesales de la ciberdelincuencia*. Madrid: Colex, 2013.

———. "Desafíos legales de las diligencias de investigación tecnológica". En *El proceso penal. Cuestiones fundamentales*, coordinado por Olga Fuentes Soriano, 303-316. Valencia: Tirant lo Blanch, 2017.

———. "Big data, vigilancias policiales y geolocalización: nuevas dimensiones de los derechos fundamentales en el proceso penal". *Diario La Ley*, n.º 9955, 18 de noviembre de 2021.

———. "Vigilancias policiales y utilización de dispositivos de seguimiento, localización y captación de la imagen". En *Reflexiones en torno al Anteproyecto de Ley de Enjuiciamiento Criminal de 2020*, coordinado por Olga Fuentes Serrano, 813-881. Valencia: Tirant lo Blanch, 2022.

Rodríguez Láinz, José Luis. "En torno al concepto de comunicación protegida por el art. 18.3 de la Constitución". *Diario La Ley*, n.º 8143, 5 de septiembre de 2013.

———. "Reflexiones sobre el tratamiento de datos personales por prestadores de servicios de comunicaciones vía internet para la lucha contra abusos sexuales de menores en línea en el Reglamento (UE) 2021/1232". *Diario La Ley*, n.º 9974, 20 de diciembre de 2021.

Snowden, Edward. *Vigilancia permanente*. Barcelona: Editorial Planeta, 2019.

LA INCIDENCIA DE LA INTELIGENCIA ARTIFICIAL EN LOS DERECHOS DE DEFENSA Y LA PRESUNCIÓN DE INOCENCIA

María Lourdes Noya Ferreiro¹

1. Derecho de defensa e Inteligencia Artificial

El derecho de defensa se configura como uno de los derechos fundamentales más relevantes, sin cuya observancia no se puede garantizar una justicia propia de un Estado de Derecho². Su respeto y valor fundamental constituye una de las premisas sobre las que se asienta nuestro proceso penal, que se erige en el instrumento necesario para la aplicación del Derecho penal y el ejercicio del *ius puniendi* del Estado. Sirve para garantizar la seguridad pública, pero también ha de preservar o garantizar los derechos de la persona investigada.

El ordenamiento jurídico ha rodeado al "imputado"³ de un conjunto de garantías que configuran un status procesal ya asentado en la cultura jurídica occidental, que nuestra Constitución consagra en su art. 24⁴. La mera imputación le confiere una pluralidad de derechos instrumentales de carácter constitucional que se engarzan en el contenido del derecho de defensa y que deben ser escrupulosamente respetados⁵.

¹ Profesora Titular de Derecho Procesal, Universidad de Santiago de Compostela, España.

² Sobre el origen de la defensa y su evolución v. Nicolás González-Cuellar Serrano, "El derecho de defensa y la marca de Caín", en *Nuevos debates en torno a la justicia española*, dirigido por Raquel Castillejo Manzanares (Valencia: Tirant lo Blanch, 2017), 289 y ss.

³ El término "imputado", utilizado para designar a aquella persona que estaba siendo objeto de una investigación judicial, ha sido sustituido por "investigado" por la Ley Orgánica 13/2015, de 5 de octubre, de modificación de la Ley de Enjuiciamiento Criminal para el fortalecimiento de las garantías procesales y la regulación de las medidas de investigación tecnológica.

⁴ V. SSTC 205/1989, de 11 de diciembre, 161/1994, de 23 de mayo, y 277/1994, de 17 de octubre, entre otras.

⁵ Víctor Moreno Catena, "La garantía de los derechos fundamentales durante la investigación penal", en *Problemas actuales del proceso penal y derechos fundamentales*, Cuadernos Penales José María Lidón, n.º 7 (Bilbao: Universidad de Deusto, 2010), 18-19.

Este conjunto de garantías o derechos que conforman el derecho de defensa se encuentran estrechamente vinculados al principio de contradicción que impera en el proceso penal, que precisa de un justo equilibrio entre las partes y, por tanto, del respeto al principio de igualdad de armas⁶. La contradicción, calificada como una de las garantías esenciales del proceso penal del Estado de Derecho⁷, es consustancial a la idea misma de proceso y constituye un principio general de la justicia penal⁸, tal y como se pone de relieve en numerosas resoluciones judiciales⁹. Su efectividad supone que las partes puedan hacer valer sus pretensiones a través de la introducción de hechos y pruebas, conociendo el material obrante en el proceso, e interviniendo en su desarrollo con el ánimo de influir en la resolución judicial¹⁰. En este sentido, la contradicción se identifica con la igualdad de las partes en el proceso, y con la facultad de reconocerles la posibilidad de oponerse a la pretensión del adversario¹¹. Se trata pues de facilitar la oposición, y el debate o discusión sobre la pretensión, a través de la comparecencia en el proceso y la introducción de hechos y pruebas¹².

No hay en la Constitución una referencia expresa al principio de contradicción, pero su reconocimiento se entiende implícito en el art. 24¹³, tanto en su primer apartado, que al declarar el derecho a la tutela judicial efectiva prohíbe la indefensión, como en el segundo, al regular el derecho de defensa y el derecho a un proceso con todas las garantías¹⁴.

Doctrina y jurisprudencia destacan el carácter poliédrico del derecho de defensa en el que se incluyen un conjunto de derechos instrumentales que conforman las garantías procesales consustanciales al Estado de Derecho, y que se recogen en el segundo apartado del art. 24 CE¹⁵. Los derechos a la asistencia de letrado, a ser informados de la acusación formulada contra ellos, a utilizar los medios de prueba pertinentes para

su defensa, a no declarar contra sí mismos, y a no confesarse culpables, conforman el contenido esencial del derecho de defensa, y se integran, además, en el contenido esencial del derecho a un proceso con todas las garantías, o al debido proceso.

Partiendo de estas consideraciones, en un proceso penal en el que se utilicen sistemas de inteligencia artificial, el derecho de defensa puede verse gravemente afectado, y ello puede comprobarse fácilmente con la simple referencia a las siguientes cuestiones. Estos sistemas de IA van a suponer un alto coste para los despachos de abogados que quieran disponer de ellos, porque desde luego no se trata de sistemas de fácil acceso para la gran mayoría de los despachos. Este simple hecho va a provocar que algunos sujetos puedan acceder a estos sistemas de apoyo a la estrategia procesal de defensa, mientras que otros, con menos recursos económicos, no se pueden beneficiar de su aplicación, lo que indirectamente va a provocar que la posición de partida de las partes en el proceso resulte ya muy desigual¹⁶.

Una segunda cuestión de interés es la aplicación de los sistemas predictivos para elaborar una estrategia de defensa, e incluso para que dicha defensa se asuma por un abogado o despacho concreto. En este sentido, el acceso a este tipo de instrumentos y al análisis que se extraiga del sistema, puede influir de forma definitiva sobre la viabilidad del caso y la decisión de aceptarlo. La complejidad del tema y, desde luego, la predicción de un mal resultado puede llevar al profesional a rechazar la defensa, colocando al sujeto investigado en una difícil situación, afectando al derecho a la defensa, pero también a la igualdad de armas y al derecho a la tutela judicial efectiva¹⁷.

No cabe duda de que la utilización de este tipo de sistemas en el proceso penal incide en el derecho de defensa. La introducción de los factores de riesgo se realiza sin contradicción, por parte de sujetos que tampoco están vinculados a ningún control oficial. Su dependencia profesional de una empresa privada y su obligación hacia el secreto empresarial impide en principio conocer su estructura y funcionamiento, por lo que hace imposible impugnar su fiabilidad¹⁸. Este problema puede solucionarse condicionando legalmente su aplicación en el proceso, a que se facilite el acceso a la información y al funcionamiento del sistema, aunque, como ya se ha apuntado, en algunas ocasiones el cruce de datos del sistema escapa incluso al entendimiento de su programador, lo que representa un serio problema¹⁹.

Por tanto, la opacidad del sistema, tantas veces criticada y atacada, pero a la que no se le ha puesto remedio, supone un grave inconveniente para el ejercicio del derecho de defensa. No sólo afecta a la posibilidad de recurrir la resolución, al desconocerse el modo en el que se ha obtenido, sino que, por el mismo motivo, también impide la proposición de medios o pruebas que la contradigan.

En el mismo orden de consideraciones, tampoco se puede obviar la entrada en el proceso penal de informes policiales basados en técnicas de IA, que adquieren la

6 V. en este sentido Vicente Gimeno Sendra, "El derecho de defensa", en *Constitución y proceso* (Madrid: Tecnos, 1988), 88. V. también las SSTs de 15 de abril de 1996 y núm. 1686/1997, de 10 de septiembre.

7 V. Eduardo de Urbano Castrillo, "El principio de contradicción en el proceso penal", *Diario La Ley*, n.º 5474, 4 de febrero de 2002, 1712.

8 V. Piero Calamandrei, *Proceso y democracia* (Buenos Aires: Editoriales Jurídicas Europa-América, 1960), 147; Andrés de la Oliva Santos, Ignacio Díez-Picazo Giménez y Jaime Vegas Torres, *Curso de derecho procesal civil* (Madrid: Editorial Universitaria Ramón Areces, 2019); Víctor Moreno Catena, "Algunos problemas del derecho de defensa", *Justicia*, n.º 3 (1990): 562; Juan Montero Aroca, *Principios del proceso penal. Una explicación basada en la razón* (Valencia: Tirant lo Blanch, 1997), 139, entre otros.

9 A título de ejemplo la sentencia 49/2008, de 25 de febrero.

10 V. Andrea Planchadell Gargallo, *El derecho fundamental a ser informado de la acusación* (Valencia: Tirant lo Blanch, 1999), 77.

11 V. Faustino Gutiérrez-Alviz y Conradi, *El derecho de defensa y la profesión de abogado* (Barcelona: Atelier, 2012), 20; v. también Francesco Carnelutti, *Derecho y proceso*, trad. Santiago Sentís Melendo (Buenos Aires: Ediciones Jurídicas Europa-América, 1971), 112 y ss.; Jaime Guasp Delgado, *Derecho Procesal Civil* (Madrid: Instituto de Estudios Políticos, 1968), 171.

12 V. a título de ejemplo las SSTC 109/1985, de 11 de julio y 155/1995, de 24 de octubre.

13 V. SSTC 178/2001, de 17 de septiembre, 179/2002, de 14 de octubre, 8/2003, de 20 de enero, 143/2004, de 13 de septiembre, 182/2009 de 7 de septiembre; y 53/1987, de 7 de mayo; 18/1995 de 24 de enero; 162/1997, de 3 de octubre; 16/2000, de 31 de enero; 79/2000, de 27 de marzo; 93/2000, de 10 de abril; 101/2001, de 23 de abril. V. también STS 806/2012, de 23 octubre. V. también las SSTs 1080/2006, de 2 de noviembre; 1238/2009, de 11 de diciembre; 383/2010, de 5 de mayo; 365/2012, de 15 de mayo; 775/2012, de 17 de octubre; 125/2012, de 29 de diciembre; 974/2012, de 5 diciembre.

14 V. Víctor Moreno Catena, *La defensa penal* (Madrid: Civitas, 1982), 25; Valentin Cortés Domínguez, *Introducción al Derecho Procesal*, con Víctor Moreno Catena (Valencia: Tirant lo Blanch, 2016), 242; Vicente Gimeno Sendra, *Introducción al Derecho Procesal* (Madrid: Colex, 2014), 276; Juan Montero Aroca, *Derecho Jurisdiccional I. Parte general*, con Juan Luis Gómez Colomer, Alberto Montón Redondo y Silvia Barona Vilar (Valencia: Tirant lo Blanch, 2014), 246.

15 V. Víctor Moreno Catena, *La defensa penal* (Madrid: Civitas, 1982), 24-25; Juan Luis Gómez Colomer, *Derecho Jurisdiccional III. Proceso Penal*, con Juan Montero Aroca, Alberto Montón Redondo y Silvia Barona Vilar (Valencia: Tirant lo Blanch, 2012), 247; Vicente Gimeno Sendra, *Los procesos penales*, con Cándido Conde-Pumpido Tourón y José Garberi Llobregat, tomo II (Barcelona: Bosch, 2000), 166; Eduardo de Urbano Castrillo, "El principio de contradicción en el proceso penal", *Diario La Ley*, n.º 5474, 4 de febrero de 2002, 1172; Verónica López Yagüe, "La condición de imputado en el proceso penal español. Formas de adquisición y status jurídico que conlleva. Ideas para su reforma", en *La reforma del proceso penal* (Madrid: La Ley, 2011); Andrea Planchadell Gargallo, *El derecho fundamental a ser informado de la acusación*, 82; Enrique Ruiz Vadillo, "Algunas breves consideraciones sobre el sistema acusatorio y la interdicción constitucional de toda indefensión en el proceso penal", *Diario La Ley*, tomo 4, 1987, 874; José Miguel Serrano Ruiz-Calderón, "El derecho de defensa según la sentencia 79/2012 de la Sala de lo Penal del Tribunal Supremo", *Diario La Ley*, n.º 7850, 3 de mayo de 2012, 5.

16 Fernando Martín Diz, "Inteligencia artificial y proceso: garantías frente a eficiencia en el entorno de los derechos procesales fundamentales", en *Justicia: ¿Garantías versus eficiencia?*, dirigido por Fernando Jiménez Conde y Rafael Bellido Penadés, coordinado por Patricia Llopis Nadal y Elena de Luis García (Valencia: Tirant lo Blanch, 2019), 292.

17 Cristina San Miguel Caso, "La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales?", *IUS ET SCIENTIA: Revista electrónica de Derecho y Ciencia*, vol. 7, n.º 1 (2021): 292.

18 Fernando Gascón Inchausti, "Desafíos para el proceso penal en la era digital: externalización, sumisión pericial e inteligencia artificial", en *La justicia digital en España y la Unión Europea*, dirigido por Jesús Conde Fuentes y Gregorio Serrano Hoyo (Barcelona: Atelier, 2019), 203.

19 Miren Josune Pérez Estrada, *Fundamentos jurídicos para el uso de la inteligencia artificial en los órganos judiciales* (Valencia: Tirant lo Blanch, 2022), 139.

categoría de prueba pericial especial o "sui géneris"²⁰, al estar basada en conocimientos técnicos y especializados, pero cuya introducción en el proceso penal conlleva un grave problema para los derechos de defensa y la contradicción²¹. No puede mantenerse que tales informes estén exentos del debate contradictorio, pero al realizarse por técnicos de alto nivel, con instrumentos complejos y en condiciones específicas, va a resultar muy complicado elaborar un contrainforme, e incluso rebatir el aportado²².

Otra cuestión estrechamente ligada al derecho de defensa está vinculada con el derecho a la información, como presupuesto básico del principio de contradicción y de su ejercicio efectivo. Es cierto que este derecho se centra ahora en la información de la acusación y de los hechos que se imputan²³. Pero, también es cierto que, el primero investigado, y después acusado, ha de conocer las diligencias de investigación practicadas y las fuentes de prueba que pueden tener acceso al proceso. Y es precisamente en esta cuestión en la que pueden derivarse problemas en la aplicación de sistemas inteligentes.

Una vez más la opacidad de los instrumentos de IA constituye una de las principales dificultades que plantea su aplicación en el sistema de justicia penal. El desconocimiento de los algoritmos, cada vez más complejos, de su funcionamiento y evolución impide trasladar la información a la defensa, colocando al acusado en una muy difícil situación, no solo a la hora de rebatir y contradecir la información facilitada por el sistema inteligente, sino también en dos campos más: el de proponer pruebas que puedan rebatir lo afirmado por la IA, y el de convencer al juzgador de la posible inexactitud en la valoración de la máquina.

Precisamente, es la dificultad de la información sobre el funcionamiento de los sistemas o aplicaciones de inteligencia artificial, lo que genera mayores dificultades. Las partes de un proceso penal han de conocer el número de decisiones o casos que se han suministrado para llevar a cabo su procesamiento; se ha de conocer el origen de esas decisiones y cómo se han elegido, y que representatividad suponen entre todas las posibles; en definitiva, ha de proporcionarse el material que se ha suministrado al sistema²⁴.

Desde otra perspectiva, otra de las cuestiones a la que ha de hacerse referencia es la relativa a las garantías que han de rodear la entrada en el proceso de la información derivada de la vigilancia policial utilizando instrumentos de IA. No es extraño que la policía realice labores de prevención a través de actuaciones de vigilancia genérica. En este contexto puede almacenarse información vinculada a una persona, sobre la que en ese momento no existe causa penal alguna²⁵. Sin embargo, cuando se produce la noticia

crimínica, y se inicia la fase de investigación del proceso, la pregunta que cabe hacerse es si ese material derivado de una vigilancia genérica anterior puede ser introducido en el proceso²⁶. Estas cuestiones exigen una regulación específica sobre la utilización de la información policial obtenida a través de sistemas de IA en labores de prevención, de igual forma a la que, en su momento, se reguló por la Ley 13/2015, en relación con los descubrimientos casuales.

2. Derecho a la presunción de inocencia e Inteligencia Artificial

La consagración de la presunción de inocencia como derecho fundamental tiene lugar en el artículo 24 de la Constitución, dando así inicio a una regulación del proceso penal informado por principios y garantías derivados de la proclamación del Estado social y democrático de Derecho. Este reconocimiento trae causa de los tratados internacionales ratificados por España una vez constituido como Estado democrático. Representa una de las principales garantías sobre la que gira el enjuiciamiento penal, erigiéndose en uno de los pilares sobre los que se apoya el derecho al proceso debido. La puesta en peligro del derecho a la presunción de inocencia implica el desconocimiento de los principios esenciales del proceso penal²⁷.

El contenido del derecho fundamental a la presunción de inocencia ha de determinarse desde una doble perspectiva. Se trata en primer lugar de una regla de tratamiento, al imponer que el investigado y encausado será considerado inocente hasta que se produzca su condena por sentencia firme. En segundo lugar, ha de considerarse como regla de juicio, al operar sobre el juzgador que fundamentará la condena en una actividad probatoria desarrollada en el proceso penal, en su fase de juicio oral, con respeto a los principios de contradicción e igualdad de armas²⁸.

Los sistemas de inteligencia artificial, que parten de los datos proporcionados previamente, podrán afectar a la presunción de inocencia cuando dichos datos se hayan introducido en función de estudios sesgados, o de patrones pasados discriminatorios, atendiendo a factores relacionados con el origen, la raza, la realidad socio-económica anterior, etcétera²⁹, que van a incidir notablemente en el juzgador, poniendo en peligro su imparcialidad y la presunción de inocencia del acusado³⁰. En estos casos, la regla de tratamiento aparece desvirtuada al suministrar datos que tienden a dirigir el funcionamiento del sistema en una determinada dirección. En este contexto, se ofrecerán unos resultados que aparecen ya viciados en su inicio, por lo que su introducción en el proceso pondrá en peligro la imparcialidad judicial, afectando a la regla de juicio³¹.

20 V. STS 783/2007, de 1 de octubre; V. también la STS 124/2009, de 13 de febrero; 290/2010, de 31 de marzo.

21 V. Federico Bueno de Mata, *Investigación y prueba de delitos de odio en redes sociales: técnicas OSINT e inteligencia policial* (Valencia: Tirant lo Blanch, 2023), 241 y ss.

22 Salvador Guerrero Palomares, "La denominada «prueba de inteligencia policial» o «pericial de inteligencia»", *Revista de Derecho y Proceso Penal*, n.º 25 (2011): 13; v. también Silvia Barona Vilar, *Algoritmización del derecho y de la justicia. De la Inteligencia Artificial a la Smart Justice* (Valencia: Tirant lo Blanch, 2021), 509.

23 STC 44/1983, de 24 de mayo. V. también las SSTC 54/1985, de 18 de abril; 141/1986, de 12 de noviembre; 17/1989, de 30 de enero; 277/1994, de 17 de octubre; 95/1995, de 19 de junio; 123/1995, de 18 de julio y 36/1996, de 11 de marzo; 19/2000, de 31 de enero; 278/2000, de 27 de noviembre y 182/2001 de 17 septiembre, entre otras. En el mismo sentido las SSTS de 4 de noviembre de 1986, 21 de abril de 1987 y 1232/2001 de 22 junio, entre otras.

24 Juan Luis Gómez Colomer, "Derechos fundamentales, proceso e Inteligencia Artificial: una reflexión", en *Inteligencia artificial legal y Administración de justicia*, dirigido por Sonia Calaza López y Mercedes Llorente Sánchez-Arjona (Cizur Menor: Aranzadi, 2022), 277.

25 V. sobre esta cuestión Ignacio Colomer Hernández, "Control y límites en el uso de la información y los datos personales por parte de la Inteligencia Artificial en los procesos penales", en *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, editado por Silvia Barona Vilar (Valencia: Tirant lo Blanch, 2021), 301 y ss.

26 V. Vicente Carlos Guzmán Fluja, "Ideas para un debate sobre la predicción del crimen", en *Inteligencia artificial legal y Administración de justicia*, dirigido por Sonia Calaza López y Mercedes Llorente Sánchez-Arjona (Cizur Menor: Aranzadi, 2022), 333.

27 Su relevancia en el proceso constitucional también es puesta de relieve en la Directiva (UE) 2016/343, del Parlamento europeo y del Consejo, de 9 de marzo de 2016 (LA LEY 3261/2016), por la que se refuerzan determinados aspectos de la presunción de inocencia y el derecho a estar presente en el juicio.

28 V. sobre la presunción de inocencia Mercedes Fernández López, *Prueba y presunción de inocencia* (Madrid: Iustel, 2005); Fernando Martín Diz, "Presunción de inocencia en procesos penales por violencia de género", *Ius et Praxis*, vol. 24, n.º 3 (2018): 26 y ss., entre otros.

29 José Ignacio Solar Cayón, "Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos", en *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, editado por José Ignacio Solar Cayón (Alcalá de Henares: Universidad de Alcalá / Defensor del Pueblo, 2020), 169.

30 Fernando Martín Diz, "Modelos de aplicación de Inteligencia Artificial en justicia: asistencial o predictiva versus decisoria", en *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, editado por Silvia Barona Vilar (Valencia: Tirant lo Blanch, 2021), 79.

31 Fernando Martín Diz, "Modelos de aplicación de Inteligencia Artificial en justicia: asistencial o predictiva versus decisoria", 79.

Se trata de datos que pertenecen al origen y al pasado del sujeto, que, sin embargo, pueden resultar determinantes para decidir sobre su futuro. Se está aceptando, que el origen, circunstancias familiares y personales, entorno familiar y otros datos sobre la personalidad del investigado pueden condicionar en gran medida la decisión judicial. Con ello, la aplicación de la ley al caso concreto puede encontrarse condicionada por la evaluación anterior del sujeto investigado, lo que desde luego va a incidir en la interpretación de la norma y en una justicia de autor, que tendrá en cuenta no sólo las circunstancias actuales del acusado, sino también su pasado personal, familiar y social³². En este sentido, se pronuncia el Reglamento Europeo de Inteligencia Artificial, en su considerando 42.

Es importante destacar en este punto algunos elementos esenciales relativos al contenido de la presunción de inocencia como son:

[...] que el acusado debe ser tratado como si no hubiera cometido ninguna infracción hasta que el Estado, a través de las autoridades responsables del ejercicio de la acción penal, presente pruebas suficientes para que un tribunal independiente e imparcial lo declare culpable; o que la presunción de inocencia requiere que los miembros del tribunal no tengan una idea preconcebida de que el acusado ha cometido la infracción de la que se le acusa.³³

Una de las cuestiones más complejas relativas a la presunción de inocencia –y su alcance como derecho fundamental– es la de determinar si ha de aplicarse antes de la incoación de un proceso penal, como parámetro delimitador de las investigaciones policiales. No se pretende, en este caso, un estudio exhaustivo sobre el alcance de la presunción de inocencia, y menos si se tiene en cuenta que las posiciones doctrinales sobre esta cuestión no son unánimes, por lo que habría que dedicarle un análisis riguroso y detallado de su naturaleza y extensión, pero sí ha de realizarse alguna reflexión.

Como ya es sabido, la actividad policial de investigación de los hechos delictivos y de las conductas criminales suele ser anterior a la incoación del proceso penal. Las actuaciones policiales de prevención, con la finalidad de evitar la comisión de conductas delictivas, marcan el día a día del trabajo policial, en el que obviamente, y cada vez con mayor frecuencia, se utilizan instrumentos de IA. En este marco, la actividad policial puede centrarse en averiguar los lugares en los que se pueden realizar conductas delictivas, y en los tipos delictivos que se pueden perpetrar, lo que en una primera aproximación no representa problema alguno en relación con la protección de la presunción de inocencia. Sin embargo, cuando las investigaciones policiales tratan de determinar personas, o grupos de personas que pueden ser posibles autores de hechos delictivos concretos, la incidencia en la presunción de inocencia parece probable.

Cabe deducir, y ello es lógico, que los datos derivados de esa actuación preprocesal puedan ser introducidos en el proceso una vez iniciado, y que se pretenda convertirlos en material probatorio con la finalidad de desvirtuar la presunción de inocencia. Sin embargo, ha de tenerse en cuenta que la obtención de estas fuentes ha podido tener lugar al margen del control judicial, por lo que la limitación de derechos fundamentales de alta relevancia, como el derecho a la intimidad, a la privacidad, o a la protección de datos, se ha producido al margen de todo control judicial, y sin efectuar el necesario

juicio de proporcionalidad³⁴.

Como ya se ha apuntado, la doctrina no es unánime respecto de la extensión o del alcance de la presunción de inocencia en la fase de investigación preprocesal, pero lo que resulta evidente es que, de un lado, no se puede impedir la necesaria función preventiva de la policía, y, de otro, que esta función no puede llevarse a cabo con vulneración de los derechos fundamentales. Será determinante, por tanto, fijar los parámetros o los mecanismos de control para que los datos obtenidos de estas investigaciones puedan acceder al proceso rodeados de las garantías necesarias que eviten la prohibición probatoria.

Es innegable que, teniendo en cuenta ciertos requisitos y condiciones, y siempre atendiendo a razones de prevención, los agentes policiales están autorizados para el tratamiento de datos personales de aquellas personas que resulten sospechosas de cometer un hecho delictivo. Así lo permite la Directiva UE 680/2016, del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos, que en el artículo 6.1 recoge entre los destinatarios o "interesados" en este tratamiento a las "personas respecto de las cuales existan motivos fundados para presumir que han cometido o van a cometer una infracción penal".

Igual sentido se desprende del artículo 1 de la LO 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales, que al establecer el ámbito de aplicación de la citada ley, incluye expresamente la protección de los datos personales de las personas físicas en las actuaciones a prevención realizadas por las autoridades competentes, básicamente por los miembros y fuerzas de la seguridad del Estado.

A la hora de establecer las reglas de control que deberían tomarse en consideración para validar una actuación policial desarrollada en el marco de una investigación preprocesal, habrá que partir de las garantías que se exigen cuando esa actuación policial se lleva a cabo en la instrucción penal, bajo control judicial. De esta forma, la investigación desarrollada por la policía en la fase de instrucción del proceso penal está supervisada y controlada judicialmente, por lo que en aquellos casos en que se procede a la limitación de derechos fundamentales, dicha actuación se hace desde el juicio de la proporcionalidad y con el control judicial correspondiente. Y aun en estos casos se protege la presunción de inocencia y el derecho de defensa mediante la oportuna contradicción.

Sin embargo, en los supuestos en que se pretenden introducir datos obtenidos preprocesalmente, estos controles judiciales no se han producido. Por tal motivo, lo primero que habrá que llevar a cabo es un control judicial sobre el material obtenido. El juzgador ha de realizar una supervisión sobre la obtención de los datos y las condiciones en las que se ha producido, además de valorar, y aquí entraría en juego el juicio de proporcionalidad, las circunstancias personales del investigado y la gravedad del hecho investigado³⁵.

32 V. Teresa Armenta Deu, *Derivas de la justicia. Tutela de los derechos y solución de controversias en tiempos de cambio* (Madrid: Marcial Pons, 2021).

33 V. Víctor Moreno Catena, *Fiscalía europea y derechos fundamentales* (Valencia: Tirant lo Blanch, 2014), 112.

34 Vicente Carlos Guzmán Fluja, "Ideas para un debate sobre la predicción del crimen", 334.

35 Víctor Moreno Catena, *Fiscalía europea y derechos fundamentales*, 135-137.

Bibliografía

Armenta Deu, Teresa. *Derivas de la justicia. Tutela de los derechos y solución de controversias en tiempos de cambio*. Madrid: Marcial Pons, 2021.

Barona Vilar, Silvia. *Algoritmización del derecho y de la justicia. De la Inteligencia Artificial a la Smart Justice*. Valencia: Tirant lo Blanch, 2021.

Bueno de Mata, Federico. *Investigación y prueba de delitos de odio en redes sociales: técnicas OSINT e inteligencia policial*. Valencia: Tirant lo Blanch, 2023.

Calamandrei, Piero. *Proceso y democracia*. Buenos Aires: Editoriales Jurídicas Europa-América, 1960.

Calaza López, Sonia y Mercedes Llorente Sánchez-Arjona (dirs.). *Inteligencia artificial legal y Administración de justicia*. Cizur Menor: Aranzadi, 2022.

Carnelutti, Francesco. *Derecho y proceso*. Traducido por Santiago Sentís Melendo. Buenos Aires: Ediciones Jurídicas Europa-América, 1971.

Colomer Hernández, Ignacio. "Control y límites en el uso de la información y los datos personales por parte de la Inteligencia Artificial en los procesos penales". En *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, editado por Silvia Barona Vilar, 287-300. Valencia: Tirant lo Blanch, 2021.

Conde-Pumpido Tourón, Cándido, Vicente Gimeno Sendra y José Garberí Llobregat. *Los procesos penales*. Barcelona: Bosch, 2000.

Cortés Domínguez, Valentín (con Víctor Moreno Catena). *Introducción al Derecho Procesal*. Valencia: Tirant lo Blanch, 2016.

De la Oliva Santos, Andrés, Ignacio Díez-Picazo Giménez y Jaime Vegas Torres. *Curso de derecho procesal civil*. Madrid: Editorial Universitaria Ramón Areces, 2019.

De Urbano Castrillo, Eduardo. "El principio de contradicción en el proceso penal". *Diario La Ley*, n.º 5474, 4 de febrero de 2002.

Fernández López, Mercedes. *Prueba y presunción de inocencia*. Madrid: Iustel, 2005.

Gascón Inchausti, Fernando. "Desafíos para el proceso penal en la era digital: externalización, sumisión pericial e inteligencia artificial". En *La justicia digital en España y la Unión Europea*, dirigido por Jesús Conde Fuentes y Gregorio Serrano Hoyo, 191-206. Barcelona: Atelier, 2019.

Gimeno Sendra, Vicente. "El derecho de defensa". En *Constitución y proceso*. Madrid: Tecnos, 1988.

———. *Introducción al Derecho Procesal*. Madrid: Colex, 2014.

———. *Los procesos penales*, con Cándido Conde-Pumpido Tourón y José Garberí Llobregat. Barcelona: Bosch, 2000.

Gómez Colomer, Juan Luis. "Derechos fundamentales, proceso e Inteligencia Artificial: una reflexión". En *Inteligencia artificial legal y Administración de justicia*, dirigido por Sonia Calaza López y Mercedes Llorente Sánchez-Arjona, 257-287. Cizur Menor: Aranzadi, 2022.

———. *Derecho Jurisdiccional III. Proceso Penal*, con Juan Montero Aroca, Alberto Montón Redondo y Silvia Barona Vilar. Valencia: Tirant lo Blanch, 2012.

González-Cuéllar Serrano, Nicolás. "El derecho de defensa y la marca de Caín". En *Nuevos debates en torno a la justicia española*, dirigido por Raquel Castillejo Manzanares, 289 y ss. Valencia: Tirant lo Blanch, 2017.

Guasp Delgado, Jaime. *Derecho Procesal Civil*. Madrid: Instituto de Estudios Políticos, 1968.

Guerrero Palomares, Salvador. "La denominada «prueba de inteligencia policial» o «pericial de inteligencia»". *Revista de Derecho y Proceso Penal*, n.º 25 (2011): 13 y ss.

Gutiérrez-Alviz y Conradi, Faustino. *El derecho de defensa y la profesión de abogado*. Barcelona: Atelier, 2012.

Guzmán Fluja, Vicente Carlos. "Ideas para un debate sobre la predicción del crimen". En *Inteligencia artificial legal y Administración de justicia*, dirigido por Sonia Calaza López y Mercedes Llorente Sánchez-Arjona, 333-354. Cizur Menor: Aranzadi, 2022.

López Yagüe, Verónica. "La condición de imputado en el proceso penal español. Formas de adquisición y status jurídico que conlleva. Ideas para su reforma". En *La reforma del proceso penal*. Madrid: La Ley, 2011.

Martín Diz, Fernando. "Inteligencia artificial y proceso: garantías frente a eficiencia en el entorno de los derechos procesales fundamentales". En *Justicia: ¿Garantías versus eficiencia?*, dirigido por Fernando Jiménez Conde y Rafael Bellido Penadés, coordinado por Patricia Llopis Nadal y Elena de Luis García, 815-827. Valencia: Tirant lo Blanch, 2019.

———. "Modelos de aplicación de Inteligencia Artificial en justicia: asistencial o predictiva versus decisoria". En *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, editado por Silvia Barona Vilar, 65-85. Valencia: Tirant lo Blanch, 2021.

———. "Presunción de inocencia en procesos penales por violencia de género". *Ius et Praxis*, vol. 24, n.º 3 (2018): 26 y ss.

Montero Aroca, Juan. *Principios del proceso penal. Una explicación basada en la razón*. Valencia: Tirant lo Blanch, 1997.

———. *Derecho Jurisdiccional I. Parte general*, con Juan Luis Gómez Colomer, Alberto Montón Redondo y Silvia Barona Vilar. Valencia: Tirant lo Blanch, 2014.

Moreno Catena, Víctor. "Algunos problemas del derecho de defensa". *Justicia*, n.º 3 (1990): 562 y ss.

———. *Fiscalía europea y derechos fundamentales*. Valencia: Tirant lo Blanch, 2014.

———. "La garantía de los derechos fundamentales durante la investigación penal". En *Problemas actuales del proceso penal y derechos fundamentales*, Cuadernos Penales José María Lidón, n.º 7, 13-54. Bilbao: Universidad de Deusto, 2010.

———. *La defensa penal*. Madrid: Civitas, 1982.

Pérez Estrada, Miren Josune. *Fundamentos jurídicos para el uso de la inteligencia artificial en los órganos judiciales*. Valencia: Tirant lo Blanch, 2022.

Planchadell Gargallo, Andrea. *El derecho fundamental a ser informado de la acusación*. Valencia: Tirant lo Blanch, 1999.

Ruiz Vadillo, Enrique. "Algunas breves consideraciones sobre el sistema acusatorio y la interdicción constitucional de toda indefensión en el proceso penal". *Diario La Ley*, tomo 4, 1987.

San Miguel Caso, Cristina. "La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales?". *IUS ET SCIENTIA: Revista electrónica de Derecho y Ciencia*, vol. 7, n.º 1 (2021): 286-303.

Serrano Ruiz-Calderón, José Miguel. "El derecho de defensa según la sentencia 79/2012 de la Sala de lo Penal del Tribunal Supremo". *Diario La Ley*, n.º 7850, 3 de mayo de 2012.

Solar Cayón, José Ignacio. "Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos". En *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, editado por José Ignacio Solar Cayón, 125-172. Alcalá de Henares: Universidad de Alcalá / Defensor del Pueblo, 2020.



FGE

FISCALÍA GENERAL DEL ESTADO



ECUADOR

Fiscalía General del Estado
Dirección de Estudios Penales
Quito - Ecuador