Cuadernos de RES PUBLICA en derecho y criminología
Nº en prensa ISSN: 2990-0697 DOI: 10.46661/respublica.13548.

# Ciberdelitos e inteligencia artificial. Desafíos en el ámbito jurídico-penal

## *Cybercrime and artificial intelligence. Challenges in the field of criminal law*

**Celia Sanfrutos Ruiz**
Universidad Pablo de Olavide. Sevilla (España)
csanrui2@alu.upo.es
ORCID: 0009-0007-0152-848X

**Resumen**

Este artículo analiza el impacto de la inteligencia artificial en el ámbito jurídico-penal, prestando especial atención a su incidencia en la configuración, comisión y persecución de los ciberdelitos, así como a los problemas de atribución de responsabilidad derivados de la creciente autonomía de los sistemas inteligentes. Tras una aproximación a la evolución, conceptualización y tipología de la inteligencia artificial, el estudio analiza el marco normativo europeo, con particular referencia al Reglamento (UE) 2024/1689, y el debate doctrinal sobre la posible consideración de estos sistemas como instrumentos del delito o, hipotéticamente, como sujetos responsables. Asimismo, se abordan las principales manifestaciones delictivas vinculadas a su utilización, las dificultades probatorias asociadas y los límites que el respeto a los derechos fundamentales, los sesgos algoritmos y la perspectiva de género imponen a la respuesta jurídico-penal.

Palabras clave: inteligencia artificial, Derecho penal, responsabilidad penal, ciberdelitos, derechos fundamentales, Reglamento de Inteligencia Artificial.

**Abstract**

This article examines the impact of artificial intelligence in the legal-criminal field, paying special attention to its impact on the configuration, commission and prosecution of cybercrimes, as well as the problems of attribution of responsibility arising from the growing autonomy of intelligent systems. After an overview of the evolution, conceptualisation and typology of artificial intelligence, the study analyses the European regulatory framework, with particular reference to Regulation (EU) 2024/1689, and the doctrinal debate on the possible consideration of these systems as instruments of crime or, hypothetically, as responsible subjects. It also addresses the main criminal manifestations linked to their use, the associated evidentiary difficulties and the limits that respect for fundamental rights, algorithmic biases and the gender perspective impose on the legal-criminal response.

Keywords: artificial intelligence, criminal law, criminal liability, cybercrime, fundamental rights, Artificial Intelligence Regulation.

***Cómo citar este trabajo:*** *Sanfrutos Ruiz, Celia*. (2026). Ciberdelitos e inteligencia artificial. Desafíos en el ámbito jurídico-penal. *Cuadernos de RES PUBLICA en derecho y criminología, (8), 01–27.* https://doi.org/10.46661/respublica.13548

**Recepción:** 01.06.2026 **Aceptación:** 18.06.2026 **Publicación:** en prensa

Este trabajo se publica bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

## 1. Introducción

El desarrollo de las tecnologías de la información y la comunicación (en lo sucesivo, TIC) y la progresiva implantación de la inteligencia artificial (en lo sucesivo, IA) han transformado profundamente la sociedad, dando paso a lo que gran parte de la doctrina cataloga como la Cuarta Revolución Industrial. El ciberespacio ha difuminado las barreras geográficas y temporales, creando un nuevo escenario de interacción humana, y, paralelamente, un nuevo ámbito de oportunidad y riesgo para la comisión de ilícitos penales conocido como cibercriminalidad. Hoy en día, la IA ha dejado de ser una mera utopía propia de la ciencia ficción para consolidarse como una tecnología omnipresente que afecta múltiples esferas de la vida cotidiana.

Esta nueva realidad tecnológica supone uno de los mayores desafíos a los que se ha enfrentado el Derecho penal moderno. Tradicionalmente, los ordenamientos penales se han construido sobre un modelo conceptual bipartito que distingue exclusivamente entre "sujetos" (fundamentalmente personas físicas dotadas de voluntad) y "objetos".

Sin embargo, la evolución ha permitido la creación de sistemas informáticos provistos de redes neuronales, algoritmos de aprendizaje (*machine learning* o *deep learning*) y niveles de autonomía funcional que les permiten procesar datos, aprender de su entorno y tomar decisiones con un margen de independencia de su programador. Esta capacidad de emular procesos cognitivos cuestiona el esquema tradicional e introduce la urgente necesidad de reformar las estructuras jurídico-penales.

Ante esta disrupción, el debate en la dogmática penal gira en torno a cómo atribuir la responsabilidad por los resultados lesivos generados mediante el uso de IA. Por un lado, se examina el uso de los sistemas inteligentes como meras herramientas o instrumentos al servicio del ser humano para la comisión de delitos, lo cual encaja de forma más evidente en las categorías de autoría directa o mediata de la persona física. Por otro lado, la aparición de una "IA fuerte" o autónoma abre la posibilidad de considerar a estos sistemas como "actuantes" dentro de la dinámica delictiva o, desde posiciones doctrinales más extremas, como eventuales autores directos del delito. Ello plantea tensiones en relación con el principio de culpabilidad y con los elementos subjetivos de la teoría del delito.

En la práctica, la IA actúa como un vector que facilita y multiplica el impacto lesivo de los ciberdelitos. En el ámbito de la cibercriminalidad económica, dota de un altísimo nivel de engaño y escalabilidad al fraude algorítmico, el *phishing* y el robo de datos. En la esfera personal, los ciberdelitos intrusivos han encontrado en tecnologías como los *deepfakes* un instrumento idóneo para la cosificación digital, afectando de manera desproporcionada los derechos al honor, a la propia imagen y la indemnidad sexual. Todo ello se desarrolla en un entorno digital que dificulta las labores de investigación y la obtención de pruebas electrónicas con plenas garantías de autenticidad e integridad.

Finalmente, el estudio del fenómeno exige contemplar la colisión de esta tecnología con el Estado social y democrático de Derecho. El uso de sistemas algorítmicos entraña riesgos éticos y jurídicos, como el efecto "caja negra", las agresiones a la privacidad y el peligro de producir sesgos que afectan especialmente a colectivos vulnerables y perpetúa brechas de género. Por ello, se erige como un hito normativo relevante la aprobación del Reglamento de Inteligencia Artificial que establece un marco jurídico escalonado basado en el riesgo, instaurando prohibiciones y garantías para asegurar que el desarrollo de la IA se centre en el bienestar y la dignidad del ser humano.

Por todo lo expuesto, el presente trabajo surge de la necesidad de examinar la compleja intersección entre la inteligencia artificial y el Derecho penal. Resulta imprescindible

analizar si las categorías dogmáticas vigentes son suficientes para dar respuesta a los ilícitos vinculados a los sistemas inteligentes y determinar los límites éticos y legales que deben regir la penetración tecnológica en la justicia, asegurando que la innovación no merme las garantías constitucionales.

## 2. Evolución y desarrollo de la inteligencia artificial

El propósito fundamental de la IA es reproducir procesos cognitivos humanos como el aprendizaje, el lenguaje, la percepción o la creatividad, valiéndose de herramientas proporcionadas por la computación, especialmente algoritmos que "aprenden" a través del procesamiento de datos. Es, en esencia, el intento de comprender e imitar la inteligencia humana, con "*el objeto de que los ordenadores hagan la misma clase de cosas que puede hacer la mente*"[1].

La historia de la IA tiene sus orígenes a mediados del siglo XX, en el contexto científico posterior a la Segunda Guerra Mundial, cuando figuras pioneras como Alan Turing y John McCarthy sentaron las bases teóricas de esta disciplina. En particular, Turing, en su célebre artículo "*Computing Machinery and Intelligence*", planteó por primera vez la posibilidad de que una máquina pudiera pensar. En este mismo escenario formuló el denominado "test de Turing", experimento consistente en la interacción entre un interrogador humano y dos interlocutores ocultos -una persona y una máquina-, de modo que, si el interrogador no lograba distinguir entre ambos a partir de sus respuestas, cabría concluir que la máquina manifestaba un comportamiento inteligente equiparable al humano[2]. Este planteamiento supuso el punto de partida del debate contemporáneo sobre la naturaleza y alcance de la IA.

Posteriormente, el término "inteligencia artificial" fue acuñado en 1956 por John McCarthy durante un seminario de la Universidad de Dartmouth (Estados Unidos), hito que marcó el inicio formal de esta disciplina científica. Desde entonces, múltiples autores han intentado definir la IA, destacando la concepción de Marvin Minsky, quien la describió como "*la ciencia de hacer que las máquinas hagan cosas que requerían de inteligencia si fueran hechas por los hombres*".

A lo largo de las décadas siguientes, el progreso científico-técnico permitió el desarrollo de sistemas cada vez más sofisticados, capaces de asumir funciones tradicionalmente reservadas a las personas, especialmente aquellas que exigen elevados niveles de cálculo, dedicación temporal o procesamiento de información. Un acontecimiento paradigmático en esta evolución tuvo lugar en 1997, cuando la supercomputadora *Deep Blue*, desarrollada por IBM, logró imponerse al campeón mundial de ajedrez Garry Kaspárov, evidenciando el tránsito de la IA desde la especulación teórica y la ciencia ficción hacia su aplicación práctica efectiva.

Si bien durante las primeras décadas de desarrollo de la inteligencia artificial se formularon predicciones optimistas según las cuales las máquinas llegarían a desempeñar funciones equiparables a las humanas en un corto plazo temporal, tales expectativas no se materializaron plenamente hasta bien entrado el siglo XXI. La consolidación definitiva de la IA se ha visto impulsada por tres factores: la disponibilidad masiva de datos (*Big Data*), el acceso a procesadores de alta capacidad a bajo coste y el desarrollo de

[1] BODEN, M. A., *Inteligencia Artificial*, Editorial Turner Publicaciones, Madrid, 2017, p. 11.

[2] GUTIÉRREZ PALACIO, J. D., "Concepto jurídico-penal de acción a partir de la inteligencia artificial", en *Estudios de Derecho penal, neurociencias e inteligencia artificial*, DEMETRIO CRESPO, E., CARO CORIA, D. C., ESCOBAR BRAVO, M. E. (eds.). Ediciones de la Universidad de Castilla-La Mancha, 2023, p. 69.

redes neuronales profundas. Este contexto ha propiciado un crecimiento tecnológico de carácter exponencial, cuya manifestación más visible se encuentra en la denominada "inteligencia artificial generativa" (en lo sucesivo, IAG).

En los últimos años, dicha evolución se ha intensificado hasta alcanzar un desarrollo tecnológico de carácter vertiginoso, reflejado paradigmáticamente en el lanzamiento de ChatGPT en noviembre de 2022 por parte de OpenAI, que logró alcanzar la cifra de un millón de usuarios en apenas cinco días y superando los cien millones en tan solo dos meses. Este acontecimiento ilustra el potencial expansivo de la IAG. Todos estos sistemas se fundamentan en algoritmos de aprendizaje profundo (*deep learning*), capaces de producir resultados originales a partir de instrucciones textuales proporcionadas por los usuarios, comúnmente denominadas "*prompts*", lo que evidencia el grado de sofisticación alcanzado por la IA en la actualidad y su creciente impacto social, económico y jurídico.

La rapidez de estos avances ha intensificado la tradicional tensión entre innovación tecnológica y capacidad de respuesta del Derecho, frecuentemente descrita bajo la metáfora de la "liebre informática" frente a la "tortuga jurídica". En efecto, la realidad tecnológica suele anticiparse a la regulación normativa, generando dilemas éticos, sociales y jurídicos cada vez más complejos. Entre ellos destacan las cuestiones relativas a la atribución de responsabilidad, la protección de bienes jurídicos en entornos digitales y la necesidad de replantear categorías dogmáticas tradicionales, especialmente a la luz de aportaciones provenientes de la neurociencia que cuestionan concepciones clásicas sobre la voluntad y la toma de decisiones humanas.

### 2.1. Conceptualización y tipología.

La conceptualización de la inteligencia artificial presenta notables dificultades, derivadas tanto de la falta de una definición universalmente aceptada como de la propia complejidad del concepto de inteligencia. El Comité Económico y Social Europeo, en su "Dictamen sobre la inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad"[3], ha reconocido expresamente la inexistencia de una definición concreta y consensuada de inteligencia artificial, lo que explica la pluralidad de aproximaciones doctrinales existentes.

Desde una perspectiva preliminar, y con independencia de las dificultades inherentes a la delimitación conceptual de la inteligencia, resulta adecuado partir de las habilidades básicas que la caracterizan. En efecto, como señala De la Cuesta Aguado[4], el concepto de inteligencia -tradicionalmente atribuido al ser humano- dista de ser pacífico tanto a lo relativo a su contenido como a su extensión. Sobre esta base, pueden identificarse como capacidades esenciales la recepción de información y la comunicación con otros agentes, su comprensión en función de objetivos, su almacenamiento, su utilización en la resolución de problemas y, finalmente, la adopción de decisiones.

La determinación de estas habilidades adquiere especial relevancia desde la perspectiva jurídica, en la medida en que permite precisar cuándo, o con qué grado de desarrollo cognitivo, podría superarse el denominado "umbral de responsabilidad", entendido como aquel punto a partir del cual

[3] COMITÉ ECONÓMICO Y SOCIAL EUROPEO: *Dictamen del Comité Económico y Social Europeo "Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad"*, Diario Oficial de la Unión Europea C 288/1, 31 de agosto de 2017, p. 3.

[4] DE LA CUESTA AGUADO, P. M., "Inteligencia artificial y responsabilidad penal", en *Revista Penal México*, núms. 16 y 17, 2019, p. 52.

las decisiones adoptadas por un agente artificial podrían llegar a serle jurídicamente imputables.

En cuanto a su definición, el Diccionario de la Real Academia Española entiende la inteligencia artificial como "*la disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico*". John McCarthy la definió como "*la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes*", mientras que Minsky la describió como "*la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran las personas*"[5].

Estas concepciones permiten entender la inteligencia artificial tanto como una rama de la ciencia de la computación dirigida al diseño de agentes inteligentes[6] como la propia inteligencia manifestada por dichos agentes, entendidos como sistemas capaces de evaluar su entorno y adoptar decisiones orientadas al éxito.

La doctrina también ha propuesto diversas tipologías de inteligencia artificial. Siguiendo la clasificación expuesta por Barona Vilar[7] a partir de Russell y Norvig, pueden distinguirse:

1) Sistemas que imitan el pensamiento humano y poseen capacidad de aprendizaje y resolución autónoma de problemas.
2) Sistemas que reproducen el comportamiento humano.
3) Sistemas basados en el razonamiento lógico racional.
4) Sistemas que emulan una conducta racional mediante estructuras propias de los sistemas expertos.

Desde otra perspectiva, Hintze[8] diferencia entre máquinas reactivas sin memoria, sistemas con memoria limitada capaces de almacenar información pasada, sistemas dotados de una teoría de la mente potencialmente capaces de comprender emociones e interactuar socialmente, y, en un estadio hipotético, máquinas con conciencia propia capaces de representarse a sí mismas y anticipar estados ajenos.

En términos teleológicos, el objetivo fundamental de la inteligencia artificial ha sido descrito como "*la automatización de comportamientos inteligentes tales como razonar, recabar información, planificar, aprender, comunicar, manipular, observar e incluso crear, soñar y percibir*"[9]. En una línea similar, Bourcier[10] señala que esta disciplina persigue tanto utilizar la máquina para probar las funciones cognitivas humanas como reproducir los mecanismos propios del pensamiento.

Finalmente, desde el plano normativo europeo, el Reglamento de Inteligencia Artificial[11] (en lo sucesivo, Reglamento de IA o RIA), define estos sistemas como "*un sistema basado en una máquina que está diseñado*

[5] HERNÁNDEZ GIMÉNEZ, M. "Inteligencia artificial y derecho penal", *Actualidad jurídica Iberoamericana*, N.º 10 bis, 2019, pp. 795.

[6] El agente inteligente se ha definido como aquel capaz de evaluar las circunstancias y condiciones de su entorno para adoptar decisiones que maximizan sus posibilidades de éxito (POOLE, D., MACKWORTH, A., y GOEBEL, R., "Computational inteligente: a logical approach", *Oxford University Press*, 1998, p. 7).

[7] BARONA VILAR, S., "Reflexiones en torno al 4.0 y la inteligencia artificial en el proceso penal", *Ius Puniendi*, vol. 7, 2018, pp. 313-336.

[8] HINTZE, A., "Understanding the four types of AI, from reactive robots to self-aware beings", *The Conversation*, 2016.

[9] COMITÉ ECONÓMICO Y SOCIAL EUROPEO, *op. cit.*, p.3.

[10] BOURCIER, D, "Inteligencia artificial y derecho", 1ª ed., Editorial UOC, Barcelona, 2003, p. 70.

[11] REGLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO Y DEL CONSEJO, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858,

*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*". Esta definición normativa resulta especialmente relevante por constituir el punto de conexión entre la conceptualización técnica de la inteligencia artificial y sus consecuencias jurídicas.

### 2.2. Marco normativo europeo: especial referencia al Reglamento de Inteligencia Artificial.

El acelerado desarrollo de la inteligencia artificial y su creciente impacto económico, social y jurídico han motivado una respuesta regulatoria progresiva en el ámbito de la Unión Europea (en lo sucesivo, UE). Las instituciones europeas han asumido un papel central en la construcción de un marco normativo destinado a compatibilizar la innovación tecnológica con la protección de los derechos fundamentales, la seguridad jurídica y el correcto funcionamiento del mercado interior. En este contexto se inscribe la aprobación del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial, primera norma de alcance general orientada a establecer un régimen jurídico integral para los sistemas de IA.

Conforme a su artículo primero, el objetivo principal del Reglamento de IA consiste en mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme que evite la heterogeneidad normativa en el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial en la Unión Europea[12]. Dicha norma promueve así una IA centrada en el ser humano y fiable, alineada con los valores consagrados en la Carta de Derechos Fundamentales de la UE, garantizando la protección de las personas físicas, las empresas, la democracia, el Estado de Derecho y la protección del medio ambiente, al tiempo que impulsa la innovación tecnológica.

Para alcanzar estos fines, adopta un enfoque preventivo basado en el riesgo, adaptando la intensidad de las obligaciones al alcance de los riesgos que puedan generar los sistemas de IA. En coherencia con ello, prohíbe determinadas prácticas inaceptables, establece requisitos estrictos para los sistemas de alto riesgo, impone obligaciones de transparencia y define las responsabilidades de los distintos operadores que intervienen en su desarrollo y utilización.

Las prácticas prohibidas, reguladas en el Capítulo II, y en particular en su artículo 5, comprenden, entre otras, el uso de técnicas subliminales destinadas a alterar significativamente el comportamiento de las personas; la explotación de situaciones de vulnerabilidad derivadas de la edad, discapacidad o situación socioeconómica; los sistemas de evaluación o clasificación sociales que generen tratos perjudiciales o desproporcionados; la evaluación del riesgo de comisión delictiva basada exclusivamente en rasgos de personalidad; la creación masiva de bases de datos de reconocimiento facial mediante extracción indiscriminada de imágenes; la inferencia de emociones en entornos laborales o educativos -salvo por razones médicas o de seguridad-; la categorización biométrica basada en datos

(UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), Diario Oficial de la Unión Europea, L 1689, de 12 de julio de 2024, pp. 46.

[12] HERNÁNDEZ LÓPEZ, J. M., *Reglamento de Inteligencia Artificial. Incluye introducción, notas, cronología, webgrafía, bibliografía e índice analítico*, 1ª ed., J. M. Bosch Editor, Barcelona, 2024, pp. 17-34.

sensibles; y la identificación biométrica en tiempo real en espacios públicos para fines de aplicación de la ley, salvo supuestos excepcionales estrictamente delimitados.

Por su parte, los sistemas de alto riesgo, regulados en el Capítulo III, son aquellos susceptibles de afectar de forma significativa a la salud, la seguridad o los derechos fundamentales. Tales sistemas -entre los que se incluyen, por ejemplo, los utilizados en identificación biométrica, educación, empleo o acceso a servicios esenciales- quedan sometidos a rigurosos requisitos de fiabilidad, gobernanza de datos, documentación técnica, supervisión humana y evaluación de conformidad previa a su introducción en el mercado o puesta en servicio. Junto a esta categoría, el Reglamento contempla sistemas de riesgo limitado, sujetos principalmente a deberes de transparencia, como la obligación de informar a los usuarios de que interactúan con IA, supuesto paradigmático en el caso de los chatbots o contenidos sintéticos (*deepfakes*); así como sistemas de riesgo mínimo, que en principio quedan fuera de regulación específica, si bien la expansión de la inteligencia artificial generativa está modificando progresivamente esta clasificación.

El Capítulo V introduce una regulación específica para los modelos de inteligencia artificial de propósito general (en lo sucesivo, GPAI), caracterizados por su elevada generalidad funcional y capacidad de integración en múltiples aplicaciones. Estos modelos pueden generar riesgos sistémicos con efectos negativos sobre la seguridad pública, los procesos democráticos o la difusión de contenidos ilícitos o discriminatorios, lo que justifica la imposición de obligaciones reforzadas de documentación, transparencia y cumplimiento del Derecho de la UE, especialmente en materia de propiedad intelectual. En particular, los proveedores de modelos GPAI deben proporcionar documentación técnica e instrucciones de uso adecuadas, cumplir la normativa europea y publicar resúmenes suficientemente detallados de los datos empleados en su entrenamiento.

El Reglamento define con precisión las distintas categorías de operadores que intervienen en la cadena de valor de la IA -proveedor, responsable del despliegue, representante autorizado, importador, distribuidor y operador- y determina las responsabilidades correspondientes a cada uno de ellos conforme a la distribución de las obligaciones prevista en su Capítulo IV. Junto a esta delimitación subjetiva de deberes, el Capítulo VI incorpora medidas de apoyo a la innovación, como la creación de entornos controlados de pruebas regulatorias, concebidos para favorecer el desarrollo tecnológico en condiciones de seguridad jurídica.

Por su parte, el Capítulo VII establece una estructura integral de gobernanza europea, que incluye la creación de una Oficina de Inteligencia Artificial en el seno de la Comisión, un Consejo Europeo de Inteligencia Artificial, un grupo de expertos científicos independientes y un foro consultivo de partes interesadas, todos ellos orientados a asegurar una aplicación coherente, coordinada y eficaz del Reglamento.

Asimismo, los Capítulos VIII a X regulan instrumentos esenciales para la supervisión y el control del mercado, tales como las bases de datos europeas de sistemas de alto riesgo, los mecanismos de vigilancia poscomercialización, el intercambio de información entre autoridades competentes y las facultades de supervisión administrativa. Finalmente, el Capítulo XI establece un régimen sancionador significativo que prevé la imposición de multas administrativas de elevada cuantía y otras medidas de ejecución frente a las infracciones cometidas por los operadores, con especial severidad en los supuestos de incumplimiento de las prácticas prohibidas.

El Reglamento refuerza igualmente la conexión entre inteligencia artificial y protección de datos personales, exigiendo el

respeto de los principios de minimización de datos y protección desde el diseño y por defecto a lo largo de todo el ciclo de vida del sistema, sin perjuicio de las competencias de las autoridades de protección de datos. De igual modo, incorpora referencias expresas al régimen europeo de derechos de autor, imponiendo a los proveedores de modelos de IA de propósito general obligaciones de cumplimiento normativo y de transparencia respecto de los datos utilizados para el entrenamiento.

Aunque el Reglamento de IA no configura directamente un sistema de responsabilidad criminal, su relevancia resulta incuestionable. En primer lugar, porque define jurídicamente el concepto de sistema de inteligencia artificial, proporcionando un marco interpretativo común para el análisis de conductas vinculadas a estas tecnologías. En segundo lugar, porque las obligaciones de diligencia, control y prevención que impone pueden incidir en la determinación de posiciones de garante, deberes de cuidado y criterios de imputación penal en supuestos de hechos delictivos cometidos mediante sistemas inteligentes. De este modo, el marco regulatorio europeo se erige en presupuesto esencial para el análisis de la responsabilidad penal derivada del uso de la IA, cuestión que será objeto de estudio en el capítulo siguiente.

## 3. Inteligencia artificial y responsabilidad penal

La irrupción de la inteligencia artificial ha transformado de manera profunda la vida en sociedad del ser humano, hasta el punto de que numerosos autores la sitúan al nivel, e incluso por encima, de la Revolución Industrial[13]. En la actualidad, se asiste a lo que se ha catalogado como Cuarta Revolución Industrial, caracterizada por la convergencia de algoritmos, una capacidad informática sin precedentes y la gestión de volúmenes masivos de datos. Este desarrollo tecnológico no solo promete importantes beneficios, sino que también plantea inmensos desafíos y riesgos que el Derecho no puede ignorar. Ante la posibilidad de que estas tecnologías afecten bienes jurídicos protegidos, resulta indispensable revisar los presupuestos tradicionales de la responsabilidad penal.

Históricamente, la doctrina mayoritaria ha partido de la premisa antropocéntrica de que solo los seres humanos pueden cometer delitos, centrando el análisis jurídico-penal exclusivamente en conductas humanas. El sistema penal se ha construido sobre una dicotomía estricta que distingue únicamente entre sujetos, capaces de tomar decisiones propias y responsables, y objetos, que carecen de dicha capacidad[14].

Sin embargo, el progreso tecnológico ha impulsado de manera significativa la creación de máquinas capaces de realizar funciones que tradicionalmente se consideraban privativas de los seres humanos. Desde la formulación del test de Turing, la evolución de la inteligencia artificial ha estado orientada no solo a reproducir comportamientos humanos, sino incluso a superarlos en determinados ámbitos.

Para comprender las implicaciones penales de esta tecnología, la doctrina ha establecido una distinción fundamental entre inteligencia artificial débil, inteligencia artificial fuerte e inteligencia artificial superinteligente[15][16]. La

[13] BLANCO CORDERO, I., "*Homo Sapiens* y ¿*Machina Sapiens*?: Un derecho penal para los robots dotados de inteligencia artificial", en MALLADA FERNÁNDEZ, C. (coord.), *Nuevos retos de la ciberseguridad en un contexto cambiante*, Madrid, 2019, p. 63.

[14] NAVARRO-DOLMESTCH, R., "Inteligencia artificial como «actuante» en el derecho penal. Una primera aproximación", *Revista de Internet, Derecho y Política*, n.º 43, 2025, p. 2.

[15] LÓPEZ DE MÁNTARAS, R., "El futuro de la IA: hacia inteligencias artificiales realmente inteligentes", *¿Hacia una nueva Ilustración? Una década trascendente*, 2019, p. 162.

[16] AGUILAR CAMPOS, P., y ALÉ MARTÍNEZ, V., "Responsabilidad penal en la era de la inteligencia artificial: De la agencia humana a la autonomía de la

IA débil hace referencia a sistemas diseñados para ejecutar tareas concretas y específicas, en ocasiones con mayor eficacia que los seres humanos, pero carentes de conciencia o estados mentales propios. Por el contrario, la IA fuerte, también denominada en la doctrina comparada como *machina sapiens*, sostiene la posibilidad de que las máquinas no solo simulen la mente humana, sino que operen de forma análoga a ésta, pudiendo llegar a desarrollar estados mentales.

Finalmente, la IA superinteligente haría referencia a entidades conscientes de sí mismas que superarían ampliamente las habilidades humanas, un escenario que, por el momento, pertenece al ámbito especulativo. Esta distinción adquiere especial relevancia en el ámbito jurídico-penal, en la medida en que la utilización de sistemas de IA débil remite, en principio, a supuestos de instrumentalización de un *software* por parte de un sujeto humano, mientras que la aparición de sistemas de IA fuerte, con un mayor grado de autonomía decisional, suscita el debate acerca de la posible atribución directa de responsabilidad penal.

La posibilidad de que las máquinas adopten decisiones de forma semejante a los seres humanos ha llevado a cuestionar los límites tradicionales de la responsabilidad penal. En este contexto, el desarrollo de las teorías neurocientíficas aplicadas al Derecho ha reabierto el debate en torno a conceptos clásicos como la voluntad y la toma de decisiones, cuestionando la idea de un control plenamente consciente de la conducta. Algunas corrientes sostienen que determinadas decisiones conductuales responden a procesos neurológicos previos a la conciencia, lo que permite establecer, al menos en un plano teórico, paralelismos con los procesos de decisión de los sistemas artificiales[17]. En esta línea, ciertos enfoques funcionalistas permiten entender tanto la mente humana como los sistemas de IA desde la lógica del procesamiento de información, lo que facilita su comparación estructural. Si bien esta aproximación no elimina una diferencia esencial, los sistemas de IA son creaciones humanas, lo que condiciona decisivamente cualquier intento de atribución de responsabilidad.

Como señala Benítez Ortúzar[18], estos sistemas actúan de manera similar a la mente humana en tanto que procesan información y generan respuestas, pero no dejan de ser productos diseñados por el propio ser humano. Esta dualidad pone de manifiesto la necesidad de determinar si las actuaciones de la inteligencia artificial pueden generar consecuencias jurídicas autónomas o si, por el contrario, deben reconducirse a la actuación de los sujetos que intervienen en su desarrollo y utilización.

El funcionamiento técnico de la inteligencia artificial -especialmente en su vertiente generativa- y de los algoritmos de aprendizaje profundo añade capas de complejidad al análisis de la responsabilidad. En términos simplificados, el ciclo de vida de estos sistemas involucra tres fases en las que pueden producirse afectaciones a bienes jurídicos protegidos: la recolección de datos, su procesamiento y la interpretación de los resultados. En un primer momento, los desarrolladores diseñan el *software*, que actúa como un sistema normativo interno compuesto por instrucciones, reglas de conducta, mandatos y prohibiciones[19]. Posteriormente, se introducen datos masivos para entrenar al sistema, permitiéndole identificar patrones mediante técnicas de *machine learning* y *deep learning*. Es en la capacidad de adaptar su comportamiento a

*machina sapiens*", *Revista de Estudios de la Justicia*, núm. 42, 2025, pp. 156-158.

[17] GUTIÉRREZ PALACIO, J. D., *op. cit.*, pp. 69-70.

[18] BENÍTEZ ORTÚZAR, I., LLEDÓ YAGÜE, F., y MONJE BALMASEDA, O. (dirs.), "La robótica y la inteligencia artificial en la nueva era de la Revolución Industrial 4.0", *Colección Inteligencia Artificial, Robots y Bioderecho*, Madrid, 2021, pp. 1-672.

[19] DE LA CUESTA AGUADO, P. M., *op. cit.*, p. 55-57.

entornos cambiantes, de manera imprevisible y más allá de los parámetros programados, donde la IA cuestiona los esquemas clásicos de imputación penal[20].

A ello se añade la pluralidad de sujetos que intervienen en el ciclo de vida de la IA, lo que incrementa la complejidad de la imputación penal. Los errores pueden originarse en distintas etapas y ser atribuibles a desarrolladores, programadores, operadores o usuarios finales, lo que dificulta la delimitación de responsabilidades. Esta realidad ha llevado a la doctrina a proponer distintos modelos explicativos. En este sentido, Quintero Olivares distingue diversos supuestos en función del grado de control y conocimiento sobre el comportamiento del sistema[21], mientras que Hallevy formula tres modelos teóricos de referencia: (i) la perpetración por medio de otro; (ii) el modelo de consecuencia natural probable; y (iii) el modelo de responsabilidad directa[22].

La relevancia práctica de estas cuestiones se evidencia en diversos supuestos en los que el funcionamiento de sistemas automatizados ha generado resultados lesivos. El accidente ocurrido en 2016 con un vehículo autónomo que no detectó un camión, o el caso de un robot industrial que provocó la muerte de un trabajador al confundirlo con un objeto de manipulación, ilustran los riesgos asociados a estos sistemas y ponen de manifiesto las dificultades para determinar la responsabilidad penal en estos escenarios. Ante la inexistencia de un Derecho penal específico aplicable a la IA, la respuesta debe articularse a partir de las categorías tradicionales del ordenamiento jurídico, que permiten atribuir responsabilidad por hechos cometidos mediante estos sistemas, tanto a personas físicas como jurídicas.

A partir de todo lo anterior, el debate puede estructurarse en torno a dos grandes posibilidades. Por un lado, la eventual consideración de los sistemas de IA como autores directos del delito. Por otro lado, su concepción como meros instrumentos utilizados por personas para la comisión de conductas delictivas. Esta tensión entre ambos planteamientos constituye el eje central del análisis que se desarrollará en los siguientes apartados.

### 3.1. Posible consideración de los sistemas de inteligencia artificial como autores directos del delito

El debate en torno a la atribución de responsabilidad directa a los sistemas de IA ha cobrado una especial intensidad a medida que se desarrollan agentes artificiales capaces de operar con un elevado grado de autonomía. Durante mucho tiempo, la idea de castigar a una máquina fue considerada una hipótesis puramente especulativa o propia de la ciencia ficción, pero el avance exponencial de la tecnología, amparado en la Ley de Moore, ha trasladado este planteamiento a un plano de realidad inminente. La cuestión central radica en determinar si la conducta desplegada por un sistema autónomo puede subsumirse en las categorías tradicionales de la Teoría del Delito, esto es, si una máquina puede realizar una acción típica, antijurídica y culpable[23].

El primer obstáculo reside en el concepto de acción que, tradicionalmente, exige una manifestación de la voluntad humana. Diversas teorías han intentado explicar este elemento estructural; mientras la teoría causalista se enfocaba en la inervación

[20] ROMEO CASABONA, C. M., y RUEDA MARTÍN, M. A. (eds.), "Derecho Penal, Ciberseguridad, Ciberdelitos e Inteligencia artificial", en *Estudios de Derecho Penal y Criminología*, Editorial Comares, Granada, 2023, p. 17.

[21] QUINTERO OLIVARES, G., "La robótica ante el derecho penal: el vacío de respuesta jurídica a las desviaciones incontroladas", *Revista Electrónica de Estudios Penales y de la Seguridad*, 2017, pp. 1-23.

[22] HALLEVY, G., "The Basic Models of Criminal Liability of AI Systems and Outer Circles", *SSRN Electronic Journal*, 2019, pp. 1-15.

[23] AGUILAR CAMPOS, P., y ALÉ MARTÍNEZ, V., *op cit*., pp. 170-172.

muscular y la teoría finalista en la dirección consciente hacia un objetivo, la teoría social introdujo la valoración de la conducta dentro del contexto social. No obstante, todas estas construcciones comparten un requisito antropocéntrico indispensable: la intervención humana. Para dar cabida a los actos de la IA, parte de la doctrina sugiere acudir a perspectivas funcionalistas de la mente, las cuales argumentan que los estados mentales son conjuntos de disposiciones relacionadas mediante entradas sensoriales y salidas conductuales, una estructura funcionalmente replicable por sistemas computacionales avanzados[24].

Desde el ámbito estrictamente jurídico-penal, algunos autores abogan por una reinterpretación de la "acción comunicativa"[25]. Bajo este prisma, el Derecho penal debe centrarse exclusivamente en aquellas acciones que puedan interpretarse como ataques a bienes jurídicos y que sean socialmente inadecuadas o significativas. Al suprimir el requisito estrictamente humano de la relevancia de la conducta, se abre la puerta para que un comportamiento desplegado por un agente no humano, capaz de dirigir su conducta en función de objetivos dentro de una práctica social, sea catalogado como acción penal.

Surgen dificultades aún mayores al examinar el elemento interno del delito, tradicionalmente vinculado a la conciencia, el dolo y la culpabilidad. Desde el ámbito del Derecho anglosajón o *common law*, autores como Hallevy proponen el modelo de responsabilidad directa, argumentando que una *machina sapiens* futura podría satisfacer los requisitos del *actus reus* y el *mens rea*[26]. Según esta postura, un sistema avanzado posee "conciencia" al absorber datos del mundo exterior a través de sensores y procesarlos en su mente artificial, y posee "voluntad" al adoptar decisiones que pueden encuadrarse en la intencionalidad, la indiferencia o la temeridad. Así, se afirma teóricamente que una IA fuerte podría obrar con una mente malintencionada y manos malhechoras, asimilando sus procesos a los cognitivos y volitivos requeridos por el Derecho.

Sin embargo, la atribución de la culpabilidad a una máquina genera un profundo rechazo en el sector mayoritario. Autores contemporáneos subrayan que los sistemas carecen de conciencia moral, empatía, motivaciones genuinas y de la capacidad de comprender el carácter ilícito de sus acciones, limitándose a ejecutar operaciones matemáticas y de procesamiento de datos. No obstante, frente a esta objeción, se ha propuesto utilizar la responsabilidad penal de las personas jurídicas como puente dogmático y base funcional[27].

Si el Derecho penal ha admitido que una corporación, entidad ficticia sin conciencia ni voluntad psicológica, puede ser considerada culpable a partir de un juicio normativo sobre su defecto de organización y su incapacidad de autorregulación, un razonamiento análogo podría aplicarse a la inteligencia artificial. Bajo este modelo, la culpabilidad se disocia de su contenido psicológico tradicional y se transforma en una valoración estructural sobre la agencia artificial, sancionando su operatividad disfuncional o antisocial en el entorno.

Incluso si se lograse sortear el obstáculo de la culpabilidad, la atribución de autoría directa a la IA enfrenta la dificultad de la punibilidad. La justificación del castigo penal gira habitualmente sobre fines de retribución, prevención o resocialización. La doctrina crítica sostiene que resulta inútil castigar a un ente que no experimenta sufrimiento, miedo o dolor, vaciando de sentido las teorías

[24] GUTIÉRREZ PALACIO, J. D., *op. cit.*, pp. 70-72.

[25] ibidem pp. 73-74.

[26] AGUILAR CAMPOS, P., y ALÉ MARTÍNEZ, V., *op. cit.*, pp. 174-175.

[27] ibidem. 169-170.

retributivas e intimidatorias. Para solucionar esta brecha, se han planteado esquemas punitivos adaptados a la naturaleza electrónica. Se propone la eliminación permanente del *software* o la destrucción del robot, la inmovilización funcional o el bloqueo o restricción de sus actividades, la reescritura de algoritmos o su reprogramación forzosa y la incautación de los activos generados por la máquina, entre otros.

A pesar de estas construcciones teóricas y del respaldo en resoluciones internacionales que sugieren la creación de una "personalidad electrónica" para los robots más avanzados a efectos de gestionar su responsabilidad, la doctrina penal dominante insiste en que, en el estado tecnológico actual, los sistemas de IA no son verdaderos agentes autónomos. Por espectaculares que sean sus capacidades de procesamiento, operan en última instancia condicionados por los fines y límites previamente establecidos por sus desarrolladores humanos, por lo que cualquier atribución de autoría directa en el presente carece de fundamentación y difumina el sentido antropocéntrico del Derecho.

Ante la opacidad de los algoritmos y la imprevisibilidad emergente de determinados sistemas, la respuesta no debe consistir en personificar a la máquina, sino en dirigir la imputación hacia los sujetos que la diseñan, implementan o instrumentalizan en el entorno social.

### 3.2. Los sistemas de inteligencia artificial como instrumentos para la comisión de delitos

Al descartarse, por el momento, la imputación de responsabilidad directa a los propios sistemas tecnológicos, el foco del análisis se dirige forzosamente hacia los agentes humanos que intervienen en su diseño, fabricación, distribución y uso. La inteligencia artificial deja de ser analizada como un sujeto culpable para pasar a ser valorada como un medio o instrumento empleado para lesionar o poner en peligro los bienes jurídicos protegidos.

Para organizar teóricamente esta problemática, la doctrina comparada ha formulado diversos modelos de atribución. Hallevy postula dos modelos que encuadran la instrumentalización de estas tecnologías: el modelo de perpetración por medio de otro, en el que el ser humano se sirve deliberadamente de la IA como un mero instrumento para ejecutar el delito; y el modelo de consecuencia natural probable, aplicable a aquellos programadores o usuarios que ponen en funcionamiento un sistema asumiendo la producción de un resultado lesivo que era previsible y natural derivado de dicha activación.

Por su parte, Quintero Olivares, sistematiza estas interacciones en cuatro escenarios concretos: (i) el uso netamente doloso de la máquina para delinquir; (ii) los casos donde la tecnología se desvía de su programación original y el sujeto conoce y acepta dicho desvío; (iii) las situaciones de desviación inesperada donde se desconoce la causa del fallo; y (iv) la utilización de sistemas expresamente prohibidos por el ordenamiento[28].

El primer y más claro supuesto de responsabilidad humana surge de la utilización directa y dolosa de la IA como medio comisivo. Estos supuestos no plantean especiales dificultades desde la perspectiva de la autoría directa. El programador que diseña un *software* malicioso dotado de aprendizaje automático para vaciar cuentas bancarias, el usuario que crea un contenido *deepfake* hiperrealista superponiendo rostros para extorsionar o dañar el honor de terceros, o el desarrollador que programa enjambres de *bots* para manipular masivamente el valor de unas acciones en el mercado bursátil, responden como autores inmediatos. En estos

[28] QUINTERO OLIVARES, G., *op. cit.*, pp. 14 y 22.

supuestos, el sistema inteligente actúa materialmente como el vehículo a través del cual la persona despliega un comportamiento delictivo que encaja perfectamente en los tipos penales vigentes, del mismo modo en que tradicionalmente se emplearía una herramienta física.

Las peculiaridades tecnológicas, sin embargo, fuerzan a los sistemas normativos a replantearse la utilidad de las categorías clásicas de participación cuando la comisión del ilícito no es inmediata. En contextos donde la creación y el uso de la tecnología se disocian, cobran relevancia los títulos de inducción, complicidad y, muy especialmente, la coautoría y la autoría mediata. La complicidad se configura cuando un fabricante diseña una herramienta algorítmica capaz de vulnerar sistemas de autenticación y la facilita a un tercero que consuma un fraude informático, aportando el desarrollador un instrumento que incrementa decisivamente las probabilidades de realización típica, aun sin existir un plan delictivo común.

La inducción tiene lugar si el experto tecnológico persuade a un tercero para que utilice un sistema inteligente en la comisión de un delito, instruyéndole en sus capacidades y garantizando la viabilidad del ilícito. Asimismo, la coautoría se afirma en aquellos casos de cibercriminalidad donde el ingeniero que diseña el *malware* actúa de manera coordinada con quienes lo distribuyen, respondiendo todos bajo un plan conjunto.

Especial tensión genera la figura de la autoría mediata cuando el instrumento empleado es un sistema de IA. La concepción clásica exige inexcusablemente que el ejecutor material, es decir, el instrumento interpuesto entre el autor mediato y el resultado sea otra persona física que actúe con un déficit de responsabilidad, ya sea por encontrarse inmersa en error, obrar bajo coacción o carecer de imputabilidad. Bajo este estricto modelo bipartito que distingue entre sujetos de derecho y meros objetos, la utilización de una máquina no permite fundamentar la autoría mediata, sino que reconduce automáticamente al sujeto de atrás a la autoría directa. Esta limitación ha sido criticada por sectores doctrinales que denuncian una asimetría entre la realidad tecnológica y el marco conceptual. Para salvar esta inoperancia de las categorías, autores como Navarro-Dolmestch proponen quebrar el rígido modelo bipartito y reconocer una tercera categoría: los "actuantes"[29].

Los actuantes serían aquellos dispositivos inteligentes que, sin reunir las condiciones para ser considerados sujetos imputables, poseen propiedades de autonomía, interactividad y adaptabilidad que superan la pasividad de los objetos tradicionales. El reconocimiento de la IA como "actuante" permitiría reinterpretar la estructura de la autoría mediata, habilitando la imputación del sujeto humano que se sirve de un sistema algorítmico para la ejecución del hecho delictivo.

El escenario se vuelve verdaderamente problemático cuando los resultados lesivos se materializan sin dolo, es decir, en el terreno de la imprudencia y la responsabilidad por negligencia. En el ciclo de vida de la IA intervienen una multitud de actores, desde los desarrolladores hasta los supervisores y los consumidores finales. Cuando un vehículo autónomo, debido a un error de cálculo en sus sensores, causa un atropello mortal, o cuando un sistema médico experto emite un diagnóstico incorrecto, se debe investigar la infracción del deber objetivo de cuidado a lo largo de una cadena de producción difusa.

La imputación penal se nutre en estos casos de los criterios de responsabilidad por producto defectuoso, evaluando si el fabricante omitió los estándares normativos de control en la fase de diseño, o si permaneció inactivo tras descubrir riesgos en

[29] NAVARRO-DOLMESTCH, R., *op. cit.*, pp. 3 y 7.

el sistema después de su puesta en circulación.

La dificultad probatoria en los delitos imprudentes radica en la opacidad inherente a ciertos modelos de aprendizaje profundo o *deep learning*, fenómeno conocido como el problema de la "caja negra". Estas redes neuronales artificiales toman decisiones a través de innumerables capas de procesamiento matemático cuya lógica exacta resulta cognitivamente inaccesible incluso para sus propios creadores.

Si el sistema genera de forma autónoma una pauta de comportamiento emergente que resulta lesiva, sin que existiera un propósito delictivo, gran parte de la doctrina considera que se excluye la responsabilidad penal del programador, al no ser el riesgo ni cognoscible ni evitable *ex ante* conforme a los estándares de previsibilidad exigibles, encuadrando el suceso en el ámbito del riesgo permitido o del caso fortuito. No obstante, algunos expertos advierten que la intrínseca imprevisibilidad de los agentes artificiales es precisamente lo que fundamenta unos deberes de cuidado reforzados.

Adicionalmente a la problemática de la imputación, la influencia de la IA transforma la estructura misma del ilícito desde el análisis de la culpabilidad en las organizaciones complejas y la determinación de la pena. La integración del trabajador o usuario en dinámicas empresariales altamente automatizadas genera escenarios donde las decisiones delictivas no derivan de una racionalidad libre y ponderada, sino de la sumisión a los resultados arrojados por el sistema tecnológico. Factores externos, como los sesgos de neutralización, la imitación de conductas validadas por el algoritmo y la presión por la eficiencia algorítmica, erosionan la capacidad del sujeto individual para motivarse normativamente. Esta afectación a la libertad volitiva dentro de entornos digitalizados debería, según un sector de la doctrina, trasladarse a la evaluación de la culpabilidad individual, operando como base para eximir o atenuar la responsabilidad penal de los escalones inferiores que simplemente acatan las inercias del sistema[30].

En contrapartida, el empleo de sistemas inteligentes por parte del delincuente incrementa cualitativamente el desvalor de la acción, lo que ha motivado propuestas para incorporar el uso de la IA como una circunstancia agravante de la responsabilidad penal. La sofisticación extrema de estos instrumentos potencia la eficacia del engaño, difumina la identidad del agente y dificulta la persecución, reduce las capacidades defensivas de la víctima y dota al delito de un efecto multiplicador capaz de vulnerar bienes jurídicos a escala masiva, automatizada y transnacional de un modo que los medios analógicos jamás podrían emular[31].

Frente a la utopía de castigar directamente a la máquina en un futuro todavía lejano, la imperiosa necesidad del presente radica en adaptar los esquemas de imputación penal, repensar el concepto de instrumento en la autoría, redefinir los márgenes de la imprudencia ante algoritmos y preservar la centralidad de la responsabilidad humana detrás del velo tecnológico.

El ordenamiento jurídico debe evolucionar con rigor para que las herramientas de la nueva era digital no se conviertan en espacios de impunidad ni para diluir la culpabilidad de quienes deciden instrumentalizarlas para vulnerar la paz social.

---

[30] VALDEZ SILVA, F. A., "La culpabilidad jurídico-penal individual en la empresa: ¿realmente somos libres dentro de una organización empresarial compleja?", en *Estudios de Derecho penal, neurociencias e inteligencia artificial*, Ediciones de la Universidad de Castilla-La Mancha, 2023, pp. 92-93.

[31] AGUILAR CAMPOS, P., y ALÉ MARTÍNEZ, V., *op. cit.*, p. 166.

## 4. Principales manifestaciones delictivas vinculadas al uso de la inteligencia artificial

El ciberespacio ha dejado de constituir una mera herramienta de comunicación para convertirse en un entorno autónomo de interacción social, económica y política, caracterizado por la desaparición de las barreras espaciales y temporales que tradicionalmente limitaban el alcance del comportamiento humano. En este contexto, la criminalidad ha experimentado un proceso de transformación que ha dado lugar a la denominada cibercriminalidad o cibercrimen.

Este fenómeno no debe entenderse exclusivamente como una categoría integrada por nuevos tipos delictivos, sino como un conjunto de conductas ilícitas en las que las tecnologías digitales actúan como medio fundamental para la comisión del ilícito[32]. El ciberespacio amplía la oportunidad delictiva, ya que la comprensión del espacio físico acerca a agresores motivados de cualquier rincón del mundo con objetivos o víctimas adecuadas, todo ello en un entorno caracterizado por la ausencia de "guardianas capaces" o mecanismos de protección eficientes. El anonimato, la transnacionalidad, la automatización y la volatilidad de los datos dotan a estas conductas de una lesividad sin precedentes en el mundo físico[33].

La inteligencia artificial, entendida como sistemas capaces de percibir su entorno, procesar datos masivos mediante algoritmos de aprendizaje (*machine learning* o *deep learning*) y tomar decisiones para alcanzar objetivos con un grado variable de autonomía, ha potenciado esta realidad criminológica de manera exponencial. La IA permite automatizar conductas delictivas, adaptar los ataques a los sistemas de defensa y ejecutar operaciones a gran escala y elevada velocidad incrementando significativamente las capacidades operativas de la cibercriminalidad.

### 4.1. La cibercriminalidad económica y el fraude algorítmico

Los ciberdelitos económicos, materializados principalmente a través del fraude informático y las estafas, representan la categoría principal de la delincuencia tecnológica, constituyendo más del ochenta por ciento de las infracciones informáticas registradas en España[34]. El objetivo final de estas conductas es siempre la obtención de un lucro o beneficio patrimonial, directo o indirecto, utilizando la IA, las redes telemáticas y los sistemas informáticos como el medio comisivo o como el objeto del ataque.

Se trata de una criminalidad dirigida contra el bien jurídico del patrimonio ajeno, pero cuya ejecución se articula a través de la informática para consumar apoderamientos ilegítimos o causar daños económicamente evaluables[35]. En el ordenamiento jurídico penal, este abanico de conductas abarca figuras como la estafa informática (art. 249 CP), los daños y sabotajes informáticos (arts. 264 y 264 bis CP), la extorsión mediante el secuestro de datos o *ransomware* (art. 243 CP), el blanqueo informático de capitales (art. 301 CP), el descubrimiento y revelación de secretos de empresa (arts. 278-280 CP) y los fraudes en las telecomunicaciones (art. 255 CP).

La IA ha aportado a la cibercriminalidad económica una capacidad de sistematización y engaño excepcional. En el ámbito de las defraudaciones, no se trata de actos aislados, sino de verdaderas maquinarias delictivas de carácter organizado. Para perpetrar un fraude, los ciberdelincuentes suelen ejecutar

[32] MIRÓ LLINARES, F., "El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio", *Revista de Derecho Penal y Criminología*, Marcial Pons, 2012, pp. 41-42.

[33] MIRÓ LLINARES, F., *op. cit.*, 2012, p. 175.

[34] MINISTERIO DEL INTERIOR, *Informe sobre la Cibercriminalidad en España 2024*, Secretaría de Estado de Seguridad, Dirección General de Coordinación y Estudios, Madrid, 2025, p. 27.

[35] MIRÓ LLINARES, F., *op. cit.*, 2012, pp. 119-121.

previamente una serie de ciberdelitos instrumentales: ataques de *hacking*, creación de *botnets* (redes de ordenadores infectados), interceptación de datos mediante *spyware* y campañas de suplantación de identidad (*phishing*, *spoofing* o *pharming*). Mediante la IA, estas fases preparatorias se han perfeccionado; por ejemplo, algoritmos entrenados pueden redactar correos de *phishing* indetectables o hiper personalizados para víctimas concretas, lo que incrementa la probabilidad de éxito en el engaño.

El nivel de sofisticación alcanza un grado especialmente elevado con el uso de sistemas generativos de IA en estafas y extorsiones. Existen casos donde, utilizando tecnologías de clonación de voz y *deep learning*, los criminales han logrado simular de manera perfecta la voz del director ejecutivo de una compañía para engañar a los empleados y ordenar la transferencia de cuantiosas sumas de dinero a cuentas en el extranjero[36].

Este tipo de prácticas constituye una modalidad avanzada de ingeniería social basada en sistemas de IAG, mediante la cual se refuerza artificialmente la apariencia de autenticidad del engaño. La utilización de técnicas de clonación de voz y generación sintética de contenido incrementa la capacidad defraudatoria de la conducta y plantea dificultades en relación con la identificación del autor y la valoración del engaño bastante exigido por el delito de estafa.

Por otro lado, la interacción de la IA en los mercados financieros globales ha propiciado la aparición de fraudes altamente complejos, como la manipulación algorítmica del mercado de valores. Programas inteligentes de inversión de alta frecuencia (*AI traders*) pueden ser diseñados o manipulados para generar información falsa, difundirla masivamente y emitir órdenes de compra y venta fantasma (*spoofing*) con el fin de alterar artificialmente la cotización de ciertos activos, práctica conocida como *scalping*. Aunque en ocasiones el desarrollador del algoritmo no haya programado expresamente la manipulación, el sistema de aprendizaje por refuerzo de la máquina descubre que alterar el mercado es la estrategia óptima para maximizar los beneficios, planteando serios desafíos para la imputación subjetiva por dolo o imprudencia en el Derecho penal económico.

La cibercriminalidad económica ha dejado de ser un ámbito exclusivo del *hacker* aislado para convertirse en una de las actividades más lucrativas del crimen organizado transnacional. Redes criminales distribuidas por todo el globo operan en el ciberespacio con estructuras horizontales, aprovechando el anonimato que confiere la red y la dificultad de rastrear criptoactivos para lavar ingentes cantidades de dinero y financiar actividades ilícitas a gran escala.

### 4.2. Ciberdelitos intrusivos: violencias en la red y la cosificación digital

La popularización de la llamada "Web 2.0" y la omnipresencia de los dispositivos móviles no transformaron los mercados, sino que abrieron el ciberespacio a la intimidad, la socialización y el libre desarrollo de la personalidad. Esta hiperconexión personal ha traído consigo los denominados "ciberdelitos sociales o intrusivos", que trasladan al ámbito digital los conflictos humanos de carácter especialmente sensible, atacando de forma directa bienes jurídicos personalísimos como el honor, la intimidad, la propia imagen, y de manera especialmente grave, la libertad e indemnidad sexual.

En la esfera de estos delitos se encuentran el ciberacoso o acoso cibernético (*cyberbullying*); el hostigamiento reiterado a través de medios digitales (*cyberstalking*), subsumible en el delito de acoso previsto en

[36] AGUILAR CAMPOS, P., y ALÉ MARTÍNEZ, V., *op. cit.*, pp. 166-167.

el art. 172 ter CP; así como las injurias y calumnias difundidas masivamente mediante redes sociales y plataformas digitales, susceptibles de encaje en los arts. 208 y ss. CP. Del mismo modo, la difusión de mensajes discriminatorios o extremistas a través de entornos virtuales puede integrar delitos de incitación al odio del art. 510 CP.

Especial relevancia adquieren los delitos de naturaleza sexual cometidos en el entorno digital, entre los que destacan el embaucamiento de menores con fines sexuales (*child grooming*), tipificado en el art. 183 CP; la extorsión mediante imágenes íntimas (*sextortion*); así como las conductas relacionadas con la elaboración, tenencia y difusión de pornografía infantil previstas en los arts. 189 y ss. CP. Asimismo, determinadas prácticas basadas en la difusión no consentida de contenidos íntimos o manipulados digitalmente pueden afectar a la intimidad, la propia imagen y la protección de datos personales, encontrando encaje en los delitos de descubrimiento y revelación de secretos regulados en los arts. 197 y ss. CP.

En todos estos supuestos, el daño a la víctima se ve intensificado por las características propias del entorno digital, especialmente por la rápida difusión de los contenidos, su permanencia en la red y la dificultad de eliminar completamente el rastro tecnológico generado.

Con la llegada de la IA, las herramientas de agresión intrusiva se han potenciado. Quizá una de las manifestaciones más preocupantes sea el auge de los denominados "*deepfakes*". Estas tecnologías, accesibles a cualquier usuario sin necesidad de conocimientos avanzados en programación, utilizan redes neuronales para crear o manipular contenidos audiovisuales de alto realismo, suplantando el rostro y la voz de personas reales en situaciones que nunca ocurrieron.

El impacto de los *deepfakes* ha recaído de manera desproporcionada sobre las mujeres y los menores de edad. Diversos informes denuncian que cerca del noventa y seis por ciento de los *deepfakes* generados por IA están destinados a crear pornografía no consentida, cosificando el cuerpo femenino para someterlo, humillarlo y ejercer violencia de género en el entorno virtual.

Esta manipulación sintética ha generado un vacío legal y una desprotección para las víctimas, quienes a menudo deben enfrentarse a un sistema jurídico lento y a discursos que las revictimizan argumentando que, al ser las imágenes artificiales, el daño no es "real". Sin embargo, la disrupción psicológica de ver una identidad íntima vulnerada públicamente es devastadora, subsumiendo estas conductas en variantes gravísimas de injurias, delitos contra la intimidad moral, e incluso extorsión.

En lo que respecta a la protección de la indemnidad de los menores, el escenario resulta especialmente grave, pues la IA no solo puede generar material de abuso sexual infantil a partir de imágenes aparentemente inocuas de menores, sino que también es utilizada por depredadores sexuales mediante *chatbots* programados para simular ser adolescentes en juegos o redes sociales. Estos *bots* conversacionales generan lazos de confianza falsos con los menores (*grooming*), a fin de inducirlos a compartir imágenes sexuales o concertar encuentros en la vida real.

Paralelamente a esta intrusión de carácter sexual, las redes sociales han sido el campo de batalla de otra modalidad delictiva a través de los *bots* automatizados: la manipulación sociopolítica y los discursos de odio. Cuentas falsas operadas por IA simulan ser personas reales (*astroturfing*) para difundir ideologías extremistas, incrementar informaciones falsas y campañas de desprestigio, promoviendo el ciberodio, exacerbando crisis sociales, e incurriendo en delitos de incitación al terrorismo o a la discriminación, con la ventaja de ocultar la "mano humana" que organiza estas masivas campañas de desinformación.

### 4.3. Desafíos probatorios en la arquitectura digital

La traslación de la criminalidad al ciberespacio y la irrupción de la IA imponen obstáculos para el sistema de enjuiciamiento criminal, poniendo a prueba la capacidad de los Estados para garantizar la tutela judicial efectiva y socavando la tradicional persecución del delito basada en coordenadas físicas. El diseño intrínseco de la arquitectura digital choca con las normas del Derecho procesal, generando lo que la doctrina denomina la "cifra negra" del cibercrimen; es decir, un abismo desproporcionado entre los delitos tecnológicos que realmente se cometen, los que se denuncian y, finalmente, la minoría que llega a recibir una sentencia condenatoria[37].

El primer desafío probatorio deriva de la comprensión del espacio-tiempo, la deslocalización y la transnacionalidad de Internet. En la delincuencia tradicional, la concreción del espacio geográfico del crimen aporta indicios para identificar al agresor; en el ciberespacio, en cambio, la acción puede originarse en un continente y desplegar efectos lesivos en múltiples jurisdicciones distintas. Esta extraterritorialidad dificulta enormemente la persecución penal, ya que requiere de una compleja, lenta y muchas veces infructuosa cooperación judicial internacional. A menudo, las investigaciones se paralizan cuando las Fuerzas y Cuerpos de Seguridad constatan que los servidores desde donde se perpetró, por ejemplo, la estafa o se alojó la pornografía infantil radican en paraísos cibernéticos o jurisdicciones sin tratados de extradición que niegan la entrega de datos.

Íntimamente ligado a la transnacionalidad se encuentra el segundo gran desafío de la arquitectura digital: el anonimato. Aunque la actividad en la red deja inevitablemente una huella electrónica, la IA y los ciberdelincuentes han perfeccionado técnicas que blindan su identidad física. La atribución de responsabilidad penal exige determinar qué persona física concreta presionó la tecla o programó el algoritmo. Sin embargo, los atacantes operan detrás de redes wifi públicas, utilizan servidores proxy o redes de anonimato (*dark web*). Además, la IA ha facilitado la propagación de redes zombis o *botnets*, infectando los ordenadores de los ciudadanos para, posteriormente, dirigir desde esos terminales ataques masivos. Esto no solo borra el rastro de la ubicación del autor real, sino que dirige las sospechas hacia víctimas tecnológicas.

Ante esta realidad, la justicia penal se ha visto obligada a transitar desde las pruebas físicas hacia las pruebas tecnológicas, entendidas como todo dato o información almacenada o transmitida por medios electrónicos que posee valor de convicción para el esclarecimiento de un hecho[38]. Siguiendo a Bujosa Vadell[39], la peculiaridad de la prueba electrónica radica en su volatilidad y facilidad de manipulación, lo que exige al órgano judicial no solo a valorar el contenido de la prueba, sino también la credibilidad del medio de prueba. La admisibilidad probatoria de estos elementos exige garantizar de forma estricta su integridad y autenticidad, lo que implica demostrar, mediante prueba pericial, que los datos no han sido alterados y que el rastro digital puede atribuirse a su presunto autor.

A tal efecto, resulta imprescindible la aplicación de técnicas específicas como la obtención de copias forenses (función *hash*), el aseguramiento de la cadena de custodia electrónica bajo normativas internacionales estrictas y la constante comparecencia de peritos informáticos que auxilien al juez en la comprensión de un entorno técnico que

[37] MIRÓ LLINARES, F., *op. cit.*, 2012, pp. 292-293.

[38] HERNÁNDEZ GIMÉNEZ, M., *op. cit.*, pp. 812-813.

[39] BUJOSA VADELL, L. M., "La valoración de la prueba electrónica", *Fodertics 3.0 (Estudios sobre nuevas tecnologías y justicia)*, en BUENO DE MATA, F. (coord.), Granada, 2015, pp. 82-85.

escapa a los conocimientos jurídicos convencionales.

La propia IA agrava estas dificultades procesales, cuestionando el tradicional axioma de que "ver para creer" constituye una garantía de verdad judicial. Las evidencias en formato de vídeo o audio, que tradicionalmente gozaban de una contundencia irrefutable, hoy pueden ser rechazadas con la simple alegación de ser *deepfakes*. La identificación de manipulaciones avanzadas resulta tan compleja que exige, paradójicamente, el recurso a herramientas forenses basadas en IA, capaces de detectar alteraciones imperceptibles a nivel de píxeles o frecuencias de voz[40]. Asimismo, el uso de sistemas algorítmicos en la toma de decisiones judiciales introduce el problema del efecto "caja negra". La opacidad del razonamiento automatizado dificulta el control contradictorio de la prueba y puede comprometer derechos fundamentales como el derecho de defensa y la tutela judicial efectiva.

Todo ello pone de manifiesto la evidente necesidad de actualizar no solo el Derecho penal sustantivo, sino todo el Derecho procesal, dotando a los operadores jurídicos de herramientas adecuadas, promoviendo mecanismos de cooperación internacional en materia de ciberseguridad y garantizando, en todo caso, que la incorporación de la IA no menoscabe las garantías constitucionales del ciudadano.

## 5. Límites del derecho penal frente al uso de la inteligencia artificial

La incorporación de la inteligencia artificial en las estructuras del Estado y, en particular, en el sistema de justicia penal y en el ámbito de la seguridad pública, plantea relevantes desafíos. Las tecnologías digitales y el análisis masivo de datos mediante sistemas inteligentes no deben considerarse un fin en sí mismas, sino instrumentos subordinados al servicio del interés general, la seguridad y el respeto a la dignidad de la persona. La progresiva integración de estas tecnologías en el ámbito punitivo ha evidenciado una tensión entre la eficiencia que aportan en la prevención e investigación de la criminalidad y los riesgos que su utilización puede generar para los principios del Estado social y democrático de Derecho. Por tanto, procede establecer límites que aseguren una "reserva de humanidad"[41] en la aplicación de la ley, garantizando que el Derecho penal siga siendo una herramienta de justicia.

### 5.1. Respeto de los derechos fundamentales

El uso de sistemas de IA por parte de los poderes públicos incide de manera directa en los derechos fundamentales consagrados en la Constitución Española (en lo sucesivo, CE) y en la Carta de los Derechos Fundamentales de la Unión Europea. La primera colisión se produce en el ámbito del derecho a la intimidad, así como en la protección de datos de carácter personal protegidos por el artículo 18 CE[42].

Estos sistemas se basan en el tratamiento masivo de datos (*Big Data*), lo que implica la recopilación, almacenamiento y procesamiento de grandes volúmenes de información relativa a los ciudadanos. Esta dinámica difumina los límites de la privacidad, al permitir la elaboración de perfiles detallados a partir de metadatos, historiales de navegación, datos de geolocalización o interacciones en redes sociales, en ocasiones sin un consentimiento plenamente informado

[40] RODRÍGUEZ SANTILLÁN, S., "Esa persona no soy yo: la inteligencia artificial como un nuevo instrumento de violencia", *American University International Law Review*, núm. 3, 2025, p. 765.

[41] LACRUZ MANTECÓN, M. L., "Panorama de la inteligencia artificial en la justicia y el derecho actuales", *Revista de derecho aragonés*, núm. 30, Zaragoza, 2024, p. 103.

[42] MIRÓ LLINARES, F., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, 3.ª Época, n.º 20, 2018, pp. 114-120.

o con escasa transparencia en el tratamiento de los datos.

El riesgo para la privacidad alcanza su máxima expresión con la implementación de sistemas de vigilancia masiva y el uso de técnicas de identificación biométrica[43]. La tecnología actual permite el reconocimiento facial en tiempo real, el análisis de expresiones faciales e incluso la inferencia de estados emocionales. En determinados contextos, como en países como China, se han desarrollado sistemas que integran redes extensivas de cámaras de vigilancia con bases de datos biométricas, permitiendo el seguimiento continuo de los ciudadanos mediante tecnologías como *Cloud Walk*[44]. Esta capacidad de supervisión plantea serias implicaciones para los derechos fundamentales y ha suscitado una notable preocupación en el ámbito de la UE.

Como respuesta, el reciente Reglamento de IA, en su artículo 5.1.h), interviene tipificando como prácticas prohibidas "*el uso de sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines de garantía del cumplimiento del Derecho*", salvo en excepciones tasadas y estrictamente necesarias, como la búsqueda de víctimas de secuestro, la prevención de amenazas terroristas inminentes o la localización de sospechosos de delitos especialmente graves, requiriendo siempre autorización judicial.

De igual manera, el artículo 5.1.b) RIA prohíbe taxativamente la creación de sistemas de puntuación ciudadana impulsados por agentes públicos o privados, evitando así que los ciudadanos reciban un trato perjudicial o desfavorable basado en la evaluación algorítmica de su comportamiento o características personales.

El segundo límite viene constituido por el principio de presunción de inocencia y el derecho a un proceso con todas las garantías, pilares del proceso penal que se ven amenazados por la irrupción de la denominada "policía predictiva" y la "justicia algorítmica". Las herramientas diseñadas para predecir el riesgo de comisión de delitos o estimar tasas de reincidencia se basan en modelos estadísticos, lo que plantea objeciones desde la perspectiva de la presunción de inocencia, en la medida en que el individuo debe ser juzgado exclusivamente por hechos reales, probados y consumados, y no por predicciones probabilísticas sobre comportamientos futuros inferidas a partir de sus características personales.

En ausencia de una valoración humana efectiva y de una sospecha razonable sustentada en hechos objetivos comprobables, el recurso a este tipo de sistemas puede desvirtuar las garantías propias del proceso penal. Consciente de estos riesgos, el legislador europeo ha optado por calificar como sistemas de "alto riesgo" aquellos sistemas de IA destinados a evaluar el riesgo de reincidencia o a asistir a las autoridades judiciales en la interpretación de los hechos, tal y como establece el artículo 6.3 RIA. Esta calificación implica la exigencia de evaluaciones previas a su puesta en funcionamiento y refuerza la necesidad de que la intervención algorítmica tenga un carácter meramente auxiliar, de modo que la decisión final siga correspondiendo al órgano judicial.

La utilización de algoritmos en el ámbito de la justicia penal plantea tensiones con el derecho a la tutela judicial efectiva y el derecho de defensa, especialmente en relación con el efecto "caja negra". Los

[43] LACRUZ MANTECÓN, M. L., *op. cit.*, pp. 115-116.

[44] *CloudWalk Technology* es una empresa china, fundada en 2015, especializada en el desarrollo de sistemas de reconocimiento facial basados en inteligencia artificial, capaces de identificar y rastrear individuos en tiempo real mediante el análisis de datos biométricos, ampliamente utilizada en proyectos de seguridad pública. Su actividad ha suscitado controversia internacional por su presunta participación en sistemas de control poblacional y por el impacto en los derechos fundamentales.

sistemas de IA basados en redes neuronales profundas (*deep learning*) operan mediante procesos de elevada complejidad que, en muchos casos, resultan difícilmente explicables incluso para sus propios programadores, lo que dificulta la trazabilidad del razonamiento que conduce a un determinado resultado. Cuando estas herramientas se emplean para sustentar decisiones relevantes, como la incriminación de un sospechoso o la valoración del riesgo de fuga, la opacidad del sistema puede impedir al abogado defensor impugnar eficazmente la prueba o cuestionar los fundamentos de la decisión adoptada.

La falta de transparencia compromete el principio de contradicción y puede derivar en situaciones incompatibles con las exigencias del proceso debido, de modo que ninguna persona pueda ser sometida a una resolución judicial basada en un procesamiento algorítmico cuyas premisas lógicas o criterios de ponderación resulten inaccesibles -opacos o no verificables- ni queden amparados en derechos de propiedad intelectual de terceros[45].

### 5.2. Sesgos algorítmicos y perspectiva de género.

Frente a la presunción de que los sistemas algorítmicos son inherentemente neutrales y objetivos, la realidad demuestra que la IA es una herramienta diseñada por seres humanos y que, en consecuencia, puede reproducir y amplificar sesgos y estereotipos históricos presentes en los datos con los que se entrena. En el ámbito del Derecho penal y la criminología, la ausencia de una perspectiva de género en el diseño, entrenamiento e implementación de estos sistemas plantean relevantes desafíos, dado que los algoritmos pueden contribuir a la perpetuación de dinámicas discriminatorias y comprometer el principio de igualdad consagrado en el artículo 14 CE[46].

Para comprender cómo se materializa la discriminación algorítmica, resulta necesario identificar las tres fases principales en las que pueden introducirse sesgos de género: (i) la recopilación de los datos; (ii) la selección de las variables y atributos que serán tenidos en cuenta por el algoritmo; y (iii) el diseño del *hardware* y de las interfaces[47].

En primer lugar, los sesgos pueden introducirse en los datos de entrenamiento, dado que los sistemas de aprendizaje automático se nutren de grandes volúmenes de información generada en contextos sociales históricamente desiguales, en los que las mujeres han estado infrarrepresentadas. Cuando un sistema de IA se entrena con bases de datos que presentan una sobrerrepresentación masculina, el algoritmo tiende a identificar y priorizar patrones asociados a dicho grupo, reproduciendo así desequilibrios preexistentes.

Un ejemplo paradigmático de discriminación directa en el ámbito laboral es el caso del sistema de selección de personal desarrollado por Amazon, que tuvo que ser retirado al constatarse que penalizaba sistemáticamente candidaturas femeninas. Al haber sido entrenado con datos de contratación correspondientes a una década en la que predominaban perfiles masculinos, el sistema infería de forma errónea que el género masculino constituía un indicador de idoneidad para el puesto.

En segundo lugar, los sesgos pueden derivar de las propias decisiones de diseño relativas a las variables que el sistema debe considerar y aquellas que deben quedar excluidas. Los algoritmos están orientados a identificar

[45] HERNÁNDEZ GIMÉNEZ, M., *op. cit.*, p. 826.

[46] MONTESINOS GARCÍA, A., "Inteligencia artificial en la justicia con perspectiva de género: amenazas y oportunidades", *Actualidad Jurídica* Iberoamericana, n.º 21, 2024, p. 572.

[47] ORTIZ DE ZÁRATE ALCARAZO, L., "Sesgos de género en la inteligencia artificial", *Revista de* occidente, vol. 502, 2023, pp. 8-13.

correlaciones y patrones en los datos, pero, al carecer de una verdadera capacidad de comprensión contextual, pueden establecer relaciones erróneas y reproducir estereotipos preexistentes. Así ocurre, por ejemplo, en determinados sistemas de procesamiento de lenguaje natural -como los traductores automáticos-, que tienden a asociar profesiones o cualidades socialmente prestigiosas a pronombres masculinos, mientras vinculan determinadas tareas de cuidado o asistencia a pronombres femeninos[48].

Asimismo, los sistemas de reconocimiento facial presentan tasas de error significativamente superiores en la identificación de mujeres racializadas en comparación con hombres blancos, lo que evidencia la persistencia de sesgos tanto en los datos empleados como en el diseño de estos sistemas. En el ámbito penal, la utilización de herramientas de reconocimiento facial sesgadas para identificar sospechosos en espacios públicos puede dar lugar a identificaciones erróneas y, en consecuencia, a detenciones injustas respecto de determinados colectivos.

El uso de sistemas de IA de alto riesgo en la Administración de Justicia abre un debate crítico sobre cómo los sesgos de género pueden viciar resoluciones judiciales y afectar a los derechos procesales. A medida que los jueces y tribunales comienzan a auxiliarse de herramientas predictivas y de valoración, surgen riesgos especialmente relevantes en relación con las víctimas de delitos de violencia de género. Así, un sistema algorítmico diseñado para valorar la credibilidad del testimonio de una víctima de agresión sexual a partir de datos jurisprudenciales históricos podría reproducir prejuicios tradicionales, penalizando circunstancias estrechamente vinculadas a los procesos psicológicos del trauma, como la demora en la denuncia, las contradicciones del relato o determinadas reacciones emocionales de la víctima[49]. Del mismo modo, el diseño de herramientas destinadas a detectar "denuncias falsas" en materia de violencia de género -asumiendo prejuiciosamente que son frecuentes-, podría generar dinámicas discriminatorias, desalentando a las víctimas a acudir al sistema judicial.

En el ámbito de la valoración del riesgo y la reincidencia, España cuenta con herramientas como el sistema VioGén, utilizado para orientar la adopción de medidas de protección frente a la violencia de género. Aunque estos sistemas pueden desempeñar una función relevante en la prevención y gestión del riesgo, también plantean interrogantes sobre los criterios utilizados y de la posible reproducción de sesgos derivados de los datos empleados en su entrenamiento.

En tercer lugar, los sesgos también pueden manifestarse en el propio diseño del *hardware* y de las interfaces de interacción. Ello resulta especialmente visible en asistentes de voz virtuales y *chatbots* -como Siri, Alexa o Cortana-, a los que de forma predominante se les han asignado nombres y voces femeninas para desempeñar funciones de asistencia y atención al usuario, reforzando determinados estereotipos de género asociados a roles de servicio y subordinación[50].

El ciberespacio no es un entorno ajeno a la violencia contra la mujer; por el contrario, la tecnología proporciona a los agresores herramientas que aumentan la accesibilidad, el anonimato y la capacidad de difusión de

[48] ORTIZ DE ZÁRATE ALCARAZO, L., y GUEVARA GÓMEZ, A., "Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España", *Fundación* Alternativas, 2021, p. 46.

[49] MONTESINOS GARCÍA, A., *op. cit.*, pp. 574-575.

[50] CERNADAS GARCÍA, E., y CALVO IGLESIAS, E., "Perspectiva de género en la inteligencia artificial, una necesidad", *Cuestiones de género: de la igualdad y la diferencia*, n.º 17, 2022, p. 114.

este tipo de conductas. El principal exponente de este fenómeno lo constituyen los *deepfakes*, contenidos audiovisuales falsos de un realismo extremo generados mediante redes adversariales generativas (*GAN*, por sus siglas en inglés)[51]. Su utilización para la creación y difusión de pornografía no consentida constituye una de las manifestaciones más graves de violencia digital de género, al permitir la superposición del rostro de una persona sobre material sexual explícito sin su consentimiento. Conductas como las ocurridas en Almendralejo (Badajoz) constatan cómo la democratización de aplicaciones basadas en IA facilita la comisión de delitos contra la intimidad y la indemnidad sexual[52].

Esta realidad ha impulsado la proposición de reformas en el Código Penal español para tipificar estas conductas como delitos de injurias agravadas (art. 208 bis CP), reflejando la necesidad de una respuesta punitiva específica ante el daño que causan. Además del daño directo a la víctima, los *deepfakes* plantean problemas probatorios, dado que permiten generar contenidos audiovisuales hiperrealistas susceptibles tanto de utilizarse para desacreditar a las víctimas como de impugnar pruebas auténticas mediante la simple alegación de manipulación digital, dificultando así la valoración judicial de la prueba[53].

La constatación de estos sesgos algorítmicos impone límites claros al uso de la IA en el ámbito penal. Para evitar que el Derecho penal y la Administración de Justicia integren algoritmos que reproduzcan dinámicas discriminatorias, resulta imprescindible incorporar una perspectiva de género desde el diseño (*gender by design*), garantizando la representatividad de las bases de datos como la explicabilidad de las decisiones automatizadas y la realización de auditorías previas y periódicas. En esta línea, la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación[54], pone de relieve la necesidad de someter los sistemas destinados a asistir a la judicatura y a las fuerzas y cuerpos de seguridad a estrictos criterios de transparencia, supervisión humana, minimización de sesgos y rendición de cuentas.

Por ello, el desarrollo de sistema de IA fiables y centrados en el ser humano (*Trustworthy AI*) constituye una exigencia para evitar que la automatización de decisiones contribuya a perpetuar desigualdades estructurales bajo una apariencia de neutralidad tecnológica. Solo desde una concepción garantista y multidisciplinar de la IA podrá lograrse que estas tecnologías actúen como herramientas útiles para combatir la criminalidad y la violencia de género.

## 6. Conclusiones

PRIMERA. La implantación de la inteligencia artificial y su convergencia con el procesamiento masivo de datos han consolidado la Cuarta Revolución Industrial, transformando el ciberespacio y generando nuevos riesgos para los bienes jurídicos protegidos por el ordenamiento penal. Esta tecnología cuestiona el esquema sobre el que históricamente se ha construido el Derecho penal, basado en la distinción entre "sujetos" dotados de voluntad y "objetos" meramente instrumentales. Los actuales sistemas informáticos, provistos de redes neuronales y algoritmos de aprendizaje profundo, poseen

[51] MONTESINOS GARCÍA, A., *op. cit.*, pp. 577-581.

[52] BORRAZ, M. Y PASTOR, A., "Deepfakes sexuales: el caso de las menores de Almendralejo consolida una nueva forma de violencia machista", *elDiario.es*, 2023, disponible en https://www.eldiario.es/sociedad/deepfakes-sexuales-caso-menores-almendralejo-consolida-nueva-forma-violencia-machista_1_10527153.html

[53] SIMÓ SOLER, E., "Retos jurídicos derivados de la Inteligencia Artificial Generativa Deepfakes y violencia contra las mujeres como supuesto de hecho", *InDret Criminología*, 2023, núm. 2, p. 501.

[54] Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, BOE núm. 167, de 13 de julio, pp. 1-39.

niveles de autonomía funcional y capacidad de decisión que imitan procesos cognitivos humanos, circunstancia que obliga a replantear las estructuras tradicionales de imputación penal.

SEGUNDA. A pesar del debate existente en torno a la IA fuerte o *machina sapiens* y la hipotética posibilidad de considerarla como autora directa del delito, la posición mayoritaria rechaza la atribución de culpabilidad a los sistemas inteligentes. La IA carece de conciencia y de capacidad para comprender la ilicitud de sus actos, operando siempre bajo los parámetros impuestos por sus desarrolladores. La respuesta jurídico-penal, por tanto, debe seguir centrada en los agentes humanos (programadores, operadores o usuarios finales), considerando a la IA como un medio o instrumento sofisticado para la comisión de delitos. Ello exige reinterpretar categorías procesales como la autoría directa, la coautoría, la negligencia ante productos defectuosos y, especialmente, la autoría mediata, donde surge la necesidad de reconocer a la IA como un "actuante" para evitar espacios de impunidad jurídica.

TERCERA. La IA se ha convertido en un medio que automatiza, facilita y multiplica la lesividad de los cibercrímenes. En el ámbito de la cibercriminalidad económica, dota de una escalabilidad y un nivel de engaño sin precedentes a prácticas como el fraude algorítmico, el *phishing*, la manipulación bursátil y la clonación de voz para consumar estafas corporativas. De igual forma, en los ciberdelitos intrusivos, tecnologías como los *deepfakes* actúan como instrumentos idóneos para vulnerar bienes personalísimos, tales como el honor, la propia imagen, la intimidad o la indemnidad sexual.

CUARTA. Las características inherentes al entorno digital, como la transnacionalidad, la volatilidad de los datos y el anonimato, dificultan considerablemente la investigación policial y judicial, incrementando la "cifra negra" de la cibercriminalidad. A esto se suma el efecto "caja negra" propio de determinados sistemas de aprendizaje profundo, cuya lógica matemática inescrutable dificulta determinar si un resultado lesivo obedece a una conducta dolosa, imprudente o fortuita. Paralelamente, el aumento de contenidos sintéticos hiperrealistas cuestiona el valor probatorio tradicional de imágenes y audios, haciendo necesaria una modernización del Derecho procesal penal y el desarrollo de herramientas forenses avanzadas capaces de verificar la autenticidad de las evidencias digitales.

QUINTA. La incorporación de sistemas en la justicia penal y la seguridad ciudadana puede comprometer principios esenciales del Estado social y democrático de Derecho, tales como la presunción de inocencia, el derecho de defensa o la tutela judicial efectiva. Frente a ello, el Reglamento de IA se erige como un hito normativo, al instaurar un marco basado en el riesgo. La prohibición de prácticas como la identificación biométrica (salvo excepciones tasadas) o el *scoring* social, junto con las exigencias de transparencia y supervisión humana para sistemas de "alto riesgo", persigue garantizar que la innovación tecnológica se desarrolle de forma compatible con las garantías constitucionales.

SEXTA. Los sistemas algorítmicos no son neutrales. Su diseño y las bases de datos masivas con las que se entrenan tienden a reproducir, e incluso amplificar, desigualdades y estereotipos estructurales, colisionando con el principio de igualdad y no discriminación. Ello puede traducirse en la criminalización errónea de determinados colectivos mediante *software* de reconocimiento facial sesgado o en valoraciones injustas hacia víctimas de violencia de género por parte de herramientas predictivas. Asimismo, la IA ha favorecido la aparición de nuevas formas de violencia de género digital, especialmente mediante la creación y difusión de pornografía no consentida a través de *deepfakes*. De esta manera, resulta necesario incorporar una perspectiva de género desde el diseño de los sistemas (*gender by design*), en consonancia

con los principios de una IA fiable (*Trustworthy AI*).

SÉPTIMA. Frente a la utopía de atribuir responsabilidad penal a la máquina, la verdadera necesidad actual reside en garantizar que el ordenamiento jurídico-penal sea capaz de evolucionar al mismo ritmo que el desarrollo tecnológico. El Derecho debe adaptar sus categorías dogmáticas, delimitar claramente las posiciones de garante de desarrolladores y operadores y redefinir los márgenes de la imprudencia ante la imprevisibilidad de determinados algoritmos, preservando en todo momento las exigencias derivadas del principio de legalidad y seguridad jurídica. La finalidad última debe ser impedir que el velo tecnológico se convierta en un espacio de impunidad, manteniendo siempre la responsabilidad humana de quienes diseñan, controlan o instrumentalizan estas tecnologías para lesionar bienes jurídicos.

### 6.1. Críticas a la investigación

La presente investigación se ha enfrentado a una serie de limitaciones derivadas, principalmente, de la propia naturaleza cambiante del objeto de estudio. El principal obstáculo radica en la ausencia de una conceptualización unívoca de la IA. La inexistencia de un marco conceptual estable obliga a que el análisis oscile entre las manifestaciones actuales de la IA débil (algoritmos de *machine learning* y *deep learning*) y los escenarios hipotéticos vinculados a la IA fuerte o *machina sapiens*. Esta dualidad provoca que gran parte del debate doctrinal relativo a la atribución directa de responsabilidad penal a la máquina permanezca, por el momento, en un plano teórico.

Por otra parte, el estudio pone de manifiesto la tensión existente entre la velocidad de la innovación tecnológica y la capacidad de respuesta del Derecho, fenómeno descrito bajo la metáfora de la "liebre informática" frente a la "tortuga jurídica". La rápida expansión de herramientas de IAG -como el auge exponencial del ChatGPT desde finales de 2022- genera un riesgo inherente de obsolescencia en los estudios jurídicos, debido a que el desarrollo tecnológico avanza a un ritmo considerablemente superior al de la respuesta legislativa.

Asimismo, el trabajo ha debido articularse en un contexto caracterizado por la inexistencia de un Derecho penal específico aplicable a la IA. Esta circunstancia obliga a adaptar las categorías tradicionales, creadas bajo un modelo antropocéntrico bipartito, para tratar de abarcar fenómenos inéditos, basándose en construcciones doctrinales que todavía no han sido contrastadas en la práctica judicial.

Finalmente, el análisis del RIA se enfrenta a las limitaciones inherentes a su reciente aprobación. Al tratarse de un marco regulatorio pionero, aún no existe un acervo jurisprudencial consolidado que permita valorar su eficacia práctica, la efectividad de su régimen sancionador o la forma en que los tribunales nacionales interpretarán las prohibiciones y obligaciones aplicables a los sistemas de "alto riesgo", especialmente en el ámbito del proceso penal.

### 6.2. Recomendaciones

Ante la insuficiencia del modelo bipartito clásico para dar respuesta a la autonomía operativa de los algoritmos, podría plantearse una actualización conceptual que permita reconocer a los sistemas de IA avanzados como "actuantes" dentro de la dinámica delictiva. Esta reinterpretación permitiría revitalizar la figura de la autoría mediata, partiendo de la idea de que el sujeto humano se sirve de un sistema algorítmico para la ejecución del hecho delictivo. Asimismo, dado el efecto multiplicador del daño que la IA introduce en la cibercriminalidad, cabría sugerir al legislador la valoración de incorporar el uso de sistemas inteligentes de alta complejidad como circunstancia agravante de la responsabilidad penal, atendiendo al mayor desvalor de la acción, a la sofisticación del engaño y al incremento de la indefensión de las víctimas.

Igualmente, podría recomendarse una tipificación más precisa de los deberes de diligencia exigibles a todos los agentes que intervienen en la cadena de valor de la IA. En los supuestos de resultados lesivos imprudentes, la respuesta penal podría inspirarse en los modelos de responsabilidad por producto defectuoso, permitiendo sancionar a aquellos desarrolladores que incumplan los estándares normativos de control o ignoren los riesgos detectados tras la puesta en circulación del sistema.

Con el fin de combatir la impunidad y reducir la "cifra negra" asociada a la arquitectura digital, resultaría conveniente una reforma procesal que otorgue garantías a la prueba electrónica. Ello implicaría la estandarización judicial del uso de funciones *hash*, el refuerzo de la cadena de custodia digital y la implantación de protocolos que aseguren la autenticidad e integridad de la prueba electrónica. Además, ante el auge de los *deepfakes* y otras formas de manipulación digital, podría sugerirse el fortalecimiento de los mecanismos de cooperación judicial internacional para facilitar el rastreo transnacional de evidencias y agilizar la persecución de conductas delictivas cometidas en entornos digitales globalizados.

Por otro lado, cabría proponer una respuesta penal más adecuada frente a las nuevas manifestaciones de violencia de género digital. Podría sugerirse al legislador la introducción de reformas en el Código Penal español orientadas a sancionar específicamente la creación y difusión de contenido sexual manipulado o pornografía no consentida mediante IA, reconociendo el daño psicológico que estas conductas ocasionan sobre bienes jurídicos como el honor, la intimidad, la propia imagen y la indemnidad sexual.

Para evitar que el Estado legitime o reproduzca dinámicas discriminatorias, resultaría aconsejable que cualquier sistema de IA empleado en la Administración de Justicia o en labores de seguridad ciudadana fuese desarrollado bajo una perspectiva de género desde su fase de diseño. Del mismo modo, podrían recomendarse auditorías previas y periódicas de las bases de datos con el objetivo de detectar y corregir posibles sesgos discriminatorios. Todo ello debería ir acompañado de mecanismos efectivos de supervisión humana, transparencia frente al efecto "caja negra" y pleno respeto al derecho a un proceso con todas las garantías.

## Referencias

AGUILAR CAMPOS, Pablo., y ALÉ MARTÍNEZ, Víctor. (2025) Responsabilidad penal en la era de la inteligencia artificial: De la agencia humana a la autonomía de la machina sapiens, *Revista de Estudios de la Justicia*, 42. https://doi.org/10.5354/0718-4735.2025.77061

BENÍTEZ ORTÚZAR, Ignacio., LLEDÓ YAGÜE, Francisco., y MONJE BALMASEDA, óscar. (dirs.), "La robótica y la inteligencia artificial en la nueva era de la Revolución Industrial 4.0", *Colección Inteligencia Artificial, Robots y Bioderecho*, Madrid, 2021, pp. 1-672.

BLANCO CORDERO, Isidoro., "*Homo sapiens* y *¿machina sapiens*?: Un derecho penal para los robots dotados de inteligencia artificial", en *Nuevos retos de la ciberseguridad en un contexto cambiante*, Madrid, 2019.

BODEN, Margaret. A., *Inteligencia Artificial*, Editorial Turner Publicaciones, Madrid, 2017, pp. 1-194.

BUJOSA VADELL, Lorenzo. Mateo., "La valoración de la prueba electrónica", *Fodertics 3.0.* 2015, pp. 75-85.

CERNADAS GARCÍA, Eva., y CALVO IGLESIAS, Encina., "Perspectiva de género en la inteligencia artificial, una necesidad", *Cuestiones de género: de la igualdad y la diferencia*, n.º 17, 2022.https://doi.org/10.18002/cg.i17.7200

CRUZ BELTRAN, José. Luis y LIZ RIVAS, Lenny. El perfil del ciberterrorista: la utilización de medios informáticos con fines terroristas, en; **"**El conflicto y su situación actual: del terrorismo a la

amenaza híbrida", coord. por Carlos Espaliú Berdud, CIVITAS, 2019 pp. 159-173. https://doi.org/10.5281/zenodo.14562806

DELGADO MORAN, Juan. José. Perspectivas Criminológicas aplicadas a las Políticas de Seguridad Pública, en Caruso Fontán/ Macías Caro (dirs.), Nuevas tendencias y modernos peligros de la política criminal. 2023. Pp. 117-153. Tirant lo Blanch.

DE LA CUESTA AGUADO, Paz. M., "Inteligencia artificial y responsabilidad penal", *Revista Penal México*, núms. 16 y 17, 2019, pp. 51-62.

GALÁN MUÑOZ, Alfonso., *Los ciberdelitos o delitos informáticos*, Ed. UOC, 2019.

GINER ALEGRIA, Cesar. Augusto., & DELGADO MORAN, Juan. José. Consideraciones criminológicas sobre el perfil del stalker y el acecho mediante ciberstalking. *Estudios en seguridad y defensa*, 12(24), 2017 19-35. https://doi.org/10.25062/1900-8325.250

GÓMEZ-DE-ÁGREDA, Ángel., FEIJÓO, Claudio., y SALAZAR-GARCÍA, Idoia. A., "Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deepfakes e inteligencia artificial", *Profesional de la información*, vol. 30, núm. 2, 2020. https://doi.org/10.3145/epi.2021.mar.16

GUTIÉRREZ PALACIO, Juan. David., "Concepto jurídico-penal de acción a partir de la inteligencia artificial", en *Estudios de Derecho penal, neurociencias e inteligencia artificial*, Ediciones de la Universidad de Castilla-La Mancha, 2023.

HALLEVY, Gabriel., "The Basic Models of Criminal Liability of AI Systems and Outer Circles", *SSRN Electronic Journal*, 2019, https://doi.org/10.2139/ssrn.3402527

HERNÁNDEZ GIMÉNEZ, María., "Inteligencia artificial y derecho penal", *Actualidad jurídica Iberoamericana*, núm. 10 bis, Valencia, 2019, pp. 792-843. https://dialnet.unirioja.es/servlet/articulo?codigo=6978830

HERNÁNDEZ LÓPEZ, José. Miguel., *Reglamento de Inteligencia Artificial. Incluye introducción, notas, cronología, webgrafía, bibliografía e índice analítico*, 1ª ed., J. M. Bosch Editor, Barcelona, 2024, pp. 1-551. https://doi.org/10.2307/jj.20522950

LACRUZ MANTECÓN, Miguel. L., "Panorama de la inteligencia artificial en la justicia y el derecho actuales", *Revista de derecho aragonés*, núm. 30, Zaragoza, 2024, pp. 97-137. https://doi.org/10.69592/978-84-1194-810-4

LIZ RIVAS, Lenny. Violencia y agresión entre iguales a través de las TICS: Cyberbulling. AlmaMater. Cuadernos de Psicosociobiología de la Violencia: Educación y Prevención, nº 5, 2024, Dykinson, 2024 pp. 89-105. https://doi.org/10.14679/3314

MAZURIER, Pablo, Andrés., DELGADO MORÁN, Juan, José & PAYA SANTOS, Claudio, Augusto. Gobernanza constructivista de la internet. *Teoría y Praxis*, 2019. 17(34), 107-130. https://doi.org/10.5377/typ.v1i34.14823

MINISTERIO DEL INTERIOR, *Informe sobre la Cibercriminalidad en España 2024*, Secretaría de Estado de Seguridad, Dirección General de Coordinación y Estudios, Madrid, 2025, pp. 1-61.

MIRÓ LLINARES, Fernando., "El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio", *Revista de Derecho Penal y Criminología*, Marcial Pons, 2012, pp. 1-337.

MIRÓ LLINARES, Fernando., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, 3.ª Época, n.º 20, 2018, pp. 87-130. https://doi.org/10.5944/rdpc.20.2018.26446

MONTESINOS GARCÍA, Ana., "Inteligencia artificial en la justicia con perspectiva de género: amenazas y oportunidades", *Actualidad Jurídica Iberoamericana*, n.º 21, 2024, p. 566-597.

NAVARRO DOLMESTCH, Roberto., "Inteligencia artificial como «actuante» en el derecho penal. Una primera aproximación", *Revista de Internet, Derecho y Política*, n.º 43, 2025, pp. 1-17. https://doi.org/10.7238/idp.v0i43.432787

ORTIZ DE ZÁRATE ALCARAZO, Lucía., "Sesgos de género en la inteligencia artificial", *Revista de occidente*, vol. 502, 2023, pp. 5-20.

ORTIZ DE ZÁRATE ALCARAZO, Lucía., y GUEVARA GÓMEZ, Ariana., "Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España", *Fundación Alternativas*, 2021, pp. 1-81.

PALACIOS GARCIA, María, Ángeles. y LIZ RIVAS, Lenny.El hostigamiento o delito de "stalking" en el trabajo. en Cuadernos de psicosociobiológia de la agresión: educación y prevención. Universidad Complutense de Madrid. 2022 Dykinson. https://doi.org/10.2307/j.ctv36k5cdb.13

PARRA SEPÚLVEDA, Dario., y CONCHA MACHUCA, Ricardo., "Inteligencia artificial y derecho. Problemas, desafíos y oportunidades", *Vniversitas*, vol. 70, 2021, https://doi.org/10.11144/Javeriana.vj70.iadp

PAYÁ SANTOS, Claudio. Augusto; RODRÍGUEZ GONZÁLEZ, Víctor; DOMÍNGUEZ PINEDA, Neidy. Zenaida; DIZ CASAL, Javier; FERNÁNDEZ RODRÍGUEZ, Juan. Carlos. & DELGADO MORÁN, Juan. José. (2025). Role of the Human Factor in the Cybersecurity Ecosystem. *Journal of Information Systems Engineering and Management*,10(4). https://doi.org/10.52783/jisem.v10i4.8983

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*, 2017, pp. 1-23.

RODRÍGUEZ SANTILLÁN, Samantha., "Esa persona no soy yo: la inteligencia artificial como un nuevo instrumento de violencia", *American University International Law Review*, núm. 3, 2025, pp. 737-768.

ROMEO CASABONA, Carlos. María., y RUEDA MARTÍN, María. Ángeles. (eds.), "Derecho Penal, Ciberseguridad, Ciberdelitos e Inteligencia artificial", *Estudios de Derecho Penal y Criminología*, Editorial Comares, 2023, pp. 1-158.

SIMÓ SOLER, Elisa., "Retos jurídicos derivados de la Inteligencia Artificial Generativa Deepfakes y violencia contra las mujeres como supuesto de hecho", *InDret Criminología*, 2023, núm. 2, pp. 493- 515. https://doi.org/10.31009/InDret.2023.i2.11

VALDEZ SILVA, Francisco. Antonio., "La culpabilidad jurídico-penal individual en la empresa: ¿realmente somos libres dentro de una organización empresarial compleja?", en *Estudios de Derecho penal, neurociencias e inteligencia artificial*, Ediciones de la Universidad de Castilla-La Mancha, 2023, pp. 87-95.