FACULTAD DE DERECHO

LA INTELIGENCIA ARTIFICIAL Y EL SUJETO DE DERECHO

Autor: Jaime Castillo Artacho

5º E-3, Grupo A

Área de Filosofía del Derecho

Madrid
Marzo de 2026

ÍNDICE

LISTADO DE ABREVIATURAS

AI Act: 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

AI HLEG: High-Level Expert Group on Artificial Intelligence (Grupo de Expertos de Alto Nivel en inteligencia artificial de la Comisión Europea)

Art./Arts.: Artículo/Artículos

BOE: Boletín Oficial del Estado

CC: Código Civil español

Cfr.: *Confert* (compárese)

CP: Código Penal (Ley Orgánica 10/1995, de 23 de noviembre)

DOUE: Diario Oficial de la Unión Europea

IA: Inteligencia artificial

Ibid.: *Ibidem* (en el mismo lugar)

LAWS: *Lethal Autonomous Weapon Systems* (sistemas de armas letales autónomos)

LPI: Texto Refundido de la Ley de Propiedad Intelectual (Real Decreto Legislativo 1/1996, de 12 de abril)

Op. cit.: *Opus citatum* (obra citada)

p./pp.: Página/Páginas

RGPD: Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de datos personales

SAE: SAE International (*Society of Automotive Engineers*)

STJUE: Sentencia del Tribunal de Justicia de la Unión Europea

STS: Sentencia del Tribunal Supremo

TJUE: Tribunal de Justicia de la Unión Europea

TRLGDCU: Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios (Real Decreto Legislativo 1/2007, de 16 de noviembre)

UE: Unión Europea

UNESCO: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Vid.: *Vide* (véase)

1. INTRODUCCIÓN

La irrupción de los sistemas de inteligencia artificial en prácticamente todos los ámbitos de la actividad humana plantea interrogantes de primer orden para la ciencia jurídica. Hoy existen sistemas capaces de adoptar decisiones con niveles crecientes de autonomía, desde la conducción automatizada de vehículos hasta la elaboración de diagnósticos médicos o la generación de contenidos creativos. Todos ellos desafían las categorías conceptuales sobre las que se ha edificado durante siglos la noción de sujeto de derecho.

La aprobación del Reglamento (UE) 2024/1689, conocido como AI Act, constituye el primer marco normativo integral destinado a regular el ciclo de vida de estos sistemas en el espacio europeo. Sin embargo, deja deliberadamente sin resolver la cuestión nuclear que vertebra el presente trabajo: si un sistema de inteligencia artificial puede, o debe, ser considerado sujeto de derecho y, en consecuencia, centro autónomo de imputación de derechos y obligaciones. La relevancia de esta pregunta no es meramente teórica. De su respuesta depende, en buena medida, la configuración del régimen de responsabilidad aplicable a los daños causados por sistemas dotados de autonomía decisoria, un problema que ni el AI Act ni los instrumentos complementarios de la Unión Europea han resuelto de forma concluyente.

La reflexión sobre la relación entre inteligencia artificial y subjetividad jurídica no es, en modo alguno, nueva. Desde las primeras propuestas formuladas en el seno de la filosofía de la mente hasta los debates más recientes suscitados por la Resolución del Parlamento Europeo de 2017, que llegó a sugerir la creación de una personalidad electrónica para los robots más sofisticados, la doctrina ha explorado de forma recurrente la posibilidad de extender a entidades no humanas las categorías reservadas tradicionalmente a las personas físicas y jurídicas.

Con todo, los avances tecnológicos registrados en los últimos años, en particular el desarrollo de modelos de aprendizaje profundo cuyo proceso de decisión resulta opaco incluso para sus propios diseñadores, han intensificado la urgencia de este debate. La insuficiencia de las herramientas dogmáticas clásicas se hace evidente cuando nos enfrentamos a fenómenos que escapan a la lógica de la imputación tradicional.

El presente escrito persigue un doble objetivo. Desde un punto de vista teórico, se propone analizar críticamente la viabilidad de reconocer a los sistemas de inteligencia artificial alguna forma de subjetividad jurídica, evaluando los fundamentos filosóficos,

dogmáticos y prácticos que podrían sustentar o desaconsejar dicho reconocimiento. Se buscará dar respuesta a la pregunta a través de casos concretos en los que los sistemas de inteligencia artificial son capaces de tomar decisiones con verdaderas consecuencias jurídicas.

Desde un punto de vista más práctico, el trabajo pretende, en primer lugar, construir un marco conceptual integrado que ponga en diálogo las definiciones institucionales de inteligencia artificial con las nociones de sujeto de derecho y responsabilidad propias de la tradición jurídica. En segundo lugar, examinar las analogías más frecuentemente invocadas, como las personas jurídicas o los menores e incapaces, para determinar su pertinencia y sus límites. Y, en tercer lugar, identificar los problemas de causalidad e imputación que la autonomía de estos sistemas genera en escenarios concretos, proponiendo criterios para su resolución.

En cuanto a la metodología, el trabajo se apoya en tres ejes complementarios de naturaleza esencialmente cualitativa. Se ha llevado a cabo, por un lado, una revisión bibliográfica de las principales aportaciones doctrinales en materia de filosofía del derecho, teoría general del derecho y filosofía de la acción buscando una respuesta en la tradición filosófica. Por otro lado, se ha realizado un análisis normativo que abarca tanto el Derecho de la Unión Europea, con particular detenimiento en el AI Act y en la ya mencionada Resolución del Parlamento Europeo de 2017, como las principales propuestas regulatorias comparadas. Finalmente, se ha recurrido al estudio de casos paradigmáticos como herramienta para contrastar las categorías teóricas con los problemas prácticos que la autonomía decisoria de estos sistemas presenta en la realidad.

2. MARCO CONCEPTUAL Y JURÍDICO

2.1. Concepto de inteligencia artificial

El término inteligencia artificial (IA) se ha convertido en un concepto cada vez más presente en el debate jurídico y político, pero su significado varía en función de la entidad o persona que lo formule. Desde la perspectiva de este trabajo, interesa especialmente cómo la IA se define en el cruce entre la informática, las instituciones internacionales y los marcos normativos recientes. De esta definición dependerá qué sistemas quedan dentro del ámbito de análisis y qué rasgos son relevantes para la agencia y la responsabilidad que más adelante se estudiarán.[1]

Es por ello que, en este primer apartado y a partir del estudio de distintas acepciones del término, buscaremos crear una definición rigurosa y práctica que pueda servir para entender mejor el papel y responsabilidad de la inteligencia artificial en el derecho moderno.

2.1.1. IA como disciplina y como sistemas

Un primer elemento común en la literatura académica y en los documentos institucionales es la distinción entre la IA como disciplina científica y la IA como conjunto de sistemas concretos.

Como disciplina, la IA se entiende como un campo de investigación que agrupa técnicas de aprendizaje automático, razonamiento automático, representación de conocimiento, búsqueda y optimización, así como robótica y sistemas ciberfísicos. La IA, en este sentido, supone un área de conocimiento que estudia y desarrolla métodos para que los sistemas puedan percibir, razonar y actuar en entornos complejos.[2]

Existe, además, una dimensión histórica y filosófica de esta disciplina. A lo largo de la historia de la IA, los científicos han desarrollado diferentes teorías sobre lo que involucra la IA, determinando distintos caminos para estudiarla y construirla.[3] Las teorías de desempeño se centran en preguntas como las siguientes: ¿Cuáles son los componentes

[1] High-Level Expert Group on Artificial Intelligence, *A Definition of AI: Main Capabilities and Scientific Disciplines*, European Commission, Bruselas, 2019, pp. 1-2; Martínez, M.V., *De qué hablamos, cuando hablamos de inteligencia artificial*, UNESCO, Montevideo, 2024, pp. 7-8.
[2] High-Level Expert Group on Artificial Intelligence, *op. cit.*, p. 6.
[3] Martínez, M.V., *De qué hablamos, cuando hablamos de inteligencia artificial*, UNESCO, Montevideo, 2024, pp. 8-9.

funcionales esenciales de un sistema capaz de exhibir inteligencia? ¿Cómo puede probarse la presencia y grado de inteligencia?

Las Teorías Estructurales o Funcionales estudian cuáles son los mecanismos por los que puede lograrse inteligencia. Las teorías existenciales buscan establecer las condiciones necesarias para que el comportamiento inteligente se produzca. Y las teorías contextuales pretenden establecer cuál es la relación entre el comportamiento inteligente y el entorno con el que la entidad inteligente debe contender.

Sin embargo, ni el legislador ni los reguladores operan con esta noción amplia de *IA como campo científico*, sino con la idea de sistemas de IA concretos. La UNESCO insiste en que, para efectos de políticas públicas, lo relevante son los sistemas concretos que simulan aspectos de la inteligencia humana, tales como la percepción, la solución de problemas o la interacción lingüística, a partir de datos, *hardware* y conectividad. El foco se desplaza así desde el proyecto científico general hacia aspectos socio-técnicos específicos que se integran en servicios, productos y procesos sociales.[4]

Esta distinción es central para un trabajo jurídico: la responsabilidad, la imputación y la regulación se refieren a sistemas desplegados y a los actores que los diseñan, comercializan o utilizan, no a la IA en abstracto. Por ello, a partir de aquí en este trabajo de investigación se hablará de sistemas de IA cuando se trate de analizar la agencia y los problemas de responsabilidad. La conceptualización de los sistemas de IA como elementos socio-técnicos específicos requiere considerarlos no solo en su dimensión funcional sino también en sus implicaciones para los derechos humanos, la seguridad, la equidad y la sostenibilidad.[5]

2.1.2. La definición técnica-institucional de la Comisión Europea

El documento del AI HLEG *A definition of AI: main capabilities and scientific disciplines* es probablemente la referencia académica-institucional más relevante en Europa para fijar el concepto de IA. En él se propone una definición actualizada que ha funcionado como base técnica para debates posteriores, incluidos los jurídicos:

> *Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from*

[4] Martínez, M.V., *op. cit.*, p. 11.
[5] Martínez, M.V., *op. cit.*, pp. 5-6.

this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions.[6]

De esta definición pueden destacarse varios elementos que serán centrales para este trabajo. En primer lugar, se habla de sistemas diseñados por humanos, lo que subraya que la IA no es un agente "natural", sino un producto de decisiones humanas de diseño, entrenamiento y despliegue. Después, se menciona la existencia de un objetivo complejo y de una actuación en el mundo físico o digital, de modo que la IA se caracteriza por estar orientada a metas y por intervenir efectivamente en entornos donde puede producir consecuencias relevantes. Por último, se enumeran tres grandes capacidades: percepción (recoger datos del entorno), interpretación o razonamiento sobre esos datos y decisión de acciones para alcanzar objetivos, a lo que se añade explícitamente la capacidad de aprender y adaptarse con la experiencia.[7]

El propio informe aclara, además, que, como disciplina científica, la IA incluye subcampos como el *machine learning* (incluyendo *deep learning* y *reinforcement learning*), el razonamiento automático (planificación, representación del conocimiento, búsqueda y optimización) y la robótica (control, percepción, sensores y actuadores) integrados en sistemas ciberfísicos. Estas referencias permiten respaldar técnicamente la idea de que no cualquier algoritmo es IA, sino solo aquellos sistemas que combinan capacidades de percepción, inferencia y acción (con posible aprendizaje) de cara a la interacción con un entorno especifico.

2.1.3. La perspectiva de la UNESCO

La UNESCO adopta una definición funcional similar a la de la Comisión Europea, pero con un énfasis particular en el carácter operativo y pragmático de los sistemas de IA. Define los sistemas de inteligencia artificial como programas y equipos informáticos diseñados por seres humanos que actúan en la dimensión física o digital mediante la percepción de su entorno para lograr un objetivo específico.[8]

Estos sistemas realizan su función a través de un ciclo completo: adquisición e interpretación de datos, razonamiento sobre el conocimiento derivado de esos datos,

[6] High-Level Expert Group on Artificial Intelligence, *op. cit.*, p. 6.
[7] *Cfr.* High-Level Expert Group on Artificial Intelligence, *op. cit.*, pp. 1-4.
[8] Martínez, M.V., *op. cit.*, p. 11.

tratamiento de la información resultante y decisión de las mejores acciones para alcanzar el objetivo establecido[9]. Un rasgo distintivo es que pueden adaptar su comportamiento mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores, diferenciándose así de sistemas meramente automatizados. Este aspecto será particularmente relevante más adelante ya que cabrá discutir qué responsabilidad tienen los sistemas dotados de inteligencia artificial en relación con esta adaptación a su entorno.

La perspectiva de la UNESCO no se agota en la descripción técnica.[10] La organización enfatiza que el desarrollo y despliegue de la IA plantea interrogantes fundamentales sobre ética, equidad y sostenibilidad. Ello refleja una concepción de la IA no como una tecnología neutral, sino como un fenómeno que genera tanto oportunidades como riesgos significativos.

La UNESCO sitúa la IA dentro de la Agenda 2030 de Desarrollo Sostenible, reconociendo que este sector impulsa la innovación tecnológica y ofrece oportunidades para abordar desafíos globales como el cambio climático, la gestión de recursos naturales y la educación[11]. Esta contextualización es relevante para el análisis jurídico; pues subraya que la IA debe ser regulada considerando sus implicaciones sociales, económicas y normativas.

Para el propósito de este trabajo, la perspectiva de la UNESCO aporta tres ideas integradas. En primer lugar, la IA está fundamentalmente ligada a un procesamiento progresivo de datos, información y conocimiento.[12] Los datos crudos se transforman en información cuando se les aplica contexto y estructura; esta información puede elevarse a conocimiento cuando se integra la capacidad de aplicarla a problemas concretos.

En segundo lugar, la IA se define funcionalmente por su capacidad operativa de realizar funciones que requerirían inteligencia humana: percepción refinada del entorno, razonamiento complejo sobre información procesada, toma de decisiones adaptativa y modificación de comportamiento en función de la experiencia[13].

[9] Martínez, M.V., *op. cit.*, p. 11.
[10] *Cfr.* Martínez, M.V., *op. cit.*, pp. 5-6.
[11] Martínez, M.V., *op. cit.*, pp. 5-6.
[12] *Cfr.* Martínez, M.V., *op. cit.*, pp. 13-14.
[13] Martínez, M.V., *op. cit.*, p. 11.

En tercer lugar, la IA debe ser entendida como un fenómeno ligado a contextos éticos, políticos y de desarrollo sostenible.[14] Sus implicaciones rebasan lo puramente tecnológico e impactan derechos humanos, procesos democráticos y trayectorias de desarrollo.

2.1.4. La definición legal de la Comisión Europea

Desde el punto de vista estrictamente jurídico, la referencia clave es el artículo 3 del Reglamento (UE) 2024/1689, conocido como AI Act, que contiene la definición legal de "sistema de IA" para el Derecho de la Unión. El precepto define el concepto de la siguiente manera:

> *AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*[15]

Aunque formulada en lenguaje normativo propio de un texto legislativo, esta definición incorpora los mismos elementos estructurales que la definición técnica del AI HLEG. A continuación, se desglosan estos elementos clave[16]. Primero, que el sistema es *machine-based* (basado en máquinas), lo que incluye tanto *software* puro como combinaciones de *hardware* y *software*. Segundo, que está diseñado para operar con distintos niveles de autonomía, reconociendo que hay sistemas fuertemente supervisados y otros con intervención humana limitada. Tercero, que puede mostrar adaptatividad tras el despliegue, es decir, modificar su comportamiento en función de nuevos datos o contextos, sin necesidad de una reprogramación completa. Por último, que a partir de las entradas que recibe, el sistema infiere cómo generar salidas que pueden tomar la forma de predicciones, contenidos, recomendaciones o decisiones con capacidad de influencia sobre entornos físicos o virtuales.

La definición del AI Act es especialmente útil para un análisis jurídico porque fija el umbral mínimo de lo que, en el ámbito de la UE, debe considerarse *sistema de IA* a

[14] *Cfr.* Martínez, M.V., *op. cit.*, pp. 5-6.
[15] 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 (Reglamento de Inteligencia Artificial), art. 3.1.
[16] *Cfr.* Reglamento (UE) 2024/1689, art. 3.1.

efectos de la aplicación del Reglamento.[17] No todo programa informático entra en esta categoría: quedan fuera, por ejemplo, *softwares* puramente deterministas que no realizan inferencias a partir de datos para generar salidas, ni muestran adaptatividad relevante.

Además, el AI Act vincula inmediatamente esta definición con un enfoque basado en el riesgo, clasificando los sistemas de IA en función de su potencial impacto sobre la seguridad, la salud, los derechos fundamentales y otros bienes jurídicos. Esto anticipa el papel central que el concepto de IA tendrá en el análisis de la responsabilidad: según cómo se defina el sistema, se le aplicarán unas u otras obligaciones en relación con su diseño, transparencia, supervisión y aseguramiento.[18]

2.1.5. Elementos comunes y delimitación frente a otros conceptos

A partir de estas tres referencias (AI HLEG, UNESCO y AI Act) se puede identificar un conjunto de rasgos comunes que permiten delimitar con cierta precisión qué se entiende por sistema de IA en el sentido relevante para este trabajo.

En cuanto al diseño humano, los sistemas de inteligencia artificial son, ante todo, artefactos concebidos y desarrollados por seres humanos, lo que los distingue de cualquier agente natural o espontáneo. Su existencia responde siempre a una voluntad deliberada de creación técnica.

Respecto al uso avanzado de datos, estos sistemas se caracterizan por su capacidad de absorber y procesar grandes volúmenes de información, ya sea estructurada o no estructurada, lo que les permite operar con una base informativa mucho más amplia y compleja que la de los sistemas de *software* tradicionales.

En lo que se refiere a la percepción e interpretación del entorno, los sistemas de IA recogen información sobre un medio, ya sea físico o digital, y la transforman en representaciones internas o en conocimiento operativo, desarrollando así una suerte de comprensión funcional del contexto en el que actúan.

En relación con la inferencia y decisión orientada a objetivos, todo lo anterior converge en la generación de salidas concretas (predicciones, recomendaciones,

[17] *Cfr.* Reglamento (UE) 2024/1689, art. 3.1.
[18] *Cfr.* Reglamento (UE) 2024/1689, arts. 3 y ss., en relación con los arts. 5 a 7 (clasificación por riesgo)..

contenidos o decisiones) dirigidas a la consecución de objetivos que pueden estar definidos de forma explícita o implícita por quienes los diseñan o utilizan.

Por lo que respecta a la autonomía relativa, estos sistemas operan con distintos grados de independencia en la selección de sus acciones: pueden ejecutar tareas sin necesidad de recibir instrucciones humanas paso a paso, aunque siempre dentro de los márgenes que su diseño les permite.

En cuanto a la adaptatividad, muchos de estos sistemas tienen la capacidad de aprender de la experiencia, ajustando y modificando su comportamiento con posterioridad a su despliegue inicial; esto les confiere una naturaleza dinámica que los diferencia de los programas informáticos convencionales.

Finalmente, sobre la capacidad de afectar a entornos y derechos, las salidas que generan estos sistemas pueden influir de manera significativa en entornos tanto físicos como virtuales, con un impacto potencial directo sobre personas, organizaciones y bienes jurídicos protegidos por el ordenamiento.

Esta caracterización permite a la vez delimitar la IA frente a conceptos próximos. Así, un programa que ejecuta reglas fijas sobre un conjunto reducido de datos, sin capacidad de aprendizaje, puede constituir una forma de automatización, pero no necesariamente inteligencia artificial en el sentido del AI Act.

Del mismo modo, un robot industrial que sigue trayectorias preprogramadas sin capacidad de adaptación representa un caso de robótica sin IA, mientras que, a la inversa, un modelo de lenguaje o un sistema de recomendación puramente digital puede ser IA sin ser robótica. Ambas categorías, por tanto, se solapan parcialmente pero no se identifican entre sí.

Para el análisis jurídico posterior interesa preservar esta frontera, porque los problemas específicos de agencia, y responsabilidad que aborda este trabajo de investigación se plantean, sobre todo, en relación con sistemas que combinan inferencia a partir de datos, cierto grado de autonomía y capacidad de adaptación.

2.1.6. Elección de una definición para esta investigación

Sobre la base de las fuentes comentadas se ha querido en este escrito elaborar una definición propia de inteligencia artificial:

> *En el contexto de este trabajo, se entenderá por sistema de inteligencia artificial todo sistema digital, diseñado por seres humanos, basado en técnicas computacionales avanzadas, que: (1) recibe y procesa datos procedentes de un entorno físico o digital, (2) infiere, a partir de esos datos, cómo generar salidas en forma de predicciones, recomendaciones, decisiones o contenidos dirigidos a la consecución de uno o varios objetivos, (3) opera con un cierto grado de autonomía en la selección de esas salidas según parámetros previamente fijados, y puede, en algunos casos, adaptar su comportamiento en función de la experiencia o de nueva información.*

Esta definición operativa retoma fielmente la estructura del AI HLEG (percepción de datos estructurados o no estructurados del entorno, interpretación y razonamiento sobre esa información, decisión de las mejores acciones para alcanzar objetivos complejos, y capacidad opcional de aprendizaje y adaptación analizando los efectos de acciones previas) al tiempo que enfatiza el diseño humano explícito como elemento fundacional.

Paralelamente, se alinea con el artículo 3 del AI Act (Reglamento UE 2024/1689), que conceptualiza los sistemas de IA como máquinas capaces de inferir salidas accionables (predicciones, contenidos, recomendaciones o decisiones) a partir de *inputs*. Esto le permite influir en entornos físicos o digitales con grados variables de autonomía y adaptatividad post-despliegue, lo que facilita su aplicación práctica en la regulación por riesgo.

Integra también la visión de la UNESCO (2023) sobre la IA como tecnología sustentada en datos masivos, algoritmos sofisticados y conectividad, diseñada para emular funciones cognitivas humanas pero siempre subordinada a propósitos humanos.

Desde la perspectiva del análisis jurídico de responsabilidad, que es el aspecto central de este trabajo, la definición de sistema de IA introduce deliberadamente el elemento de diseño humano. Esta característica que ancla la cadena inicial de responsabilidad en los diseñadores, proveedores y desplegadores humanos.

Sin embargo, esta delimitación no excluye la posibilidad de que la IA misma pueda, en ciertos contextos, asumir alguna forma de responsabilidad jurídica derivada de su capacidad de generar *outputs* autónomos con impacto causal.

La autonomía funcional del sistema (limitada a parámetros predefinidos pero no por ello menos generadora de decisiones impactantes) exige que se plantee la cuestión de si esa capacidad de actuar de manera independiente con respecto a la intervención humana inmediata justificaría atribuirle responsabilidad como sujeto jurídico, al menos de forma parcial o funcional.

Al centrar la atención en la generación de *outputs* con impacto causal en el mundo real se posibilita la evaluación de imputación de daños, la clasificación de niveles de riesgo (prohibidos, alto riesgo, riesgo mínimo), y la asignación de obligaciones de transparencia y explicabilidad. Estas cuestiones que este trabajo analiza sin prejuzgar si esa cadena de responsabilidad debe distribuirse únicamente entre actores humanos o si la IA podría ser también titular de obligaciones y, eventualmente, de responsabilidad.

Con todo esto, la definición elaborada se considerará eficiente para el propósito académico: excluye artefactos computacionales simples no orientados a objetivos complejos (como hojas de cálculo o *scripts*), pero resulta inclusiva para sistemas relevantes en litigios (sistemas de decisión autónoma, vehículos con IA embarcada, etc.). Así, proporciona coherencia transversal en capítulos posteriores para examinar si estos sistemas califican como *agentes* en sentido jurídico y determinar modelos óptimos de responsabilidad para los distintos agentes implicados en las decisiones tomadas por la IA.

A partir de este punto, toda mención a *sistema de IA* en este estudio remitirá explícitamente a esta definición operativa, asegurando un marco conceptual unificado para el estudio comparado de personalidad jurídica y regímenes de responsabilidad civil o penal en el contexto europeo, con especial atención a las implicaciones del AI Act y principios éticos del HLEG.

2.2. Sujeto de derecho y responsabilidad

En el apartado anterior se ha visto que la *inteligencia artificial* puede entenderse como un conjunto de sistemas concretos que procesan información, toman decisiones y actúan en entornos físicos o digitales. A partir de aquí, el objetivo de este epígrafe es aclarar con qué categorías jurídicas y filosóficas se van a analizar su posible responsabilidad: qué significa ser sujeto de derecho, qué quiere decir exactamente ser responsable y cómo se han utilizado históricamente estos conceptos en el derecho. Todo ello enfocado a aquellos aspectos que puedan interesar al objeto de este análisis.

Este apartado pretende, por tanto, construir un marco conceptual que sirva de base a todo el trabajo de investigación. En primer lugar, se revisará la idea de sujeto de derecho en la tradición jurídica, mostrando cómo el ordenamiento ha ido extendiendo esta categoría desde las personas naturales a las personas jurídicas y otros entes no humanos, y qué rasgos comunes comparten todos ellos como centros de imputación de derechos y obligaciones.

En segundo lugar, se van a analizar distintas concepciones de responsabilidad jurídica y moral, distinguiendo entre la capacidad básica de ser responsable y las consecuencias (sanciones, deber de reparar, reproche) que el derecho asocia a determinadas conductas. Finalmente, se incorporarán los aportes de la filosofía del derecho y de la acción sobre agencia colectiva y responsabilidad de entes artificiales, que permiten comparar el caso de la IA con otros "sujetos" creados por el derecho, como las personas jurídicas.

En definitiva, la idea es que, una vez fijado este marco, resulte más claro qué opciones hay cuando nos preguntamos si la IA puede ser responsable: si debe ser tratada como mero objeto de imputación indirecta a través de humanos, si puede asimilarse a una persona jurídica o si requiere un modelo distinto de agencia y responsabilidad.

2.2.1. El sujeto de derecho en la tradición iusfilosófica

La noción de sujeto de derecho es uno de los pilares fundamentales sobre los que se asienta el ordenamiento jurídico. En su acepción más básica, sujetos jurídicos son aquellos individuos cuya conducta es regulada por alguna regla jurídica, es decir, aquellos

a quienes el Derecho se dirige para permitirles, ordenarles o prohibirles determinados actos[19].

Sin embargo, esta definición aparentemente sencilla acaba por ser de una complejidad considerable, pues no solo los seres humanos han sido reconocidos históricamente como destinatarios de las normas. El Derecho Positivo, tanto en el ámbito español como en otros ordenamientos, ha extendido la condición de sujeto jurídico a entidades de diversas naturalezas: asociaciones, sociedades mercantiles, fundaciones, municipios, universidades e incluso patrimonios autónomos como las herencias yacentes[20].

Esta ampliación del concepto se debe a la necesidad práctica de regular la convivencia en una sociedad donde los seres humanos actúan no solo de manera individual, sino también colectivamente y a través de organizaciones complejas. Ya en el Derecho Romano clásico se reconocían las *universitas personarum* como entes formados por agrupaciones de hombres, si bien los derechos y deberes no se predicaban del ente en sí mismo, sino de sus integrantes[21]. Fue con las constituciones imperiales y la aparición de la Iglesia como sujeto jurídico capaz de administrar patrimonio cuando se amplió el elenco de posibles titulares de derechos más allá de la persona física[22].

Un antecedente decisivo de esta evolución se encuentra en la teoría del contrato social formulada por Thomas Hobbes en el *Leviatán* (1651). Hobbes distinguió entre personas naturales, cuyas palabras y acciones se consideran como propias, y personas artificiales, cuyas palabras y acciones representan las de otro. Sobre esta distinción construyó su teoría del pacto social: los individuos, incapaces de garantizar su propia supervivencia en el estado de naturaleza, acuerdan mediante un pacto mutuo autorizar y transferir su derecho de gobernarse a sí mismos a una sola persona o asamblea, generando así el Estado como persona artificial que actúa en nombre de todos.[23]

El gran salto conceptual se produjo con la pandectística germana del siglo XIX. Friedrich Karl von Savigny, partiendo de la premisa de que solo el ser humano individual es verdadero sujeto de derechos, articuló la teoría de la ficción: las personas jurídicas

[19] Hernández Marín, R., "Sujetos jurídicos, capacidad jurídica y personalidad jurídica", p. 96.
[20] *Ibid.*, pp. 96-97.
[21] Cubillos Garzón, C. E., "La persona jurídica. De Savigny a la jurisprudencia", *Revist@ e-Mercatoria*, vol. 22, n.º 1, 2023, p. 98.
[22] *Ibid.*, pp. 98-99.
[23] Hobbes, T., *Leviatán, o la materia, forma y poder de un Estado eclesiástico y civil*, trad. de C. Mellizo, Alianza Editorial, Madrid, 2018, caps. XVI-XVII, pp. 220 y ss.

serían seres creados artificialmente por el Derecho positivo, que no existen naturalmente de la misma manera que las personas físicas, sino únicamente para fines jurídicos[24]. Frente a esta posición, autores como Otto von Gierke sostuvieron que los grupos humanos organizados poseen una realidad propia que no puede desvanecerse en una mera ficción, dando lugar a las denominadas teorías de la personalidad real[25].

La depuración de este concepto llegó con Hans Kelsen y su Teoría Pura del Derecho. Para Kelsen, tanto la persona natural como la jurídica carecen de una existencia real independiente de las normas. La persona no es sino un centro ideal de imputación de derechos y deberes, un modo especial de designar unitariamente una pluralidad de normas que atribuyen situaciones jurídicas subjetivas[26]. Desde esta perspectiva normativista, el ser humano solo se transforma en elemento del contenido de las normas jurídicas cuando convierte algunos de sus actos en objeto de obligaciones, responsabilidades o derechos subjetivos[27]. Con ello, Kelsen elaboró la identificación del concepto de persona con el de centro de imputación normativa, eliminando toda referencia a una esencia previa al Derecho.

La tensión entre estas concepciones no es puramente histórica. Tradicionalmente se ha criticado a Kelsen por reducir a la persona jurídica exclusivamente a su dimensión formal. Esto impide comprenderla en su totalidad existencial; pues detrás del marco lógico-formal se mueven y actúan seres humanos que vivencian valores[28]. Para este autor, la personalidad jurídica posee una naturaleza tridimensional en la que interactúan conductas humanas intersubjetivas, valores y normas jurídicas, de modo que prescindir de cualquiera de estas dimensiones ofrece solo una visión fragmentada de la institución[29].

Esta concepción tiene consecuencias directas para el objeto del presente trabajo. Si la personalidad jurídica no es una mera etiqueta formal que el ordenamiento adhiere a cualquier entidad, sino una categoría que presupone conductas humanas reales y un sustrato axiológico que las orienta, su extensión a los sistemas de inteligencia artificial no puede resolverse mediante una simple decisión legislativa. Los sistemas de IA no

[24] Cfr. Savigny, F. K. von, *Sistema del Derecho Romano Actual*, T. I, p. 273 y T. II, p. 57; citado en Cubillos Garzón, C. E., *op. cit.*, p. 97.
[25] Fernández Sessarego, C., "Naturaleza tridimensional de la «persona jurídica»", p. 253.
[26] Kelsen, H., *La Teoría Pura del Derecho*, Losada, Buenos Aires, 1946, p. 83; citado en Fernández Sessarego, C., "Naturaleza tridimensional de la «persona jurídica»", p. 254.
[27] Figueroa Rubio, S., "Sobre la relación entre responsabilidad y normas jurídicas en el esquema kelseniano", *Revista Ius et Praxis*, Año 23, n.º 2, 2017, p. 386.
[28] Fernández Sessarego, C., *op. cit.*, pp. 255-256.
[29] *Ibid.*, pp. 263-264.

participan de relaciones intersubjetivas ni son portadores de valores, estos carecen de la capacidad de experimentarlos y de orientar su actuación conforme a ellos. Atribuirles personalidad jurídica supondría vaciar la categoría de dos de sus tres dimensiones constitutivas, reduciéndola a la dimensión puramente formal que se consideraba insuficiente. Ello no cierra necesariamente la puerta a toda forma de subjetividad para la IA, pero sí obliga a asumir que no se trataría de una extensión natural de la categoría existente, sino de la creación de una figura distinta.

2.2.2. Conceptos de responsabilidad jurídica y moral

Se ha visto en el punto anterior que la noción de sujeto de derecho admite lecturas divergentes. Para Kelsen, se trata de un centro de imputación normativa cuya configuración depende enteramente del ordenamiento. Otros autores critican la postura defendiendo que esa dimensión formal es necesaria pero insuficiente si no se integra con las conductas humanas y los valores que la sustentan. Esta tensión, lejos de ser un obstáculo, proporciona el punto de partida adecuado para examinar el otro gran pilar sobre el que descansa cualquier sistema de atribución jurídica: el concepto de responsabilidad. Solo comprendiendo qué significa ser responsable, en sus dimensiones jurídica y moral, podrá determinarse si alguna de estas concepciones del sujeto de derecho permite, o impide, trasladar tales categorías a entidades carentes de voluntad y conciencia como los sistemas de inteligencia artificial.

El término responsabilidad es, probablemente, uno de los más utilizados y, al mismo tiempo, más ambiguos del lenguaje jurídico. Su significado varía según se emplee en el terreno de las relaciones causales, de la moral, de la política o del Derecho, sin que pueda hallarse una definición clara, precisa y unívoca que abarque todas sus manifestaciones[30]. Hart puso de relieve esta pluralidad al detectar cuatro sentidos distintos: la responsabilidad como rol (*role-responsibility*), como relación causal (*causal-responsibility*), como capacidad o estado mental (*capacity-responsibility*) y como sancionabilidad propiamente dicha (*liability-responsibility*).[31] La clasificación de Hart es útil para este trabajo por dos razones. En primer lugar, porque revela que no toda conexión entre un sujeto y

[30] *Cfr.* Sanz Encinar, A., «El concepto jurídico de responsabilidad en la Teoría General del Derecho», *AFDUAM*, n.º 4, 2000, pp. 27-29.
[31] *Vid.* Hart, H. L. A., *Punishment and Responsibility: Essay in the Philosophy of Law*, Clarendon Press, Oxford, 1968, p. 211 y ss. Recogido en Sanz Encinar, A., *op. cit.*, pp. 39-42.

un daño deviene en responsabilidad jurídica: la mera relación causal, por ejemplo, no basta por sí sola para fundar un reproche. En segundo lugar, porque permite situar con precisión la posición de Kelsen dentro de ese mapa conceptual.

Desde la perspectiva normativista de Kelsen, la responsabilidad se identifica con el cuarto sentido de Hart: la sancionabilidad. Kelsen la define por la relación entre la sanción y el individuo contra el cual esta se dirige: es jurídicamente responsable quien debe soportar las consecuencias coactivas del ilícito. Esta formulación permite distinguir con nitidez entre obligación y responsabilidad: obligado es el individuo cuya conducta puede configurar el ilícito. Responsable, en cambio, es aquel sobre quien recae la sanción, pudiendo ambas posiciones recaer en personas distintas, como sucede en la responsabilidad vicaria o indirecta[32].

El análisis del Derecho positivo revela dos grandes sistemas: civil y penal. Dentro del civil, este se puede desgranar entre derecho civil subjetivo y objetivo. La responsabilidad penal y la responsabilidad civil subjetiva comparten un rasgo decisivo: ambas parten de la infracción de una norma y expresan un reproche jurídico que presupone la culpabilidad del agente. En el ámbito penal ese reproche se manifiesta de mayor manera, pues para imputar responsabilidad es preciso considerar al sujeto un agente moral, esto es, un ser dotado de capacidad de libre decisión[33]. La responsabilidad civil objetiva, por el contrario, prescinde de la infracción y de la culpabilidad, imponiéndose con independencia de la diligencia del agente. De ahí que diversos autores sostengan que este sistema no establece supuestos de auténtica responsabilidad, sino una obligación legal de reparar daños derivados de actividades lícitas generadoras de riesgo[34].

No obstante, frente a esta lectura cabe sostener que la responsabilidad objetiva constituye responsabilidad jurídica en sentido pleno, si bien su fundamento no es la culpa sino una decisión normativa de distribución del riesgo: quien crea el peligro o se beneficia de la actividad soporta sus consecuencias. Esta segunda lectura es la que subyace al régimen europeo de responsabilidad por productos defectuosos y, como se verá, al tratamiento de los daños causados por sistemas de inteligencia artificial.

[32] *Cfr.* Figueroa Rubio, S., *op. cit.*, pp. 390-393.
[33] *Cfr.* Sanz Encinar, A., *op. cit.*, pp. 29-30.
[34] *Cfr.* Sanz Encinar, A., *op. cit.*, pp. 31-32.

Precisamente en la frontera entre la responsabilidad jurídica y la moral se sitúa uno de los nudos conceptuales más relevantes para este trabajo. La responsabilidad, en cuanto juicio de reproche, es un concepto común al lenguaje moral y al jurídico que designa la condición de quien es objeto apropiado de una reprobación: solo puede reprocharse una conducta a quien es capaz de conocer el significado de sus actos y de actuar voluntariamente[35].

Ahora bien, el esquema kelseniano, al definir la responsabilidad desde la sanción y no desde la voluntad, abre la puerta a formas de imputación en las que la acción del sujeto responsable deja de ser condición necesaria, bastando un acontecimiento vinculable a una persona por su posición jurídica[36]. Esta tensión entre un concepto de responsabilidad arraigado en la agencia moral y otro construido sobre la pura imputación normativa constituirá el trasfondo teórico que los filósofos del derecho y de la acción, objeto del siguiente apartado, contribuirán a perfilar con mayor profundidad.

2.2.3. Aporte crítico de filósofos del derecho y de la acción

La tensión detectada al cierre del apartado anterior entre una responsabilidad fundada en la agencia moral y otra construida sobre la pura imputación normativa crea el punto de partida idóneo para incorporar al análisis las herramientas que ofrecen la filosofía del derecho y la filosofía de la acción. Su contribución es esencial porque, más allá de las categorías dogmáticas, obliga a preguntarse qué condiciones deben concurrir para que el juicio de responsabilidad conserve sentido.

Garzón Valdés sitúa el problema definiendo el enunciado de responsabilidad como una imputación de autoría de una relación causal que conduce a un estado de cosas con implicaciones normativas[37]. Para que dicho enunciado pueda formularse sería preciso un agente moral, es decir, un sujeto capaz de actuar voluntariamente tras un proceso de deliberación que presupone un ámbito de libertad enmarcado entre lo imposible y lo necesario[38]. Si se acepta esta premisa, la responsabilidad queda conceptualmente vinculada a la intencionalidad y a la voluntariedad del acto, de forma que un sistema incapaz de

[35] *Cfr.* Sanz Encinar, A., *op. cit.*, pp. 47-49
[36] *Cfr.* Figueroa Rubio, S., *op. cit.*, pp. 396-402
[37] *Cfr.* Garzón Valdés, E., «El enunciado de responsabilidad», *DOXA. Cuadernos de Filosofía del Derecho*, n.º 19, 1996, p. 262.
[38] *Cfr. Ibid.*, pp. 261-262

deliberar difícilmente satisfará las exigencias del enunciado de responsabilidad en su versión tipo.

Strawson profundiza en esta línea desde un ángulo distinto. A su juicio, la práctica de atribuir responsabilidad descansa en lo que denomina actitudes reactivas (resentimiento, gratitud, indignación), que son reacciones naturales de los seres humanos ante la buena o mala voluntad ajena[39]. Frente a quien se percibe como incapaz de participar en relaciones interpersonales ordinarias, como el niño o el demente, adoptamos una actitud objetiva, tratándolo como objeto de tratamiento o control, y no como destinatario de reproche[40]. Trasladada al ámbito de la inteligencia artificial, esta distinción plantea una pregunta relevante: ¿tiene sentido dirigir actitudes reactivas hacia una máquina, o su naturaleza nos conduce inevitablemente a la actitud objetiva?

Frankfurt introduce un matiz relevante al demostrar que la responsabilidad moral no exige necesariamente la posibilidad de haber actuado de otro modo[41]. Mediante sus contraejemplos muestra que un agente puede ser plenamente responsable aun cuando circunstancias inadvertidas le impidiesen elegir una alternativa, siempre que haya actuado por sus propias razones[42]. La tesis posee interés directo para los sistemas deterministas: el hecho de que un algoritmo no pueda desviarse de su programación no zanjaría por sí solo la cuestión de la responsabilidad si se adoptara un criterio frankfurtiano. No obstante, la aplicación de dicho criterio presupone la existencia de razones propias del agente, lo que reconduce el debate al problema de la intencionalidad.

Es aquí donde las posiciones de Searle y Dennett se contraponen con especial nitidez. Searle sostiene, a través de su argumento de la Habitación China (un experimento mental en el que una persona que desconoce el chino produce respuestas correctas en ese idioma limitándose a seguir un manual de reglas formales, demostrando que la manipulación sintáctica de símbolos no genera comprensión), que la mera ejecución de un programa informático no genera comprensión ni intencionalidad genuina; el sistema manipula símbolos formales sin acceder a su significado[43]. Dennett, en cambio, propone la

[39] *Cfr.* Strawson, P. F., «Freedom and Resentment», *Proceedings of the British Academy*, vol. 48, 1962, p. 9.
[40] *Cfr. Ibid.*, p. 10.
[41] *Cfr.* Frankfurt, H. G., «Alternate Possibilities and Moral Responsibility», *The Journal of Philosophy*, vol. 66, n.º 23, 1969, pp. 829-830.
[42] *Cfr. Ibid.*, pp. 835-836.
[43] *Cfr.* Searle, J. R., «Minds, Brains, and Programs», *Behavioral and Brain Sciences*, vol. 3, n.º 3, 1980, pp. 417-418.

actitud intencional como estrategia predictiva: podemos atribuir creencias y deseos a cualquier sistema complejo cuyo comportamiento resulte predecible mediante esa atribución[44]. Desde esta perspectiva, tratar a una IA como si fuera un agente intencional sería legítimo en la medida en que resulte instrumentalmente útil, sin necesidad de afirmar que posee una capacidad mental real.

Ambas posiciones ilustran el dilema central que este trabajo pretende abordar: o bien la responsabilidad requiere una intencionalidad genuina y entonces los sistemas de IA quedan excluidos del reproche, o bien basta una atribución funcional de agencia y entonces se abre la puerta a formas novedosas de imputación. Determinar cuál de estas vías es la más coherente con las categorías del ordenamiento jurídico exigirá examinar, en el apartado siguiente, las analogías que el propio Derecho ya ha construido para atribuir responsabilidad a entidades carentes de voluntad en sentido estricto, como las personas jurídicas o los menores de edad.

2.3. Analogías conceptuales aplicadas a la responsabilidad de la IA

El análisis desarrollado anteriormente ha puesto de manifiesto una tensión irresuelta en la teoría jurídica: mientras que la tradición filosófica vincula la responsabilidad a la agencia moral, exigiendo intencionalidad, voluntariedad y capacidad de deliberación, el Derecho positivo conoce desde hace tiempo formas de imputación que prescinden de tales requisitos y operan sobre la base de criterios estrictamente normativos. La pregunta que queda abierta es, por tanto, si alguna de esas construcciones jurídicas puede servir de modelo para entender la responsabilidad de los sistemas de inteligencia artificial, entidades que, como se ha visto, no satisfacen las condiciones clásicas de la agencia moral.

Para responder a esa pregunta se considera oportuno recurrir al método analógico, instrumento clásico del razonamiento jurídico. Como señala Atienza, la analogía permite resolver uno de los problemas básicos de cualquier ordenamiento: la innovación del sistema conservando su estructura, es decir, la adecuación de un conjunto de normas fijas a un medio social en constante transformación[45]. Su uso se fundamenta en el principio de igualdad o regla formal de justicia, conforme al cual deben recibir un tratamiento igual

[44] *Cfr.* Dennett, D. C., «Three Kinds of Intentional Psychology», en *The Intentional Stance*, MIT Press, Cambridge (Mass.), 1987, pp. 49-50.
[45] Atienza, M., «Algunas tesis sobre la analogía en el Derecho», *DOXA. Cuadernos de Filosofía del Derecho*, n.º 2, 1985, p. 224.

los casos que son semejantes en los aspectos relevantes[46]. La analogía no presupone necesariamente una laguna normativa, sino que puede operar también como procedimiento discursivo que extrae principios de una norma o de un grupo de normas para aplicarlos a supuestos nuevos[47].

En los apartados que siguen se examinarán dos analogías concretas del Derecho positivo que pueden arrojar luz sobre la cuestión. En primer lugar, se analizará la responsabilidad civil y penal de las personas jurídicas, entes a los que el ordenamiento atribuye responsabilidad pese a carecer de voluntad propia. A continuación, se explorará la comparación con la responsabilidad atribuida a menores y a sujetos con capacidad limitada. Finalmente, se evaluará la pertinencia y los límites de estas analogías para el caso específico de la inteligencia artificial.

2.3.1. Responsabilidad civil y penal de las personas jurídicas

La primera analogía que conviene explorar es la de la figura de la persona jurídica. El ordenamiento español atribuye responsabilidad, tanto civil como penal, a entidades que, por definición, carecen de conciencia, voluntad y capacidad de sufrimiento. Examinar cómo se ha construido dogmáticamente esa atribución es imprescindible para evaluar si las mismas técnicas podrían extenderse, *mutatis mutandis*, a los sistemas de inteligencia artificial.

En el ámbito penal, la evolución ha sido especialmente significativa. Durante largo tiempo rigió en España el principio *societas delinquere non potest*, conforme al cual solo las personas físicas podían ser sujetos activos de un delito. La razón dogmática era clara: si la responsabilidad penal exige acción voluntaria y culpabilidad individual, un ente colectivo carente de psique no puede satisfacer tales requisitos[48]. Sin embargo, la necesidad político-criminal de combatir la criminalidad económica organizada, cometida frecuentemente a través de estructuras societarias, llevó al legislador español a abandonar progresivamente aquel principio[49].

---

[46] *Ibid.*, p. 224.
[47] *Cfr.* Atienza, M., *op. cit.*, p. 228.
[48] Zugaldía Espinar, J. M., «Aproximación teórica y práctica al sistema de responsabilidad criminal de las personas jurídicas en el Derecho penal español», ponencia, Ministerio de Justicia, pp. 1-2.
[49] De la Cuesta Arzamendi, J. L., «Responsabilidad penal de las personas jurídicas en el Derecho español», *Revue électronique de l'AIDP*, 2011, A-05, p. 1.

El paso definitivo se produjo con la Ley Orgánica 5/2010, de 22 de junio, que introdujo el artículo 31 *bis* del Código Penal, estableciendo que las personas jurídicas serán penalmente responsables de los delitos cometidos en su nombre o por su cuenta, y en su provecho, por sus representantes legales y administradores de hecho o de derecho, así como de los cometidos por quienes, estando sometidos a la autoridad de estos, hubieran podido actuar por no haberse ejercido sobre ellos el debido control[50]. El precepto configuró un sistema de doble vía: una más grave cuando el hecho de referencia procede de los directivos de la entidad, y otra de menor gravedad cuando lo realizan empleados sobre los que no se ha ejercido la vigilancia debida[51].

La posterior reforma operada por la Ley Orgánica 1/2015 reforzó este régimen al introducir expresamente la posibilidad de exención de responsabilidad penal de la persona jurídica cuando esta hubiera adoptado, con carácter previo a la comisión del delito, modelos de organización y gestión eficaces para la prevención de delitos (los denominados programas de *compliance*).

En el plano doctrinal, la cuestión más debatida ha sido la fundamentación material de esta responsabilidad. Un sector de la doctrina sostiene que el modelo español responde a un esquema de heterorresponsabilidad o sistema vicarial, según el cual la responsabilidad penal de la persona jurídica se basa en la transferencia del hecho delictivo cometido por la persona física.

Frente a esta lectura, otro sector defiende un modelo de autorresponsabilidad, que localiza el fundamento de la responsabilidad penal corporativa en un injusto propio: el defecto de organización, es decir, la infracción del deber de la entidad de configurar su estructura interna de modo que prevenga la comisión de delitos en su seno[52]. La doctrina del defecto de organización, formulada por Tiedemann y de amplia aceptación, sitúa la culpabilidad de la persona jurídica en la omisión de las medidas de cuidado necesarias para garantizar un desarrollo lícito de su actividad[53].

En el ámbito civil, el ordenamiento dispone de mecanismos de imputación que prescinden de la culpa del agente y que resultan particularmente relevantes para el análisis

[50] *Vid.* art. 31 *bis* del Código Penal, introducido por la LO 5/2010, de 22 de junio, de reforma del Código Penal, y modificado por la LO 1/2015, de 30 de marzo.
[51] Zugaldía Espinar, J. M., *op. cit.*, pp. 10 y 17.
[52] De la Cuesta Arzamendi, J. L., *op. cit.*, pp. 14-16.
[53] *Cfr.* Zugaldía Espinar, J. M., *op. cit.*, p. 3.

de la responsabilidad por daños causados por sistemas de IA. La responsabilidad civil extracontractual, regulada con carácter general en los artículos 1902 y siguientes del Código Civil, exige en principio la concurrencia de culpa o negligencia. No obstante, la jurisprudencia ha evolucionado hacia una progresiva objetivación del sistema ya que, en casos caracterizados por una asimetría de información entre el causante del daño y la victima, exigir a esta última la carga de la prueba podría equivaler en la práctica a la privación de su tutela. La legislación especial ha consagrado supuestos de responsabilidad objetiva en los que basta acreditar el daño y el nexo causal[54].

El ejemplo más significativo es el régimen de responsabilidad por productos defectuosos, regulado en los artículos 135 a 146 del Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios, que traspone al Derecho español la Directiva 85/374/CEE[55]. Conforme a este régimen, el productor o fabricante responde de los daños causados por los defectos de sus productos con independencia de su culpa, debiendo el perjudicado probar únicamente el defecto, el daño y la relación de causalidad entre ambos.

La aplicación de esta normativa a los sistemas de IA, sin embargo, plantea ciertas dificultades: entre ellas, el obstáculo que supone calificar un algoritmo intangible como producto en el sentido legal, la imposibilidad de fijar un momento único de puesta en circulación cuando el sistema se transforma continuamente mediante aprendizaje automático, y la compleja delimitación de la excepción de riesgos del desarrollo en un ámbito marcado por la opacidad algorítmica[56].

La constatación de estas dificultades condujo a que, en el plano institucional europeo, se explorase la posibilidad de extender la lógica de la personalidad jurídica a los robots autónomos más complejos. La Resolución del Parlamento Europeo de 16 de febrero de 2017 propuso la creación de una "personalidad electrónica" específica para aquellos robots con capacidad de tomar decisiones autónomas o de relacionarse con terceros

[54] Laín Moyano, G., «Responsabilidad en inteligencia artificial: Señoría, mi cliente robot se declara inocente», *Ars Iuris Salmanticensis*, vol. 9, 2021, pp. 206-207.
[55] *Vid.* arts. 135 a 146 del Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios, que traspone la Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, sobre responsabilidad por productos defectuosos.
[56] Laín Moyano, G., *op. cit.*, pp. 211-212.

de forma independiente, de modo que pudieran ser directamente responsables de reparar los daños causados[57].

La propuesta, sin embargo, fue finalmente descartada. Más de doscientos expertos de catorce países se opusieron públicamente, advirtiendo de que la personalidad electrónica sobrevaloraría las capacidades reales de los sistemas de IA y podría servir de pretexto para eximir a los fabricantes de su responsabilidad. El Grupo de Expertos en Responsabilidad y Nuevas Tecnologías de la Comisión Europea concluyó en 2019 que no era necesario otorgar personalidad jurídica a los dispositivos autónomos, puesto que los daños que causan pueden y deben ser atribuidos a personas u organismos ya existentes[58].

El recorrido trazado muestra, en suma, que el Derecho ha desarrollado técnicas sofisticadas para atribuir responsabilidad a entes sin voluntad propia, tanto mediante la construcción de una culpabilidad organizativa en el ámbito penal como a través de la objetivación de la responsabilidad en el civil. No obstante, la analogía con la persona jurídica presenta límites estructurales que no pueden ignorarse: las personas jurídicas actúan siempre a través de personas físicas identificables, mientras que la autonomía creciente de los sistemas de IA difumina precisamente ese vínculo humano.

2.3.2. Comparación con la responsabilidad atribuida a menores o sujetos con capacidad limitada

Señalada la primera analogía, la de la persona jurídica, procede ahora examinar una segunda vía que el ordenamiento ofrece para pensar la responsabilidad de la inteligencia artificial: la que parte de los sujetos que, aun gozando de personalidad jurídica, ven restringida su capacidad de obrar por carecer del grado de madurez o discernimiento exigido para responder plenamente de sus actos. El régimen del menor de edad y, tras la reforma operada por la Ley 8/2021, el de las personas con discapacidad que precisan medidas de apoyo, son los referentes para realizar esta comparación[59]

En el Derecho civil español, la responsabilidad por los daños que causa un menor se articula fundamentalmente a través del artículo 1903 del Código Civil, que impone a

[57] Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). *Cfr.* Laín Moyano, G., *op. cit.*, pp. 204 y 217.
[58] Laín Moyano, G., *op. cit.*, pp. 221-222.
[59] *Cfr.* Ley 8/2021, de 2 de junio, por la que se reforma la legislación civil y procesal para el apoyo a las personas con discapacidad en el ejercicio de su capacidad jurídica (*BOE* n.º 132, de 3 de junio de 2021).

padres, tutores y guardadores una responsabilidad directa basada en la denominada *culpa in vigilando*: se presume que el daño ocasionado por el menor obedece a un defecto de supervisión o educación de quien debía velar por él.[60] La jurisprudencia del Tribunal Supremo ha endurecido progresivamente esta presunción hasta configurar, en la práctica, una responsabilidad prácticamente objetiva de los progenitores, de modo que la exoneración termina por ser extraordinariamente difícil[61].

A ello se añade la posibilidad de que el propio menor responda directamente *ex* artículo 1902 del Código Civil cuando, siendo civilmente imputable (esto es, cuando posea la madurez intelectual y volitiva suficiente para comprender el alcance de sus actos), haya obrado con culpa; la diligencia exigible se gradúa entonces comparando su conducta con la que cabría esperar de un menor normalmente desarrollado de su misma edad y en idénticas circunstancias.[62]

El sistema se completa, en el ámbito penal, con el régimen de la Ley Orgánica 5/2000, reguladora de la responsabilidad penal de los menores, cuyo Título VIII atribuye la responsabilidad civil solidaria a los padres, tutores, acogedores y guardadores del menor infractor[63]. Este aspecto resulta interesante porque los padres, tutores o guardadores no son penalmente responsables. Se les impone el deber de reparar el daño precisamente porque ocupan una posición de garante respecto al sujeto (menor) que actúa con autonomía pero sin pleno discernimiento.

La doctrina civilista ha explorado la posibilidad de trasladar esta estructura a la inteligencia artificial, concibiendo al sistema autónomo como una especie de "menor digital" cuyo fabricante, operador o usuario respondería en calidad de guardador. La idea atrae porque capta un aspecto esencial del problema: al igual que el menor, la IA actúa con cierta autonomía, pero carece de plena capacidad de discernimiento, lo que justificaría imputar la responsabilidad a quien la controla o se beneficia de su actividad.

Sin embargo, como advierte Laín Moyano, la capacidad de obrar no se deriva de la mera autonomía funcional, sino que exige un grado de madurez o discernimiento del

[60] *Vid.* art. 1903, párrafos segundo y tercero, del Código Civil español.
[61] Gómez Calle, E., «La responsabilidad civil del menor», *Derecho Privado y Constitución*, n.º 7, 1995, pp. 96 y ss.
[62] *Ibid.*, pp. 95-96.
[63] *Cfr.* Ley Orgánica 5/2000, de 12 de enero, reguladora de la responsabilidad penal de los menores, Título VIII (arts. 61 y ss.).

que los sistemas de inteligencia artificial carecen en sentido jurídico estricto.[64] En términos similares, Cerdeira Bravo de Mansilla recoge el rechazo de Lacruz Mantecón a la aplicación analógica del artículo 1903 del Código Civil a los robots. Este entiende que las relaciones paternofiliales, laborales o de parentesco en que se insertan los sujetos contemplados por la norma impiden una utilización racional de ese modelo para resolver los problemas que suscita la IA. Técnicamente, al no tratarse de personas, no cabría predicar del dueño del robot una culpa *in educando* o *in vigilando* respecto de un ente que no es susceptible de educación ni de vigilancia en sentido propio.[65]

Con todo, este rechazo merece alguna matización. Es cierto que el artículo 1903 presupone relaciones humanas que no pueden darse con una máquina. Pero lo que la norma protege, en el fondo, es una situación más básica: alguien controla la actividad de otro, se beneficia de ella y por eso responde cuando causa un daño. Esa situación sí se da entre el operador y el sistema de IA. En cuanto a la imposibilidad de una *culpa in educando*, el argumento es formalmente correcto, pero pasa por alto que el operador sí tiene deberes de configuración, entrenamiento y supervisión del sistema que cumplen una función comparable. No se educa a una máquina como a un hijo, pero sí se la entrena, se la prueba y se decide cómo y dónde desplegarla. La analogía, por tanto, no funciona de forma mecánica, pero el principio que la inspira sigue siendo útil: quien pone en circulación un ente autónomo incapaz de responder por sí mismo debe asumir las consecuencias de su actividad.

Parte de la doctrina ha propuesto, como alternativa o complemento, la analogía con los animales prevista en el artículo 1905 del Código Civil, que establece una responsabilidad objetiva del poseedor. Lacruz Mantecón advierte cierta semejanza entre robots y animales por tratarse en ambos casos de entes semovientes que actúan con cierta independencia (según sus instintos el animal, según sus algoritmos el robot), aunque finalmente rechaza la equiparación porque la norma no cubre los daños sufridos por el propio usuario y porque no existe en los animales una inteligencia análoga a la artificial.[66] La responsabilidad objetiva del art. 1905 CC funciona porque se presupone que el poseedor del animal es capaz de anticipar su comportamiento, algo que , debido a los problemas de

[64] Laín Moyano, G., *op. cit.*, p. 203
[65] Cerdeira Bravo de Mansilla, G., «Entre personas y cosas: animales y robots», *Actualidad Jurídica Iberoamericana*, n.º 14, 2021, pp. 39-40, recogiendo la posición de Lacruz Mantecón, M. L., *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*, Editorial Reus, Madrid, 2020, p. 131.
[66] *Cfr.* Cerdeira Bravo de Mansilla, G., «Entre personas y cosas…», *op. cit.*, p. 40, exponiendo y valorando la posición de Lacruz Mantecón.

caja negra, no se puede predicar en los casos de IA. Cerdeira matiza, con razón, que la diferencia más relevante radica en la capacidad de los animales para sentir, de la que los robots carecen, lo que impide considerarlos seres vivos o sintientes.[67] La diferencia es relevante para la responsabilidad ya que el poseedor puede razonablemente prever el comportamiento del animal, mientras que la IA, cuya autonomía es adaptativa y creciente, desborda este presupuesto.

Una tercera vía, quizás más rebuscada, es la que conecta los robots con los esclavos del Derecho romano clásico, carentes de libertad y personalidad jurídica propia, cuya actuación producía siempre efectos para el *dominus*; Lacruz Mantecón llega a predecir que "nuestra futura sociedad robótica va a ser semejante a una sociedad esclavista", proponiendo dotar a los robots de un patrimonio afecto (análogo al *peculium*) mediante un seguro obligatorio que cubra los daños que puedan causar.[68] La propuesta, si bien desplaza el foco a la reparación para la victima (que en términos prácticos es lo más importante), sigue sin resolver quién estaría obligado a aportar al fondo para constituirlo. Esto nos lleva otra vez al punto de partida de este escrito.

En definitiva, la analogía con el menor y, más ampliamente, con los sujetos de capacidad limitada permite iluminar un rasgo central de la IA, la combinación de autonomía operativa y ausencia de discernimiento propio, pero tropieza con un límite estructural que no puede ignorarse: el menor crece y adquiere progresivamente plena capacidad; el animal posee una naturaleza biológica jurídicamente reconocida; la IA, en cambio, no encaja perfectamente en ninguna de esas categorías. Siguiendo a Atienza, la analogía constituye un instrumento de la justicia formal que permite adecuar un conjunto de normas fijas a un medio en constante transformación, pero solo en la medida en que los casos comparados sean iguales en los aspectos que se estiman relevantes.[69] Determinar si las similitudes señaladas bastan para fundar un régimen de responsabilidad o si, por el contrario, las diferencias lo impiden, es precisamente la cuestión que habrá de abordarse a continuación.

---

[67] *Ibid.*, pp. 40-41.
[68] Cerdeira Bravo de Mansilla, G., «Entre personas y cosas…», *op. cit.*, pp. 36-37, glosando a Lacruz Mantecón, M. L., *Robots y personas…*, *op. cit.*, p. 132. *Vid.* también Díaz Alabart, S., *Robots y Responsabilidad Civil. Derecho Español Contemporáneo*, Editorial Reus, Madrid, 2018, *passim*.
[69] Atienza, M., *op. cit.*, p. 228.

2.3.3. Pertinencia y límites de estas analogías para el caso de la IA

Las páginas precedentes han examinado dos construcciones del Derecho positivo que permiten atribuir responsabilidad a entes desprovistos, total o parcialmente, de voluntad autónoma: la persona jurídica, el menor de edad y el animal. Procede ahora evaluar conjuntamente la pertinencia y los límites de ambas analogías cuando se proyectan sobre los sistemas de inteligencia artificial.

Las semejanzas que justifican el recurso analógico no son pocas. De la persona jurídica se extrae la idea de que la responsabilidad puede edificarse normativamente sobre un defecto de organización, sin necesidad de acreditar una voluntad psicológica en el ente al que se imputa el hecho. De acuerdo con Zugaldía Espinar, la culpabilidad por defecto de organización (construcción formulada por Tiedemann y de aceptación prácticamente generalizada en la doctrina) considera que la persona jurídica será culpable siempre que, a través de sus órganos o representantes, haya omitido tomar las medidas de cuidado necesarias para garantizar un desarrollo ordenado y no delictivo de la actividad de empresa[70]. En la misma línea, De la Cuesta Arzamendi sitúa el defecto de organización en el plano del injusto típico propio de la persona jurídica, vinculándolo a la infracción del deber de garantía de la entidad en la evitación de la comisión de delitos en su seno[71].

Trasladada a la IA, esta construcción ofrece un punto de partida valioso: la responsabilidad del desarrollador o del operador podría fundamentarse en la omisión de las medidas organizativas y de supervisión adecuadas para prevenir que el sistema cause daños. En ambos casos hay un ente que actúa con cierta autonomía operativa, en ambos casos puede producir resultados lesivos y en ambos casos existe un sujeto humano. Existe además el elemento introducido por la reforma de 2015 (los programas de *compliance*): si el administrador o, análogamente, el fabricante, programador o usuario, hubieran tomado las medidas necesarias para que no se produzca un resultado negativo causado por la inteligencia artificial, estos podrían quedar exentos de culpa. Esta es la misma lógica que subyace en el enfoque basado en riesgo del AI Act.

Del régimen de responsabilidad del menor se obtiene, a su vez, el principio de que cabe hacer responder a quien ejerce el control o se beneficia de la actividad de un sujeto

[70] Zugaldía Espinar, J. M., *op. cit*., pp. 3-4, recogiendo la doctrina de Tiedemann, K., «Die Bebüssung von Unternehmen nach dem 2. Gesetz zur Bekämpfung der Wirtschaftskriminalität», *Neue Juristische Wochenschrift*, n.º 41, 1988, pp. 1169 *ss*.
[71] *Cfr.* De la Cuesta Arzamendi, J. L., *op. cit.*, p. 8

carente de pleno discernimiento, conforme a la *culpa in vigilando* que fundamenta el artículo 1903 del Código Civil[72]. De esta analogía se extrae además un criterio de graduación que puede ser operativamente viable. En el régimen del menor, la responsabilidad de los padres es inversamente proporcional al grado de madurez del hijo. Trasladado a la IA, este criterio sugiere que la responsabilidad del operador debería intensificarse en proporción a la autonomía del sistema (cuanto más opaco es el funcionamiento de la inteligencia artificial, más diligente deberá ser el operador). Ahora bien, a diferencia del menor, que crece y pasa a responder por sí mismo, la responsabilidad de la IA siempre permanecerá en actores humanos.

Sin embargo, la analogía tropieza con límites estructurales que ninguna de las dos situaciones logra salvar. El primero y más profundo es la ausencia de agencia moral. Como ha subrayado Garzón Valdés, todo enunciado de responsabilidad retrospectivo condenatorio requiere un agente susceptible de ser reprochado por sus actos (un agente moral), y la admisión de la posibilidad de deliberación presupone conceptualmente que el sujeto puede alterar el curso de los acontecimientos dentro de un ámbito de libertad enmarcado por lo imposible y lo necesario[73].

Este requisito se satisface en la persona jurídica de modo indirecto, pues esta actúa siempre a través de personas físicas identificables que sí deliberan y en el menor se satisface de modo prospectivo, pues la imputabilidad civil se gradúa en función de la edad precisamente porque se presupone un horizonte de plena capacidad de entender y querer[74]. La IA, en cambio, carece estructuralmente de deliberación y voluntariedad: ni actúa a través de personas físicas cuyas decisiones puedan reconstruirse, especialmente cuando el aprendizaje automático genera resultados opacos incluso para su diseñador[75], ni experimenta un proceso de maduración análogo al del menor. Ante la IA, de acuerdo con lo argumentado previamente, solo es posible la actitud objetiva de Strawson, lo que supone un impedimento para la aplicación de toda analogía.

Tampoco las restantes comparaciones exploradas por la doctrina resuelven la dificultad. La equiparación con los animales, apoyada en el artículo 1905 del Código Civil,

[72] *Cfr.* Gómez Calle, E., *op. cit.*, pp. 95-96.
[73] Garzón Valdés, E., *op. cit.*, pp. 261-262.
[74] *Cfr.* Gómez Calle, E., «La responsabilidad civil del menor», *op. cit.*, pp. 95-96, donde se vincula la imputabilidad civil a la capacidad de entender y querer, y se explica la graduación de la diligencia conforme a la edad del menor.
[75] *Cfr.* Laín Moyano, G., *op. cit.*, p. 214.

fue desestimada por Lacruz Mantecón, entre otras razones, porque los animales poseen una capacidad de sentir de la que los robots carecen[76] y por el hecho de que su comportamiento instintivo permanece dentro de unos márgenes predecibles. La analogía con los esclavos del Derecho romano, pese a su atractivo, parte de un sujeto que, aun privado de personalidad jurídica, era persona humana[77].

Cerdeira recoge la posición mayoritaria de la opinión civilista española al concluir que los robots son cosas, aunque singulares, cuya singularidad les viene dada por una inteligencia y conciencia que son solo aparentes[78]. Desde esa premisa, la propuesta de configurarlos como un *tertium genus* entre personas y cosas resulta, como advirtió Díaz Alabart, inconducente a efectos prácticos, al dejar sin resolver qué régimen jurídico les sería aplicable[79].

Las limitaciones de estas analogías, si bien no anulan su utilidad como herramientas de comprensión, explican que las propuestas más recientes apunten hacia la construcción de un régimen específico. Lacruz Mantecón propone combinar la responsabilidad por productos defectuosos, para los vicios originarios del sistema, con un esquema inspirado en la responsabilidad por accidentes de circulación, complementado por un seguro obligatorio[80]. En el plano institucional, el propio legislador europeo ha descartado la vía de la personalidad electrónica propuesta por el Parlamento Europeo en 2017[81] y ha optado, en el Reglamento (UE) 2024/1689 (el denominado AI Act), por un enfoque basado en la gestión del riesgo que distribuye obligaciones entre proveedores, distribuidores y usuarios en función del nivel de peligrosidad del sistema.

En definitiva, las analogías examinadas son un instrumento valioso para iluminar aspectos parciales del problema, pero ninguna de ellas ofrece, por sí sola, una solución completa. La persona jurídica enseña que la responsabilidad puede prescindir de la voluntad individual y construirse sobre el defecto de organización; el menor muestra que cabe graduar la imputación conforme a la capacidad del agente; los animales y los

[76] *Cfr.* Cerdeira Bravo de Mansilla, G., *op. cit*., pp. 40-41.
[77] *Cfr.* Cerdeira Bravo de Mansilla, G., «Entre personas y cosas: animales y robots», *op. cit.*, pp. 36-38.
[78] Cerdeira Bravo de Mansilla, G., «Entre personas y cosas: animales y robots», *op. cit.*, p. 36, citando a Lacruz Mantecón, M. L., *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*, Editorial Reus, Madrid, 2020.
[79] *Cfr.* Cerdeira Bravo de Mansilla, G., «Entre personas y cosas: animales y robots», *op. cit.*, p. 34, recogiendo la advertencia de Díaz Alabart, S., *Robots y Responsabilidad Civil. Derecho Español Contemporáneo*, Editorial Reus, Madrid, 2018.
[80] *Cfr.* Cerdeira Bravo de Mansilla, G., «Entre personas y cosas: animales y robots», *op. cit.*, p. 42.
[81] Resolución del Parlamento Europeo, de 16 de febrero de 2017 (2015/2103(INL)), apartado 59, letra f).

esclavos romanos subrayan, cada uno a su modo, los límites de toda equiparación con un ente artificial. Como ha recordado Zugaldía Espinar, las categorías dogmáticas deben adaptarse a las nuevas necesidades sociales[82], y el fenómeno de la inteligencia artificial plantea exigencias que desbordan las construcciones pasadas.

Este hecho obliga a desplazar el análisis desde el plano teórico de las categorías dogmáticas al terreno de los problemas aplicados: los contextos específicos (vehículos autónomos, diagnósticos médicos, armas letales autónomas, creaciones generadas por IA) en los que la cuestión de la responsabilidad se plantea con mayor urgencia y en los que la insuficiencia de las analogías se manifiesta de forma más acusada serán objeto del capítulo siguiente.

3. LA IA COMO AGENTE Y EL PROBLEMA DE LA RESPONSABILIDAD

3.1. Casos paradigmáticos

El capítulo precedente ha dejado claro que las categorías tradicionales del Derecho (sujeto de derecho, imputabilidad, responsabilidad) fueron pensadas para un mundo en el que toda acción jurídicamente relevante podía reconducirse a una voluntad humana. Las analogías con las personas jurídicas y con los menores han servido para iluminar aspectos parciales del problema, pero ninguna ofrece una solución completa, porque la inteligencia artificial carece de intencionalidad genuina, no puede ser destinataria de actitudes reactivas y no posee un horizonte de maduración moral equiparable al del menor. Sentadas estas premisas, el presente capítulo da el paso de la abstracción al análisis aplicado: se toman las categorías del capítulo anterior y se confrontan con situaciones concretas en las que un sistema de IA opera con un grado de autonomía que dificulta, cuando no impide, la atribución de responsabilidad conforme a los esquemas clásicos. Como ha señalado Coeckelbergh, estas preocupaciones no son solo filosóficamente interesantes sino también muy prácticas y urgentes, puesto que los sistemas de IA ya permean la vida cotidiana y, en esa medida, todos los ciudadanos son pacientes morales con derecho a exigir explicaciones sobre las decisiones que les afectan[83].

[82] *Cfr.* Zugaldía Espinar, J. M., «Aproximación teórica y práctica…», *op. cit.*, p. 4
[83] *Cfr.* Coeckelbergh, M., *op. cit.*, p. 2066.

Se han seleccionado cinco casos paradigmáticos que responden a un doble criterio. Por un lado, representan sectores diversos (transporte, sanidad, defensa, propiedad intelectual y personalidad jurídica) que permiten una visión panorámica de los problemas de responsabilidad. Por otro, todos comparten un hilo conductor: la opacidad del sistema y su capacidad para adoptar decisiones no predeterminadas erosionan los presupuestos de la imputación tradicional[84]. Este efecto de caja negra no es un problema meramente técnico. En los términos aristotélicos que emplea Coeckelbergh, la opacidad algorítmica es una forma de ignorancia que compromete simultáneamente la condición de control y la condición epistémica de la responsabilidad: quien no comprende su instrumento difícilmente gobierna sus efectos, y quien delega una decisión en un sistema opaco no sabe, en rigor, qué está haciendo.[85]

3.1.1. Coches autónomos

El vehículo autónomo es, probablemente, el caso en el que mejor se aprecian las tensiones conceptuales analizadas en el capítulo segundo. No se trata ya de una hipótesis de laboratorio: existen coches con niveles intermedios de automatización circulando por ciudades como San Francisco o Hamburgo, se han producido accidentes mortales con implicación de estos sistemas y varios ordenamientos han empezado a legislar sobre la cuestión[86]. Todo ello justifica que sea el primero de los casos paradigmáticos y el que reciba un tratamiento más detenido.

Para entender el problema jurídico que plantea el coche autónomo es necesario partir de la escala de niveles de automatización de la SAE International, que va del nivel 0 (sin automatización alguna) al nivel 5 (automatización completa)[87]. En los niveles más bajos, del 0 al 2, el conductor sigue siendo quien conduce: recibe ayudas como el control de crucero o el frenado de emergencia, pero mantiene el control del vehículo en todo momento. El salto relevante se produce a partir del nivel 3, donde el sistema ya toma las decisiones de conducción de forma continuada y el humano pasa a ser un conductor de

[84] Ayo Ferrándiz, C.; Seijo Bar, Á.; Garre Anguera de Sojo, I.; González Guillén, P., «Responsabilidad civil e inteligencia artificial», *Actualidad Jurídica Uría Menéndez*, n.º 67, mayo 2025, p. 30.
[85] *Cfr.* Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», *op. cit.*, pp. 2054-2056 y 2058-2060
[86] Barrio Andrés, M., «Consideraciones jurídicas acerca del coche autónomo», *Actualidad Jurídica Uría Menéndez*, n.º 52, 2019, pp. 101-108, en p. 101.
[87] SAE International, *Recommended Practice J3016. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, junio de 2018. *Vid.* Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», *Revista de Derecho Civil*, vol. VII, n.º 5, 2020, pp. 175-223, en pp. 177-179.

reserva que solo interviene si el vehículo se lo pide[88]. Lo que ocurre es que esta premisa es en buena medida irreal, porque numerosos estudios demuestran que las personas no son capaces de mantener la atención supervisando tareas automatizadas que funcionan correctamente[89]. En los niveles 4 y 5 el problema se resuelve de raíz eliminando al conductor: en el nivel 4 el vehículo conduce solo en entornos acotados y en el nivel 5 lo hace en cualquier circunstancia[90]. Es aquí donde surge el verdadero problema jurídico, porque al desaparecer la figura del conductor desaparece también el sujeto sobre el que tradicionalmente ha recaído la responsabilidad por los accidentes de circulación[91].

Coeckelbergh cita lo que Matthias denominó "brecha de responsabilidad" a esta situación, recogiendo la formulación del autor: cuando los humanos carecen de control suficiente sobre máquinas que aprenden y actúan autónomamente, la condición de control aristotélica queda en entredicho[92]. En el nivel 5 no hay un humano que decida ni uno que supervise, lo que obliga a preguntarse no solo quién responde, sino si debe permitirse que desaparezca por completo el espacio para la agencia humana[93].

En el Derecho español, esa responsabilidad se canaliza a través del Real Decreto Legislativo 8/2004, que parte de la existencia de un conductor identificable[94]. Cuando no lo hay, la víctima debe acudir al régimen de responsabilidad por productos defectuosos del Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios (arts. 135-146)[95]. La doctrina mayoritaria admite que el *software* de conducción es producto a estos efectos[96], y la nueva Directiva (UE) 2024/2853 ha zanjado esta discusión al incluir expresamente los sistemas de IA en dicha categoría[97].

Resuelta esa cuestión, queda otra de mayor importancia: la imprevisibilidad inherente al *machine learning*. Los vehículos autónomos funcionan con algoritmos que aprenden y se modifican a medida que procesan datos nuevos, de manera que sus decisiones

[88] Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», *op. cit.*, pp. 178-179.
[89] *Ibid.*, p. 197.
[90] Barrio Andrés, M., «Consideraciones jurídicas acerca del coche autónomo», *op. cit.*, pp. 102-103.
[91] *Cfr.* Barrio Andrés, M., «Consideraciones jurídicas acerca del coche autónomo», *op. cit.*, p. 107.
[92] *Cfr.* Coeckelbergh, M., *op. cit.,* p. 2055
[93] *Ibid*., pp. 2055-2056.
[94] Real Decreto Legislativo 8/2004, de 29 de octubre, por el que se aprueba el texto refundido de la Ley sobre responsabilidad civil y seguro en la circulación de vehículos a motor.
[95] Real Decreto Legislativo 1/2007, de 16 de noviembre (TRLGDCU), arts. 135-146. *Vid.* Directiva 85/374/CEE del Consejo, de 25 de julio de 1985.
[96] Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», *op. cit.*, pp. 180-182.
[97] Ayo Ferrándiz, C.*et al*., *op. cit.*, pp. 33-36.

no siempre son previsibles ni siquiera para quien los diseñó[98]. La opacidad del *machine learning* se manifiesta aquí llegando al punto de que esta imprevisibilidad facilite al fabricante la invocación de la eximente de riesgos de desarrollo del artículo 140.1.e) del TRLGDCU Esta es la razón por la cual la doctrina más autorizada ha propuesto excluirla para los vehículos automatizados[99].

La exclusión resulta necesaria, el legislador no puede permitir que circulen vehículos cuyo comportamiento puede ser impredecible y, además, otorgar al fabricante una eximente basada en esa misma impredecibilidad. Permitirlo supondría un incentivo para opacar todo lo que se pueda el sistema, de forma que sea más fácil eludir la responsabilidad.

La pluralidad de agentes implicados (fabricante, desarrollador del *software*, proveedores de sensores, operador del servicio y usuario) actualiza lo que la literatura filosófica denomina el "problema de las muchas manos", al que Coeckelbergh añade una dimensión temporal: el *software* tiene una larga historia causal con múltiples desarrolladores en momentos distintos[100]. La propuesta de que respondiese el propio sistema fue descartada con amplio consenso en 2020[101]. Como se argumentó al examinar a Strawson en el apartado 2.2.3, ante un sistema de IA solo cabe la actitud objetiva; y si, como sostiene Coeckelbergh, estas tecnologías no satisfacen los criterios de agencia moral plena, la responsabilidad debe mantenerse en los humanos[102].

En el plano comparado, destaca la reforma alemana de la Straßenverkehrsgesetz de 2017, que incorporó la obligación de instalar una *caja negra* para registrar los traspasos de control, instrumento que sirve precisamente a la trazabilidad que la filosofía identifica como presupuesto de la explicabilidad[103]. Este sistema es acertado al dejar de depender

[98] Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», *op. cit.*, pp. 191-193.
[99] Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», *op. cit.*, pp. 215-216.
[100] *Cfr.* Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», *op. cit.*, pp. 2056-2057
[101] Resolución del Parlamento Europeo de 16 de febrero de 2017 (2015/2103(INL)). *Vid.* Resolución del Parlamento Europeo de 20 de octubre de 2020 (2020/2014(INL)), apartado 7. Cfr. Barrio Andrés, M., «Consideraciones jurídicas acerca del coche autónomo», *op. cit.*, p. 107.
[102] *Cfr.* Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», *op. cit.*, pp. 2054-2055.
[103] *Cfr.* Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», *op. cit.*, pp. 198-200.

del fabricante al reconstruir la cadena de decisión y convertirlo en una obligación legal verificable.

El vehículo autónomo pone de manifiesto, en definitiva, que basta con que la IA irrumpa en un sector regulado para que entren en crisis conceptos aparentemente consolidados. Pero bajo esta crisis técnico-jurídica late, como se ha visto, un desajuste más profundo: desde la perspectiva relacional de Coeckelbergh, la víctima de un accidente no es solo alguien que reclama una indemnización, sino un paciente moral con derecho a exigir razones de un agente humano responsable[104]. Estas tensiones reaparecerán en el siguiente apartado, donde la opacidad algorítmica se enfrenta ya no a la figura del conductor sino a la del profesional sanitario.

3.1.2. Diagnósticos médicos

La aplicación de sistemas de inteligencia artificial al ámbito del diagnóstico médico constituye uno de los escenarios donde con mayor nitidez se manifiestan las tensiones jurídicas analizadas en el capítulo anterior. La IA permite hoy analizar ingentes volúmenes de datos clínicos, interpretar imágenes médicas y detectar patrones patológicos con una precisión que, en determinadas especialidades como la radiología o la dermatología, iguala o supera las capacidades del juicio humano[105]. No obstante, estas herramientas no operan en un vacío normativo: el Reglamento (UE) 2024/1689 (el AI act) clasifica como de alto riesgo los sistemas de IA destinados a ser utilizados como componentes de seguridad de productos sanitarios sometidos a evaluación de conformidad por un organismo independiente, lo que incluye la práctica totalidad del *software* con fines de diagnóstico, pronóstico o elección terapéutica[106]. Esta clasificación impone al proveedor una serie de requisitos de obligado cumplimiento, entre los que destaca la supervisión humana efectiva prevista en el artículo 14 del Reglamento, conforme al cual dichos sistemas deben diseñarse de modo que puedan ser vigilados por personas físicas durante su uso.

En este contexto, el médico asume la posición de lo que podríamos denominar un guardián (*gatekeeper*) del proceso diagnóstico. Es él quien conserva la responsabilidad

[104] *Cfr.* Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», *op. cit.*, pp. 2061-2062 y 2066.
[105] *Cfr.* Lazcoz Moratinos, G., «Sistemas de inteligencia artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos (v.2)», *Revista de Derecho y Genoma Humano*, núm. 61, 2026, pp. 204-205.
[106] *Cfr.* art. 6.1 del Reglamento (UE) 2024/1689, en relación con el Anexo I, que incluye los Reglamentos (UE) 2017/745 y 2017/746 sobre productos sanitarios. *Vid.* Lazcoz Moratinos, G., «Sistemas de inteligencia artificial en la asistencia sanitaria…», *op. cit.*, pp. 245-246.

última sobre la decisión clínica, con independencia de que un algoritmo haya sugerido el resultado. La Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente, reconoce en su artículo 3 la figura del médico responsable como interlocutor principal del paciente en todo lo referente a su atención e información[107]. Este reconocimiento legal, lejos de verse debilitado por la irrupción de la IA, se refuerza: la Carta de Derechos Digitales declara expresamente que el empleo de sistemas digitales de asistencia al diagnóstico no limitará el derecho al libre criterio clínico del personal sanitario[108]. Sin embargo, esta posición de garante se enfrenta a un obstáculo considerable: la opacidad inherente a los sistemas de aprendizaje automático (las ya descritas cajas negras). El profesional sanitario debe, por tanto, ejercer un juicio crítico sobre resultados cuya fundamentación no puede conocer plenamente, lo que compromete la calidad epistémica de su supervisión.

El régimen jurídico aplicable a un eventual resultado lesivo derivado de un diagnóstico erróneo asistido por IA presenta una complejidad considerable, pues obliga a deslindar dos posibles focos de responsabilidad. Por un lado, la responsabilidad del profesional sanitario se articula en torno a la *lex artis ad hoc*, criterio valorativo de la corrección del acto médico que atiende a las circunstancias concretas del caso[109], y que encuentra su fundamento normativo tanto en los artículos 1902 y 1101 del Código Civil como, en el ámbito penal, en la construcción dogmática de la imprudencia profesional. La irrupción de la IA no suprime este estándar, sino que lo transforma: como ha señalado Gómez Rivero, el empleo de estos sistemas eleva los deberes de cuidado exigibles al sanitario, puesto que a los saberes tradicionales se suman los relativos al manejo del sistema y al juicio crítico sobre sus resultados[110].

Por otro lado, cuando el fallo provenga del propio sistema y no de su utilización por el profesional, entra en juego el régimen de responsabilidad por producto defectuoso, articulado en el Texto ya citado previamente. La coexistencia de ambos regímenes genera una tensión que el ordenamiento aún no ha resuelto de manera satisfactoria. La solución

[107] Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, art. 3.
[108] Gobierno de España, *Carta de Derechos Digitales*, 2021, Título XXIII, punto 4. *Vid.* Lazcoz Moratinos, G., «Sistemas de inteligencia artificial en la asistencia sanitaria…», *op. cit.*, p. 221.

[109] STS de 11 de marzo de 1991, núm. rec. 245/1987 (Roj: STS 13345/1991), Fundamento de Derecho segundo. *Vid.* Lazcoz Moratinos, G., «Sistemas de inteligencia artificial en la asistencia sanitaria…», *op. cit.*, p. 224.
[110] Gómez Rivero, M.ª C., «Inteligencia artificial aplicada a la salud…», *op. cit.*, pp. 156-157.

podría pasar por establecer una presunción legal clara por la cual, cuando el profesional haya seguido las indicaciones del sistema sin apartarse de ellas, la responsabilidad recaiga sobre el proveedor.

Un aspecto que merece atención diferenciada es el del consentimiento informado. La Ley 41/2002 reconoce en sus artículos 4 y 8 el derecho del paciente a conocer toda la información disponible sobre cualquier actuación en el ámbito de su salud como presupuesto para la prestación de un consentimiento válido. Cabe entonces preguntarse si el paciente debe ser informado de que un sistema de IA interviene en la elaboración de su diagnóstico. La respuesta, a la luz del artículo 22 del Reglamento General de Protección de Datos y del artículo 86 del propio AI Act, parece necesariamente afirmativa: el interesado tiene derecho a no ser objeto de decisiones basadas exclusivamente en tratamientos automatizados que le afecten significativamente, así como a obtener explicaciones sobre el papel del sistema en la toma de decisiones[111]. La información sobre la intervención algorítmica deviene así un componente necesario del consentimiento informado en la medicina asistida por IA.

Desde la perspectiva filosófica desarrollada en el primer apartado de este capítulo, el paciente médico es, además, un paciente moral en el sentido relacional del término empleado por Coeckelbergh: un sujeto que, inserto en una red de relaciones, tiene derecho a exigir razones y explicaciones sobre las decisiones que afectan a su salud. La opacidad algorítmica supone aquí una forma cualificada de aquella "ignorancia sobre el instrumento" que, como se analizó en el apartado anterior, compromete la condición epistémica de la responsabilidad. Esta circunstancia no invalida el recurso a la IA diagnóstica, pero sí exige rodear su uso de las garantías normativas ya señaladas: supervisión humana significativa, información al paciente y cotejo permanente de los resultados algorítmicos con las reglas de la experiencia clínica.

El ámbito sanitario ofrece así un escenario donde la brecha de responsabilidad identificada en los apartados precedentes se manifiesta con particular intensidad, si bien la persistencia de la figura del médico responsable permite, al menos por el momento, evitar que esa brecha se convierta en un vacío. Distinto es, como se verá a continuación,

[111] *Cfr.* art. 22 del Reglamento (UE) 2016/679 (RGPD), Considerando 71; art. 86.1 del Reglamento (UE) 2024/1689 (AI Act). *Vid.* Gómez Rivero, M.ª C., «Inteligencia artificial aplicada a la salud…», *op. cit.*, pp. 141-142.

el caso de los sistemas letales autónomos, donde la eliminación del factor humano en la cadena de decisión plantea problemas cualitativamente diferentes.

3.1.3. Sistemas de armas letales autónomos

Si en el ámbito sanitario la presencia del médico responsable permite amortiguar la brecha de responsabilidad generada por la opacidad algorítmica, los sistemas de armas letales autónomos (en adelante, LAWS) representan el escenario en que dicha brecha alcanza su máxima expresión. Se trata de sistemas de armas que, una vez activados, pueden seleccionar y atacar objetivos sin intervención humana adicional[112]. A diferencia del diagnóstico asistido por IA, donde el profesional conserva la última palabra sobre la decisión clínica, en los LAWS el ser humano es eliminado del bucle de decisión, de modo que la determinación de usar fuerza letal queda confiada íntegramente a la máquina[113]. La doctrina especializada describe este paradigma como el modelo percibir-pensar-actuar, en el que la máquina recopila datos del entorno, los procesa mediante algoritmos y ejecuta autónomamente la decisión adoptada[114]. Conviene señalar, además, que el artículo 2.3 del AI Act excluye parcialmente de su ámbito de aplicación los sistemas desarrollados con fines exclusivamente militares, lo que agrava la insuficiencia del marco regulatorio europeo frente a esta categoría de sistemas[115].

Desde la perspectiva del derecho internacional humanitario, la viabilidad jurídica de los LAWS tropieza con los principios de distinción y proporcionalidad. El principio de distinción exige discriminar en todo momento entre objetivos militares legítimos y población civil[116]. Como señaló Heyns, los LAWS presentan graves limitaciones para realizar esta distinción en los contextos de conflicto asimétrico y guerra urbana, donde la identificación del combatiente depende de la interpretación de conductas e intenciones inaccesibles a los sensores de una máquina[117]. González Calvache subraya que, si bien las redes neuronales son el mecanismo más prometedor para perfeccionar la distinción entre objetivos, son también el principal factor de impredecibilidad, pues aumentan el riesgo de que el sistema realice conexiones aleatorias no previstas por el programador[118].

---

[112] *Cfr.* Heyns, C., *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, A/HRC/23/47, Naciones Unidas, 2013, párr. 38.
[113] Ibid., párrs. 27 y 39-41.
[114] *Cfr.* González Calvache, L., «La responsabilidad del Estado por el uso de armas autónomas letales», *Revista Española de Derecho Militar*, núm. 118, 2022, p. 15
[115] *Cfr.* art. 2.3 del Reglamento (UE) 2024/1689 (AI Act).
[116] Protocolo I adicional a los Convenios de Ginebra, 1977, arts. 51 y 57.
[117] *Cfr.* Heyns, C., op. cit., párrs. 67-68.
[118] *Cfr.* González Calvache, L., op. cit., pp. 33-34.

Esto indica que la técnica y la controlabilidad operan en direcciones opuestas, algo paradójico a efectos de esta investigación y que requeriría mejorar técnicamente los sistemas de control.

El principio de proporcionalidad, por su parte, requiere una ponderación cualitativa entre el daño esperado a civiles y la ventaja militar anticipada, dependiente de nociones como el *sentido común* y el estándar del *comandante militar razonable*[119]. La citada autora insiste en que, incluso incorporando herramientas como la *Collateral Damage Estimate Methodology*, dicha ponderación es inherentemente subjetiva y no puede ser delegada en la máquina[120]. El propio Informe Heyns concluye que las decisiones sobre la vida y la muerte no deben delegarse en máquinas, y recomienda a los Estados la adopción de moratorias nacionales[121].

La cuestión de la responsabilidad penal se enfrenta a lo que Sparrow ha formulado como un trilema irresoluble: no cabe imputar la responsabilidad a la máquina, que carece de agencia moral, ni al programador, cuya conexión causal queda rota por la autonomía del sistema, ni al comandante, cuyas órdenes no determinan las acciones concretas del arma[122]. González Calvache añade que esta disolución se agudiza en el plano estatal, pues aunque el artículo 91 del Protocolo I Adicional permite atribuir al Estado los actos de sus fuerzas armadas, existe controversia doctrinal sobre si dicha norma resulta aplicable a decisiones adoptadas autónomamente por una máquina[123].

El trilema de Sparrow, sin embargo, se puede solucionar si no se busca un responsable único. La cuestión deja de ser quién es responsable y pasa a ser qué obligación incumplió cada actor en la cadena. El programador puede ser responsable por defectos en el algoritmo, el comandante por desplegar el sistema en un contexto inadecuado y el Estado por haber garantizado su uso sin las garantías suficientes para su correcta utilización. Esta solución sigue teniendo que lidiar con los problemas de caja negra, por lo que una cadena de responsabilidad similar a la legislada en Alemania debería implementarse para asegurar que se conoce la actuación de cada actor.

[119] *Cfr.* Heyns, C., op. cit., párrs. 70-72.
[120] *Cfr.* González Calvache, L., op. cit., pp. 35-36
[121] *Cfr.* Heyns, C., op. cit., párrs. 89-94 y 113.
[122] *Cfr.* Sparrow, R., «Killer Robots», *Journal of Applied Philosophy*, vol. 24, núm. 1, 2007, pp. 67-75.
[123] *Cfr.* González Calvache, L., op. cit., pp. 38-39; vid. también Geiss, R., «The International-Law Dimension of Autonomous Weapon Systems», *Friedrich Ebert Stiftung*, 2015, p. 22.

Es precisamente aquí donde el marco filosófico de Coeckelbergh adquiere su mayor fuerza explicativa. En los LAWS ambas condiciones aristotélicas de la responsabilidad se quiebran simultáneamente. La condición de control se desvanece porque la velocidad de procesamiento de la máquina supera con creces la capacidad de reacción humana, de modo que el operador no puede intervenir a tiempo ni corregir la decisión autónoma[124]. La condición epistémica fracasa igualmente: nadie, ni el programador ni el comandante, conoce de antemano lo que el sistema hará en una situación concreta, pues su comportamiento emerge de procesos de aprendizaje cuyo resultado es impredecible[125].

Desde la perspectiva relacional, la quiebra es aún más profunda. La responsabilidad, entendida como *answerability*, exige que un agente sea capaz de responder ante el agente moral, ofreciéndole razones que justifiquen lo ocurrido[126]. Cuando no hay ser humano en el bucle, no existe un agente identificable ante quien las víctimas puedan dirigirse como entes legítimos de esa relación: ni el agente puede dar razones, ni el paciente moral puede obtenerlas[127].

Los LAWS representan, en suma, el punto límite en que la automatización de decisiones sobre la vida humana hace colapsar el entero edificio de la responsabilidad, tanto en su vertiente jurídica como moral. Ello plantea la pregunta de si existen decisiones que, por su naturaleza, no deben ser delegadas en ninguna máquina, lo que equivale a postular la existencia de límites absolutos a la automatización. En este escrito la pregunta quedará sin resolverse ya que la búsqueda de una respuesta tendría una complejidad que, por si misma, podría tratarse por separado en otro trabajo de investigación.

3.1.4. Creatividad y autoría: el problema del copyright en las obras generadas por IA.

Si los sistemas de armas letales autónomos representan el escenario en que la brecha de responsabilidad alcanza su máxima expresión, el ámbito de la propiedad intelectual plantea una tensión estructuralmente análoga en un terreno distinto: la atribución de la creatividad. Cuando un sistema de IA generativa produce una composición musical, una obra plástica o un texto literario cuya calidad llega a ser indistinguible de la creación humana, el ordenamiento se ve obligado a responder una pregunta para la que no fue

[124] *Cfr.* Coeckelbergh, M., *op. cit.*, pp. 2055-2056.
[125] *Cfr.* Matthias, A., «The Responsibility Gap», *Ethics and Information Technology*, vol. 6, 2004, pp. 175-183, citado por Coeckelbergh, M., op. cit., p. 2055.
[126] *Cfr.* Coeckelbergh, M., op. cit., pp. 2061-2062.
[127] *Ibid.*, p. 2062.

diseñado: ¿puede existir una obra sin un autor humano que la sustente? La relevancia para el problema de la responsabilidad es directa, pues si no puede identificarse un autor, tampoco podrá determinarse un sujeto responsable de las eventuales infracciones que la obra genere.

El punto de partida normativo es claro. El artículo 5 del Texto Refundido de la Ley de Propiedad Intelectual establece que se considerará autor a la persona natural que cree una obra literaria, artística o científica, excluyendo así tanto a los animales como a las máquinas[128]. En el plano europeo, la sentencia del TJUE en el asunto Infopaq International (2009) consolidó un estándar de originalidad conforme al cual la obra protegible debe constituir una creación intelectual propia del autor, presupuesto derivado del Convenio de Berna y de las Directivas 91/250, 96/9 y 2006/116[129]. Como ha señalado Saiz García, de esta premisa se deriva que ni siquiera la perfecta emulación del cerebro humano por un sistema de IA permitiría calificar el resultado producido exclusivamente por la máquina como obra de ingenio[130]. En la misma línea, Valdezate Pelegrín subraya que las máquinas carecen de conciencia y emociones, facultades vinculadas intrínsecamente a la noción de autoría[131].

El marco normativo europeo vigente no ofrece una respuesta específica a este vacío. El Reglamento de Inteligencia Artificial (Reglamento (UE) 2024/1689) impone obligaciones de transparencia sobre los datos de entrenamiento, pero se centra en la fase de *input* y no en la titularidad ni en la responsabilidad sobre el *output* generado[132]. La Directiva sobre Derechos de Autor en el Mercado Único Digital (Directiva 2019/790) regula las excepciones de minería de textos y datos, pero tampoco aborda la autoría de las obras resultantes[133]. En junio de 2025, el Parlamento Europeo publicó un proyecto de informe instando a la Comisión a establecer un régimen de licencias, aunque sin resolver la

[128] Art. 5 del Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el Texto Refundido de la Ley de Propiedad Intelectual (en adelante, LPI).
[129] STJUE de 16 de julio de 2009, asunto C-5/08, *Infopaq International A/S c. Danske Dagblades Forening*, ECLI:EU:C:2009:465, apartados 34-37.
[130] *Cfr.* Saiz García, C., «Las obras creadas por sistemas de inteligencia artificial y su protección por el derecho de autor», *InDret*, núm. 1, 2019, p. 15.
[131] *Cfr.* Valdezate Pelegrín, P., «La autoría en creaciones generadas por Inteligencia Artificial», *Derecom*, núm. 37, 2024, pp. 28-29.
[132] Reglamento (UE) 2024/1689. *Vid.* considerando 105 y art. 53.
[133] Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital. Vid. arts. 3 y 4.

cuestión de fondo[134]. Se trata de avances significativos que, no obstante, siguen sin responder a la pregunta central: quién responde cuando no hay intervención humana alguna.

Ante esta laguna, la doctrina ha explorado diversas vías para identificar un sujeto responsable. Saiz García propone reconducir los supuestos con intervención humana organizativa relevante a la figura de la obra colectiva del artículo 8 LPI[135]. Navas Navarro amplía el abanico de soluciones partiendo de la distinción de Boden entre creatividad combinatoria, exploratoria y transformacional, y sostiene que los algoritmos genéticos pueden emular los dos últimos modelos hasta replicar casi idénticamente el proceso creativo humano[136]. De esta premisa extrae varias propuestas orientadas a localizar un responsable. En primer lugar, sugiere la ficción legal de un autor jurídico (la persona que encargó o utilizó el algoritmo) diferenciado del autor material, apoyándose en precedentes como el *work made for hire*[137]. En segundo lugar, propone un derecho *sui generis* análogo al del fabricante de bases de datos que compense la inversión realizada[138]. Finalmente, contempla un estatuto jurídico independiente que sustituya categorías como "obra" o "creación" por otras más neutras como "resultado" o "producción"[139]. Valdezate Pelegrín, por su parte, distingue entre autoría y titularidad: la primera correspondería al usuario que introduce las instrucciones, mientras que el desarrollador recibiría una compensación como titular de la licencia[140]. Esta distinción es especialmente relevante, pues permitiría señalar al usuario como potencial responsable de los contenidos y al programador como garante del sistema.

De las propuestas examinadas, la formulada por Valdezate Pelegrín es la más operativa a efectos de este escrito ya que permite señalar al responsable sin necesidad de crear ficciones sobre la creatividad de la máquina. El usuario que introduce las instrucciones responde por el contenido generado y el programador, por su parte, responde como garante de acuerdo con el régimen de productos defectuosos. Esta distribución es coherente con el modelo de responsabilidad escalonada ya explicado.

---

[134] Parlamento Europeo, Proyecto de informe sobre «Copyright and generative artificial intelligence – opportunities and challenges», junio de 2025.
[135] *Vid.* Saiz García, C., op. cit., pp. 25-26.
[136] *Vid.* Navas Navarro, S., «Obras generadas por algoritmos. En torno a su posible protección jurídica», *Revista de Derecho Civil*, vol. V, núm. 2, 2018, pp. 281-282.
[137] *Vid. ibid*., pp. 285-286.
[138] *Vid. ibid*., p. 286
[139] *Vid. ibid*., pp. 287-288.
[140] *Cfr.* Valdezate Pelegrín, P., op. cit., pp. 29-30.

En el plano comparado, la sección 9(3) del *Copyright, Designs and Patents Act* británico atribuye la autoría de las *computer-generated works* a quien realizó los trabajos preliminares necesarios, si bien, como advierte Navas Navarro, dicha expresión no aclara si tales arreglos deben ser realizados por un humano[141]. Esta ambigüedad no ayuda a resolver la cuestión de esta investigación al no resolver si ese titular es responsable del contenido.

Es aquí donde el marco de Coeckelbergh adquiere de nuevo su fuerza explicativa. La opacidad de los sistemas de aprendizaje automático, analizada en términos de quiebra de la condición epistémica, se proyecta sobre el proceso creativo: si ni los desarrolladores pueden explicar cómo la red neuronal llegó a una composición determinada, difícilmente podrá sostenerse que existe una creación intelectual propia en el sentido de la doctrina Infopaq[142]. La caja negra no solo impide atribuir responsabilidad por el daño, sino que socava la posibilidad de identificar un acto creativo imputable a un sujeto concreto.

Se reproduce así la misma tensión que recorre todo el capítulo: el sistema jurídico presupone un sujeto humano cuya identificación resulta cada vez más problemática. El régimen vigente carece de instrumentos para determinar quién es el titular de los derechos sobre una obra generada autónomamente y, sobre todo, quién debe responder por las infracciones que ocasione. La necesidad de una intervención legislativa que clarifique simultáneamente la atribución de derechos y la imputación de responsabilidades se presenta como una exigencia ineludible de seguridad jurídica.

3.2. Responsable de los daños: fabricante, programador, usuario o IA

Los casos examinados en el apartado anterior comparten un denominador común: en todos ellos existe un daño jurídicamente relevante, existe un sistema de inteligencia artificial cuya intervención ha sido causalmente decisiva para producirlo, y sin embargo no existe un sujeto al que el ordenamiento pueda imputar con plenitud la responsabilidad. La pregunta que vertebra este epígrafe es, precisamente, la que ha quedado pendiente a lo largo de todo el capítulo: ¿quién responde? Para intentar contestarla es necesario

[141] *Cfr. Copyright, Designs and Patents Act* 1988, secciones 9(3) y 178.
[142] *Vid.* Coeckelbergh, M., *op. cit*, pp. 2059-2060.

examinar a cada posible candidato, contrastando su posición con las categorías filosóficas y dogmáticas construidas en los capítulos anteriores.

3.2.1. Fabricante

El primer candidato es el fabricante o desarrollador del sistema. Su posición es, en principio, la más sólida, porque es quien toma las decisiones de diseño, selecciona los datos de entrenamiento y define los parámetros dentro de los cuales el sistema operará. Desde la perspectiva de Garzón Valdés, el fabricante satisface la condición básica del enunciado de responsabilidad: es un agente moral capaz de deliberar y de alterar el curso de los acontecimientos dentro de su ámbito de libertad[143]. La Directiva (UE) 2024/2853, sobre responsabilidad por los daños causados por productos defectuosos, refuerza esta vía al ampliar la noción de producto para incluir expresamente los programas informáticos y los sistemas de inteligencia artificial, manteniendo un régimen de responsabilidad objetiva en el que el perjudicado solo debe acreditar el defecto, el daño y el nexo causal[144].

Sin embargo, la imputación al fabricante tropieza con la dificultad filosófica de Coeckelbergh que ya se ha formulado previamente: la quiebra de la condición epistémica. Cuando un sistema de aprendizaje profundo evoluciona tras su despliegue y genera resultados que ni sus propios diseñadores pueden explicar ni anticipar, el vínculo entre la decisión originaria de diseño y el daño concreto se debilita hasta un punto en que la imputación deviene, cuando menos, problemática[145]. Si, como exige el propio Garzón Valdés, la responsabilidad retrospectiva condenatoria presupone que el agente pudo alterar el curso causal que condujo al estado de cosas dañoso[146], resulta difícil sostener esa imputación frente a un desarrollador que, por la propia naturaleza del *machine learning*, carece de control sobre los *outputs* concretos del sistema una vez desplegado.

La Directiva 2024/2853 intenta paliar este problema incorporando presunciones de defectuosidad y facilitando la carga probatoria del perjudicado[147], pero no resuelve la cuestión de fondo: el fabricante responde por un producto cuyo funcionamiento posterior escapa, al menos parcialmente, a su dominio.

---

[143] *Cfr.* Garzón Valdés, E., *op. cit.*, pp. 261-262.
[144] *Cfr.* Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos, arts. 4 y 8.
[145] *Cfr.* Coeckelbergh, M., *op. cit.*, pp. 2055-2056.
[146] *Cfr.* Garzón Valdés, E., «El enunciado de responsabilidad», *op. cit.*, p. 262.
[147] *Cfr.* Directiva (UE) 2024/2853, art. 10, apartados 2 a 4

3.2.2. Programador

El segundo candidato es el programador, entendido como la persona o equipo que escribe el código y diseña la arquitectura del algoritmo. Su posición presenta una dificultad adicional respecto del fabricante, pues la distancia entre la programación originaria y el daño es aún mayor. Como se ha señalado a propósito de los sistemas de armas letales autónomos, Sparrow ya advertía que cuando el sistema toma decisiones de forma genuinamente autónoma, el programador no puede ser considerado responsable de acciones que no pudo prever ni controlar[148]. En términos filosóficos, la cuestión puede plantearse desde el marco de Frankfurt: la responsabilidad moral no exige necesariamente la posibilidad de haber actuado de otro modo, pero sí requiere que el agente haya actuado por sus propias razones[149]. Pues bien, el programador actúa por sus propias razones al escribir el código, pero el daño no lo causa el código en abstracto, sino el comportamiento emergente del sistema tras un proceso de aprendizaje que, como ya se ha señalado, rompe la trazabilidad entre la instrucción humana y la acción de la máquina[150].

La condición epistémica se quiebra aquí de manera aún más acusada que en el caso del fabricante, porque el programador opera en un nivel de abstracción que no le permite anticipar las interacciones concretas del sistema con su entorno. Ello no significa que el programador quede exento de toda obligación, pues el deber de diligencia en la fase de diseño puede fundamentar una responsabilidad por culpa cuando se acredite que el código contenía un error o un sesgo evitable. Pero cuando el daño no deriva de un defecto de programación sino de la autonomía adaptativa del sistema, el nexo causal se desvanece.

3.2.3. Usuario

El tercer candidato es el usuario u operador. Su responsabilidad podría articularse a través de la *culpa in vigilando* o *in eligendo*, figuras que este escrito ha analizado a propósito de la analogía con los menores de edad. Sin embargo, la comparación revela sus propios límites. En el régimen del artículo 1903 del Código Civil, la *culpa in vigilando* presupone que el guardador tiene capacidad efectiva de supervisar la conducta del menor; la jurisprudencia la ha objetivado hasta el punto de presumirla casi irrefutablemente, pero

[148] *Cfr.* Sparrow, R., *op. cit.*, pp. 67-75
[149] *Cfr.* Frankfurt, H. G., *op. cit.*, pp. 835-836.
[150] *Cfr.* Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», *op. cit.*, pp. 2058-2060.

el presupuesto material sigue siendo la posibilidad de vigilancia[151]. En el caso de la inteligencia artificial, este presupuesto se cumple solo de forma parcial: un usuario que confía en la recomendación de un sistema de diagnóstico médico o en la sugerencia de un asistente legal carece, por definición, de los conocimientos técnicos necesarios para evaluar si el *output* del sistema es correcto o defectuoso.

La opacidad algorítmica que Coeckelbergh sitúa en el centro de su análisis no afecta únicamente al diseñador, sino también, y quizá de modo más agudo, al usuario, cuya condición epistémica es todavía más precaria[152]. Es cierto que, en contextos profesionales, como la sanidad o el ámbito militar, cabe exigir un deber de supervisión cualificado precisamente porque el profesional asume una posición de garante frente al paciente o al destinatario de los efectos. Pero incluso en estos supuestos, como se ha visto en el análisis de los diagnósticos médicos, la confianza excesiva en la máquina y la dificultad de contradecir una recomendación algorítmica cuyo fundamento se desconoce erosionan la efectividad de la supervisión humana. La figura del guardián, por tanto, mantiene su validez siempre que el profesional pueda contrastar el resultado a través de una explicación clínica de este. Este concepto se podría extrapolar a otros casos en los que la supervisión humana de los resultados del sistema sea posible.

3.2.4. Inteligencia artificial

El cuarto y último candidato es la propia inteligencia artificial. La posibilidad de atribuirle directamente la responsabilidad fue explorada, como ya se ha señalado, por la Resolución del Parlamento Europeo de 2017 a través de la figura de la personalidad electrónica[153]. La propuesta fue rechazada con contundencia tanto por la carta abierta suscrita por más de doscientos expertos como por el Grupo de Expertos en Responsabilidad y Nuevas Tecnologías de la Comisión, que consideró innecesario otorgar personalidad jurídica a dispositivos cuyos daños pueden y deben ser atribuidos a personas ya existentes[154]. Desde la óptica filosófica, el rechazo se justifica con mayor profundidad. Como ya se ha comentado, Strawson defiende que ante la IA solo cabe una actitud objetiva incompatible con el reproche. Searle, por su parte, mostró que la ejecución de un programa no genera comprensión ni intencionalidad genuina, sino mera manipulación formal de

[151] *Cfr.* Gómez Calle, E.,*op. cit*, pp. 96 y ss.
[152] *Cfr.* Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», *op. cit.*, pp. 2061-2062.
[153] Resolución del Parlamento Europeo, de 16 de febrero de 2017 (2015/2103(INL)), apartado 59, letra f).
[154] *Cfr.* Laín Moyano, G., *op. cit.*, pp. 221-222.

símbolos[155]. Incluso si se adopta la actitud intencional propuesta por Dennett como estrategia predictiva útil, ello no equivale a afirmar que el sistema posea estados mentales reales que sustenten un juicio de responsabilidad[156]. En el estado actual, la IA carece de patrimonio propio, de capacidad de sufrimiento y de los presupuestos mínimos de agencia moral que el ordenamiento exige para fundar la imputación.

El recorrido efectuado muestra que ninguno de los candidatos examinados satisface por sí solo las condiciones que la tradición filosófica y jurídica impone para una imputación plena de responsabilidad. El fabricante responde por el producto, pero la autonomía del sistema desborda su control; el programador diseña la arquitectura, pero el aprendizaje automático rompe el nexo causal con el daño; el usuario opera el sistema, pero carece de la condición epistémica para supervisarlo eficazmente; la IA misma no reúne los atributos de un agente moral.

La insuficiencia de cada solución individual explica que la doctrina y las instituciones europeas se orienten hacia modelos de responsabilidad escalonada por fases del ciclo de vida del sistema, combinados con mecanismos de socialización del riesgo como los seguros obligatorios y los fondos de compensación[157]. La propuesta de Directiva sobre responsabilidad en materia de inteligencia artificial, aún en fase de negociación, apunta en esta dirección al contemplar presunciones de causalidad y reglas de aligeramiento probatorio que faciliten al perjudicado el acceso a la reparación sin necesidad de identificar un único culpable[158]. No se trata, en suma, de inventar un nuevo sujeto de derecho, sino de distribuir la responsabilidad entre los agentes humanos que intervienen en cada fase, asumiendo que la opacidad y la autonomía de estos sistemas impiden perpetuar la ficción de un responsable único y plenamente identificable.

Este modelo de responsabilidad no está exento de riesgos. El principal problema de la dilución es concretar las obligaciones que corresponden a cada agente en cada fase y que se prevean, en su caso, mecanismos de responsabilidad solidaria para supuestos en que la opacidad del sistema imposibilite conocer las contribuciones de cada uno.

[155] *Cfr.* Searle, J. R., *op. cit.*, pp. 417-418.
[156] *Cfr.* Dennett, D. C., *op. cit.*, pp. 49-50.
[157] *Cfr.* Navarro-Michel, M., *op. cit.*, pp. 215-216.
[158] *Cfr.* Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (COM/2022/496 final), arts. 3 y 4.

3.3. Causalidad y responsabilidad en sistemas autónomos: distinción y consecuencias jurídicas.

El apartado anterior ha puesto de manifiesto que ninguno de los candidatos examinados satisface por sí solo las condiciones necesarias para una imputación plena de responsabilidad. Antes de extraer conclusiones definitivas conviene, sin embargo, detenerse en un problema conceptual previo que subyace a todas las dificultades analizadas: la distinción entre causalidad y responsabilidad. Confundir ambas categorías conduce a atribuir responsabilidad allí donde solo hay una conexión fáctica, o a negarla allí donde el ordenamiento debería construirla normativamente.

Hart advirtió que la responsabilidad como relación causal (*causal-responsibility*) y la responsabilidad como sancionabilidad (*liability-responsibility*) constituyen sentidos distintos de un mismo término, y que únicamente el segundo posee un significado jurídico autónomo[159]. En el ámbito de la causalidad, la pregunta es descriptiva: ¿qué cadena de acontecimientos produjo el daño? En el de la responsabilidad, la pregunta es normativa: ¿a quién debe el ordenamiento imputar sus consecuencias? Kelsen llevó esta distinción a sus últimas consecuencias al separar la imputación de la causalidad: mientras que esta describe una ley natural, la imputación es un nexo establecido por la norma que vincula un hecho con una consecuencia coactiva sin necesidad de que medie una conexión natural entre ambos[160].

Los sistemas de inteligencia artificial tensionan ambos planos de un modo sin precedentes. En el plano causal, el problema es epistémico, con la ya conocida opinión de Coeckelbergh acerca de la opacidad de los sistemas. La brecha de responsabilidad de Mathias se agudiza aquí al comprometer la cadena entre acción humana y resultado[161]. Sparrow proyectó esta misma lógica sobre las armas letales autónomas: si el robot selecciona sus propios objetivos, ni el programador ni el comandante han causado, en sentido estricto, la muerte[162].

Ahora bien, que la causalidad resulte difusa no significa que la responsabilidad deba quedar vacía. Si la imputación es un acto normativo y no una constatación fáctica, el Derecho puede construir mecanismos de atribución que prescindan de la reconstrucción

[159] *Vid.* Hart, H. L. A., *op. cit.*, p. 211 y ss. Recogido en Sanz Encinar, A., *op. cit.*, pp. 39-42.
[160] *Cfr.* Kelsen, H., *op. cit.*, p. 83. *Vid.* Figueroa Rubio, S., *op. cit.*, pp. 386-390.
[161] *Cfr.* Matthias, A., *op. cit.*, pp. 175-183, citado por Coeckelbergh, M., *op. cit.*, p. 2055.
[162] *Cfr.* Sparrow, R., *op. cit.*, pp. 67-75.

completa de la cadena causal. En esta dirección apuntan las presunciones iuris tantum introducidas por la Directiva (UE) 2024/2853 en materia de productos defectuosos[163] y por la propuesta de Directiva sobre responsabilidad en materia de IA, cuyo artículo 4 permite al juez presumir el nexo causal entre la culpa del demandado y el resultado producido por el sistema[164]. El legislador europeo asume que, en un entorno de opacidad algorítmica, exigir al perjudicado la prueba completa del nexo causal equivaldría a privarle de toda tutela.

Sin embargo, las presunciones no resuelven el problema de fondo, sino que lo desplazan. Alivian la carga probatoria, pero no determinan sobre quién debe recaer la responsabilidad. Y aquí reaparece la tensión filosófica central del trabajo. Garzón Valdés exige, para todo enunciado de responsabilidad condenatorio, un agente moral capaz de deliberar y de alterar el curso de los acontecimientos[165]. Si ningún agente humano satisface plenamente esa condición respecto del daño concreto, y la IA carece de agencia moral, nos encontramos ante un déficit estructural de imputación: el ordenamiento dispone de herramientas para facilitar la prueba, pero carece de un sujeto al que dirigir con plenitud el reproche. La consecuencia es que la responsabilidad solo puede articularse de forma distribuida, escalonando obligaciones de diligencia entre los agentes humanos de cada fase del ciclo de vida del sistema y complementando ese esquema con mecanismos de socialización del riesgo, como seguros obligatorios y fondos de compensación, que garanticen la reparación incluso cuando la identificación de un responsable individual resulte imposible[166].

La distinción entre causalidad y responsabilidad permite, en definitiva, superar el aparente callejón sin salida al que conduce la opacidad algorítmica. Que no podamos reconstruir completamente la cadena causal no significa que debamos renunciar a la imputación; significa que debemos construirla sobre bases normativas y no fácticas. El Derecho ya lo hace en otros ámbitos, desde la responsabilidad vicaria hasta la objetiva por riesgo. La novedad de la IA no está en el mecanismo, sino en la escala del problema y en la necesidad de combinar varios mecanismos simultáneamente.

[163] *Cfr.* Directiva (UE) 2024/2853, art. 10, apartados 2 a 4.
[164] *Cfr.* Propuesta de Directiva (COM/2022/496 final), art. 4.1.
[165] *Cfr.* Garzón Valdés, E., *op. cit.*, pp. 261-262.
[166] *Cfr.* Navarro-Michel, M., *op. cit.*, pp. 215-216.

4. CONCLUSIONES

A lo largo de este trabajo se ha intentado responder a una pregunta que, formulada en términos sencillos, admite una respuesta igualmente directa: ¿puede un sistema de inteligencia artificial ser sujeto de derecho? En el estado actual de la tecnología y del ordenamiento jurídico, la respuesta es negativa. Pero la relevancia de la pregunta no reside tanto en la respuesta como en lo que su formulación revela sobre las insuficiencias del propio sistema jurídico.

El Derecho fue construido sobre la premisa de que toda acción jurídicamente relevante puede reconducirse a una voluntad humana identificable. Las categorías de sujeto, imputación y responsabilidad descansan sobre esa premisa, y durante siglos ha funcionado razonablemente bien, incluso cuando el ordenamiento extendió la subjetividad a entidades sin voluntad propia como las personas jurídicas, porque detrás de ellas siempre era posible encontrar personas físicas cuyas decisiones podían reconstruirse. La inteligencia artificial quiebra esa premisa de un modo cualitativamente distinto. No se trata simplemente de que el sistema actúe sin voluntad, como ocurre con cualquier máquina, sino de que actúa con una autonomía adaptativa cuyas consecuencias resultan opacas incluso para quienes lo diseñaron. La brecha de responsabilidad que ello genera no es un problema técnico que pueda resolverse con más transparencia o mejores algoritmos de explicabilidad; es un problema estructural que afecta a los fundamentos mismos de la imputación jurídica.

La tentación de resolver ese problema creando un nuevo sujeto de derecho, una personalidad electrónica que permita dirigir la responsabilidad directamente contra la máquina, es comprensible pero profundamente equivocada. Como se ha argumentado a lo largo del trabajo, la responsabilidad jurídica en su sentido más robusto presupone un agente al que quepa dirigir un reproche, y el reproche solo tiene sentido frente a quien puede participar en las relaciones interpersonales que, según Strawson, sustentan la práctica misma de la responsabilidad. Atribuirle personalidad jurídica no resolvería el problema de la víctima, porque un ente sin patrimonio autónomo ni capacidad de sufrimiento no ofrece ninguna garantía real de reparación. Y, lo que es más grave, podría servir de coartada para diluir la responsabilidad de quienes sí pueden y deben responder: los seres humanos y las organizaciones que diseñan, comercializan, despliegan y se benefician de estos sistemas.

La solución, por tanto, no pasa por inventar ficciones jurídicas, sino por asumir con honestidad que el modelo clásico de responsabilidad individual es insuficiente para este fenómeno. La opacidad algorítmica fractura la cadena causal y la autonomía adaptativa desborda el control de cualquier agente singular. El camino más coherente, tanto con la tradición filosófica examinada como con las iniciativas normativas europeas más recientes, es el de una responsabilidad distribuida que escalone las obligaciones de diligencia entre los distintos agentes humanos que intervienen en cada fase del ciclo de vida del sistema. El fabricante debe responder por las decisiones de diseño y por los datos de entrenamiento; el programador, por los defectos evitables de la arquitectura; el operador profesional, por el incumplimiento de su deber de supervisión cualificado. De modo análogo a los mecanismos de *compliance,* la adopción de medidas preventivas deberían actuar como causa de atenuación de la responsabilidad. Allí donde la opacidad del sistema impida identificar al responsable concreto, los mecanismos de socialización del riesgo, seguros obligatorios y fondos de compensación, deben garantizar que la víctima no quede desprotegida. A estos mecanismos debe sumarse la imposición de obligaciones de trazabilidad que permitan reconstruir la cadena de decisiones del sistema y facilitar la identificación del responsable.

Este modelo no es perfecto. Exige del legislador un esfuerzo de concreción considerable para determinar qué obligaciones corresponden a cada agente, cómo se articula la prueba en un entorno de opacidad y cómo se financian los mecanismos de compensación colectiva. Pero tiene una virtud que ninguna otra propuesta ofrece: mantiene la responsabilidad firmemente anclada en los seres humanos, que son los únicos agentes morales del sistema, y evita la ilusión peligrosa de que la tecnología pueda responder por sí misma de las consecuencias que produce. El reto último no es solo jurídico ni técnico, sino filosófico: aceptar que hemos creado entidades capaces de actuar en el mundo con consecuencias reales y, al mismo tiempo, asumir que la responsabilidad por esas consecuencias sigue siendo, y debe seguir siendo, enteramente nuestra.

5. BIBLIOGRAFÍA

Legislación

Circular 1/2016 de la Fiscalía General del Estado, sobre la responsabilidad penal de las personas jurídicas conforme a la reforma del Código Penal efectuada por Ley Orgánica 1/2015.

Código Civil español (Real Decreto de 24 de julio de 1889).

Copyright, Designs and Patents Act 1988 (Reino Unido), secciones 9(3) y 178.

Dictamen del Comité Económico y Social Europeo, de 31 de mayo de 2017, sobre inteligencia artificial.

Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital.

Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos y por la que se deroga la Directiva 85/374/CEE del Consejo. DOUE L, 2024/2853, 18.11.2024.

Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros en materia de responsabilidad por los daños causados por productos defectuosos.

Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos.

Gobierno de España, Carta de Derechos Digitales, 2021, Título XXIII, punto 4.

Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, arts. 3, 4 y 8.

Ley 8/2021, de 2 de junio, por la que se reforma la legislación civil y procesal para el apoyo a las personas con discapacidad en el ejercicio de su capacidad jurídica (BOE n.º 132, de 3 de junio de 2021).

Ley Orgánica 1/2015, de 30 de marzo, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal.

Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal.

Ley Orgánica 5/2000, de 12 de enero, reguladora de la responsabilidad penal de los menores.

Ley Orgánica 5/2010, de 22 de junio, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal.

Parlamento Europeo, Proyecto de informe sobre «Copyright and generative artificial intelligence – opportunities and challenges», junio de 2025.

Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (COM/2022/496 final).

Protocolo I adicional a los Convenios de Ginebra de 1949 relativo a la protección de las víctimas de los conflictos armados internacionales, 8 de junio de 1977.

Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el Texto Refundido de la Ley de Propiedad Intelectual.

Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el Texto Refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias.

Real Decreto Legislativo 8/2004, de 29 de octubre, por el que se aprueba el texto refundido de la Ley sobre responsabilidad civil y seguro en la circulación de vehículos a motor.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales (RGPD).

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 (Reglamento de Inteligencia Artificial). DOUE L, 2024/1689, 12.7.2024.

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

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

Straßenverkehrsgesetz (Alemania), reformada por la Ley de 17 de junio de 2017 (BGBl. I S. 2421), §§ 1a, 1b y 63a.

Jurisprudencia

STJUE de 16 de julio de 2009, asunto C-5/08, Infopaq International A/S c. Danske Dagblades Forening, ECLI:EU:C:2009:465.

STS de 11 de marzo de 1991, núm. rec. 245/1987 (Roj: STS 13345/1991), Fundamento de Derecho 2.º

Thaler v. Perlmutter, 687 F. Supp. 3d 140 (D.D.C. 2023), confirmada por el D.C. Circuit en 2025.

Obras doctrinales

Atienza, M., «Algunas tesis sobre la analogía en el Derecho», DOXA. Cuadernos de Filosofía del Derecho, n.º 2, 1985, pp. 223-229.

Ayo Ferrándiz, C.; Seijo Bar, Á.; Garre Anguera de Sojo, I.; González Guillén, P., «Responsabilidad civil e inteligencia artificial», Actualidad Jurídica Uría Menéndez, n.º 67, mayo 2025.

Barrio Andrés, M., «Consideraciones jurídicas acerca del coche autónomo», Actualidad Jurídica Uría Menéndez, n.º 52, 2019, pp. 101-108.

Cerdeira Bravo de Mansilla, G., «Entre personas y cosas: animales y robots», Actualidad Jurídica Iberoamericana, n.º 14, febrero 2021, pp. 14-53.

Coeckelbergh, M., «Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability», Science and Engineering Ethics, vol. 26, 2020, pp. 2051-2068.

Cubillos Garzón, C. E., «La persona jurídica. De Savigny a la jurisprudencia», Revist@ e-Mercatoria, vol. 22, n.º 1, 2023, pp. 93-113.

De la Cuesta Arzamendi, J. L., «Responsabilidad penal de las personas jurídicas en el Derecho español», Revue électronique de l'AIDP / Electronic Review of the IAPL / Revista electrónica de la AIDP (ISSN 1993-2995), 2011, A-05, pp. 1-29.

Dennett, D. C., «Three Kinds of Intentional Psychology», en The Intentional Stance, MIT Press, Cambridge (Mass.), 1987, pp. 43-68.

Díaz Alabart, S., Robots y Responsabilidad Civil. Derecho Español Contemporáneo, Editorial Reus, Madrid, 2018.

Fernández Martín-Granizo, M., Los daños y la responsabilidad objetiva en el derecho positivo español, Aranzadi, Pamplona, 1972.

Fernandez Sessarego, C., «Naturaleza tridimensional de la "persona jurídica"», Derecho PUCP, núm. 52, 1999, pp. 251-269.

Figueroa Rubio, S., «Sobre la relación entre responsabilidad y normas jurídicas en el esquema kelseniano», Revista Ius et Praxis, Año 23, n.º 2, 2017, pp. 383-412.

Frankfurt, H. G., «Alternate Possibilities and Moral Responsibility», The Journal of Philosophy, vol. 66, n.º 23, 1969, pp. 829-839.

Garzón Valdés, E., «El enunciado de responsabilidad», DOXA. Cuadernos de Filosofía del Derecho, n.º 19, 1996, pp. 259-286.

Geiss, R., «The International-Law Dimension of Autonomous Weapon Systems», Friedrich Ebert Stiftung, Alemania, octubre de 2015.

Gómez Calle, E., «La responsabilidad civil del menor», Derecho Privado y Constitución, n.º 7, septiembre-diciembre 1995, pp. 87-134.

Gómez Rivero, M.ª C., «Inteligencia artificial aplicada a la salud: viejas y nuevas cuestiones para el Derecho penal», Revista de Derecho y Genoma Humano, núm. 61, 2026, pp. 135-170.

González Calvache, L., «La responsabilidad del Estado por el uso de armas autónomas letales», Revista Española de Derecho Militar, núm. 118, julio-diciembre 2022, pp. 9-54.

Grupo de Expertos de Responsabilidad y Nuevas Tecnologías de la Comisión Europea, Liability for Artificial Intelligence and other Emerging Digital Technologies, noviembre de 2019.

Hart, H. L. A., Punishment and Responsibility: Essay in the Philosophy of Law, Clarendon Press, Oxford, 1968.

Hernández Marín, R., «Sujetos jurídicos, capacidad jurídica y personalidad jurídica», Persona y Derecho: Revista de fundamentación de las Instituciones Jurídicas y de Derechos Humanos, núm. 36, 1997, pp. 95-126.

Heyns, C., Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, A/HRC/23/47, Naciones Unidas, 9 de abril de 2013.

High-Level Expert Group on Artificial Intelligence, A Definition of AI: Main Capabilities and Scientific Disciplines, European Commission, Bruselas, 2019.

Hobbes, T., *Leviatán, o la materia, forma y poder de un Estado eclesiástico y civil*, trad. de C. Mellizo, Alianza Editorial, Madrid, 2018.

ICRC, Autonomy, artificial intelligence and robotics: technical aspects of human control, Ginebra, agosto de 2019.

Kelsen, H., La Teoría Pura del Derecho, Losada, Buenos Aires, 1946.

Lacruz Mantecón, M. L., Robots y personas. Una aproximación jurídica a la subjetividad cibernética, Editorial Reus, Madrid, 2020.

Laín Moyano, G., «Responsabilidad en inteligencia artificial: Señoría, mi cliente robot se declara inocente», Ars Iuris Salmanticensis, vol. 9, 2021, pp. 197-232.

Lazcoz Moratinos, G., «Sistemas de inteligencia artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos (v.2)», Revista de Derecho y Genoma Humano, n.º 61, 2026, pp. 195-250.

Martínez, M. V., De qué hablamos cuando hablamos de inteligencia artificial, Policy Brief CILAC, MTD/SC/2024/PI/06, Montevideo, UNESCO / CLACSO, 2024.

Matthias, A., «The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata», Ethics and Information Technology, vol. 6, núm. 3, 2004, pp. 175-183.

Navarro-Michel, M., «Vehículos automatizados y responsabilidad por producto defectuoso», Revista de Derecho Civil, vol. VII, n.º 5, 2020, pp. 175-223.

Navas Navarro, S., «Obras generadas por algoritmos. En torno a su posible protección jurídica», Revista de Derecho Civil, vol. V, núm. 2, 2018, pp. 273-291.

SAE International, Recommended Practice J3016. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, junio de 2018.

Saiz García, C., «Las obras creadas por sistemas de inteligencia artificial y su protección por el derecho de autor», InDret, núm. 1, 2019.

Sanz Encinar, A., «El concepto jurídico de responsabilidad en la Teoría General del Derecho», AFDUAM, n.º 4, 2000, pp. 27-55.

Savigny, F. K. von, Sistema del Derecho Romano Actual, T. I y T. II.

Schmitt, M. y Thurnher, J., «Out of the loop: autonomous weapon systems and the law of armed conflict», Harvard National Security Journal, vol. 4, núm. 2, 2013, pp. 231-281.

Searle, J. R., «Minds, Brains, and Programs», Behavioral and Brain Sciences, vol. 3, n.º 3, 1980, pp. 417-457.

Sparrow, R., «Killer Robots», Journal of Applied Philosophy, vol. 24, núm. 1, 2007, pp. 62-77.

Strawson, P. F., «Freedom and Resentment», Proceedings of the British Academy, vol. 48, 1962, pp. 1-25.

Tiedemann, K., «Die Bebüssung von Unternehmen nach dem 2. Gesetz zur Bekämpfung der Wirtschaftskriminalität», Neue Juristische Wochenschrift, n.º 41, 1988, pp. 1169 ss.

Valdezate Pelegrín, P., «La autoría en creaciones generadas por Inteligencia Artificial», Derecom, núm. 37, 2024, pp. 28-30.

Valls Prieto, J., «Sobre la responsabilidad penal por la utilización de sistemas inteligentes», Revista Electrónica de Ciencia Penal y Criminología, núm. 24-27, 2022, pp. 1-35.

Zugaldía Espinar, J. M., «Aproximación teórica y práctica al sistema de responsabilidad criminal de las personas jurídicas en el Derecho Penal español», ponencia, Madrid, Ministerio de Justicia, s. f. [ca. 2010-2011].