

Derecho de la neurología: Libro blanco para el desarrollo de la
Carta de Derechos Digitales en el ámbito de la neurotecnología
Beatriz Gregoraci, Jorge Castillo Abella y Francisco Velasco Caballero (dirs.)
ISBN: 979-13-87913-93-3
Madrid, 2026
pp. 771-792
DOI: 10.37417/derecho-de-la-neurotecnologia/19
Marcial Pons Ediciones Jurídicas y Sociales
Editado bajo licencia Creative Commons Attribution 4.0 International License

CAPÍTULO XIX

RETOS Y CARENCIAS DE LA REGULACIÓN DE IA EN EL DERECHO PENAL

Javier VALLS PRIETO
Universidad de Granada
jvalls@ugr.es

1. INTRODUCCIÓN

Esta aportación establece una metodología de análisis de los sistemas de IA para establecer la responsabilidad penal de los diferentes actores implicados en su uso. La metodología que se propone crea una herramienta para analizar el grado de implicación y responsabilidad de los diferentes sujetos que intervienen en el uso de la inteligencia artificial, desarrolladores, profesionales y usuarios finales, dependiendo de sus obligaciones, del impacto que tienen sus actuaciones en los bienes jurídicos y en qué fase del ciclo de la vida del sistema inteligente influyen sus decisiones.

En un primer momento la idea detrás de este concepto de IA es la realización de actividades consideradas como inteligentes porque requieren algún grado de inteligencia incluyendo la creación de las herramientas y de los programas que puedan desarrollarla¹. El Reglamento

¹ JANSEN, P. *et al.* (2019), «SIENNA D4.1: State-of-the-art Review: Artificial Intelligence and robotics». Zenodo. pp. 12. «we establish definitions and demarcations for both fields that will be used in subsequent SIENNA research on AI and robotics. The report discusses the fields

européo sobre IA (AI Act) la define como un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales². La regulación añade un concepto interesante para el momento de su redacción que es el de IA de propósito general, entendido como un modelo de IA, incluyendo aquellos entrenados con una gran cantidad de datos mediante auto-supervisión a escala, que muestra una generalidad significativa y es capaz de realizar competentemente una amplia gama de tareas distintas, independientemente de cómo se comercialice, y que puede integrarse en una variedad de sistemas o aplicaciones posteriores³.

Además, la Comisión ha creado las guías éticas de uso de IA en la que se define el concepto de IA confiable que consiste en un desarrollo y uso centrado en los humanos y respetuoso con los derechos fundamentales, que establece un marco sobre el que va a construir el concepto de uso fiable de la inteligencia artificial, en el que nos vamos a centrar para determinar qué son buenas praxis⁴. De igual forma, la OCDE define un sistema de IA como un sistema basado en máquinas que, por objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los distintos sistemas de IA varían en sus niveles de autonomía y adaptabilidad tras la implantación⁵.

Existen ya multitud de ejemplos en los que el uso de los sistemas inteligentes lleva aparejado un impacto negativo en los derechos fundamentales. La vida se ve afectada por los vehículos autónomos, los sistemas crediticios discriminan a sectores poblacionales, los sistemas de reconocimiento facial discriminan a negros, la utilización de datos afecta a la privacidad, etc. Ante todos estos casos el derecho penal no puede quedarse al margen de las lesiones a los bienes jurídicos afectados por el uso de esta nueva tecnología. El determinar qué derechos

in terms of their backgrounds, positions, challenges, and present and expected applications. This report includes a socio-economic impact assessment (SEIA)».

² Artículo 3.1 del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) núm. 300/2008, (UE) núm. 167/2013, (UE) núm. 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial) (Texto pertinente a efectos del EEE).

³ Exposición de motivos núm. 63 del Reglamento de Inteligencia Artificial.

⁴ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HLEG) (2019), *A definition of Artificial Intelligence: main capabilities and scientific disciplines*.

⁵ OECD (2021), «Recommendation of the Council on Artificial Intelligence». OECD. pp. 7.

fundamentales se ven afectados no es sencillo como se puede observar de los estudios realizados hasta la fecha en los que se se aprecia que el espectro de derechos afectados es muy amplio y, por extensión, el de bienes jurídicos penales.

2. DESARROLLO

2.1. Principios éticos del Grupo de expertos de alto nivel de la Comisión Europea en IA

La utilización de sistemas inteligentes lleva consigo dilemas, problemas y conflictos sociales, muchos de ellos que atentan contra los derechos fundamentales de los ciudadanos, que pueden prevenirse con un uso ético de esta tecnología. Es por ello que la Comisión Europea creó un grupo de expertos para el diseño de una guía de uso ético de la IA con el fin de garantizar los tres elementos que considera que deben tenerse en cuenta en el futuro desarrollo de esta tecnología: que sea lícita, ética y robusta⁶.

La base sobre la que se sustenta este documento son los tratados de la Unión Europea, la Carta de la Unión Europea y las normas internacionales de derechos humanos⁷. Este grupo de expertos ha destacado en su análisis el respeto por la dignidad humana, la libertad individual, el respeto por la democracia, la justicia y el Estado de derecho, la igualdad, no discriminación y solidaridad y, finalmente, los derechos ciudadanos⁸. No hay que olvidar que tres de estos principios se hallan en la discusión teórica de la democracia, transparencia, justicia y rendición de cuentas, que dan una legitimación al uso en nuestras sociedades⁹.

De aquí surgen cuatro principios éticos principales: el respeto a la autonomía de los humanos, el principio de prevención de daños,

⁶ AI HLEG (2019), *op. cit.*, pp. 18.

⁷ *Ibid.*, pp. 9.

⁸ *Ibid.*, pp. 10 y 11.

⁹ ERMAN, E., FURENDAL, M. (2022), «The Global Governance of Artificial Intelligence: Some Normative Concerns». *Moral Philosophy and Politics* 9, pp. 267-291. pp. 273 y ss., «political philosophers can make valuable contributions by addressing such questions. Currently, AI development and application are to a large extent regulated through non-binding ethics guidelines penned by transnational entities. Assuming that the global governance of AI should be at least minimally democratic and fair, this paper sets out three desiderata that an account should satisfy when theorizing about what this means. We argue, first, that an analysis of democratic values, political entities and decision-making should be done in a holistic way; second, that fairness is not only about how AI systems treat individuals, but also about how the benefits and burdens of transformative AI are distributed; and finally, that justice requires that governance mechanisms are not limited to AI technology, but are incorporated into a range of basic institutions. Thus, rather than offering a substantive theory of democratic and fair AI governance, our contribution is metatheoretical: we propose a theoretical framework that sets up certain normative boundary conditions for a satisfactory account».

el principio de trato justo y el principio de explicación¹⁰ que van a servir de base para la determinación de los requisitos que se deben seguir para conseguir que la IA sea fiable.

Estos se dividen en siete requisitos que son: la gestión y supervisión humana; la robustez técnica y seguridad del sistema; la privacidad y gestión de datos; la transparencia; la diversidad, la no discriminación y el trato justo; el bienestar social y ambiental; y la rendición de cuentas¹¹.

Con el primero se persigue que los usuarios de sistemas basados en IA puedan tomar sus decisiones autónomas relacionadas con estos sistemas. Deben tener el conocimiento y las herramientas para comprender e interactuar de manera satisfactoria y, dentro de lo posible, autoasesorarse o enfrentarse al sistema. Los sistemas de inteligencia artificial deben ayudar a los individuos a mejorar y tomar decisiones informadas de acuerdo con sus objetivos, sin minar la autonomía humana o causar otros efectos adversos.

La robustez técnica y la seguridad interna hacen referencia a que los sistemas deben ser creados con una percepción preventiva de los riesgos de tal forma que se comporten de forma fiable y evitar los riesgos considerados como inaceptables. Al mismo tiempo, se debe asegurar la integridad mental y física de los humanos. Desde el punto de vista de la seguridad ha de demostrar su fiabilidad ante ataques que pueden afectar a los datos, al modelo o a la infraestructura. Hay que garantizar que estas vulnerabilidades no van a ser utilizadas para atacar al sistema, principalmente cuando se trata de sistemas basados en *software*. También se han de tomar las medidas adecuadas para prevenir y mitigar aplicaciones inintencionadas o usos abusivos por actores maliciosos para que sea considerado seguro. Hay que asegurarse, mediante un plan de impacto, que el sistema actúa bajo los objetivos con los que se creó. Las medidas de seguridad variarán dependiendo de cada sistema concreto y los fines para los que aplica. Es crucial que se tomen medidas que sean desarrolladas y probadas de forma proactiva para evitar los riesgos altos. La exactitud de las decisiones consiste en la habilidad de hacer predicciones correctas por parte del sistema. Esta robustez se basa en la capacidad para hacer predicciones, recomendaciones o tomar decisiones correctas. Para ello es necesario desarrollar un proceso de evaluación que mitigue los errores. Finalmente, la fiabilidad y repetición de los resultados es vital para considerar un sistema robusto.

La protección de datos y de la privacidad debe estar garantizada en todo el ciclo de uso de la IA, en sus tres fases, obtención de datos para alimentar el sistema, el procesamiento de estos y los resultados que el sistema ofrece. Hay que preocuparse de que la calidad e integri-

¹⁰ AI HLEG (2019), *op. cit.*, pp. 12 y 13.

¹¹ *Ibid.*, pp. 14-21.

dad de los datos que hacen funcionar la IA sea la adecuada. Así, en la fase de recogida de los datos se puede encontrar que estos tienen sesgos sociales, son inexactos o erróneos. Esta integridad de los datos se debe resolver antes de entrenar al sistema inteligente y eliminar la información errónea o que pueda generar discriminaciones. Además, es necesario que los conjuntos de datos sean seguros. Alimentar con malos datos puede cambiar el comportamiento de los sistemas de inteligencia artificial. El acceso a los conjuntos de datos debe de estar restringido y solo personal cualificado con competencia y una necesidad clara puede justificar el acceso a los datos individuales.

Relacionado con el requisito de explicación está el de transparencia que viene referido a los elementos principales de los sistemas inteligentes: los datos, el sistema y los sistemas de negocios. El grupo de expertos de alto nivel pone su foco en tres puntos. El primero consiste en la trazabilidad de los datos, que hace referencia a los conjuntos de datos y a los procesos que sirven para la decisión de los sistemas inteligentes, entre los que se encuentra la recogida de estos, su etiquetado, así como los algoritmos utilizados. Este seguimiento permite identificar las razones de por qué una decisión es errónea y lo que permitirá prevenir futuros errores. La explicación del proceso técnico de un sistema de IA y de las decisiones relacionadas con los humanos, segundo elemento en el que centran su interés es básica para entender la decisión tomada. La explicación se debe adaptar a la experiencia de la persona interesada, con un lenguaje que sea entendible. Finalmente, la comunicación del sistema inteligente con el usuario ha de ser clara y el primero debe de presentarse como no humano, así como sus capacidades y limitaciones.

Para conseguir el objetivo de crear una IA fiable es necesario observar todo el ciclo de vida de un sistema basado en IA y esto implica involucrar a todas las partes interesadas en el proceso, permitiendo obtener información de cada una de ellas con el fin de conseguir cumplir con el principio de trato justo.

La importancia jurídica de estos principios es vital para entender casos en los que se intenta atribuir la responsabilidad de la toma de decisiones a la máquina¹², ya que establecen un control de humano sobre el sistema inteligente y las decisiones que toma.

¹² SALVADORI, I. (2021), «Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal». *Cuadernos de política criminal* pp. 137-174. pp. 150.

2.2. Derechos fundamentales afectados con el uso de la inteligencia artificial

Asentados y determinados los principios sobre los que se debe basar el desarrollo y uso de la IA podemos derivar qué derechos fundamentales y qué bienes jurídicos se ven reforzados o estarían en riesgo. Si analizamos los tres trabajos más importantes en el análisis del impacto en los derechos fundamentales por el uso de esta tecnología podremos tener un mapa de en qué áreas el derecho penal tendrá que actuar. Poder determinar qué derechos fundamentales se ven afectados y de qué manera es difícil por la falta de una definición clara de IA y por la diversidad de usos y ámbitos de utilización en los que se pueden emplear, lo que deriva en diferentes formas de impacto.

El estudio realizado por el centro Berkman Klein se centra en el impacto en los derechos fundamentales¹³ en seis casos de uso: sistema judicial, sistema financiero, sanidad, moderación de contenido, recursos humanos y educación¹⁴. La selección de los casos tiene como objetivo centrarse en los derechos políticos, civiles, económicos, sociales y culturales.

Inician su análisis con el supuesto de la Administración de Justicia, en el que se produce un impacto en los derechos procesales del ciudadano importante cuando se utilizan sistemas inteligentes. Dado que el sistema de justicia en un Estado democrático es una institución que puede restringir o limitar los derechos más básicos de la ciudadanía, la utilización de sistemas inteligentes dentro del sistema judicial va a tener un impacto alto en este grupo. En el análisis realizado se detecta un impacto negativo en el derecho a una audiencia pública justa, el derecho a ser considerado inocente hasta que se demuestre lo contrario y el derecho a la privacidad. Existe un segundo nivel de impacto que podemos considerar como neutro, porque tiene un impacto positivo y negativo dependiendo de su uso, lo podemos observar entre los que se encontrarían el principio de igualdad ante la ley y el derecho a la no discriminación, el derecho a la vida, la libertad o la seguridad personal y el derecho a no ser arrestado, detenido o expulsado arbitrariamente¹⁵. No destacan un impacto positivo en este sector.

En el análisis del impacto en los bienes jurídicos de ámbito penal, se puede observar la lesión de los bienes jurídicos relacionados

¹³ Considerados como tales los recogidos en la Declaración Universal de Derechos Humanos, la Convención Internacional de Derechos Civiles y Políticos y la Convención Internacional de Derechos Económicos, Sociales y Culturales.

¹⁴ RASO, F. *et al.* (2018), «Artificial Intelligence & Human Rights: Opportunities & Risks». pp. 19.

¹⁵ *Ibid.*, pp. 20.

con los funcionarios públicos y con la Administración de Justicia y en la privacidad. También pueden verse afectados los bienes jurídicos relacionados con la defensa de la Constitución, en concreto, con el derecho a la igualdad. La responsabilidad penal de los funcionarios por la utilización de sistemas inteligentes en la toma de decisiones en la Administración de Justicia es clara en los casos en los que se lesionen bienes jurídicos¹⁶.

Dentro del sistema sanitario destaca el beneficio positivo por el uso de la IA en los derechos de los pacientes, en concreto, al derecho a la vida —las mejoras en el diagnóstico y que puede llegar a más personas consiguen una mejora en la salud de muchos sectores de población que no podrían tener acceso a esos tratamientos— un mejor nivel de vida, el derecho a conseguir el trabajo deseado —la mejora de la salud permitiría reducir el número de personas que son excluidas de un trabajo digno por problemas de salud— y el derecho a la educación —mejor salud, menos casos de personas que no pueden tener una educación por problemas médicos¹⁷. Dentro del sistema sanitario está claro que los bienes jurídicos vida e integridad física pueden verse afectados con el uso de la IA. En este punto es clave el principio de control humano de estos sistemas.

El control de los contenidos en las redes sociales, desde las noticias falsas pasando por contenidos injuriosos, vejatorios, etc., se realiza mediante sistemas inteligentes. Los derechos a la vida, la libertad y la seguridad de las personas se ven afectados positivamente si son los sistemas inteligentes los que moderan los contenidos en la red. Puede tener un impacto positivo y negativo en cuestiones de discriminación, puesto que puede reproducir o evitar los sesgos de las personas, tal y como se ha expuesto antes. Los impactos negativos se ven claros en el derecho a la privacidad, puesto que se gestiona información personal de las personas, incluyendo ideas políticas, religiosas o personales en espacios que pueden ser considerados como no públicos y que son escaneados por máquinas, y en los derechos de libertad de opinión, expresión e información, ya que los sistemas automáticos cometen más errores que los humanos a la hora de borrar contenidos legales¹⁸. Los delitos relativos a la Constitución, como defensores del buen funcionamiento de las instituciones democráticas, están siendo afectados por los sistemas inteligentes, siendo necesaria una adaptación de los tipos penales para los nuevos ataques a los bienes jurídicos.

¹⁶ VALLS PRIETO, J. (2017), *Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial*. Madrid: Dykinson. pp. 107 y ss.

¹⁷ RASO, F. *et al* (2018), *op. cit.*, pp. 32.

¹⁸ *Ibid.*, pp. 37.

Dentro de los sistemas de contratación y selección de personal mediante sistemas inteligentes nos encontramos, en el análisis realizado, dos derechos que se pueden ver afectados: el derecho a la no discriminación y el derecho a conseguir el trabajo deseado. En el primero, la posibilidad de que una máquina pueda tomar una decisión sin el sesgo de una persona a la hora de contratar puede evitar discriminaciones por motivos de raza o sexo. De la misma manera, no solo incidirá en la posibilidad de que la persona pueda conseguir el puesto de trabajo sino también en la posibilidad de recibir un salario más justo si la decisión depende de la IA, si está diseñada e implementada de forma adecuada. Al mismo tiempo y de forma contraria, si los datos con los que se alimenta el sistema son sesgados puede repetir o incluso aumentar la situación de discriminación que se produce por los humanos en el mundo real en materia salarial, como ocurre con las mujeres. En lo que sí tiene un efecto negativo es en los derechos de libertad de opinión, expresión e información, así como en el derecho de asociación. Nos podemos encontrar con que las personas que estén buscando trabajo restrinjan sus opiniones políticas para que no afecten a sus posibilidades de conseguir un empleo ya que, si el empleador realiza un análisis de datos en, por ejemplo, sus redes sociales puede decidir no contratarlo por ser muy activo sindicalmente. Esta ingente cantidad de datos que se van a utilizar, obviamente, va a tener un impacto negativo en el derecho a la privacidad¹⁹. El impacto en los bienes jurídicos de los trabajadores en estos casos es de gran importancia.

Finalmente, se analiza el impacto en el sistema educativo. El ejemplo que se utiliza para el sistema educativo en este análisis es la evaluación de trabajos mediante sistemas inteligentes. Tendría un impacto medio en la libertad de opinión, expresión e información, la habilidad de escribir puede ser mejorada por los sistemas, pero al mismo tiempo la captura de la escritura por estos sistemas puede perjudicarla ya que el sistema se ve alimentado con opiniones que pueden ser utilizadas en otros contextos en su contra. Los sistemas automáticos, al mejorar la escritura, van a tener un impacto positivo en la posibilidad de mejorar a la hora de conseguir trabajo, al igual que puede mejorar la escritura de los alumnos mejorando su derecho a la educación. Por el contrario, todo este intercambio de datos influye en el derecho a la privacidad, viéndose mermado²⁰. No existen como tales bienes jurídicos que protejan la educación de los ciudadanos, sin embargo, la exclusión de sectores al sistema educativo afecta claramente a bienes jurídicos como la igualdad²¹, la libertad de los ciudadanos o el libre desarrollo de la personalidad.

¹⁹ *Ibid.*, pp. 42.

²⁰ *Ibid.*, pp. 47.

²¹ GÓMEZ MARTÍN, V. (2016), «Incitación al odio y género. Algunas reflexiones sobre el nuevo art. 510 CP y su aplicabilidad al discurso sexista». *Revista electrónica de ciencias penales y criminología* 18. pp. 18.

El segundo informe que es de nuestro interés es el de la Agencia Europea para los Derechos Fundamentales con el fin de detectar cuál es el impacto en los derechos fundamentales del uso que se da a la IA en cuatro áreas: ayudas sociales, predicción del crimen, servicios de salud y publicidad personalizada²². Del análisis del texto podemos extraer una selección de derechos afectados como son la dignidad humana, la privacidad y la protección de datos, la igualdad y no discriminación, el acceso a la justicia, el derecho a la seguridad social y a la asistencia social, los relacionados con los consumidores y el de la ciudadanía a una buena administración²³. Todos estos derechos tienen un reflejo en el Código Penal con un bien jurídico afectado que deriva en un tipo penal que le da protección. Este análisis realizado por la Agencia Europea busca no solo realizar el análisis, sino que se acerca a la defensa de estos derechos fundamentales mediante el sistema jurídico.

El art. 1 de la Carta establece la dignidad humana como uno de los pilares de la Unión Europea y esto implica que todo desarrollo de IA ha de estar supeditado a los humanos, como señalan expresamente las guías éticas que se han estudiado anteriormente. Además, en el art. 2 se estipula el derecho a la vida y en el art. 3 la integridad de las personas. Pero no solo cubre estos casos más extremos en la protección de derechos. Se ha de entender también que se ha de evitar que las personas se encuentren sujetas a la utilización de IA sin su conocimiento y sin su consentimiento expreso²⁴. Hay que aclarar que la dignidad humana no se define como bien jurídico en el Código Penal²⁵. Se puede proteger su lesión mediante los artículos 173.3 y 174, el 177 *bis*, el 232 cuando se trate de menores y el 311 y 312 del CP, sin olvidar los delitos dedicados a la protección de la Constitución, en donde podemos englobar las lesiones a determinados derechos fundamentales.

El derecho a la privacidad y la protección de datos, recogidos en los artículos 7 y 8 de la Carta, se encuentran fuertemente afectados. Esta utilización de información personal puede afectar a otros derechos de forma indirecta como pueden ser la libertad de expresión y de información (art. 11) o el derecho de asociación (art. 12) con la importancia que puede tener su afectación en conexión con las ideas políticas cuando esa asociación va referida a sindicatos o partidos políticos, o a la libertad religiosa (art. 10)²⁶. El bien jurídico protegido en los arts. 197 y

²² EUROPEAN UNION AGENCY FOR HUMAN RIGHTS (2020), *Getting the future right. Artificial intelligence and fundamental rights*. Luxemburgo: Publications Office of the European Union. pp. 1.

²³ *Ibid.*, pp. 66 y ss.

²⁴ *Ibid.*, pp. 60.

²⁵ ALONSO ÁLAMO, M. (2011), «Fundamentación pre-positiva de los bienes jurídico-penales y derecho penal mínimo de los derechos humanos». *Revista General de Derecho Penal*. pp. 15.

²⁶ EUROPEAN UNION AGENCY FOR HUMAN RIGHTS (2020), *op. cit.*, pp. 61.

ss. nos ofrece una herramienta para trasladar el principio de privacidad de la guía de altos expertos en IA. Su anclaje en la protección de bienes jurídicos en el sistema penal es sencillo.

Los artículos 20 y 21 de la Carta de Derechos Fundamentales recogen los derechos de igualdad y no discriminación. La discriminación ha de entenderse como el trato menos favorable de una persona respecto del trato que recibe o recibiría otra persona en la misma situación. Un elemento interesante es el trato que se les da a los grupos vulnerables como puedan ser los grupos raciales, étnicos, sexo, religión, etc., los menores (art. 24), mayores (art. 25) y las personas con discapacidad (art. 26). En el Código Penal podríamos englobarlos en los apartados relativos a la defensa de la Constitución.

El derecho a una tutela judicial efectiva (art. 47), como hemos señalado anteriormente, también se puede ver afectado por los sistemas inteligentes cuando son los encargados de determinar si se acepta o no un determinado requerimiento o cuando la decisión final de un determinado caso la toma un sistema inteligente. La opacidad de los sistemas hace que el ciudadano recele de las decisiones tomadas al no poder comprender la justificación de la resolución judicial²⁷.

Desde una perspectiva europea destacan los derechos colectivos o sociales. El derecho a la seguridad y asistencia social es un derecho social recogido en el art. 34 de la Carta. Es un derecho que no pone un límite mínimo de asistencia, pero debe entenderse como un límite a la hora de restringir el estado actual de intervención social por parte de los Estados. Afectaría, por ejemplo, a la no discriminación por nacionalidad, dentro de la libertad de movimientos en la Unión, en cuestiones de derechos laborales y a las pensiones que se deben recibir después de los años cotizados, independientemente de la nacionalidad²⁸. Aquí es difícil encontrar un bien jurídico que proteja estos derechos.

Caso diferente es la protección de los consumidores, que como tal no es un derecho recogido en la Carta, no ha sido tenido en cuenta como derecho fundamental pero el informe lo tiene en cuenta ya que deriva del art. 169 del TFUE, con una especial importancia en las técnicas de márketing dirigido mediante IA que utilizan determinadas empresas, o la vigilancia de los mercados que afectan a la libertad de los consumidores, manipulando sus decisiones o los precios mediante los perfiles de los usuarios de plataformas de comercio electrónico²⁹. La protección penal del consumidor cobra aquí un especial rol de protección³⁰.

²⁷ *Ibid.*, pp. 75.

²⁸ *Ibid.*, pp. 79.

²⁹ *Ibid.*, pp. 80.

³⁰ VALLS PRIETO, J. (2023), «La inteligencia artificial frente a la colectividad: Una protección de los bienes jurídicos colectivos frente al auge de los sistemas inteligentes». *Estudios penales y criminológicos* pp. 11. pp. 9.

Desde el punto de vista penal, los bienes jurídicos colectivos nos van a permitir tener una base para proteger bienes jurídicos como el medio ambiente, la salud pública, los derechos de los trabajadores, la seguridad en el tráfico, etc. El principio de conseguir el bienestar común se puede trasladar fácilmente a estos derechos colectivos y por ende a los bienes jurídicos colectivos.

El último de los derechos a los que hace referencia el informe es el referido al art. 41 de la Carta referido a la buena administración que debe de recibir un ciudadano por parte del Estado. Este punto hace referencia al derecho que tiene el administrado de tener acceso a los datos que maneja la administración y a la obligación de justificar la decisión que se ha tomado respecto al ciudadano. En ambos casos la incidencia de la utilización de sistemas inteligentes por parte de la administración lo puede poner en serio peligro³¹. Al igual que en el caso de las decisiones judiciales, los delitos relativos a los funcionarios públicos cubren la protección mediante los bienes jurídicos relativos a la Administración de Justicia para los casos más importantes de mala administración pública.

El último informe que vamos a analizar es el del Consejo de Europa, en el que se estudia cómo regular la IA desde una perspectiva de los derechos humanos, la democracia y el Estado de derecho³². En este trabajo el primer principio es el respeto al valor humano que se refleja en varios derechos: la libertad y la seguridad (art. 5 CEDH), el derecho a un juicio justo (art. 6 CEDH), el derecho a no ser penado sin una ley (art. 7 CEDH), el derecho a la privacidad y la integridad física y mental (art. 8 CEDH)³³. Un segundo punto de estudio por el Consejo de Europa son las libertades afectadas. Aquí destaca la libertad de expresión (art. 10 CEDH) y las libertades de reunión y de asociación (art. 11 CEDH)³⁴. El tercer punto de estudio es el principio de igualdad y no discriminación (art. 14 CEDH y el protocolo 12). Pone como ejemplos ilustrativos las formas de contratación en las empresas en las que se discrimina a las mujeres mediante la selección de trabajadores por sistemas de inteligencia artificial. En conexión con el mercado laboral el informe del Consejo de Europa da una especial importancia a los derechos sociales, económicos y los relacionados al puesto de trabajo recogidos en la Carta Social Europea, resaltando que el uso de la IA puede ser muy beneficioso para las condiciones laborales de los trabajadores, no es menos cierto que, mal utilizado, puede afectar a los

³¹ EUROPEAN UNION AGENCY FOR HUMAN RIGHTS (2020), *op. cit.*, pp. 81.

³² BEN-ISRAEL, I. *et al* (2020), «Towards regulation of AI systems Global perspectives on the development of a legal framework on Artificial Intelligence (AI) systems based on the Council of Europe's standards on human rights, democracy and the rule of law». CAHAI Secretariat. pp. 1.

³³ *Ibid.*, pp. 29 y ss.

³⁴ *Ibid.*, pp. 30 y ss.

derechos de los trabajadores, las condiciones de seguridad y salud en el trabajo, la dignidad del trabajo (arts. 2 y 3 CSE) así como al derecho a la organización en el trabajo (art. 5 CSE). También existe un riesgo de no respetar el derecho a la igualdad de oportunidades y de trato en cuestiones de empleo y ocupación (art. 20 CSE)³⁵.

La IA también puede afectar al sistema democrático, que no es un derecho per se, pero que, sin duda, es el entorno político que mejor sirve para el respeto y eficacia de los derechos fundamentales que estamos viendo. En relación con el sistema democrático se destacan tres puntos principales: el discurso social y político, acceso a la información e influencia en los votantes, desigualdad y segregación y fallo sistémico y disrupción de la democracia³⁶.

El informe del Consejo de Europa finaliza su trabajo planteando nuevos derechos y adaptaciones de los derechos existentes que deben considerarse para analizar jurídicamente el uso de los sistemas inteligentes. El primero, que ya ha sido resaltado en las guías éticas, es el de la autonomía humana, control y supervisión por humanos de la IA, que podría incluirse en el bien jurídico libertad o dignidad. Además, se le añade un derecho a la transparencia y a recibir una explicación de los resultados de los sistemas inteligentes siempre que afecte a intereses humanos, aunque no utilice datos personales. Este derecho, como tal, no tiene la debida protección en un bien jurídico ni en un tipo penal. Y es importante porque sin una debida protección de la transparencia y a recibir una explicación, va a ser difícil conseguir los elementos necesarios para poder determinar los sujetos responsables de las actuaciones de los sistemas inteligentes. El informe continúa con un derecho a la integridad física, psíquica y moral a la luz de los perfiles mediante IA, en concreto, aquellos que afecten al reconocimiento. Aumentar el alcance de la privacidad cuando se trate de vigilancia masiva mediante IA, así como dar protección contra el rastreo indiscriminado y de alcance social de individuos³⁷. Estos derechos dirigidos a evitar un control exhaustivo de los sujetos por parte de los sistemas inteligentes no están prohibidos por nuestro Código Penal.

2.3. Metodología para comprender los problemas jurídicos que surgen del uso de los sistemas inteligentes

Ya han quedado establecidos dos elementos importantes para el derecho penal a la hora de determinar la responsabilidad derivada del uso de la IA, la existencia de un uso apropiado de la IA que debe cumplir

³⁵ *Ibid.*, pp. 33 y ss.

³⁶ *Ibid.*, pp. 34.

³⁷ *Ibid.*, pp. 41.

con los siete principios éticos a nivel europeo, lo que nos determina las medidas que se van a considerar como diligencia debida y se ha determinado un grupo amplio de derechos fundamentales, tanto individuales como colectivos, que se pueden ver afectados de forma negativa por el uso de esta tecnología, con la consiguiente lesión de los bienes jurídicos que protegen esos derechos.

Sin embargo, desde el punto de vista penal tenemos que enfrentarnos a otro dilema, la determinación de quién va a ser el responsable de la lesión del bien jurídico protegido por el sistema penal, principalmente por la falta de nexo causal por la existencia de las *black box* que hace imposible establecer un nexo causal claro entre la lesión del bien jurídico y el proceso de toma de decisiones por parte del sistema inteligente. Para evitar esta situación se debe realizar una evaluación en dos momentos cruciales: antes de empezar a diseñar el artefacto y antes de su lanzamiento al público en general. Esta evaluación quedará registrada de tal manera que permitirá al juez determinar si se ha cumplido con las medidas adecuadas para mitigar los riesgos conocidos y, en algunos casos, de los desconocidos. Sólo mediante un control de la actuación de los sistemas inteligentes se podrá determinar la relación entre la toma de decisiones y el accidente.

Con la intención de superar esta incertidumbre en la determinación de la responsabilidad aparejada al uso de los sistemas inteligentes propongo una metodología de análisis con el objetivo de comprender dónde se producen los fallos, dónde debería centrarse la imputación y la responsabilidad de los diferentes sujetos que intervienen en todo el proceso. Estos sujetos se pueden dividir en tres grupos dependiendo de la fase en la que actúan. El primero sería el compuesto por desarrolladores y fabricantes del producto que lo diseñan y mandan al mercado. El segundo estaría constituido por los profesionales que utilizan estos sistemas inteligentes para realizar una parte de su trabajo y que van a interaccionar con la máquina de forma diferente a como lo realizarían los desarrolladores de estos. El tercer grupo lo constituirían los usuarios finales, en muchos casos consumidores del mismo. Cada uno de ellos tiene una relación y control diferente del sistema inteligente, lo que implica una responsabilidad distinta.

La delimitación y clasificación del uso de la IA es una labor complicada. Afortunadamente, no partimos de cero ya que desde la filosofía de la tecnología se han hecho avances en esta materia con el fin de realizar la evaluación ética del uso de la tecnología mediante un análisis por fases³⁸. La metodología propuesta por Brey sobre la ética anticipatoria en tecnología propone tres fases de estudio para resolver de forma

³⁸ BREY, P.A. E. (2012), «Anticipatory Ethics for Emerging Technologies». *NanoEthics*, 6, pp. 1-13. pp. 3 y ss.

adecuada este estudio: el nivel de evaluación de la tecnología, en donde se define qué tecnología en particular se va a utilizar, entendiendo por tecnología una colección de técnicas que están relacionadas entre sí por el propósito común, dominio o características formales o funcionales; el nivel de desarrollo del artefacto, que es la configuración física que opera de forma y en el entorno adecuados y produce el desarrollo deseado; y el nivel de aplicación, que se centra en el uso del artefacto y los problemas que puede derivar de ese específico uso³⁹.

Determinados estos puntos, proponemos una metodología de análisis que nos va a permitir determinar dónde nos vamos a encontrar los fallos de implementación y su responsable, influenciada por este análisis de predicción de los problemas éticos, en tres diferentes niveles y habiendo determinado la necesidad de la utilización de un sistema inteligente para realizar una labor en vez de un humano, del campo en el que se va a aplicar y los objetivos que se pretenden.

Así, podemos distinguir tres importantes momentos para determinar un uso inadecuado e irresponsable desde el punto de vista del uso concreto que se le vaya a dar a la IA: en la recolección de datos, en el procesamiento y en la interpretación de los resultados finales. Como hemos visto el gran avance en los últimos años en IA se ha producido en aquellos sistemas inteligentes que se alimentan con datos, lo que implica que debemos conocer cómo se han conseguido los datos, qué datos se van a utilizar, con qué fin y por cuánto tiempo se van a utilizar, quién tiene acceso a los datos y cómo se aseguran los mismos. En la siguiente fase, la de procesamiento, se debe de entender cómo funciona el sistema (algoritmo y los modelos sobre los que se toman las decisiones), quiénes tienen acceso al mismo, medidas de seguridad que existen, entre otras. En el caso en que el procesamiento no pueda explicarse, se debe entender cuáles han sido los procesos y procedimientos en la creación y la instauración del uso de esos sistemas. Finalmente, en la fase tercera de análisis de los resultados extraídos de la máquina, hay que analizar si los resultados son de calidad, si muestran algún impacto negativo en los derechos y si justifican la utilización de los sistemas inteligentes. Los resultados no siempre son fiables entre otras cosas porque la IA no es una bola de cristal⁴⁰. Estas tres fases nos van a dar un acercamiento al entendimiento del sistema para poder entender dónde va a surgir el ataque al bien jurídico.

Además, es preciso analizar estas tres fases en la utilización de la tecnología desde la perspectiva de la relación entre el sujeto y el sistema inteligente. Los errores de los sujetos implicados en todo el ciclo de vida de la IA se pueden producir por los creadores y desarrolladores,

³⁹ *Ibid.*, pp. 7 y 8.

⁴⁰ DIGNUM, V. (2019), *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Cham: Springer. pp.28.

por los profesionales que la utilizan y, finalmente, por los destinatarios/usuarios finales. Limitarlo exclusivamente a los desarrolladores o fabricantes no cubre todo el espectro de posibilidades⁴¹. Al mismo tiempo, será necesario analizar la utilidad práctica que le asigne el usuario profesional, ya que dependiendo del uso que se le atribuya puede afectar o no a los bienes jurídicos que hemos analizado. Imaginemos un sistema de predicción del crimen que funciona bien, ha sido probado en un entorno cerrado y seguro y da una fiabilidad del 80 por ciento, y el usuario profesional decide solo utilizarlo para determinados barrios de una ciudad o solo para determinados sujetos o con bases de datos basadas en la «intuición» policial que determina los datos de qué sectores de población van a ser utilizados. O un sistema de inversión de banca que se centra en modificar precios con el fin de eliminar a la competencia. Es el usuario profesional quien utiliza la IA con un determinado fin que puede derivar en la comisión de un delito, pero el sistema en sí no tiene ningún fallo que derive en la lesión del bien jurídico. Es una máquina que desde el punto de vista del producto no es defectuosa. Finalmente, hay que estudiar al tercer grupo de sujetos, los usuarios finales, que mostrará si son estos los que, derivado de su mal uso, han producido los fallos en el sistema. La responsabilidad se puede producir por una mala utilización por parte de pacientes del dispositivo inteligente, por ejemplo, pero también cuando un producto está diseñado y probado solo con determinados sectores de población, por ejemplo, los sistemas de conducción automática se testean dentro del sector de 65 a 80 años solo con hombres, puede producir fallos en el diseño que afecten a los derechos de otros sectores de población, en el ejemplo puesto, a las mujeres de 65 a 80 años. O con los dispositivos inteligentes para la medición de determinadas constantes mediante sistemas inteligentes si son desarrollados solo para determinados sectores puede que no sirvan adecuadamente para otros sectores de población, dando resultados erróneos solo a este grupo poblacional. Al mismo tiempo, el diseño ha de permitir que la generalidad de los usuarios finales tengan la facilidad de utilizarlos sin riesgo. Un ejemplo claro es en los coches automáticos cuando detectan que el piloto se ha quedado dormido. En el caso del coche de Tesla emite un pitido repetidamente hasta que la persona vuelve a tomar el control del vehículo. En el caso del consorcio automovilístico alemán, el coche cuando detecta esta anomalía conduce inmediatamente, reduciendo la velocidad hasta el arcén donde detiene el vehículo. El análisis de este tercer grupo de sujetos es clave para determinar si los dos grupos anteriores tienen responsabilidad o no.

Desde esta metodología podremos realizar un sistema de gestión de riesgos adecuado para determinar si se han tomado las precauciones

⁴¹ SALVADORI, I. (2021), *op. cit.*, pp. 157.

adecuadas o no tras la creación del sistema de gestión de riesgos que va a imponerse como estándar en la Unión Europea, mediante un sistema de impacto en tres niveles: bajo, medio o alto.

Teniendo clara la fase y el nivel de implicación de los sujetos podremos analizar la responsabilidad derivada del uso de los sistemas inteligentes que nos va a servir de ayuda para determinar cuándo ha existido un incumplimiento de las normas de cuidado para determinar su responsabilidad. Esta metodología está alineada con el sistema de gestión de riesgos que establece el Reglamento sobre IA para los casos de alto riesgo.

2.4. El sistema gestión de riesgos del Reglamento de la inteligencia artificial como determinante de la responsabilidad

Existen cuatro posibilidades de responsabilidad penal en el uso de tecnologías inteligentes: el uso doloso de la máquina para cometer un delito que no plantea especiales problemas jurídicos, cuando su uso se desvía de su tarea y se sabe y acepta el desvío en su actuación, los casos en que se desvía sin tener conocimiento del porqué de la reacción fuera del comportamiento esperado, y el uso de una tecnología que ha sido prohibida⁴². El centro del debate se va a centrar en el grado de conocimiento y la posibilidad de determinar si el comportamiento es imprudente o si se trata de un comportamiento fortuito sin responsabilidad penal.

El primer paso es determinar qué se considera un uso no adecuado. Es por lo que la Comisión Europea ha establecido las guías éticas con una serie de principios para crear y utilizar sistemas inteligentes y en el Reglamento sobre IA (art. 27) que establece un sistema de prevención de riesgos en la utilización de sistemas inteligentes con el fin de asegurar a los europeos que se puedan beneficiar del desarrollo y uso de estas nuevas tecnologías de acuerdo con los valores, principios y derechos fundamentales.

Si el *Libro Blanco sobre Inteligencia Artificial de la Comisión Europea*⁴³ establecía las opciones políticas que se pueden derivar del uso de los sistemas inteligentes y se esbozaban los riesgos que se podían producir, el Reglamento desarrolla este segundo objetivo con el fin de proporcionar confianza a los usuarios de estas tecnologías basadas en IA, tomando como base la protección de los principios éticos. Por la

⁴² QUINTERO OLIVARES, G. (2017), «La robótica ante el derecho penal: El vacío de respuesta jurídica a las desviaciones incontroladas». *Revista Electrónica de Estudios Penales y de la Seguridad: REEPS*, pp. 1.

⁴³ COMISIÓN EUROPEA (2020), *White Paper on Artificial Intelligence: a European approach to excellence and trust*.

regulación queda claro que la ausencia de un plan ético de prevención de riesgos derivará en algún tipo de responsabilidad, al establecerlo como obligatorio cuando se trate de IA de alto riesgo y deseable en el resto de los sistemas inteligentes.

En el art.9 del Reglamento se señala que sistemas de gestión de riesgos deben establecerse, implementarse, documentarse y mantenerse. Estos sistemas de gestión de riesgo consisten en un proceso reiterativo y continuo durante todo el ciclo de vida del sistema inteligente. El análisis de gestión de riesgos tiene cuatro puntos a recoger: la identificación y análisis de los riesgos conocidos y previsibles asociados con cada IA, la estimación y evaluación de los riesgos que pueden emerger cuando el sistema de alto riesgo es utilizado dentro de sus fines previstos y bajo las condiciones de un uso indebido previsible, la evaluación de otros riesgos que surjan del análisis de datos recolectados después de su puesta en el mercado según lo regulado en el art.61⁴⁴, y la adopción de medidas aplicables para la gestión de riesgos.

Estas medidas han de ser de tal naturaleza que cualquier riesgo residual asociado a cada situación individual, así como el riesgo residual global de los sistemas de IA de alto riesgo⁴⁵, se pueda considerar aceptable de acuerdo con su finalidad o dentro de lo razonablemente previsible en aquellos casos en que se use de forma indebida. Para identificar las medidas el Reglamento estipula tres puntos que deben asegurarse: *a)* la eliminación o reducción de riesgos en la medida de lo posible a través del desarrollo y diseño adecuados, en su caso, *b)* aplicación de medidas adecuadas de mitigación y control en relación con los riesgos que no pueden eliminarse, y *c)* el suministro de información adecuada con arreglo al art. 13, en particular por lo que respecta a los riesgos mencionados en la letra *b)* del apartado 2 de este artículo⁴⁶, y, en su caso, la formación de los usuarios profesionales y consumidores.

Además, hay que añadir que los sistemas han de ser probados con el fin de identificar las medidas más apropiadas para eliminar o mitigar los riesgos. Estas pruebas han de ser adecuadas para conseguir el fin

⁴⁴ El art. 61 hace referencia al control de la IA una vez se ha puesto en el mercado. Se exige el requerimiento de establecer y documentar un sistema de seguimiento de forma proporcional a la naturaleza de la IA y de los riesgos detectados anteriormente. Este sistema recopilará la información suministrada por los usuarios o recogida por otros medios en el uso de estos sistemas durante todo el ciclo de vida.

⁴⁵ Anexo III del RGPD.

⁴⁶ El suministro de información sobre cómo funciona el sistema y, en particular, para la eliminación de riesgos la finalidad que se pretende con su uso, el nivel de precisión, robustez y seguridad del sistema, los malos usos o aplicaciones que pueden tener, en especial, un impacto en la salud, seguridad y derechos fundamentales, su rendimiento con respecto a las personas o grupos de personas a los que se destina el sistema, y cuando proceda, las especificaciones de los datos de entrada, o cualquier otra información pertinente en cuanto a los conjuntos de datos de entrenamiento, validación y prueba utilizados, teniendo en cuenta la finalidad prevista del sistema de IA.

por el que se utiliza el sistema inteligente, dentro del fin que se persigue con su uso y deben realizarse durante el proceso de desarrollo en los momentos que se considere necesario y antes de poner el producto en el mercado. En particular, ha de tenerse un especial cuidado en el acceso e impacto en niños.

La ausencia de este sistema de gestión de riesgos va a ser un elemento clave a la hora de determinar la responsabilidad penal del desarrollador o usuario profesional. Así, Romeo Casabona presenta diferentes sistemas de determinación de la responsabilidad criminal. El primero de ellos es considerar que el sistema inteligente autónomo es responsable por los crímenes cometidos por sus actos⁴⁷. Entrando en una disquisición sobre si la máquina puede tener intención o no⁴⁸, deja la solución en el aire sobre si en un futuro, bastante lejano, la máquina podría tener esta intención de la misma forma que un humano⁴⁹.

Una segunda forma de responsabilidad se centraría en la imprudencia por actos inintencionales u omisivos de estos sistemas que derivan en un diseño defectuoso del mismo. Esto añade otro problema por la ausencia de conocimiento por parte del sistema del contenido de la obligación de cuidado, dejando sin posibilidad de evaluar el elemento normativo, y la valoración como tal, para la cual, a diferencia de los humanos, no está capacitado. Por tanto, no podrá ir en contra de la norma, elemento necesario para la aplicación a la máquina de esta figura jurídica⁵⁰.

Puesto este marco jurídico, como se establece con el Reglamento de la inteligencia artificial, nos llevaría a un sistema de *compliance* similar al de las personas jurídicas, entendido como un sistema en el que se valora la existencia de un sistema de organización y gestión de riesgos que prevenga de la utilización de conclusiones no validadas y rutinarias que salgan de la máquina⁵¹. Además de la necesidad de un examen general y detallado externo por expertos.

Siguiendo los puntos del art. 27 del Reglamento de IA, lo primero es determinar qué tipo de tecnología se va a utilizar. Hemos analizado diferentes tipos de IA en el apartado dedicado a su definición. Por ejemplo, el análisis para la privacidad no es lo mismo un aprendizaje de máquina basada en datos, que otro tipo de IA basada en razonamiento ló-

⁴⁷ ROMEO CASABONA, C. M. (2020), «Criminal responsibility of robots and autonomous artificial intelligent systems?». *Comunicaciones en propiedad industrial y derecho de la competencia* pp. 167-187. pp. 170.

⁴⁸ SALVADORI, I. (2021), *op. cit.*, pp. 154.

⁴⁹ BRUNDAGE, M. *et al* (2018), «The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation». Oxford. pp. 2.

⁵⁰ ROMEO CASABONA, C. M. (2020), *op. cit.*, pp. 174.

⁵¹ *Ibid.*, pp. 173.

gico. El siguiente paso es determinar el contexto en el que se va a usar, ámbito sanitario, educativo, etc., que nos va a dar un espacio, con unas características determinadas, sobre el que se tiene que determinar la responsabilidad y que nos va a permitir conocer si se está ante un uso de inteligencia artificial adecuado o no. En una segunda fase, hay que añadir que pueden existir diferentes niveles de responsabilidad en el cumplimiento de la gobernanza de la IA; en la creación, el uso profesional y uso de un sistema inteligente. Vamos a encontrarnos con un nivel en el desarrollo y puesta en el mercado de la máquina inteligente, que requerirá de un control de riesgos en su creación. Se podrían diferenciar dos fases, una en la creación y otra en la producción de este, pero para nuestro análisis ahora, no es relevante. Además, es necesario otro control de riesgo en su uso profesional. Una vez puesto en el mercado, la utilización del sistema inteligente requerirá de un análisis de riesgos añadido. Ningún fabricante, después de entregar el producto finalizado, va a asumir los riesgos derivados de su uso. Por ejemplo, un sistema inteligente que ha sido probado con éxito en su entorno controlado de uso, entrenado con determinados datos, cuando pase al uso profesional se puede alimentar con bases de datos que impliquen que genere riesgos hacia los derechos fundamentales y, por tanto, a los bienes jurídicos de las personas.

Los sistemas inteligentes en prevención del crimen pueden funcionar muy bien en el desarrollo, pero cuando se alimenten con los datos de diferentes operaciones policiales reales pueden afectar a derechos diferentes y de forma diferente. En el caso de los coches autónomos, el coche está perfectamente diseñado para captar los datos con sus sensores y comunicarse con otros autos y se produce una ausencia de datos provenientes de las agencias de control de tráfico. Además, hay que analizar el uso, riesgos e impacto de los usuarios finales/consumidores de estos sistemas inteligentes, lo cual implica un análisis de riesgos en una tercera capa.

Diferenciar entre tres sectores a la hora de determinar la responsabilidad penal, la evaluación de impacto y de riesgos, nos sirve para detectar dónde se ha producido el fallo y determinar si la responsabilidad va a ser dolosa, imprudente o que no exista. En cualquier caso el responsable siempre será un humano. La máquina solo operará en relación con los parámetros y la moralidad que le infunda el creador, usuario profesional o consumidor y es en estas personas donde tenemos que buscar el grado de responsabilidad que se dé.

3. CONCLUSIONES

La relación de la tecnología con las personas siempre ha sido disruptiva. Para bien generando una mejora de las condiciones de vida mediante la innovación, para mal, cambiando nuestra forma de vivir

y nuestras libertades. La implementación de la IA no está exenta de estos riesgos y, por ello, las instituciones europeas están trabajando para asegurar que se respetan los valores democráticos sobre los que basan nuestra calidad de vida.

La definición de IA es clave para poder centrar la responsabilidad penal. Las definiciones aportadas por la industria, desarrolladores o filósofos no llegan a ser definiciones que aporten un contenido jurídico interesante. Muchas se centran en la tecnología que se desarrolla o en la definición filosófica de la inteligencia o lo humano que aporta poco contenido para resolver problemas jurídicos. Ante esta situación, la importancia de la aportación de la definición del grupo de expertos en IA de la Comisión Europea que nos ofrece, por un lado, una definición clara y lo suficientemente amplia para abarcar a las diferentes tecnologías inteligentes y, por otro, nos va a permitir determinar los elementos, los principios éticos, que han de ser considerados como claves para entender qué elementos son imprescindibles para analizar la responsabilidad penal.

Estos principios éticos nos van a permitir determinar qué bienes jurídicos se van a ver afectados con el impacto negativo de los sistemas inteligentes. Como hemos visto, hay un gran número de bienes jurídicos individuales y colectivos que se ven afectados. Algunos de forma muy directa y clara (privacidad, discriminación, libertad) otros de forma más indirecta (derechos laborales, libre mercado, medio ambiente). Con la propuesta de análisis podemos determinar, mediante los principios éticos los derechos fundamentales afectados y, por tanto, los bienes jurídicos puestos en peligro. Detectados los bienes jurídicos afectados, que como hemos visto son muy variados, nos permite entrar en el análisis de la responsabilidad penal teniendo claro el marco de estudio desde el que partir con el análisis dogmático penal.

Los análisis realizados por la doctrina nos ofrecen diferentes soluciones de las cuales la que aparece con más fuerza es la de la responsabilidad por imprudencia y la de los modelos de *compliance*. Ambos tienen dos problemas al llevarlos a la realidad: la determinación del sujeto responsable y la determinación de en qué fase de la vida de la IA se encuentran los elementos necesarios de la responsabilidad.

Con el fin de resolver este problema tenemos que llegar a la conclusión de que solo es posible analizar la responsabilidad penal centrándonos en diferentes fases. La primera es la de determinar qué tecnología inteligente se está utilizando, para lo cual es imprescindible tener clara la definición de sistema inteligente. Después determinar el contexto en el que se produce, que nos va a permitir estudiar los principios éticos para determinar qué es lo aceptado en cada entorno de uso y qué no. Esta fase nos va a permitir determinar qué se considera como un uso adecuado por la sociedad y empezar a delimitar las normas de cuidado.

En este punto, la detección de qué bienes jurídicos se pueden ver afectados es una tarea tremendamente complicada. Es por ello por lo que aportamos una metodología de análisis de los derechos fundamentales afectados para realizar un acercamiento a los bienes jurídicos que se ven lesionados, mediante una metodología de análisis basada en las tres fases de funcionamiento que tiene un sistema inteligente y los tres grupos de sujetos a los que se les puede imputar los fallos del sistema inteligente. Sin un análisis de este tipo resulta extremadamente complicado determinar la responsabilidad de los sujetos que intervienen en el ciclo de vida de los sistemas inteligentes dentro de las diferentes fases.

Aun así, como ha dejado claro la doctrina, la imputación de la responsabilidad penal no es sencilla. Es por ello por lo que es necesario estudiar cuáles son las normas de cuidado que se tienen que respetar y que son exigibles a los ciudadanos en el uso de sistemas inteligentes para poder determinar quién ha cometido el delito. Dentro de los diferentes códigos éticos que existen, la guía ética de la Comisión Europea es la más completa y, para nuestro entorno democrático, la que se centra en el respeto de las personas y los derechos fundamentales. La responsabilidad penal en los casos de dolo no parece difícil determinarla. Es más complicado en los casos de imprudencia. Es por ello por lo que se propone una metodología de análisis para poder facilitar la labor de los juristas cuando se tienen que resolver los casos concretos que se dan con el uso de los sistemas inteligentes hoy día.

4. BIBLIOGRAFÍA

- ALONSO ÁLAMO, M. (2011), «Fundamentación pre-positiva de los bienes jurídico-penales y derecho penal mínimo de los derechos humanos». *Revista General de Derecho Penal*.
- BEN-ISRAEL, I. *et al.* (2020), «Towards regulation of AI systems Global perspectives on the development of a legal framework on Artificial Intelligence (AI) systems based on the Council of Europe's standards on human rights, democracy and the rule of law». CAHAI Secretariat.
- BREY, P.A. E. (2012), «Anticipatory Ethics for Emerging Technologies». *NanoEthics* 6, pp. 1-13.
- BRUNDAGE, M. *et al.* (2018), «The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation». Oxford.
- COMISIÓN EUROPEA (2020), *White Paper on Artificial Intelligence: a European approach to excellence and trust*. Accesible en: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en [16/04/25].
- DIGNUM, V. (2019), *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Cham: Springer.
- ERMAN, E., FURENDAL, M. (2022), «The Global Governance of Artificial Intelligence: Some Normative Concerns». *Moral Philosophy and Politics* 9, pp. 267-291.

- EUROPEAN UNION AGENCY FOR HUMAN RIGHTS (2020), «Getting the future right. Artificial intelligence and fundamental rights». Luxemburgo: Publications Office of the European Union.
- GÓMEZ MARTÍN, V. (2016), «Incitación al odio y género. Algunas reflexiones sobre el nuevo art. 510 CP y su aplicabilidad al discurso sexista». *Revista electrónica de ciencias penales y criminología* 18.
- HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HLEG) (2019), *A definition of Artificial Intelligence: main capabilities and scientific disciplines*. Accesible en: <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> [26/05/23].
- JANSEN, P. et al. (2019), «SIENNA D4.1: State-of-the-art Review: Artificial Intelligence and robotics». Zenodo.
- OECD (2021), «Recommendation of the Council on Artificial Intelligence». OECD.
- QUINTERO OLIVARES, G. (2017), «La robótica ante el derecho penal: El vacío de respuesta jurídica a las desviaciones incontroladas». *Revista Electrónica de Estudios Penales y de la Seguridad: REEPS* pp. 1.
- RASO, F. et al. (2018), «Artificial Intelligence & Human Rights: Opportunities & Risks».
- ROMEO CASABONA, C. M. (2020), «Criminal responsibility of robots and autonomous artificial intelligent systems?». *Comunicaciones en propiedad industrial y derecho de la competencia*, pp. 167-187.
- SALVADORI, I. (2021), «Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal». *Cuadernos de política criminal*, pp. 137-174.
- VALLS PRIETO, J. (2017), *Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial*. Madrid: Dykinson.
- (2023), «La inteligencia artificial frente a la colectividad: Una protección de los bienes jurídicos colectivos frente al auge de los sistemas inteligentes». *Estudios penales y criminológicos*, pp. 11.