Revista Electrónica de Ciencia Penal y Criminología
ISSN 1695-0194

Sobre la responsabilidad penal por la utilización de sistemas inteligentes

Javier Valls Prieto
Universidad de Granada

VALLS PRIETO, JAVIER. Sobre la responsabilidad penal por la utilización de sistemas inteligentes. *Revista Electrónica de Ciencia Penal y Criminología*. 2022, núm. 24-27, pp. 1-35.
http://criminet.ugr.es/recpc/24/recpc24-27.pdf

RESUMEN: En el ámbito penal ha habido varios intentos de determinar la responsabilidad que puede surgir por el uso de sistemas inteligentes. Las soluciones aportadas por la doctrina son buenas desde un punto de vista jurídico-técnico, sin embargo, su aplicación práctica dentro de la realidad de la inteligencia artificial puede presentar problemas, dejando la determinación de la responsabilidad penal en zonas grises que diluya la responsabilidad. Con el fin de solucionar este problema se propone una definición de inteligencia artificial que sirva para el mundo jurídico, con el objetivo de permitir entender las características de los sistemas inteligentes de tal manera que clarifique en qué fase de la vida del sistema inteligente se encuentran las esferas de responsabilidad en cada uno de los tres sujetos que interviene. Una vez determinada una definición se delimita el marco de principios que deben regir el desarrollo y uso de la inteligencia artificial con el fin de delimitar el contenido de las normas de debido cumplimiento aceptadas por los expertos. Al mismo tiempo, estos principios éticos nos van a suministrar herramientas para determinar los criterios para determinar la responsabilidad penal, como el control humano, la transparencia o la rendición de cuentas. Por último, estos principios nos van a servir de guía para determinar qué derechos fundamentales se encuentran en riesgo y qué bienes jurídicos son los que se van a ver lesionados. A partir de este estudio, el artículo ofrece una metodología de análisis de la responsabilidad que complete las aportaciones doctrinales ya existentes, facilitando a los operadores jurídicos determinar la responsabilidad por la creación, su uso profesional y su utilización por los usuarios finales de los sistemas inteligentes de forma que ofrezca seguridad jurídica a la industria y respete los derechos de los ciudadanos.

PALABRAS CLAVE: Responsabilidad penal, Inteligencia artificial, principios éticos, compliance, Derechos fundamentales.

TITLE: **On criminal liability for the use of intelligent systems**

ABSTRACT: Within the criminal field there have been several attempts to determine the different possibilities of liability that may arise from the use of intelligent systems. The solutions provided by the doctrine are very good legal-technical, however, their practical application in the reality of artificial intelligence may present problems, leaving the determination of liability in gray areas that dilute liability. To solve this problem, we are going to determine a definition of artificial intelligence that is useful for the legal world, with the aim of allowing us to understand the characteristics of intelligent systems in such a way that clarifies where the spheres of responsibility lie. Once a definition has been determined, we will limit the framework of principles that should govern the development and use of artificial intelligence to delimit the content of the rules to be complied with and accepted by professionals. At the same time, ethical principles will give us tools to determine the criteria that must be present when determining criminal liability, such as human control, transparency, or accountability. Finally, these principles will serve as a guide to determine which fundamental rights are at risk and to be able to determine which legal assets will be harmed. This article aims, based on this study, to offer a methodology for liability analysis that complements existing doctrinal contributions, making it easier for legal operators to determine liability for the creation, professional use and use by end users of intelligent systems in a way that offers legal certainty to the industry and respects the rights of citizens.

KEYWORDS: Criminal liability, artificial intelligence, ethical principles, compliance, fundamental rights.

Fecha de recepción: 15 mayo 2022

Fecha de publicación en RECPC: 31 agosto 2022

Contacto: jvalls@ugr.es

SUMARIO: I. Introducción. II. Definición de inteligencia artificial para el mundo jurídico. III. Principios éticos del Grupo de expertos de alto nivel de la Comisión Europea. IV. Derechos fundamentales afectados con el uso de la inteligencia artificial. V. Metodología para comprender los problemas jurídicos que surgen del uso de los sistemas inteligentes. VI. El sistema gestión de riesgos de la Propuesta de Reglamento de la inteligencia artificial como determinante de la responsabilidad. VII. Conclusiones. Bibliografía

---

I. **Introducción**

Este artículo pretende aportar una metodología para determinar la responsabilidad penal de los diferentes sujetos que intervienen en la vida del sistema inteligente. Los estudios realizados hasta ahora por la doctrina analizan, desde el punto de vista jurídico penal, la responsabilidad centrándose en tres puntos centrales: responsabilidad de la máquina autónoma, la responsabilidad por imprudencia y los sistemas de cumplimento normativo.[1] El objetivo principal de esta aportación es la de acercar estas aportaciones doctrinales a la realidad de la inteligencia artificial, ofreciendo una metodología de análisis para determinar las personas que van a ser responsables penalmente con la utilización de las diferentes tecnologías de inteligencia artificial en el contexto concreto en el que se implementa. Así, la utilización de reconocimiento de imágenes utilizada en medicina para analizar radiografías de tórax para diagnóstico médico difiere mucho la utilización de esta tecnología en prevención del crimen mediante reconocimiento facial. La base de la tecnología es igual, sin embargo, el fin y el contexto en el que se utilicen influirán en que el análisis penal de los sujetos implicado sea diferente. El médico se rige por unos principios diferentes de la policía, como sujetos profesionales que van a utilizar esta tecnología, y al mismo tiempo la ponderación de principios afectados implica un análisis jurídico diferenciado. De esta forma, el debate libertad del individuo/salud del individuo tiene un resultado diferente al que se produce en el entorno de la seguridad entre libertad del individuo/seguridad, que suele estar más presente en materias relativas al combate de la criminalidad. La metodología que se propone pretende crear 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.

El origen del concepto de inteligencia artificial surge en los años 50 del siglo XX de la mano de diferentes áreas de conocimiento que van desde la filosofía, la computación, las matemáticas, psicología o biología. En un primer momento la idea detrás

[1] DE LA CUESTA AGUADO, 2017, p. 1 y ss.; QUINTERO OLIVARES, 2017, p. 1 y ss.; ROMEO CASABONA, 2020, p. 1 y ss.

de este concepto 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 desarrollarla.[2] La evolución de esta tecnología en estos 70 años no ha estado libre de picos y valles en su desarrollo, lo que se ha denominado dentro del sector como los "inviernos de la inteligencia artificial",[3] siendo de los primeros los más importantes a principios de los 70, finales de los 80 y finales de los 90, este último en relación a la aplicación de redes neuronales en investigación espacial. Tal amalgama de disciplinas, de tecnologías y de definiciones han hecho difícil el tener un concepto claro y definitivo de lo que se considera inteligencia artificial. Encontrar una definición que pueda ser utilizada por los juristas de forma clarificadora del fenómeno es más importante que una el tener una definición técnica. Lo que nos interesa es que los encargados de aplicar las normas relativas a la inteligencia artificial sepan cuándo están ante un sistema inteligente y cuando no.

El actual auge en el que se encuentra esta tecnología se debe al aumento de la capacidad de procesamiento de datos y al abaratamiento del almacenaje de los mismos.[4] Para cuando llega la popularización de su uso nos encontramos, al principio de la segunda década del siglo XXI, con una creación de datos horizontal (realizada por los ciudadanos al utilizar los dispositivos móviles), una muy fuerte capacidad de procesamiento de estos y un buen desarrollo de ingenieros informáticos y de datos, tres factores que determinan la gran eclosión de la utilización de inteligencia artificial[5] en diferentes campos de la vida. La velocidad a la que se han producido los acontecimientos ha derivado en multitud de usos y de contextos en los que se desarrollan, dando tal cantidad de conceptos referidos a la inteligencia artificial que es difícil poder definirla de forma unívoca lo que implica dificultades a la hora de abordar los problemas que han derivado de uso a los derechos fundamentales de los ciudadanos, como ha ocurrido con el derecho a la igualdad o la privacidad. Por ello, la Comisión Europea ha tomado la iniciativa a la hora de determinar una definición que sirva para determinar de qué se está hablando y establecer el estándar de uso adecuado de esta tecnología, con la creación de las guías éticas sobre inteligencia artificial del grupo de altos expertos. Este documento se basa en un desarrollo y uso centrado en los humanos y respetuoso con los Derechos Fundamentales, que estable 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 praxis.

Existen ya multitud de ejemplos en los que el uso de los sistemas inteligentes lleva

[2] JANSEN/BROADHEAD/RODRIGUES/WRIGHT/BREY/FOX/WANG, 2018, p. 12.
[3] LEE, 2018, p. 6.
[4] MAYER-SCHONBERGER/CUKIER, 2013, p. 6.
[5] LEE, 2018, p. 14.

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

Una vez determinada la importancia penal del uso de la inteligencia artificial nos vamos a centrar en el estudio que ha realizado la doctrina sobre la responsabilidad penal. Podemos sintetizar el análisis en tres ámbitos: la responsabilidad de la máquina autónoma por sus propias decisiones, la responsabilidad por imprudencia de los humanos que la crean o la utilizan, teniendo que determinar cuáles son las normas de cuidado y, por último, los mecanismos de cumplimiento normativo, similares a los de las personas jurídicas.[6] Los tres ámbitos van a ser analizados desde la perspectiva real centrada en el ciclo de vida de la inteligenciar artificial, que va desde los desarrolladores e industria, hasta los usuarios finales, pasando por su uso profesional, con el fin de no limitarlo únicamente al estudio teórico. Desde la experiencia de los proyectos europeos ePOOLICE, COPKIT y SIENNA se ha creado una metodología de análisis del impacto en los derechos fundamentales para determinar quién tiene la responsabilidad sobre la lesión del bien jurídico.

Como hemos señalado, los análisis de responsabilidad penal hasta este momento se han centrado en un modelo dogmático clásico. A la par que han aparecido estas aportaciones por parte de la doctrina, la Unión Europea ha desarrollado mucho material en relación con la inteligencia artificial. De todos estos trabajos de la Comisión, nos interesan para nuestro estudio dos, que tienen un interés particular para centrar el estudio de la responsabilidad penal: las guías éticas sobre inteligencia artificial del grupo de altos expertos y el borrador de Reglamento sobre la inteligencia artificial que han de ser utilizados a la hora de ver la responsabilidad penal para contextualizar el análisis penal de esta tecnología.

Basándonos en estos dos documentos vamos a determinar los conceptos que debe utilizar el jurista, en general, a la hora de realizar un análisis y al penalista, en particular, para determinar qué bienes jurídicos se van a lesionar y quiénes son los responsables penales de este daño.

[6] DE LA CUESTA AGUADO, 2017, p. 1 y ss.; QUINTERO OLIVARES, 2017, p. 1 y ss.; ROMEO CASABONA, 2020, p. 1 y ss.

II. **Definición de inteligencia artificial para el mundo jurídico**

La inteligencia artificial es considerada como la capacidad de la máquina de pensar, de razonar y actuar con inteligencia. Sin entrar de lleno en la definición de inteligencia, podemos considerar que la inteligencia artificial es un sistema que puede realizar tareas como si las realizasen humanos.[7] En determinados contextos, encontramos casos en los que realizan la tarea designada con un rendimiento mejor que el de los humanos, como pueda ser el ejemplo de la búsqueda de información, reconocimiento de tumores en radiografías o jugando al ajedrez.

Encontrar una definición de inteligencia artificial es complicado, no ya por las disquisiciones filosóficas sobre que es inteligencia, lo cual es complicado por sí mismo, sino porque, además, la definición de inteligencia artificial cambia dependiendo de la disciplina que trata de abordarla y de los objetivos que se pretenden conseguir con la definición.[8]

Hay algún consenso en que la definición se puede encontrar mediante dos caminos, por qué hace o por cómo funciona, como conjunto de mecanismo causales que dan a un resultado inteligente. Si nos centramos en la primera se puede observar que existen diferentes técnicas que pueden ser denominadas inteligencia artificial. Así, si puede ver e identificar lo que ve, podemos considerarlo como visión computación y procesamiento de imagen. Si puede escuchar y responde de manera útil y sensible, podemos hablar de procesamiento del lenguaje natural. Este último concepto puede incluir a aquellos algoritmos que pueden leer lo que se está escribiendo o leer párrafos de texto analizándolo por patrones. Si se puede mover por sí mismo mediante lo que ve y escucha, se puede denominar un dispositivo inteligente. Finalmente, si puede razonar mediante el análisis de gran cantidad de datos buscando patrones que permitan decidir, entonces lo podemos denominar *machine learning*.[9]

Desde el segundo enfoque nos vamos a encontrar varios intentos de definición. El primero que vamos a ver es el realizado por Russel y Norvig que distinguen cuatro definiciones[10] dependiendo de diferentes claves que nos van a servir para clasificarlas. Estas son: procesar pensamiento y razonamiento, comportamiento, actuación humana y actuación ideal. Siguiendo este análisis para determinar cuándo un sistema inteligente va a realizar una terea humana, podremos considerar que cualquier máquina que actúa como un humano podría ser considerada como inteligencia artificial. Esta primera definición se basa en el test de Turing que, resumidamente, lo que pretende es ver si una máquina puede ser confundida por una persona. El test se supera si contestadas las preguntas establecidas la persona no puede distinguir si está inter-

[7] Inf. Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe, 2017, p. 2.
[8] MIRÓ LLINARES, 2018, p. 91 y ss.
[9] HAO, 2018, p. 1.
[10] RUSSEL/NORVIG, 2010, p. 2 y ss.

accionado con un ordenador o un humano. El planteamiento de Turing en la actualidad ha sido superado por maquinas en muchas ocasiones y, por este motivo, los científicos, hoy día, han considerado más interesante conocer lo que hay detrás de la actuación humana más que en replicarla.

Desde el punto de vista del comportamiento se intenta dar una definición basada en la habilidad de la máquina de pensar/comportarse como un humano desde el punto de vista cognitivo. El fin principal es observar cómo funciona internamente el conocimiento humano y aplicar los mismos procedimientos a los algoritmos a la hora de realizar actividades consideradas como inteligentes.

En los años 70 el desarrollo de una definición se centra en la actuación humana, el pensamiento racional, tomando las leyes de la lógica para construir razonamientos lógicos/inteligentes equivalentes a los humanos. Se encontraron límites en su desarrollo. Por un lado, no es fácil integrar pensamiento informal en un sistema de reglas lógicas, sobre todo cuando la información de la que se nutre no es completamente certera y, por otro, no es lo mismo resolver problemas lógicos-teóricos que problemas reales. Esta situación se presentó con los correctores ortográficos de los programas de procesamiento de texto. La ortografía es un conjunto lógico de reglas con algunas excepciones que permite escribir sin errores. Si se conocen las reglas se escriben bien. En el desarrollo del procesador de textos Word, que empezó funcionando bajo estas premisas lógicas y con unos resultados bastante buenos, sin embargo, los ingenieros descubrieron, tras ser superado por el corrector de su competidos Google docs, que los sistemas de corrección basados en el análisis de textos (básicamente en muchos datos) y no en las reglas de la ortográfica tenían mejor porcentaje de éxito.

Por último, la actuación ideal. Como señalan Russel y Norvig en su cuarto punto, actuar es realizar algo y son los agentes los que realizan estas acciones. Los agentes inteligentes son los que realizan acciones de la mejor manera posible. No sólo esperamos que realicen una acción, sino que esta sea la mejor posible, incluso superando las habilidades de las personas. Aquí se puede ver como las definiciones se pueden solapar. Así, el mejor resultado en muchas ocasiones depende del razonamiento lógico y está directamente relacionado con el pensamiento racional. La deducción correcta no siempre es la racional, como muestra el ejemplo de la reacción refleja cuando uno se quema. Si tuviésemos que hacer un razonamiento lógico ante la situación de coger un cazo ardiendo, muy probablemente, la solución no sería la menos lesiva que la que nos depara la irracional, que es soltar el objeto cuanto antes, aunque se pueda romper o provocar otro tipo de daños.

Siguiendo con el intento de obtener una definición de inteligencia artificial basada en cómo funciona se han desarrollado varios documentos por diferentes instituciones internacionales con descripciones técnicas sobre lo que es la inteligencia artificial. Sistematizar un campo tan extenso y que evoluciona tan rápido plantea un gran reto de difícil solución.

Para tener una idea del reto al que nos enfrentamos, el estudio más importante sobre la definición de la inteligencia artificial es el realizado por el *Joint Research Center* sobre Inteligencia Artificial[11], analizando las definiciones a nivel europeo, nacional e internacional, desde una perspectiva de la investigación científica generada en el área, de las definiciones generadas por el mercado y por la industria, dando diferentes perspectivas para analizar el concepto, con la finalidad de ofrecer una definición operacional. El principal problema al abordar tal tarea es que la inteligencia artificial es un campo dinámico que está en continuo cambio lo cual dificulta su análisis. Después de analizar 55 documentos, cada uno con una definición propia, se puede encontrar elementos comunes dentro de cuatro características de la inteligencia artificial: percepción del entorno, incluyendo la consideración de la complejidad del mundo real; procesamiento de información, recopilando e interpretando inputs en forma de datos; toma de decisiones (que implica razonamiento y aprendizaje), tomando acciones, realización de tareas con cierto niveles de autonomía; y lograr determinados objetivos específicos, considerado como la última razón de los sistemas de inteligencia artificial.[12] Además de análisis de las definiciones, se centra en las tareas que puede realizar, estructurando en seis grupos lo que se puede entender por inteligencia artificial: aprendizaje máquina, visión computacional, procesamiento del leguaje natural, vehículos conectados y automáticos, robótica y servicios de inteligencia artificial.[13] La conclusión final es que no puede darse una definición concreta.

Para los objetivos de esta investigación y con el fin de simplificar el problema vamos a centrarnos en dos de estos documentos que destacan por su importancia como referentes. El primero es el realizado por la Asociación por el avance de la inteligencia artificial (*Association for the Advancement of Artificial Intelligent*, AAAI), que la definió como "la disciplina científico-técnica que se ocupa de la comprensión de los mecanismos subyacentes en el pensamiento y la conducta inteligente y su incorporación en las máquinas"[14] concepto que, desde el punto de vista del mundo de la ingeniería, puede ser suficiente para abarcar todos los diferentes aspectos que van surgiendo en la disciplina, pero que, desde el punto de vista jurídico, no cubre las necesidades que nos van a surgir a la hora de resolver los problemas legales que puedan surgir ya que es una definición extremadamente ambigua. El segundo documento es el realizado por OCDE, que considera un sistema de inteligencia artificial es un sistema basado en máquina que puede, para un conjunto de objetivos

[11] SOMOILI, 2020, p. 1 y ss.
[12] SOMOILI, 2020, p. 8.
[13] SOMOILI, 2020, p. 10.
[14] Inf. ASSOCIATION FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE, 2016, p. 1 [Consultado 27/02/2020] https://ai100.stanford.edu/2016-report/section-i-what-artificial-intelligence/defining-ai

definidos por humanos, realizar predicciones, recomendaciones o decisiones influidas por entornos virtuales o reales. Los sistemas de inteligencia artificial están diseñados para operar con diferentes niveles de autonomía.[15] Esta definición, al menos, nos va a permitir tener algunos de los puntos clave para su análisis jurídico que veremos más adelante. Como se puede apreciar hay una gran diferencia entre las conceptualizaciones que interesan a los creadores y las que interesan a los juristas.[16]

Recientemente, desde la Unión Europea se han promovido dos documentos que nos van a servir para determinar nuestra posición sobre la definición que estamos estudiando. El grupo de expertos de alto nivel en inteligencia artificial de la Comisión Europea ha redactado las guías éticas para el uso fiable de la inteligencia artificial, intentando enmarcar el concepto de inteligencia artificial. Uno de los primeros puntos que resalta es la dificultad de definir lo que es la inteligencia, tanto en humanos como en máquinas, ya que se trata de un concepto vago. Es por ello por lo que, en un cambio de enfoque, prefieren centrarse, aunque no exclusivamente, en el concepto de racionalidad, entendida como la habilidad de elegir la mejor opción para conseguir un determinado objetivo, dados determinados criterios y los recursos disponibles. Otro punto interesante de partida es que no hablan de inteligencia artificial *per se,* sino que prefieren utilizar el término sistema de inteligencia artificial para definir a cualquier componente basado en inteligencia artificial, ya sea software o hardware. Esta decisión se basa en la inteligencia artificial se encuentra como un componente incluido en sistemas mayores más que como un componente único.[17] Continúa el documento con algunas definiciones que hemos tratado anteriormente (aprendizaje de máquina, inteligencia, inteligencia artificial general o específica, etc.) para, finalmente, aportar su definición de inteligencia artificial: "Sistemas de inteligencia artificial son sistemas de software (y posiblemente, también, de hardware) diseñados por humanos que, dado un objetivo complejo, actúa en la dimensión física o digital mediante la percepción de su entorno, adquiriendo de datos, interpretando datos estructurados y desestructurados, razonando el conocimiento o procesando la información derivados de estos datos y decidiendo la/s mejores acción/es para conseguir el fin perseguido. Los sistemas de inteligencia artificial pueden usar reglas simbólicas, aprender un modelo numérico o pueden adaptar su comportamiento analizando como el entorno se ve afectado por sus acciones previas."[18] Esta

[15] Inf. OECD, 2019, p. 7.

[16] El único intento de definir la inteligencia artificial por parte de un jurista es el realizado por Hallewy que parte de una visión muy cercana a la de la AAAI para intentar solucionar la responsabilidad penal de la inteligencia artificial. HALLEVY, 2010, p. 175.

[17] Inf. INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, 2019, p. 1 [Consultado 02/03/2020] https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[18] Inf. INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, 2019, p. 6 [Consultado 02/03/2020] https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

decisión ha sido la mejor considerada por el JRC en su análisis y, para nosotros, también es la que cumple mejor las funcionalidades para nuestro trabajo.

El último de los documentos que vamos a analizar es el de 2021 del Parlamento Europeo y el Consejo, cuando lanzaron una propuesta de regulación de la inteligencia artificial. En su artículo 3, párrafo 1, define la inteligencia artificial como un software que puede generar resultados como contenido, predicciones, recomendaciones o decisiones por el análisis de los entornos con los que interactúa en base a un conjunto de datos dado. Este software es desarrollado con una o más técnicas de las referidas en el Anexo I de la propuesta. En realidad, no se define la inteligencia artificial sino las técnicas usadas como inteligencia artificial. Siendo realistas, la propuesta comienza con la inteligencia artificial basada en técnicas de *machine learning*, incluyendo el aprendizaje reforzado, supervisado y no supervisado, usando una variedad muy amplia de métodos, entre los que se incluyen el *Deep learning*. Como podemos observar, la primera definición es el uso más actual de esta tecnología, pero no es la única, como hemos podido ver con anterioridad. En el punto b de la definición se centra en los enfoques basados en lógica y conocimiento, incluyendo representación de conocimiento, programación (lógica) inductiva, bases de conocimiento, motores deductivos e inferenciales, razonamiento simbólico y sistemas expertos. La novedad de esta definición radica en la inclusión de enfoques estadísticos, estimación bayesiana y métodos de búsqueda y métodos de optimización.

En todas las definiciones de inteligencia artificial se pueden diferenciar los dos tipos de inteligencia artificial anteriormente descritos, la basada en el conocimiento y la basada en datos, pero incidiendo en un factor que hemos señalado anteriormente: el auge de la inteligencia artificial en la tercera revolución es debido al auge de procesamiento de datos por los factores anteriormente descritos.

De este análisis podemos destacar que las maquinas tienen la posibilidad de realizar comportamientos automáticos inteligentes en la medida en que pueden realizados de forma parecida a los humanos dentro de un contexto concreto en el que se tienen que elegir entre diferentes posibilidades y que, en determinadas facetas de la vida, aportan soluciones útiles para los humanos.

Para entender cómo va a afectar al mundo del Derecho y cómo se va a tener que lidiar en el sistema judicial cuando se planteen los problemas judiciales, nos podemos quedar con algunos puntos interesantes. El primero y muy importante es entender que la inteligencia artificial es una herramienta creada por las personas. Esto implica que, aunque pueda crear cosas (entre ellas más inteligencias artificiales) siempre va a ser mediante unos patrones impuestos por los humanos, que son el origen de la actividad que desarrollen y, por tanto, la máquina no va a realizar ninguna tarea que no sea predeterminada por un humano.[19] El segundo es que los algoritmos de toma de decisiones de los sistemas de inteligencia artificial se alimentan de datos. Es por

[19] SALVADORI, 2021, p. 150

esto por lo que la privacidad se encuentra en grave riesgo por la capacidad de procesamiento y el cruzado de bases de datos que se desarrollan con estos sistemas inteligentes. Y de ahí el mercado que hay con los *datatraders*, vendedores y compradores de datos, que obtienen sustanciales ganancias a consta de los datos de los usuarios de diferentes servicios y aplicaciones. Un tercer punto de interés para los juristas es que cada sistema de inteligencia artificial tiene un objetivo determinado en su función. Así, el acotar el ámbito de aplicación práctico, olvidando las concepciones de una inteligencia artificial general que supera las capacidades generales de los humanos, permite concentrase en medidas concretas para evitar un impacto negativo en los derechos fundamentales de los ciudadanos, facilitando la concreción de las medidas exigidas de protección al estar el ámbito de actuación muy limitado. Por último, la definición del grupo de altos expertos, aunque no lo numera expresamente, indica que los sistemas inteligentes no son capaces de adivinar lo imprevisto ni de crear algo de la nada. Todo aprendizaje, intento de predecir el futuro o realización de tareas se basa en precedentes que pueden venir de los datos o de situaciones que dan información de supuestos anteriores similares, pero no es capaz de actuar de forma efectiva ante situaciones completamente nuevas. Hay que descartar, en el análisis jurídico, cualquier intento de equiparar las acciones de los humanos con las de los sistemas inteligentes por muy eficaces que puedan ser. Hoy por hoy, no existe un sistema inteligente capaz de reaccionar como un humano, aunque puede realizar tareas humanas con mayor eficacia. Por poner un ejemplo, los coches autónomos tienen serias dificultades de identificar señales de tráfico cuando las situaciones no son ideales. Así, que una señal esté doblada o que tenga pegatinas en su superficie hace que el vehículo no sea capaz de identificarla de forma apropiada.[20] Lo que para un humano es un problema que se resuelve sin pensar para una máquina puede no tener solución que, en la vida real, sí puede tener un impacto jurídico importante.

La definición de inteligencia artificial que nos va a permitir resolver los problemas desde el punto de vista jurídico es la del grupo de altos expertos, a la que nos adherimos. Por un lado, la definición de inteligencia artificial, con la que comienza el documento, es lo suficientemente clara y precisa para poder determinar a qué nos vamos a referir. Por otro, el desarrollo que hacen del concepto de inteligencia artificial fiable, que es la que consideran que es la aceptable para la sociedad, ha de respetar los derechos humanos y cumplir con los siete requisitos que enumera. Los siete principios tienen una importancia a la hora de determinar la responsabilidad penal. Así, el primero de ellos, la acción y supervisión humanas, va a trasladar las formas de responsabilidad siempre hacia una persona ya que ningún sistema debe tomar decisiones autónomas sin el control por parte de un humano, siendo este el último responsable. Dentro de estos principios se exige una solidez técnica y de seguridad para

[20] BROUSSARD, 2018, pp. 125 y ss.

impedir que el sistema sea defectuoso ni que pueda ser dañado o alterado en su funcionamiento lo que implica tomar e implementar unas medidas de cuidado a la hora de desarrollarlos. Un tercer punto es que, debido a que, os sistemas inteligentes en su gran mayoría se alimentan de datos, se va a exigir que se gestionen de forma adecuada los datos y se respete la privacidad de las personas en su uso. La exigencia de transparencia implica que la toma de decisiones sea explicable, se pueda entender por qué el sistema da una determinada solución, que se sepa cuál es su finalidad y la capacidad que posee al igual que un deber de informar adecuadamente del funcionamiento del sistema. El requisito de que el sistema inteligente debe ser equitativo, no discriminatorio y pensar en la diversidad de grupos de una sociedad nos lleva al respeto de las minorías y que en su desarrollo y utilización se eviten las situaciones discriminatorias. Los bienes colectivos de medio ambiente, sostenibilidad, impacto social y democracia deben tenerse en cuenta en su desarrollo y uso. Finalmente, ha de tenerse una rendición de cuentas que implica que el sistema ha de ser auditable, que minimice los efectos negativos y en el caso en que se produzcan notificarlos, buscar equilibrios entre su uso y los riegos que van aparejados y, para terminar, gestionar las compensaciones que deriven de su uso sabiendo quién es el responsable de las medidas tomadas[21].

Esta definición de las guías éticas de la Comisión Europea es la más acertada para el mundo jurídico, permitiendo incluir no sólo a los sistemas inteligentes basados en *software* sino también aquellos cuyos componentes principales son de *hardware* como es el caso de los robots guiados por inteligencia artificial. Es por ello por lo que denominan sistemas inteligentes a los sistemas de *hardware* o *software* que tiene una autonomía. La clave de esta definición, desde nuestro punto de vista, es el hecho de que el concepto de inteligencia artificial fiable implica el cumplimento de unos principios, que van a constituir la base de la utilización adecuada de los mismos y el respeto a los derechos fundamentales. El decantarnos por esta definición, para su uso en el mundo jurídico, nos va a permitir adentrarnos en el contenido de los principios que van a determinar qué ha de entenderse como comportamiento debido en el desarrollo y uso de estos sistemas y poder centrarnos en qué elementos son importantes para delimitar la responsabilidad penal.

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

En 2018 la Comisión Europea creó un grupo de altos expertos en inteligencia artificial con el fin de realizar unas guías éticas 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 robusta. La utilización de sistemas inteligentes lleva

[21] COMISIÓN EUROPEA, 2019, p. 18.

consigo dilemas, problemas y conflictos sociales, muchos de ellos que atentan contra los derechos fundamentales, que se pueden prevenir cumpliendo con una serie de principios éticos. El documento ha sido un referente dentro de la comunidad de inteligencia artificial y robótica para la estandarización de las guías a nivel mundial.

La base detrás de este proyecto son los tres pilares sobre los que se van a desarrollarla inteligencia artificial, que como ya hemos señalado, se espera que sea legal, con un compromiso con los derechos fundamentales y el respeto al Estado de Derecho; que sea ética, una inteligencia artificial fiable, con siete requisitos que vamos a desarrollar; y, por último, que sea robusta, que no cause ningún daño imprudente, es decir, que sea segura internamente, que sea segura externamente y que se pueda confiar en el sistema.[22]

La base de acercamiento de la ética en la inteligencia artificial se basa en los derechos fundamentales en los tratados de la Unión Europea, la Carta de la Unión Europea y las normas internacionales de Derechos Humanos.[23] El grupo de expertos de alto nivel han destacado 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 ciudadanos.[24]

De aquí surgen cuatro principios éticos principales: el respeto a autonomía de los humanos, el principio de prevención de daños, el principio de trato justo y el principio de explicación.[25] Estos cuatro principios generales van a servir de base para la determinación de los requisitos que se deben seguir para conseguir que la inteligencia artificial sea fiable. Se han destacado siete requisitos por parte del grupo de expertos que no constituyen un numerus clausus de por sí. Estos 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 cuentas.[26]

Con el primero se persigue que los usuarios de sistemas basados en inteligencia artificial 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. En referencia a la supervisión humana se trata de asegurar que los sistemas basados en inteligencia artificial no minan la autonomía humana o causa otros efectos adversos.[27]

La robustez técnica y la seguridad interna hacen referencia a que los sistemas deben

[22] COMISIÓN EUROPEA, 2019, p. 18.
[23] COMISIÓN EUROPEA, 2019, p. 9.
[24] COMISIÓN EUROPEA, 2019, pp. 10 y 11.
[25] COMISIÓN EUROPEA, 2019, pp. 12 y 13.
[26] COMISIÓN EUROPEA, 2019, p. 14.
[27] COMISIÓN EUROPEA, 2019, p. 18.

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 variaran 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 presiones, 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 considera un sistema robusto.[28]

Los elementos sobre privacidad y gobierno de datos[29] implican tres pilares. La protección de datos y de la privacidad debe estar garantizada en todo el ciclo de uso de la inteligencia artificial, en sus tres fases, obtención de datos para alimentar el sistema, el procesamiento de los mismos y los resultados que el sistema ofrece. Hay que preocuparse de que la calidad e integridad de los datos que hacen funcionar la inteligencia artificial sea la adecuada. Así, en la fase de recogida de los datos se puede encontrar que estos tienen sesgos sociales,[30] son inexactos o erróneos. Esta integridad de los datos se debe de 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 sólo personal cualificado con competencia y una necesidad clara puede justificar el acceso a los datos individuales. La privacidad es quizás el tema estrella para los juristas cuando entran a abordar la inteligencia artificial ya que, aparte de ser uno de los derechos que han sido más afectados por las técnicas de Big Data,[31] está debidamente protegida en el Código penal y ha sido estudiado ampliamente por la doctrina.

[28] COMISIÓN EUROPEA, 2019, p. 18 y 19.

[29] Para el contexto de esta obra se va a entender gobierno de datos (data governance) como las estructuras en la organización, propietarios de datos, políticas, reglas, proceso terminaos de negocios y métricas en todo el proceso de la vida de los datos, desde su recolección, almacenamiento, uso, protección, archivo y borrado). Mientras que por gestión de datos se entenderá la implementación técnica del gobierno de datos.

[30] VIDA FERNÁNDEZ, 2018, p. 219.

[31] Cfr. VALLS PRIETO, 2017, p. 1 y ss.

Relacionado con el requisito de explicación está el de transparencia[32] 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 cosiste en que 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 inteligencia artificial y de las decisiones relacionada a 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 inteligencia artificial fiable es necesario observar todo el ciclo de vida de un sistema basado en inteligencia artificial 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.[33] Los supuestos de discriminación por datos históricos que están sesgados o incompletos y malos modelos de gobernanza han creado sistemas inteligentes que han aumentado la desigualdad, la discriminación y disminuido la calidad de vida de muchos de los interesados.[34] Es este intento de crear sistemas inteligentes supeditados al principio de no discriminación, trato justo y diversidad plantea un modelo sobre el que trabajar con tres elementos. El primero es evitar los sesgos injustos. La identificación de sesgos discriminatorios va a permitir eliminarlos de la fase de obtención de datos, por ejemplo. Esto se conseguirá con la supervisión, segundo elemento, del proceso de análisis y concretar el fin, los requerimientos, limitaciones y decisiones que permita de manera clara y transparente el control. El tercero consiste en tener en cuenta las opiniones de los diferentes grupos de afectados por su uso. Si se pasa por el escrutinio de grupos culturales y con diferentes antecedentes, así como un enfoque multidisciplinar permitirá encontrar las perspectivas de diferentes grupos. El diseño universal y la accesibilidad permiten que todos los grupos de usuarios puedan utilizar los sistemas inteligentes de forma adecuada. Todos los grupos de edad, de sexo, de habilidades o de características deberían estar incluidos en el diseño de los sistemas. Consultar a los interesados en su utilización, producción y los afectados por sus decisiones es beneficioso para todos.

La implementación y desarrollo de la inteligencia artificial debe tener en mente la

[32] COMISIÓN EUROPEA, 2019, p. 18.
[33] COMISIÓN EUROPEA, 2019, pp. 18 y 19.
[34] El caso de las ayudas sociales en Estados Unidos es lo suficientemente clarificativo.

responsabilidad ecológica y la sostenibilidad[35] para mantener el bienestar medioambiental. La ubicuidad de los sistemas inteligentes en todas las áreas de nuestra vida puede alterar la concepción de la acción social. La inteligencia artificial puede facilitar las actividades humanas, pero también puede perjudicarlas. Al mismo tiempo debe tenerse en cuenta el impacto en las instituciones, democracia y sociedad en sentido amplio. Se debe observar la influencia de la inteligencia artificial en el proceso democrático, no sólo en la decisión política sino también en los contextos electorales.

El último de los principios es el de rendición de cuentas.[36] La auditoría implica capacitación en la evaluación de algoritmos, datos y procesos diseñados. No debe tener información sobre el sistema de negocio y sobre la propiedad intelectual, pero los sistemas deben estar disponibles en abierto para una auditoría interna y externa. La minimización y la información de los impactos negativos sirven para responder a las consecuencias del resultado obtenido. La obligada rendición de cuentas debe estar disponible para informadores, ONGs, sindicatos y otras entidades que documenten sobre legítimas preocupaciones sobre los sistemas basados en inteligencia artificial. Al implementar todas estas medidas se van a producir conflictos y discusiones que deberán resolverse con concesiones por las partes. En este punto, estas concesiones deben hacerse teniendo en cuenta los riesgos para los principios éticos y los derechos humanos. Cuando se produzca un impacto adverso se deben prever los mecanismos que aseguren la adecuada compensación, teniendo en especial consideración a los grupos y personas vulnerables.

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 máquina,[37] ya que establecen un control de humano sobre el sistema inteligente y las decisiones que toma.

IV. **Derechos fundamentales afectados con el uso de la inteligencia artificial**

Asentado y determinados los principios sobre los que se debe basar el desarrollo y uso de la inteligencia artificial podemos derivar qué Derechos fundamentales y qué bienes jurídicos se ven reforzados, en el caso de los primeros, o estarían en riesgo, en ambos casos. 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. Ante la dificultad de poder determinar claramente el qué Derechos fundamentales se ven afectados y de qué manera, por un lado, por la falta de una definición clara de inteligencia artificial

[35] COMISIÓN EUROPEA, 2019, p. 19.
[36] COMISIÓN EUROPEA, 2019, p. 19-20.
[37] SALVADORI, 2021, p. 150

y, por otro, por la diversidad de usos y campos de utilización en los que se pueden emplear que deriva en diferentes forma de impacto, estos estudios se centran en una metodología mixta de análisis, por un lado, se estudiarán casos de uso por sector con el fin de observar cómo esta tecnología varía en su impacto dependiendo de contexto, y, por otro, estudios sociológicos jurídicos que van a ofrecer información relevante sobre los peligros jurídicos que plantea la implementación de los sistemas jurídicos, que serán de gran utilidad para nuestro análisis sobre las situaciones en las que se pueda derivar una responsabilidad penal.

El primero de estos estudios, realizado por el centro Beckman Klein, se centra en el impacto en los derechos fundamentales (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) en seis casos de uso: sistema judicial, sistema financiero, sanidad, moderación de contenido, recursos humanos y educación.[38] Esta selección de casos tiene como objetivo el que sirvan de referencia a 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 estos derechos. 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 –la naturaleza y complejidad inherente al proceso, así como la inescrutabilidad de los resultados- y el derecho a la privacidad –la necesidad de grandes cantidades de datos que genera un problema para la privacidad. Existe un segundo nivel de impacto, que podemos considerar como neutro, porque tiene un impacto positivo y negativo que dependerá de su uso, entre los que se encontrarían el principio de igualdad ante la ley y el derecho a la no discriminación –el sistema puede evitar la discriminación que tienen los jueces como humanos pero al mismo tiempo por los datos, por el diseño o por la interpretación de los resultados, puede generar sesgos en algunos grupos de ciudadanos-, el derecho a la vida, la libertad o la seguridad personal –los casos de menor riesgo pueden verse beneficiados por unas sentencias más rápidas y con penas más cortas y la sociedad puede beneficiarse de una mayor efectividad de la justicia al tener resoluciones más rápidas- el derecho a no ser arrestado, detenido o expulsado arbitrariamente –los algoritmos pueden clasificar a ciertos individuos como de alto riesgo, dando la posibilidad de arrestos pre/post juicio.[39] En los casos de impacto negativo tenemos una lesión de

[38] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 19.
[39] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 20.

los bienes jurídicos relacionados con los funcionarios públicos y con la administración de justicia. También pueden verse afectados los bienes jurídicos relacionados con la defensa de la Constitución, en concreto, con el derecho a la igualdad.[40] 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 lesiones bienes jurídicos.

El siguiente caso de uso que se analiza es el de los sistemas crediticios que toman sus decisiones basadas en sistemas de inteligencia artificial. Este uso puede tener un efecto positivo en el respeto del derecho a la igualdad ante la ley –grupos de población que no conseguirían acceso a un crédito mediante el análisis por humanos pueden verse beneficiados del análisis de crédito por parte de la inteligencia artificial- y el derecho a un estándar de vida –al aumentar los sectores de población que pueden acceder al crédito se permite que estos sectores puedan utilizar el dinero para mejorar su vida. A un doble nivel, positivo y negativo, se encontrarían los derechos de no discriminación y el derecho a la educación. El primero, al igual que en el sistema judicial, va a depender de si los datos en los que se basan las decisiones están sesgados o si son erróneos, pudiendo mejorar las decisiones de los humanos. El segundo, el acceder a crédito, que con humanos no se conseguirían, permitiría mejorar la educación de los beneficiarios. Desde el punto de vista de un impacto negativo afectaría a la privacidad, a la libertad de expresión, opinión e información y al derecho de asociación –como todos los datos pueden ser utilizados como datos para obtener el crédito los solicitantes de los mismos pueden tener inquietud a expresar sus opiniones para no perjudicar sus posibilidades de obtención de crédito- y el derecho a conseguir el trabajo deseable –la falta de financiación puede frustrar las aspiraciones laborales de determinadas personas.[41] Además, del respeto a la igualdad y la no discriminación, nos encontramos que se ven afectados los bienes jurídicos relacionados con el patrimonio y el mercado, con especial relevancia a la protección de los consumidores.

El análisis que realizan del sistema sanitario destaca el beneficio positivo por el uso de esta tecnología 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 consigue una mejora en la salud de muchos sectores de población que, principalmente en EE.UU., 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 es excluida 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 médicos.[42] Dentro del sistema sanitario está claro que los bienes

[40] RODRÍGUEZ YAGÜE, 2007, p. 5
[41] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 26.
[42] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 32.

jurídicos vida e integridad física pueden verse afectados con el uso de la inteligencia artificial. 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 comenten más errores que los humanos a la hora de borrar contenidos legales.[43] Los delitos relativos a la Constitución, como defensores del buen funcionamiento de las instituciones democráticas, están siendo afectados por los sistemas inteligentes, siendo necesario un adaptación de los tipos penales para los nuevos ataques a los bienes jurídicos.

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 sólo 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 inteligencia artificial, 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 que las personas que estén buscando trabajo restrinjan sus opiniones políticas para que no afecte a sus posibilidades de conseguir un empleo ya que, si el empleador realiza un análisis de datos en, por ejemplo, sus redes sociales pueden 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 privacidad.[44] El impacto en los bienes jurídicos de los trabajadores en estos casos es de gran importancia.

[43] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 37.
[44] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 42.

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 mermada.[45] No existen como tales bienes jurídicos que protejan la educación de los ciudadanos, sin embargo, la exclusión de sectores a sistema educativo afecta claramente a bienes jurídicos como la igualdad,[46] la libertad de los ciudadanos o el libre desarrollo de la personalidad.

El segundo informe que es de nuestro interés es el de la Agencia de la Unión Europea para la Derechos Fundamentales que utiliza más de 100 entrevistas a personas del sector público y privado, relacionados con la inteligencia artificial en Europa, con el fin de detectar cuál es su consideración con respecto al impacto en los derechos fundamentales del uso que se da a la inteligencia artificial en cuatro áreas: ayudas sociales, predicción del crimen, servicios de salud y publicidad personalizada.[47] Toma como base la Carta de Derechos Fundamentales de la Unión Europea, los Tratados de la Unión Europea -en donde se pueden encontrar otros derechos y principios además de los de la Carta- y las normas de derecho derivado. Ejemplo de los Derechos incluidos en los Tratados podemos encontrar el de no discriminación, que viene en la Carta (artículos 20 y 21), en el Tratado de la UE y los Tratados del Funcionamiento de la Unión Europea (artículos 2 y 10 respectivamente) y en el Derecho derivado (Directivas 2000/78/CE, 2000/43/CE, 2004/113/CE y 2006/54/CE) consiguiendo una mayor determinación conforme se deriva al Derecho secundario.

Del análisis del texto podemos detraer 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 Justica, el derecho a la seguridad social y a la asistencia social, los relacionados con los consumidores y el de la ciudadanía a una buena administración.[48] Todos estos derechos tienen un reflejo en el código penal con un bien jurídico afectado que deriva a un tipo penal que le da protección. Este análisis realizado por la Agencia europea busca no sólo realizar el análisis, sino que se acerca a la defensa de estos derechos fundamentales mediante el sistema jurídico.

El artículo 1 de la Carta establece la dignidad humana como uno de los pilares de

[45] RASO/HILLIGOSS/KRISHNAMURTHY/BAVITZ/KIM, 2018, p. 47.
[46] GÓMEZ MARTÍN, 2016, p. 18
[47] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 1.
[48] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 60 y ss.

la Unión Europea y esto implica que todo desarrollo de inteligencia artificial ha de estar supeditado a los humanos, como señala expresamente las guías éticas que se han estudiado anteriormente. Pero no sólo eso, en el artículo 2 se estipula el derecho a la vida y el en artículo 3 la integridad de las personas. Se debe evitar el uso dañino de la inteligencia artificial para evitar la violación de estos derechos. Pero no sólo 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 sujetos a la utilización de inteligencia artificial sin su conocimiento y sin su consentimiento expreso.[49] Aunque, la dignidad humana es complicada definirla como bien jurídico,[50] se puede proteger su lesión mediante el sistema penal en los artículos 173.3 y 174, el 177 bis, el 232 cuando se trate de menores y el 311 y 312. Sin olvidar los preceptos 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 en el centro de la discusión cuando hablamos de sistemas de inteligencia artificial regida por datos. Esta utilización de información personal puede afectar a otros derechos de forma indirecta como pueden ser la libertad de expresión y de información (artículo 11) o el derecho de asociación (artículo 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 (artículo 10).[51] El bien jurídico protegido en los artículos 197 y ss. nos ofrece una herramienta para trasladar el principio de privacidad de la guía de altos expertos en inteligencia artificial. Es probablemente uno de los principios que se protege de forma más sencilla en la normativa penal.

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 punto interesante es el trato de los grupos vulnerables como pueda ser los grupos raciales, étnicos, sexo, religión, etc., los menores (artículo 24), mayores (artículo 25) y las personas con discapacidad (artículo 26). Estos últimos grupos tienen un especial interés en los casos de utilización de inteligencia artificial en cuestiones de cuidados y en medicina. Pero no únicamente, a la hora de asignar un seguro se tiene muy en cuenta la edad del sujeto que lo solicita.[52]

El principio de tutela judicial efectiva (artículo 47) 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

[49] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 60.
[50] ALONSO ÁLAMO, 2011, p. 15.
[51] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 61.
[52] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 68.

recele de las decisiones tomadas al no poder comprender la justificación de la resolución judicial.[53] Aquí, los bienes jurídicos relativos a los funcionarios públicos cobran una nueva dimensión para proteger a los ciudadanos de la toma de decisiones mediante sistemas inteligentes, entre ellas, la resolución de juicios.

A diferencia del estudio anterior, realizado en un contexto más americano de defensa de los derechos ciudadanos, en el análisis europeo destacan los derechos colectivos o sociales. El derecho a la seguridad y asistencia social es un derecho social recogido en el artículo 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 nacionalidad.[54]

Siguiendo con este análisis a los derechos colectivos, la Agencia tiene en cuenta la protección de los consumidores, que como tal no es un derecho recogido en la Carta pero el informe lo tiene en cuenta ya que deriva del artículo 169 del TFUE, con una especial importancia en las técnicas de márquetin dirigido mediante inteligencia artificial 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 electrónico.[55] Desde el punto de vista penal, los bienes jurídicos colectivos nos van a permitir tener una base para protegerlos. 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 artículo 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 respeto al ciudadano. En ambos casos la incidencia de la utilización de sistemas inteligentes por parte de la administración lo puede poner en serio peligro.[56] 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 informe del Consejo de Europa, en el que se analiza cómo regular la inteligencia artificial desde una perspectiva de los Derechos Humanos, la democracia y el

[53] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 75.
[54] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 79.
[55] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 80.
[56] Inf. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2020, p. 81.

Estado de Derechos el tercero de los informes en los que nos vamos a centrar.[57] 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).[58] Un segundo punto de estudio por el Consejo de Europea son las libertades afectadas. Aquí destaca la libertad de expresión (art. 10 CEDH) y las libertades de reunión y de asociación (art. 11 CEDH).[59] 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 Carta Social Europea, resaltando que el uso de la inteligencia artificial puede ser muy beneficiosa para las condiciones laborales de los trabajadores, no deja de ser menos cierto, que mal utilizada, puede afectar a los derechos de los trabajadores, las condiciones de seguridad y salud en el trabajo, la dignidad del trabajo (art 2 y 3 CSE) y 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).[60]

La inteligencia artificial 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 democracia.[61]

El informe del Consejo de Europea finaliza su trabajo planteando nuevos derechos y adaptaciones de los derechos existentes que deben de 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 inteligencia artificial, que podría incluirse en el bien jurídico libertad o dignidad.[62] Además, se le añade un derecho a la trasparencia 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

[57] BEN-ISRAEL, 2020, p. 1.
[58] COUNCIL OF EUROPE, 2019, p. 29 y ss.
[59] COUNCIL OF EUROPE, 2019, p. 30 y ss.
[60] COUNCIL OF EUROPE, 2019, p. 33 y ss.
[61] COUNCIL OF EUROPE, 2019, p. 34.
[62] FUENTES OSORIO, 2017, p. 19.

protección de la trasparencia 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 inteligencia artificial, en concreto, aquello que afecten al reconocimiento. Aumentar el alcance de la privacidad cuando se trate de vigilancia masiva mediante inteligencia artificial, así como, dar protección contra el rastreo indiscriminado y de alcance social de individuos.[63] Estos derechos dirigidos a evitar un control exhaustivo de los sujetos por parte de los sistemas inteligentes no están prohibidos por nuestro código penal.

V. **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 inteligencia artificial. Por lado, la existencia de un uso apropiado de la inteligencia artificial que debe de cumplir con los siete principios éticos a nivel europeo, lo que nos determina las medidas que se van a considerar como diligencia debida. Por otro, 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. En inteligencia artificial nos encontramos con un principio, similar al principio de incertidumbre de la mecánica cuántica, por el que cuanto más preciso es un análisis realizado por esta tecnología, más difícil es conseguir una explicación de por qué la máquina ha tomado esa decisión. Ni siquiera los creadores del sistema inteligente son capaces de encontrar una explicación a la decisión tomada. Este oscurantismo, que se describe con el concepto de *black box*, 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 han 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

[63] COUNCIL OF EUROPE, 2019, p. 41.

análisis con el objetivo de comprender dónde se produce 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 este. 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 inteligencia artificial es una labor complicada. Afortunadamente, no partimos de cero ya que desde la filosofía de la tecnología se han hecho avances en esta materia, anteriormente, con el fin de realizar la evaluación ética del uso de la tecnología mediante un análisis por fases.[64] La metodología propuesta por Brey sobre la ética anticipatoria en tecnología propone tres fases de estudio para resolver de forma adecuada este estudio: el nivel 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 uso.[65] Así, el reconocimiento de imágenes es una tecnología de inteligencia artificial desarrollada en nuestro mundo que, puede ser utilizada para analizar imágenes de rayos X para diagnóstico de enfermedades de tórax o puede ser utilizada en un aeropuerto para detectar sospechosos de terrorismo. En estos dos ejemplos podemos observar cómo el desarrollo en el primer caso es mucho más elevado y das más acierto que en el segundo, donde el número de errores es inferior. En la última fase podemos detectar que los problemas planteados en el primer caso son diferentes al del segundo. En el primer caso los falsos negativos pasan al mismo sistema que antes de la aplicación de la inteligencia artificial, siendo los positivos los que van por un sistema de análisis más rápido. En el segundo caso el nivel de incidencia en ciudadanos no terroristas es muy alto siendo un sistema muy invasivo en la privacidad. Además, los falsos positivos implican una restricción muy alta de los derechos fundamentales de los ciudadanos erróneamente catalogados como terroristas.

Determinado estos puntos, que nos van a permitir tener un primer acercamiento de estudio a la tecnología, nos tenemos que centrar en otros elementos para poder concretar dónde se producen los errores del sistema y quién es el responsable de los

[64] BREY, 2012, p. 3 y ss.
[65] BREY, 2012, p. 7 y 8.

mismos. Para este fin, 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 del uso concreto que se le vaya a dar a la inteligencia artificial: 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 inteligencia artificial 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 puede explicarse, se debe de entender cuáles han sido los procesos y procedimientos en la creación y la instauración del uso de esos sistemas. Finalmente, 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 inteligencia artificial no es una bola de cristal.[66] 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 sujeto y el sistema inteligente. Los errores de los sujetos implicados en todo el ciclo de vida de la inteligencia artificial se pueden producir por los creadores y desarrolladores, 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 posibilidades.[67] Está comprobado que el mundo tecnológico está dominado por hombres, blancos, con ingresos económicos bastante más altos de la media que producen, consciente o inconscientemente, que los sistemas inteligentes produzcan sesgos en la toma de decisiones que pueden afectar a los tres niveles, selección de datos, sistema de procesamiento y en cómo se interpretan los resultados. Al mismo tiempo, suponiendo que la máquina demuestre un funcionamiento correcto en su diseño, lo que podría eliminar

[66] DIGNUM, 2019, p. 28.
[67] SALVADORI, 2021, p. 157

la responsabilidad de los fabricantes y desarrolladores, en un segundo paso, será necesario analizar la utilidad práctica que le asigne el usurario 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%, y el usuario profesional decide sólo utilizarlo para determinados barrios de una ciudad o sólo 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 al fin de eliminar a la competencia. Es el usuario profesional quien utiliza la inteligencia artificial con un determinado fin que puede derivar en la comisión de un delito, pero el sistema en sí, no tiene ningún fallo que derive 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 sólo con determinados sectores de población, por ejemplo, los sistemas de conducción automática se testean dentro de sector de 65 a 80 años sólo 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 en con los dispositivos inteligentes para la medición de determinadas constantes mediante sistemas inteligentes si son desarrollados sólo para determinados sectores pueden que nos sirvan adecuadamente para otros sectores de población, dando resultados erróneos sólo a este grupo poblacional. Al mismo tiempo, el diseño ha de permitir que la generalidad de los usuarios finales tienen 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.

El análisis de todos estos componentes nos daría una matriz de análisis similar a esta:

	Obtención de datos	Procesamiento de datos	Resultados del sistema
Desarrollador			
Profesionales			
Usuarios finales			

Desde esta metodología podremos realizar un sistema de gestión de riesgos adecuado para determinar si se han tomados las precauciones 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 poco, medio o alto.

Teniendo claro 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 alienada con el sistema de gestión de riesgos que propone la propuesta de Reglamento sobre inteligencia artificial que proponen las instituciones europeas.

VI. **El sistema gestión de riesgos de la Propuesta de Reglamento de la inteligencia artificial como determinante de la responsabilidad**

Para Quintero Olivares existen cuatro posibilidades de responsabilidad penal en el uso de tecnologías inteligentes:[68] 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 es que se desvía sin tener conocimiento del porqué de la reacción fuera del comportamiento esperado, y el uso de una tecnología que ha sido prohibida. El centro del debate[69] 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 para solventar este problema es determinar qué se considera por un uso no adecuado. Es por lo que la Comisión Europea ha establecido las guías éticas con una serie de principios que se han de tener en cuenta para crear y utilizar sistemas inteligentes y la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadoras sobre inteligencia artificial y se modifica ciertos actos legislativos de la Unión.[70] En este último documento se 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[71] establecía las opciones políticas que se pueden derivar del uso de los sistemas inteligentes y se esbozaban los riesgos que se podían producir, la propuesta de Reglamento intenta

[68] Aunque centrado en robots, es cierto que todos los ejemplos que utiliza, controlados por humanos en mayor o menor medida, tienen integrados sistemas inteligentes en su funcionamiento, y su análisis nos vale para nuestro estudio. Cfr. QUINTERO OLIVARES, 2018, p 22.

[69] PALMA HERRERA, 2020, p. 56 y ss.

[70] Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadoras sobre inteligencia artificial y se modifica ciertos actos legislativos de la Unión https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52021PC0206 [Consultado 18 de abril de2022]

[71] https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

cubrir este segundo objetivo con el fin de proporcionar confianza a los usuarios de estas tecnologías basadas en inteligencia artificial, tomando como base la protección de los principios éticos. Se puede debatir sobre si los principios éticos tienen que estar protegidos por el Derecho Penal. En principio, podemos entender que algunos sí y otros no. Unos ya tienen un bien jurídico en nuestros sistemas penales como la privacidad, libertad o seguridad de los productos, mientras que otros no. Por la regulación que se plantea, lo que sí está claro es que la ausencia de un plan ético de prevención de riesgos pueda derivar en algún tipo de responsabilidad, al establecerlo como obligatorio cuando se trate de inteligencia artificial de alto riesgo y deseable en el resto de los sistemas inteligentes. Este modelo de gobernanza se encuentra limitado a los requerimientos mínimos para detectar los riesgos y problemas derivados del uso de la inteligencia artificial.[72] En el artículo 9 de la propuesta se señala que sistemas de gestión de riesgos deben establecerse, implementarse, documentarse y mantenerse. Estos sistemas de gestión de riesgo consisten en 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 asociado con cada inteligencia artificial, 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 artículo 61,[73] 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 riesgo, 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 la propuesta de reglamento estipula tres puntos que deben de 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 artículo 13, en particular por lo que respecta a los riesgos mencionados en la letra b) del apartado 2 de este artículo,[74] y, en su caso, la formación de los usuarios profesionales y consumidores.

[72] PROPUESTA DE REGLAMENTO, 2020, p. 4.

[73] El artículo 61 hace referencia al control de la inteligencia artificial 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 inteligencia artificial 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.

[74] 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

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 adecuados para conseguir el fin por el que se utiliza el sistema inteligente, dentro del uso que se persigue con su uso y deben de 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. Desde este punto de vista, Romeo Casabona, presenta diferentes sistemas de determinación de la responsabilidad criminal. La primera de ellas es considerar que el sistema inteligente autónomo es responsable por los crímenes cometidos por sus actos.[75] Entrando en una disquisición sobre si la máquina puede tener intención o no,[76] deja la solución en el aíre sobre si en un futuro, bastante lejano, la máquina podría tener esta intención de la misma forma que un humano.[77] Esto implicaría, como bien señala el referido autor, la necesidad de modificar el actual sistema de responsabilidad penal para adaptarlo a los entes autónomos, con un alto precio para el sistema penal que tenemos en la actualidad.[78]

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.[79] Esto, como el autor señala, trae otro problema añadido por la falta de conocimiento, por parte del sistema, de 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 jurídica.[80]

El autor nos conduce hacía el análisis de los riesgos permitidos, que abre el debate sobre la labor preventiva del Derecho Penal y de la regulación administrativa de los sistemas de inteligencia artificial. Puesto este marco jurídico, como se pretende hacer con la propuesta de Reglamento de la inteligencia artificial, nos llevaría a un sistema de *compliance* similar al de las personas jurídicas en nuestro actual sistema penal,

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.

[75] ROMEO CASABONA, 2020, p. 170.

[76] SALVADORI, 2021, p. 154.

[77] Cuestionable desde nuestro punto de vista. Cfr. BRYSON, 2018, p. 2.

[78] ROMEO CASABONA, 2020, p. 171.

[79] ROMEO CASABONA, 2020, p. 172.

[80] Como señala el autor en la ROMEO CASABONA, 2020, p. 174.

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 máquina.[81] A esto habría que añadir la necesidad de un examen general y detallado externo por expertos. Y es que hay que dejar claro que el Derecho penal se encuentra en una posición inmejorable para construir un sistema de control interno con estas características.[82]

La figura del cumplimiento normativo de las personas jurídicas, traída al mundo de los sistemas inteligentes, presenta algunos problemas. Los programas de *compliance* son realizados por personas jurídicas de forma obligatoria y, en caso de cumplir, les exime de responsabilidad jurídico-penal. En nuestro caso, el sistema inteligente no puede realizar por sí mismo este cumplimiento normativo, será un externo el que tomará la decesión de realizarlo no. Mientras que dentro de la empresa se pueden tomar este tipo de decisiones, en el sistema inteligente no es posible. Siempre será una persona externa al sistema el que tomará las decisiones. Luego el paralelismo, en este punto falla. Hay que añadir que el cumplimiento normativo implica la necesidad de un marco jurídico que regule el incumplimiento. Sin embargo, al partir de los principios éticos (no normativizados) los sistemas de prevención de los riesgos son más amplios que los sistemas de *compliance*. Desde este punto, las herramientas de gobernanza de la inteligencia artificial implicarían un análisis: la parte normativa, principalmente los Derechos Fundamentales y las normas regulatorias, y la parte ética, que funciona a modo de *Soft Law*, que nos permiten definir otras figuras jurídicas, como la determinación de la responsabilidad.

Lo primero es determinar qué tipo de tecnología se va a utilizar. Hemos analizado diferentes tipos de inteligencia artificial en el capítulo 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 inteligencia artificial basada en razonamiento lógico. 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 inteligencia artificial; en la creación, el uso profesional y uso de un sistema inteligente. Vamos a encontrarnos con uno nivel en el desarrollo y puesta en el mercado de la máquina inteligente, que requerirá de un control de riesgos en su creación. Se podría diferenciar dos fases, una en la creación y otra en la producción de este, pero para nuestro análisis ahora, no es relevante. Y, 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

[81] ROMEO CASABONA, 2020, p. 173.
[82] DE LA CUESTA AGUADO, 2016, p. 183.

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 probada con éxito en su entorno controlado de uso, alimentado con determinados datos. Cuando pase al uso profesional se puede alimentar con datos diferentes lo que implique que genere riesgos hacia los derechos fundamentales y, por tanto, a los bienes jurídicos de las personas. Voy a exponer dos ejemplos que pueden ser significativos. Los sistemas inteligentes en prevención del crimen pueden funcionar muy bien en el desarrollo, pero cuando se alimenten con los datos de diferentes operaciones reales pueden afectar a derechos diferentes y de forma diferente. Otro ejemplo puede ser el 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.

Diferenciando 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, como mantenemos en esta obra, el responsable siempre será un humano. La máquina sólo 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.

Este modelo de análisis de riesgos va a constituir un modelo estándar para analizar la tecnología dentro de la Unión Europea. Ya en el Reglamento General de Protección de Datos[83] y la Directiva relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos establecen sistemas de análisis de riesgos en la privacidad de las personas[84] que han servido de modelos. Además, el Reglamento de Servicios Digitales de 2022 sigue el mismo camino de evaluación de riesgos.[85]

No quisiera finalizar sin comentar que estos cumplimientos normativos no pueden limitarse exclusivamente a la aplicación de las normas reguladoras. Para determinar la intención también hay que evaluar el sistema de gobernanza de los sistemas inteligentes, lo que conlleva un análisis normativo pero también el respeto de los principios éticos.[86] La ausencia del principio de transparencia o de rendición de cuentas,

[83] El artículo 35 obliga a la realización de una evaluación de impacto relativa a la protección de datos.

[84] El artículo 27 tiene la misma finalidad, pero en el sistema judicial penal. Cfr. VALLS PRIETO, 2016, p. 1.

[85] Digital Services Act. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en [Consultado 03/05/2022]

[86] FLORIDI, 2018, p. 2.

por ejemplo, no es ni más ni menos que un intento de diluir la posibilidad de determinar quién es el responsable por los actos derivados del sistema inteligente[87] y, por tanto, su incumplimiento impediría la posibilidad de determinar la responsabilidad penal.

VI. **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 inteligencia artificial 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 inteligencia artificial 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 interesantes. 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 inteligencia artificial 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 que bienes jurídicos se van a ver afectados con el impacto negativo de los sistemas inteligentes. Cómo 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

[87] PASQUALE, 2020, p. 229.

vida de la inteligencia artificial se encuentran los elementos necesarios de la responsabilidad.

Con el fin de resolver este problema tenemos que llegar a la conclusión de que sólo es posible analizar la responsabilidad penal centrándonos en diferentes fases. El primero es el 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 tienen 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 quién ha cometido el delito. Dentro de los diferentes códigos éticos que existen, los principios expuestos por las guías éticas 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.

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, 15, pp. 1-33

Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe, (2017).

BEN-ISRAEL, I., CERDIO, J., EMA, A., FRIEDMAN, L., IENCA, M., MANTELERO, A., MATANIA, E., MULLER, C., SHRIOYAMA, H., y VAYENA, E. (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, Council of Europe Study.
BREY, P. A. E. (2012), "Anticipatory Ethics for Emerging Technologies", NanoEthics, 6, 1, pp. 1-13.
BROUSSARD, M. (2018), Artificial unintelligence: how computers misunderstand the world, Cambridge, Massachusetts.
COMMISSIONER FOR HUMAN RIGHTS, C. fo E. (2019), Unboxing Artificial Intelligence: 10 steps to protect Human Rights.
DE LA CUESTA AGUADO, P. M. (2016), "La ambigüedad no es programable: racionalización normativa y control interno en inteligencia artificial", Revista de derecho y proceso penal, 44, pp. 165-194.
DIGNUM, V. (2019), Responsible artificial intelligence: how to develop and use AI in a responsible way, Artificial Intelligence foundations, theory, and algorithms, Cham.
FLORIDI, L. (2018), "Soft Ethics and the Governance of the Digital", Philosophy & Technology, 31, 1, pp. 1-8.
FUENTES OSORIO, J. L. (2017), "El odio como delito", Revista Electrónica de Ciencias Penales y Criminología, 19, pp. 1-52.
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. 1-25.
HALLEVY, G. (2010), "The criminal liability of artificial intelligence entities from science fiction to legal social control", Akron Intellectual Property Journal, 4, 29, pp. 171-201.
HAO, K. (2018), "What is AI? We drew you a flowchart to work it out", MIT Technology Review, 03/05/2021, < https://www.technologyreview.com/2018/11/10/139137/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/>.
HENLEY, J. y BOOTH, R. (2020), "Welfare surveillance system violates human rights, Dutch court rules", The Guardian, 05/02/2020, <https://www.theguardian.com/technology/2020/feb/05/welfare-surveillance-system-violates-human-rights-dutch-court-rules>.
COMISIÓN EUROOPEA (2019), HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, Ethics Guidelines for Trustworthy AI.
JANSEN, P., BROADHEAD, S., RODRIGUES, R., WRIGHT, D., BREY, P., FOX, A., y WANG, N. (2019), SIENNA D4.1: State-of-the-art Review: Artificial Intelligence and robotics, <https://doi.org/10.5281/zenodo.4066571>.
LEE, K.-F. (2018), AI superpowers: China, Silicon Valley, and the new world order, Bostón.
MIRO LLINARES, F. (2018), "Inteligencia artificial y Justicia Penal: Más allá de los resultados lesivos causados por robots", Revista de Derecho Penal y Criminología, 20, pp. 87-130.
MAYER-SCHÖNBERGER, V. y CUKIER, K. (2014), Big data: a revolution that will transform how we live, work, and think, First Mariner Books edition. Bostón.
OECD (2021), Recommendation of the Council on Artificial Intelligence, OECD Legal Instrument, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
PALMA HERRERA, J. M. (2020), "Inteligencia artificial y ciencias penales. Aproximación a las bases de una compleja relación", en Galán Muñoz, Mendoza Calderón, Martínez González (coords.): Derecho penal y política criminal en tiempos convulsos: libro homenaje a la Profa. Dra. María Isabel Martínez González, Valencia, pp. 37-60.
PASQUALE, F. (2020), New Laws of Robotics: Defending Human Expertise in the Age of AI, Cambridge, Massachusetts.

Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión, (2021), <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52021PC0206>.

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, 1, pp. 1-23.

RASO, F., HILLIGOSS, H., KRISHNAMURTHY, V., BAVITZ, C., y LEVIN, K. (2018), "Artificial Intelligence & Human Rights: Opportunities & Risks", pp. 1-62,<https://dash.harvard.edu/handle/1/38021439>.

RODRÍGUEZ YAGÜE, C. (2007), "Una propuesta de clasificación de los delitos de discriminación en el Código Penal español", Dos mil-tres mil, 11, pp. 1-25.

ROMEO CASABONA, C. M. (2020), "Criminal responsability of robots and autonomous artificial intelligent systems?", Comunicaciones en propiedad industrial y derecho de la competencia, 91, pp. 167-187.

RUSSELL, S. y NORVIG, P. (2019), Artificial Intelligence: A Modern Approach, 4.a ed., Harlow, <https://www.prioritytextbook.com/artificial-intelligence-a-modern-approach-4th-edition-stuart-russell-and-peter-norvig-global-edition/>.

SALVADORI, I. (2021), "Agentes artificiales, opacidad tecnológica y distribución de la responsabilidad penal", Cuadernos de política criminal, 133, pp. 137-164.

SAMOILI, S., LOPEZ, C. M., GOMEZ, G. E., DE, P. G., MARTINEZ-PLUMED, F., y DELIPETREV, B. (2020), "AI WATCH. Defining Artificial Intelligence", JRC Publications Repository, <https://doi.org/10.2760/382730>.

VALLS PRIETO, J. (2016), "Nuevas formas de combatir el crimen en Internet y sus riesgos", Revista Electrónica de Ciencias Penales y Criminología, 18, pp. 1-36.

VIDA FERNÁNDEZ, J. (2018), "Los retos de la regulación de la inteligencia artificial: algunas aportaciones desde la perspectiva europea", 1. ed. en Quadra-Salcedo y Fernández del Castillo y Piñar Mañas (eds.): Sociedad digital y Derecho, Madrid, pp. 203-224.