Cuadernos de RES PUBLICA en derecho y criminología
Nº en prensa ISSN: 2990-0697 DOI: 10.46661/respublica.11635

Derechos humanos y derecho penal en la era de la inteligencia artificial. Retos y propuestas

Human rights and criminal law in the age of artificial intelligence. Challenges and proposals

Giuseppe Kodjack Gangi-Guillen
Universidad Nebrija, Madrid, España
ggangi@nebrija.es
ORCID: 0000-0001-5627-3874

Resumen

El desarrollo de la inteligencia artificial (IA) ha transformado el derecho penal y la administración de justicia. Desde la predicción del crimen hasta la toma de decisiones judiciales automatizadas, la IA ofrece oportunidades para mejorar la eficiencia del sistema judicial. Sin embargo, su implementación plantea serios desafíos en términos de transparencia, sesgos algorítmicos y vulneración de derechos fundamentales. Este artículo analiza los impactos de la IA en el derecho penal y los derechos humanos, destacando la necesidad de un marco regulador ético y garantista que equilibre innovación y justicia. Se propone un modelo de supervisión y auditoría de algoritmos, así como el uso de IA como herramienta de apoyo y no como único criterio en la toma de decisiones judiciales.

Palabras clave: *Inteligencia artificial, derecho penal, derechos humanos, regulación ética, justicia automatizada*.

Abstract

The development of artificial intelligence (AI) has transformed criminal law and the administration of justice. From crime prediction to automated judicial decision-making, AI offers opportunities to improve the efficiency of the judicial system. However, its implementation poses serious challenges in terms of transparency, algorithmic biases and infringement of fundamental rights. This article analyses the impacts of AI on criminal law and human rights, highlighting the need for an ethical regulatory framework that balances innovation and justice. It proposes a model for monitoring and auditing algorithms, as well as the use of AI as a support tool and not as the sole criterion in judicial decision-making.

Keywords: Artificial intelligence, criminal law, human rights, ethical regulation, automated justice.

Cómo citar este trabajo: *Gangi-Guillen, Giuseppe Kodjack.* (2025). Derechos humanos y derecho penal en la era de la inteligencia artificial. Retos y propuestas. *Cuadernos de RES PUBLICA en derecho y criminología, (en prensa),* 01–14. https://doi.org/10.46661/respublica.11635

Recepción: 14.02.2025 **Aceptación:** 02.06.2025 **Publicación:** en prensa

1. Introducción

La IA ha permitido avances significativos en el análisis predictivo del crimen y en herramientas de identificación biométrica. Por ejemplo, puede procesar grandes volúmenes de datos para detectar patrones delictivos y predecir posibles crímenes, lo que permite una mejor asignación de recursos policiales y la optimización de estrategias de prevención (Ferguson, 2017). En el caso del reconocimiento facial, ha sido implementado en varios países para identificar sospechosos y resolver casos complejos, aunque su uso ha suscitado preocupaciones sobre la privacidad y los derechos fundamentales (Garvie, Bedoya, & Frankle, 2016).

Sin embargo, el uso de la IA en el derecho penal también conlleva riesgos significativos como la falta de transparencia en la toma de decisiones de los algoritmos, ya que muchos de estos sistemas funcionan como "cajas negras", dificultando la comprensión de cómo llegan a determinadas conclusiones (Pasquale, 2015). En el ámbito judicial, esto es un desafío para la garantía del debido proceso, ya que las decisiones en los procedimientos penales deben ser justificadas y comprensibles para los ciudadanos. De igual manera, la automatización de decisiones judiciales sin supervisión humana adecuada puede llevar a errores que vulneren derechos fundamentales como la presunción de inocencia y el derecho a un juicio justo.

Otro riesgo emergente es su uso indebido para la comisión de delitos, tales como la generación de *deepfakes* y la suplantación de identidad digital; herramientas que pueden ser utilizadas para manipular pruebas, difundir desinformación o cometer fraudes informáticos de difícil detección (Chesney & Citron, 2019). La sofisticación de estos ataques hace que la persecución penal de estos delitos sea un desafío, ya que se requiere la adaptación de normativas para abordar estas nuevas formas de criminalidad. En este sentido, resulta crucial establecer mecanismos regulatorios que equilibren el desarrollo tecnológico con la protección de los derechos humanos, evitando que la IA se convierta en un instrumento de vulneración de garantías fundamentales.

La implementación de la IA en diversos ámbitos plantea desafíos significativos para los derechos humanos, especialmente en términos de privacidad, debido proceso y discriminación. Uno de los principales riesgos es la vulneración del derecho a la privacidad, donde las tecnologías de IA para el reconocimiento facial y la vigilancia masiva permiten la recopilación y análisis de grandes volúmenes de datos personales sin el consentimiento adecuado, lo que puede conducir a una vigilancia intrusiva y no autorizada de individuos y comunidades. La Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH) ha señalado que las deducciones y actividades de seguimiento realizadas por herramientas de IA plantean graves interrogantes sobre la privacidad y la protección de datos. (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2021).

En cuanto al debido proceso, la opacidad de muchos sistemas de IA dificulta la comprensión de cómo se toman las decisiones algorítmicas, lo que puede llevar a comprometer la capacidad de los individuos para impugnar decisiones que afectan sus derechos, ya que no pueden entender ni cuestionar adecuadamente el proceso.

La discriminación es otro desafío crítico asociado con la IA; los algoritmos pueden perpetuar o incluso amplificar sesgos existentes si se entrenan con datos que reflejan prejuicios sociales. Esto puede resultar en decisiones discriminatorias en áreas como la contratación laboral, la concesión de créditos o la aplicación de la ley. Un artículo en *Ius et Scientia* señala que uno de los riesgos más importantes del aprendizaje automático es la posibilidad de amplificar la discriminación y los sesgos existentes contra ciertos grupos (González Fuster, 2021).

Para mitigar estos riesgos, es fundamental establecer marcos regulatorios que promuevan la transparencia, la rendición de cuentas y la equidad en el desarrollo y la implementación de sistemas de IA. Esto incluye la realización de evaluaciones de impacto en derechos humanos antes de la implementación de tecnologías de IA, la supervisión continua de su funcionamiento y la garantía de que existan mecanismos efectivos para que los individuos puedan impugnar decisiones automatizadas que afecten sus derechos.

El objetivo del artículo es ofrecer una reflexión crítica sobre la incorporación de la IA en el ámbito penal, identificando sus beneficios y riesgos desde una perspectiva de derechos humanos. Se busca explorar cómo la IA puede contribuir a la mejora del sistema de justicia penal sin comprometer garantías fundamentales, donde se propongan soluciones que permitan equilibrar innovación y protección de derechos. Para ello, se analizarán los principales desafíos asociados a su uso, incluyendo la falta de transparencia en los algoritmos, el riesgo de discriminación algorítmica y la necesidad de establecer marcos normativos adecuados.

Asimismo, se plantearán estrategias para mitigar estos problemas, tales como la creación de auditorías independientes de sistemas de IA, la implementación de principios de transparencia, la explicación de su diseño, y el fortalecimiento de la supervisión judicial en los procesos automatizados.

2. El papel de la IA en la justicia penal

La integración de la IA en la justicia penal ha generado un debate académico significativo centrado en sus aplicaciones y las implicaciones éticas y legales que conlleva. Según Miró (2018), la IA se emplea en labores policiales para la prevención del crimen mediante el análisis de grandes volúmenes de datos que permiten identificar patrones delictivos y predecir posibles conductas criminales. Esta capacidad predictiva facilita una asignación más eficiente de los recursos policiales y una respuesta proactiva ante el delito.

Sin embargo, la aplicación de la IA en el ámbito penal plantea desafíos significativos en cuanto a la protección de los derechos fundamentales. Hernández (2019) advierte sobre los riesgos de discriminación algorítmica, donde los sesgos presentes en los datos utilizados para entrenar los sistemas de IA pueden perpetuar o incluso amplificar desigualdades existentes, lo que afecta desproporcionadamente a ciertos grupos sociales, dificulta la transparencia y la rendición de cuentas en la toma de decisiones judiciales (Liz, 2024; 2018).

Ante estos desafíos, es imperativo establecer marcos regulatorios que guíen el uso ético y responsable de la IA en la justicia penal. Balcells (2020) propone la implementación de auditorías independientes de los algoritmos y la creación de organismos de supervisión que velen por la equidad y la transparencia en la aplicación de estas tecnologías; y enfatiza también la necesidad que la IA actúe como una herramienta de apoyo para los operadores jurídicos, sin sustituir el juicio humano.

2.1. Vigilancia y reconocimiento facial

El uso de la IA en la vigilancia y el reconocimiento facial ha experimentado un crecimiento notable en los últimos años, especialmente en el ámbito de la justicia penal. Estas tecnologías permiten la identificación y seguimiento de individuos en tiempo real, facilitando la labor de las fuerzas de seguridad en la prevención y resolución de delitos. Sin embargo, su implementación plantea desafíos en términos de privacidad, precisión y equidad; lo que evidencia la existencia de beneficios, pero también grandes limitaciones.

Un aspecto crítico es la precisión de los sistemas de reconocimiento facial, investigaciones previas han demostrado que

estos sistemas pueden presentar tasas de error más altas en la identificación de personas pertenecientes a minorías étnicas y mujeres, lo que puede conducir a detenciones injustas y perpetuar sesgos existentes en la sociedad.

Por ejemplo, el estudio publicado en *Science Advances* encontró que los algoritmos de reconocimiento facial tenían tasas de error significativamente más altas al identificar a personas de piel más oscura en comparación con individuos de piel más clara (Buolamwini & Gebru, 2018).

Además de las preocupaciones sobre precisión y sesgo, el uso de tecnologías de vigilancia basadas en IA plantea interrogantes sobre la erosión de la privacidad y las libertades civiles. La capacidad de monitorear y registrar movimientos de individuos en espacios públicos y privados sin su consentimiento explícito puede generar la sensación de vigilancia constante, lo que incide en el comportamiento y autonomía de las personas.

Un artículo publicado por *Big Data & Society* argumenta que la implementación de sistemas de vigilancia masiva puede llevar a una normalización de la vigilancia estatal, reduciendo la resistencia pública y la conciencia sobre las implicaciones para los derechos civiles (Brayne, 2017).

Para mitigar estos desafíos, es esencial establecer marcos regulatorios y éticos que guíen el desarrollo y uso de tecnologías de vigilancia y reconocimiento facial, la implementación de evaluaciones de impacto en la privacidad, la transparencia en los algoritmos utilizados y la garantía de mecanismos de rendición de cuentas.

La Unión Europea ha propuesto regulaciones que restringen el uso de sistemas de reconocimiento facial en espacios públicos, enfatizando la necesidad de salvaguardar los derechos fundamentales de los ciudadanos (European Commission, 2021), iniciativa que busca equilibrar los beneficios de la tecnología con la protección de los derechos humanos.

2.2. Predicción del crimen y toma de decisiones judiciales

En la predicción del crimen, la IA se utiliza para analizar grandes volúmenes de datos históricos y actuales con el fin de identificar patrones y tendencias delictivas. Mediante el empleo de algoritmos de aprendizaje automático, es posible prever áreas y momentos con mayor probabilidad de actividad criminal, lo que permite a las fuerzas del orden optimizar la asignación de recursos y planificar estrategias preventivas más efectivas.

Un estudio reciente de Mandalapu et al (2023) destaca que la aplicación de técnicas de aprendizaje profundo ha mejorado la precisión en la predicción de delitos específicos, como robos y asaltos, al considerar variables como ubicación geográfica, hora del día y datos socioeconómicos.

Sin embargo, la implementación de estos sistemas no está exenta de críticas; uno de los principales desafíos es el riesgo de perpetuar sesgos existentes en los datos utilizados para entrenar los algoritmos.

Si los datos históricos reflejan prácticas policiales discriminatorias o prejuicios sociales, los modelos predictivos pueden reforzar estas disparidades, generando una vigilancia desproporcionada hacia comunidades marginadas; en palabras de Scantamburlo et al (2018) la dependencia excesiva en la predicción algorítmica puede llevar a la estigmatización de ciertas áreas o grupos, sin abordar las causas subyacentes del crimen.

2.3. Toma de decisiones judiciales

La IA también ha sido empleada para asistir en la evaluación de riesgos y la determinación de sentencias. Herramientas como *COMPAS* en Estados Unidos analizan factores relacionados con el historial criminal y características personales para estimar la probabilidad de reincidencia de un acusado. Estas evaluaciones pueden influir en decisiones

sobre libertad condicional, fianzas y sentencias, esto facilita una administración de justicia más objetiva y consistente. La investigación de Bertalan & Ruiz (2022) ha explorado el uso de modelos de atención en el procesamiento del lenguaje natural para predecir resultados judiciales, analizando textos legales y precedentes para identificar patrones decisionales.

No obstante, la delegación de decisiones judiciales a sistemas de IA plantea preocupaciones sobre la transparencia y la rendición de cuentas; ya que los algoritmos utilizados a menudo son poco transparentes y dificulta comprender cómo se llegan a ciertas conclusiones. Esto puede socavar la confianza en el sistema judicial y plantea interrogantes sobre la equidad, especialmente si los modelos incorporan sesgos implícitos; por tanto, puede reducir la consideración de circunstancias atenuantes o contextuales que son fundamentales en la justicia penal (Chén, 2022).

2.4. Oportunidades y riesgos: eficiencia vs. vulneración de derechos

La integración de la IA en el ámbito del derecho penal también ha generado un debate significativo en torno a su eficiencia y las posibles vulneraciones de los derechos humanos. Por un lado, la IA promete optimizar procesos judiciales, mejorar la precisión en la toma de decisiones y reducir la carga de trabajo en los tribunales, y por otro, surgen preocupaciones sobre la equidad, la transparencia y el respeto a las garantías procesales fundamentales.

En el proceso penal plantea riesgos significativos para los derechos humanos por el desconocimiento en la generación del algoritmo, lo que dificulta la comprensión de los criterios utilizados en la toma de decisiones, lo que vulnera el derecho al debido proceso y a una defensa efectiva. La dependencia de datos históricos para entrenar estos sistemas también puede perpetuar o incluso amplificar sesgos y discriminaciones existentes, lo que puede incidir desproporcionadamente en los procesos aplicados a ciertos grupos sociales. Al respecto, el Parlamento Europeo ha destacado el potencial de sesgo y discriminación derivado del uso de aplicaciones de IA, especialmente en lo que respecta a los algoritmos en los que se basan dichas aplicaciones .

La tensión entre eficiencia y garantías procesales es evidente; mientras que la IA puede agilizar procedimientos, existe el riesgo que dicha celeridad comprometa derechos fundamentales como la presunción de inocencia y el derecho a un juicio justo.

Para mitigar estos riesgos, es esencial establecer marcos regulatorios que aseguren la transparencia, la rendición de cuentas y la supervisión humana en la aplicación de la IA en el ámbito penal. La Alta Comisionada de las Naciones Unidas para los Derechos Humanos ha enfatizado que las tecnologías de IA deben estar fundamentadas en los derechos humanos, y aquellas que no puedan operar en cumplimiento con la normativa internacional deben ser prohibidas o suspendidas hasta que se implementen protecciones adecuadas.

3. Desafíos para los Derechos Humanos

3.1. Privacidad: vigilancia masiva y protección de datos

Respecto a los sistemas de vigilancia masiva y la gestión de datos personales, la IA ha generado preocupaciones significativas en relación con la protección de los derechos humanos; aunque ofrecen beneficios en términos de seguridad y eficiencia, plantean críticos que requieren un análisis profundo y una regulación adecuada (Gómez,2013).

Uno de los principales desafíos radica en la capacidad de la IA para realizar una vigilancia masiva sin precedentes, ello a través de sistemas avanzados de reconocimiento facial y análisis de comportamiento que le permiten monitorear y analizar las actividades de individuos en tiempo real, lo que puede

conducir a una erosión de la privacidad y a la restricción de libertades fundamentales.

La recopilación y el procesamiento de grandes volúmenes de datos personales por parte de estos sistemas plantean riesgos significativos para la protección de datos, la falta de transparencia en los algoritmos y la opacidad en el manejo de la información posiblemente dificulten que los individuos comprendan cómo se utilizan sus datos y con qué fines.

Esto puede derivar en una vulneración del derecho a la privacidad y también posibles abusos por una discriminación algorítmica. Estudios de Brayne (2017; Delgado, Jímenez y Cremades, 2019; Gangi 2023) destacan la implementación de sistemas de vigilancia masiva como un motivo de normalización de la vigilancia estatal, por lo que es necesario abordar los desafíos que plantean estas tecnologías desde una perspectiva de derechos humanos, ello para robustecer la protección de los derechos humanos en la era digital.

3.2. Debido proceso: transparencia y explicaciones en decisiones automatizadas

La automatización de decisiones judiciales mediante algoritmos puede comprometer principios esenciales como la presunción de inocencia, el derecho a un juicio justo y la igualdad ante la ley. La opacidad de estos sistemas dificulta la comprensión de los criterios utilizados en la toma de decisiones, lo que también puede vulnerar el derecho a una defensa efectiva.

La falta de transparencia en los algoritmos de IA impide que los acusados y sus defensores comprendan cómo se han evaluado las pruebas o se ha determinado la culpabilidad; lo cual socava la posibilidad de impugnar decisiones basadas en errores o sesgos algorítmicos contrario al principio de contradicción y al derecho a un recurso efectivo.

La implementación de la IA en el sistema judicial plantea interrogantes adicionales sobre la responsabilidad y la rendición de cuentas. En el caso de errores judiciales derivados de decisiones algorítmicas, resultaría complejo determinar responsabilidades: el desarrollador del algoritmo, la entidad que lo implementa o el juez que confía en sus resultados. Esta ambigüedad puede generar una falta de confianza en el sistema judicial y vulnerar el derecho de las personas a ser juzgadas por tribunales competentes e imparciales.

La utilización de la IA en procedimientos judiciales puede afectar la igualdad de armas entre las partes; las instituciones con mayores recursos pueden acceder a tecnologías más avanzadas, lo que les permite obtener ventajas indebidas en litigios, lo cual contraviene el principio de igualdad ante la ley y puede resultar en decisiones injustas, lo que incrementa las brechas de poder y recursos en el sistema judicial.

Para mitigar estos desafíos es esencial desarrollar marcos legales y éticos que regulen el uso de la IA en el sistema judicial, la implementación de evaluaciones de impacto en los derechos humanos, la garantía de transparencia en los algoritmos y la supervisión humana en la toma de decisiones. La Unión Europea ha propuesto regulaciones que restringen el uso de sistemas de IA en procedimientos judiciales, enfatizando la necesidad de salvaguardar los derechos fundamentales de los ciudadanos.

3.3. Discriminación algorítmica: sesgos en la aplicación de la IA.

Como se ha señalado, uno de los principales problemas derivados de la aplicación de la IA es el riesgo de perpetuidad y amplificación de las desigualdades ya existentes, esto parte del hecho que, la IA, es entrenada con datos históricos que pueden internalizar y reproducir prejuicios presentes y/o preexistentes en dichos datos.

Un estudio realizado por la agencia de noticias de New York *Pro-Publica* reveló que el software COMPAS, utilizado en Estados Unidos para evaluar la probabilidad de

reincidencia, tendía a sobreestimar el riesgo en individuos afroamericanos y a subestimar el mismo en individuos caucásicos. Este tipo de sesgos no solo compromete la equidad del sistema judicial, sino que también vulnera derechos fundamentales como la presunción de inocencia y el derecho a un juicio justo.

La falta de transparencia en los procesos de toma de decisiones automatizadas impide que los afectados comprendan las razones detrás de ciertas determinaciones, lo que vulnera el derecho a la información y limita la posibilidad de impugnar decisiones injustas.

Es crucial reconocer que los sesgos en la IA no son meramente fallos técnicos, son reflejos de desigualdades estructurales ya existentes en la sociedad, por lo tanto, abordar estos desafíos requiere una aproximación multidisciplinaria que combine soluciones técnicas con reformas sociales y jurídicas para garantizar una gobernanza adecuada, cooperación para defender los derechos y reducir las desigualdades en el contexto de la IA (Ricaurte, 2024).

4. Regulación y propuestas de solución

4.1. Vacíos legales en la normativa Española y Europea

La rápida evolución de la IA ha superado la capacidad de las normativas españolas y europeas para regular adecuadamente su desarrollo y aplicación, generando vacíos significativos que afectan la seguridad jurídica y la protección de derechos fundamentales. Uno de los principales problemas radica en la falta de una definición clara y precisa de lo que constituye la IA.

Al respecto, la propuesta de Reglamento de la Unión Europea sobre IA ofrece una definición amplia que podría dar lugar a interpretaciones diversas, dificultando la identificación y regulación de sistemas específicos. Esta ambigüedad afecta la toma de decisiones en relación con la responsabilidad civil y penal, lo que es especialmente preocupante en sectores como el transporte autónomo y la automatización industrial (Xavier, 2023).

Si un vehículo autónomo provoca un accidente, por ejemplo, no existe una normativa clara que determine si la responsabilidad recae en el fabricante, el propietario del vehículo o el desarrollador del software, lo que puede dificultar la compensación de las víctimas y la correcta asignación de responsabilidades (El-Kadi, 2025). También la creciente capacidad de la IA para influir en el comportamiento humano mediante técnicas de persuasión subliminal o manipulación de la información plantea serias preocupaciones éticas y legales. Conceptos emergentes como los neuroderechos, que buscan proteger la integridad mental y la autonomía individual, aún no están recogidos en la legislación europea, dejando un vacío en la protección contra posibles abusos que podrían derivarse de tecnologías avanzadas como el reconocimiento facial y la vigilancia masiva (Alonso, 2021).

El aspecto de la protección de datos y la privacidad es un vacío, un ámbito donde la actual normativa, como el Reglamento General de Protección de Datos (RGPD), resulta insuficiente ante los desafíos que plantea la IA. Los sistemas de IA procesan grandes volúmenes de datos personales con el fin de optimizar su funcionamiento, muchas veces sin el consentimiento explícito de los individuos, lo que podría vulnerar gravemente su privacidad. En palabras de Mandalapu et al (2023), la capacidad de la IA para generar *deepfakes* complica aún más la protección de la imagen y la reputación de las personas, especialmente en contextos judiciales y de seguridad pública (Delgado, 2023).

En el ámbito de la vigilancia, el uso de sistemas de reconocimiento facial por parte de entidades públicas y privadas ha generado otra controversia, ya que estas tecnologías operan en un vacío normativo que no establece límites claros sobre el equilibrio entre la seguridad pública y la protección de derechos individuales (Delgado, Mazurier y Paya, 2019).

La falta de regulación específica en este campo podría llevar a prácticas invasivas que

vulneren libertades fundamentales, como la libertad de expresión y el derecho a la intimidad. La Unión Europea ha comenzado a debatir regulaciones que limiten el uso de estas tecnologías en espacios públicos, pero la falta de consenso en torno a su aplicación efectiva sigue representando un obstáculo para la implementación de normativas adecuadas.

4.2. ¿Quién es responsable en delitos facilitados por la IA?

La atribución de responsabilidad en delitos facilitados por la IA constituye un desafío jurídico de creciente complejidad. Tradicionalmente, el derecho penal se ha centrado en la imputación de conductas delictivas a personas físicas o jurídicas con capacidad de acción y voluntad. Sin embargo, la irrupción de sistemas de IA capaces de operar de manera autónoma y de aprender de su entorno ha generado situaciones en las que no existe una intervención humana directa en la comisión del delito, complicando la identificación del sujeto responsable (El Kadi, 2025).

Un ejemplo ilustrativo es el caso ocurrido en Hungría en 2019, donde se utilizó una IA para simular la voz de un CEO y solicitar un depósito urgente a una empresa para que transfiriera una suma considerable de dinero (Hildebrandt, 2022). Este incidente plantea interrogantes sobre quién debe ser considerado responsable: el desarrollador de la IA, el usuario que la implementó o la propia entidad que sufrió el engaño por falta de medidas de seguridad adecuadas.

De acuerdo con El-Kadi (2025), la creciente autonomía de los algoritmos dificulta la aplicación de principios tradicionales del derecho penal, pues la imprevisibilidad de la IA rompe con el concepto clásico de dolo y culpa, elementos esenciales para la imputación de responsabilidad.

Desde una perspectiva jurídica, la responsabilidad penal en estos casos podría analizarse desde diferentes enfoques; por un lado, se podría considerar la responsabilidad de los desarrolladores de la IA si se demuestra que diseñaron el sistema con deficiencias que facilitaron su uso delictivo; sin embargo, esta imputación requeriría probar una negligencia grave o dolo en el diseño, lo cual es complejo debido a la naturaleza autónoma y evolutiva de la IA. Por otro lado, los usuarios que emplean la IA con fines delictivos podrían ser considerados autores directos del delito, especialmente si se evidencia una intención clara de utilizar la tecnología para cometer ilícitos. Un ejemplo de ello es el uso de *deepfakes* para crear contenido sexual no consentido, como ocurrió en España en 2023, donde menores manipularon imágenes de compañeras de colegio utilizando IA, generando un debate sobre la tipificación penal de estas conductas (El País, 2024).

Según González-Cuellar (2023), la proliferación de estos delitos evidencia la necesidad de reformular el concepto de autoría en el derecho penal, ya que la IA ha cambiado la relación entre el individuo y la acción delictiva, diluyendo los límites tradicionales de la responsabilidad.

La cuestión se complica aún más cuando la IA actúa de manera impredecible, sin una programación específica para cometer delitos. En estos casos, la imputación de responsabilidad se difumina, ya que ni los desarrolladores ni los usuarios pueden prever o controlar completamente las acciones de la IA. Algunos juristas proponen la creación de un marco normativo que establezca una responsabilidad objetiva para los operadores de sistemas de IA, similar a la responsabilidad de las personas jurídicas, donde se les exigiría implementar medidas de control y supervisión adecuadas para prevenir conductas delictivas (Revista IUS, 2022).

Este enfoque busca equilibrar la promoción de la innovación tecnológica con la necesidad de proteger a la sociedad de los posibles riesgos asociados con la IA. En este sentido, Wachter et al (2018;2021) han defendido la importancia de introducir auditorías obligatorias para algoritmos de IA en sectores críticos, a fin de garantizar que los sistemas no

sean utilizados de manera malintencionada o con fines delictivos.

Otro aspecto relevante es la consideración de la IA como posible sujeto de derecho; aunque actualmente la IA carece de personalidad jurídica y, por ende, no puede ser penalmente responsable, se ha debatido la posibilidad de otorgarle un estatus jurídico especial que permita atribuirle cierta responsabilidad. Hildebrandt (2022) señala que este planteamiento enfrenta desafíos significativos, ya que implicaría reconocer a la IA como un ente con capacidad de acción y voluntad, lo cual contradice su naturaleza como herramienta creada y controlada por humanos (Morán, 2021).

Sancionar a una IA carecería de sentido práctico, ya que no experimenta sufrimiento ni puede ser disuadida mediante penas. Por lo tanto, la responsabilidad debería recaer en las personas o entidades que diseñan, implementan y supervisan estos sistemas. La idea de conceder autonomía jurídica a la IA es una falacia, ya que, en última instancia, cualquier decisión que tome un algoritmo está condicionada por las reglas establecidas por su creador (Luque, Payá y Rodríguez, 2024).

La regulación actual en muchos países, incluido España, no contempla de manera específica la responsabilidad penal por delitos facilitados por la IA. El Código Penal español, por ejemplo, no prevé la imputación de responsabilidad a sistemas de IA, lo que genera un vacío legal en situaciones donde la tecnología desempeña un papel central en la comisión del delito (El Derecho, 2024).

Esta laguna normativa plantea la necesidad de actualizar el marco legal para abordar los desafíos que presenta la IA en el ámbito penal, estableciendo criterios claros para la imputación de responsabilidad y las medidas de prevención necesarias. Como señala Pagallo (2020), la clave para abordar este problema radica en el diseño de normativas que no solo establezcan sanciones, sino que también incentiven buenas prácticas en el desarrollo de IA, fomentando la implementación de mecanismos de control y supervisión que minimicen riesgos.

La determinación de responsabilidad en delitos facilitados por la IA requiere un enfoque multidisciplinario que combine el análisis jurídico con el entendimiento técnico de la tecnología. Es esencial desarrollar un marco normativo que establezca obligaciones claras para los desarrolladores, operadores y usuarios de sistemas de IA, garantizando que se implementen medidas de seguridad y supervisión adecuadas (Mazurier, Delgado y Payá, 2019) .

Asimismo, es fundamental fomentar la investigación y el debate académico sobre este tema, promoviendo la colaboración entre juristas, ingenieros y expertos en ética para abordar de manera integral los desafíos que plantea la IA en el ámbito penal. Como concluyen Zarsky y Weerts (2023), la regulación efectiva de la IA no solo debe centrarse en la asignación de responsabilidad ex post, sino que debe priorizar la prevención y la rendición de cuentas en todas las fases del desarrollo y uso de estos sistemas (Martino y Merenda, 2021).

4.3. Propuesta de supervisión y auditoría de algoritmos

La supervisión y auditoría de algoritmos en sistemas de IA son procesos esenciales para garantizar que estas tecnologías operen de manera ética, transparente y conforme a la normativa vigente. La auditoría algorítmica implica una evaluación exhaustiva de los algoritmos, abarcando desde el tipo de datos que utilizan hasta el impacto de sus decisiones en distintos grupos sociales.

Este proceso no solo verifica el cumplimiento legal, sino que también identifica riesgos asociados a su aplicación, permitiendo implementar correcciones cuando sea necesario.

Un ejemplo concreto de auditoría algorítmica se encuentra en el ámbito de los motores de búsqueda de hoteles, donde la Comisión Australiana de Competencia y Consumo

(ACCC) llevó a cabo una auditoría de un popular motor de búsqueda de hoteles y descubrió que el algoritmo favorecía injustamente a los hoteles que pagaban comisiones más altas en su sistema de clasificación. Este hallazgo permitió a la ACCC intervenir y exigir modificaciones en el algoritmo para garantizar una competencia justa y proteger los derechos de los consumidores.

En España, se han dado pasos significativos hacia la supervisión de la IA con la creación de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA). Esta entidad autónoma tiene como objetivo supervisar, asesorar y formar en el uso y desarrollo adecuado de los sistemas de IA, con especial énfasis en los algoritmos (Payá y Delgado, 2017; Guthrie, 2017). Además de funciones de inspección y sanción, la AESIA busca minimizar los riesgos asociados al uso de estas tecnologías, asegurando su desarrollo ético y responsable

4.4. Propuesta de uso de la IA como herramienta de apoyo y no como único criterio judicial

Es ampliamente aceptado que la IA debe servir como una herramienta de apoyo para los sistemas judiciales, complementando su labor sin reemplazar el juicio humano. Esta perspectiva se basa en la necesidad de mantener la imparcialidad, la equidad y la justicia en los procesos judiciales, valores que son intrínsecamente humanos y que la IA, por sí sola, no puede garantizar.

Un ejemplo que ilustra los riesgos de depender exclusivamente de la IA en decisiones judiciales se aprecia en el estudio de Guevara et al. (2024), donde señala que un juez en Colombia utilizó ChatGPT, una herramienta de IA, para resolver una tutela relacionada con los derechos de un niño con autismo. Aunque la decisión final fue favorable para el demandante, el uso de la IA sin una supervisión adecuada generó controversia y llevó a la Corte Constitucional a enfatizar que los jueces no deben delegar sus responsabilidades en sistemas de IA y deben transparentar su uso en los procesos judiciales.

Este incidente subraya la importancia de utilizar la IA como un complemento en el proceso judicial, aportando eficiencia en el análisis de grandes volúmenes de información y en la identificación de patrones que pueden ser útiles para los jueces.

Sin embargo, la decisión final debe recaer siempre en el juez, quien posee la capacidad de interpretar la ley con sensibilidad y comprensión del contexto humano, elementos que la IA aún no puede replicar. Además, es esencial establecer marcos regulatorios claros que guíen el uso de la IA en el ámbito judicial, asegurando que su implementación respete los derechos fundamentales y mantenga la integridad del sistema de justicia (Payá, 2023; Rodríguez, Payá y Peña, 2023; Sanz et al, 2024).

5. Conclusiones

La irrupción de la IA en el derecho penal ha representado un punto de inflexión en la evolución de los sistemas jurídicos contemporáneos. Su aplicación ha permitido avances sin precedentes en la predicción del crimen, la gestión de pruebas digitales y la optimización de los procesos judiciales, lo que ha transformado profundamente la administración de justicia

La automatización de ciertos procedimientos y la capacidad analítica de los sistemas de IA han agilizado la toma de decisiones y han mejorado la capacidad de los operadores jurídicos para gestionar grandes volúmenes de información de manera eficiente.

No obstante, la creciente dependencia de estos sistemas plantea desafíos estructurales que trascienden el ámbito tecnológico y afectan directamente la configuración misma del derecho penal.

El principio de culpabilidad, la presunción de inocencia y la proporcionalidad de las penas, pilares fundamentales de la teoría penal, podrían verse erosionados si la IA no es

implementada dentro de un marco de garantías y con los debidos controles institucionales (Thamer et al ., 2024).

La celeridad y eficiencia que prometen estos sistemas no pueden sustituir el rigor y la prudencia inherentes a la labor judicial, ya que el derecho penal no solo se fundamenta en el análisis racional de los hechos, sino también en una dimensión ética y humanista que la IA, por su propia naturaleza, hasta ahora, no es capaz de replicar.

Resulta ineludible reflexionar sobre la imperiosa necesidad de establecer un equilibrio entre el desarrollo tecnológico y la protección de los derechos humanos (Grigore, 2022).

La IA, aunque dotada de capacidades predictivas y analíticas extraordinarias, no debe convertirse en un mecanismo que vulnere las garantías procesales de los individuos ni que comprometa la igualdad ante la ley. La automatización de ciertos procedimientos, si no se somete a un control adecuado, podría generar decisiones discriminatorias o arbitrarias, afectando de manera desproporcionada a determinados colectivos vulnerables y erosionando la confianza en la administración de justicia. La justicia penal no puede convertirse en un sistema mecanicista donde los algoritmos determinen el destino de las personas sin considerar la complejidad de cada caso y la singularidad de las circunstancias individuales.

Los avances tecnológicos deben estar subordinados a un marco ético que garantice que la IA no sea utilizada como un instrumento de exclusión ni como una herramienta que reproduzca sesgos estructurales presentes en la sociedad. La equidad y la proporcionalidad deben prevalecer sobre cualquier criterio de eficiencia algorítmica, pues el derecho penal no es únicamente una cuestión de cálculo y optimización, sino un sistema normativo que debe garantizar la dignidad y la libertad de todos los ciudadanos.

La necesidad de una regulación ética y garantista no es un mero imperativo técnico, sino un mandato esencial para la preservación del Estado de derecho. La incorporación de la IA en el ámbito penal no debe interpretarse como un fenómeno inevitable que deba ser aceptado sin reservas, sino como un proceso que exige una construcción normativa rigurosa y un debate profundo sobre sus implicaciones.

La transparencia de los algoritmos, la auditabilidad de los sistemas y la responsabilidad de sus desarrolladores deben formar parte del andamiaje normativo que rija su implementación; de lo contrario, permanecerá latente el riesgo de consolidar un modelo de justicia en el que las decisiones judiciales sean opacas e inapelables, lo que atentaría directamente contra los principios fundamentales del debido proceso.

El derecho penal, como expresión máxima del poder punitivo del Estado, no puede delegar su función en sistemas que carecen de sensibilidad moral ni de capacidad crítica, pues el juicio penal no es solo una cuestión de determinación factual, sino un acto de deliberación que debe ponderar valores, contextos y principios.

Por ello, la regulación de la IA en el derecho penal debe partir de una visión garantista, en la que la tecnología sea un complemento y no un sustituto del razonamiento humano, y en la que los derechos fundamentales sean el eje rector de cualquier innovación jurídica.

Referencias

ALONSO SALGADO, Cristina (2021). Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad. *IUS ET SCIENTIA*. Vol. 7 Nº 1. pp. 25 - 36 https://dx.doi.org/10.12795/IETSCIENTIA.2021.i01.03

BALCELLS, Marc. (2020). Luces y sombras del uso de la inteligencia artificial en el sistema de Justicia penal. En A. Cerrillo I Martínez y M. Peguera Poch (Eds.). Retos

jurídicos de la inteligencia artificial. Cizur Menor (Navarra): Aranzadi.

BRAYNE, Sarah. (2017). Big Data Surveillance: The Case of Policing. American Sociological Review, 82(5), 977-1008. https://doi.org/10.1177/0003122417725865

BUOLAMWINI, Joy., & GEBRU, Timnit. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Science Advances, 4(1), eaas8576. Proceedings of Machine Learning Research. 81:77-91 Available from https://proceedings.mlr.press/v81/buolamwini18a.html

CHESNEY, Robert., & CITRON, Danielle. (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. Foreign Affairs, 98(1), 147-155. Available at: https://scholarship.law.bu.edu/shorter_works/76

CHÉN, Oliver. (2022). Uniting Machine Intelligence, Brain and Behavioural Sciences to Assist Criminal Justice. arXiv preprint arXiv:2207.01511. https://doi.org/10.48550/arXiv.2207.01511

DELGADO MORAN, Juan. José, MAZURIER, Pablo. Andrés. & PAYA SANTOS, Claudio. Augusto. (2019). The race to securitize the arctic in a post-cold War scenario. *Revista de Pensamiento Estratégico y Seguridad CISDE*, 4(1), 59-64. http://hdl.handle.net/10272/17180

El Derecho. com (2024). Responsabilidad penal y la inteligencia artificial en España. Editorial Jurídica El Derecho.

El País. (2024). El caso de deepfakes en España reabre el debate sobre la regulación de la IA. Recuperado de https://www.elpais.com

El País. (2024). Una demanda de un niño con autismo abre el debate para regular el uso de la IA en la justicia colombiana. Recuperado de https://elpais.com/america-colombia/2024-08-25/una-demanda-de-un-nino-con-autismo-abre-el-debate-para-regular-el-uso-de-la-ia-en-la-justicia-colombiana.html

European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. COM/2021/206 final.

EL KADY, Ramy. (2025). "Challenges of Criminal Liability for Artificial Intelligence Systems." In Exploration of AI in Contemporary Legal Systems, edited by Halim Bajraktari, 1-42. Hershey, PA: IGI Global,. https://doi.org/10.4018/979-8-3693-7205-0.ch001

GUTHRIE FERGUSON, Andrew. (2017). The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. New York: NYU Press. https://doi.org/10.18574/nyu/9781479854608.001.0001

GANGI GUILLEN, Giuseppe. Kodjack. (2023). Dinámicas migratorias en la frontera colombo-venezolana y su relación con la criminalidad transnacional. Revista Científica General José María Córdova, 21(44), 907–924. https://doi.org/10.21830/19006586.984

GARVIE, Clare., BEDOYA, Alvaro. & FRANKLE, Jonathan. (2016). The Perpetual Line-Up: Unregulated Police Face Recognition in America. Georgetown Law Center on Privacy & Technology. https://www.perpetuallineup.org/

GÓMEZ ROMERO, Miriam Judit. (2023). El Metaverso como forma de trabajo. Riesgos, y oportunidades: una mirada al derecho internacional. *Cuadernos de RES PUBLICA en derecho y criminología*, n.º 2 (junio):112-21. https://doi.org/10.46661/respublica.8223

GONZÁLEZ TRIGO, Norberto Aser. (2023). El agente encubierto ante la criminalidad organizada transnacional. *Cuadernos de RES PUBLICA en derecho y criminología*, n.º 1 (mayo): 85-94. https://doi.org/10.46661/respublica.8062

GRIGORE, Andrea. Elena. (2022). Derechos humanos e inteligencia artificial. *IUS ET SCIENTIA*, 8(1), 164–175. https://doi.org/10.12795/IETSCIENTIA.2022.i01.10

HERNÁNDEZ GIMÉNEZ, María. (2019). Inteligencia artificial y Derecho penal.. *Actualidad Jurídica Iberoamericana,* 10(bis), 792-843. https://revista-aji.com/wp-content/uploads/2019/06/792-843.pdf

HILDEBRANDT, Mirelle. (2022). Law for Computer Scientists and Other Folk. Oxford University Press. , https://doi.org/10.1093/oso/9780198860877.001.0001

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

LIZ RIVAS, Lenny. (2018). Algunas bases neurológicas sobre la violencia y la agresión, en ;"Conflictos y diplomacia, desarrollo y paz, globalización y medio ambiente " coord. Por Emilio José García Mercader, Claudio Payá Santos; César Augusto Giner Alegría (dir.), Juan Jose Delgado Morán (dir.), Thomson Reuters/Aranzadi, pp. 943-955. https://doi.org/10.5281/zenodo.14559664

LUQUE JUÁREZ, José. María., SANZ GONZÁLEZ, Roger., PAYÁ SANTOS, Claudio. Augusto & RODRÍGUEZ GONZÁLEZ, Víctor. (2024). Aplicaciones de la inteligencia artificial en la criminología y ciencias policiales. https://doi.org/10.13140/RG.2.2.11118.63041.

MANDALAPU, Varun., ELLURI, Lavanya., VYAS, Piyush., & ROY, Nirmalya. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future. https://doi.org/10.48550/arXiv.2303.16310 https://doi.org/10.1109/ACCESS.2023.3286344

MARTINO, Luigi. (2024). Cybersecurity in Italy. Governance, Policies and Ecosystem. Springer Nature. https://doi.org/10.1007/978-3-031-64396-5

MARTINO, Luigi. (2024). International Law, State Sovereignty and Competition in the Digital Age. *Rivista di filosofia del diritto internazionale e della politica globale*, Vol. 21, Nº. 2, 2024. https://dialnet.unirioja.es/descarga/articulo/10098952.pdf

MARTINO, Luigi. & MERENDA, Federica. (2021). Artificial intelligence: A paradigm shift in international law and politics? Autonomous weapon systems as a case study. in: Giampiero Giacomello & Francesco N. Moro & Marco Valigi (ed.), Technology and International Relations, chapter 5, pages 89-107, Edward Elgar Publishing. https://doi.org/10.4337/9781788976077.00012

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

MIRÓ LLINARES, Fernando. (2018). Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. *Revista de Derecho Penal y Criminología*, (20), 87–130. https://doi.org/10.5944/rdpc.20.2018.26446

MORÁN ESPINOSA, Alejandra. (2021). Responsabilidad penal de la Inteligencia Artificial (IA). ¿La próxima frontera? *Revista del Instituto de Ciencias Jurídicas de Puebla*, México. vol. 15, No. 48. https://doi.org/10.35487/rius.v15i48.2021.706

Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (2021). Los riesgos de la inteligencia artificial para la privacidad exigen una acción urgente. Recuperado de https://goo.su/TegmwK

PAGALLO, Ugo (2013). The Laws of Robots: Crimes, Contracts, and Torts. Dordrecht: Imprint: Springer. https://doi.org/10.1007/978-94-007-6564-1

PASQUALE, Frank. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge: Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

PAYÁ SANTOS, Claudio. Augusto. (2023). El desempeño de la inteligencia en España en el ámbito público, empresarial y académico. *Revista Científica General José María Córdova*, *21*(44), 1029–1047. https://doi.org/10.21830/19006586.1222

PAYÁ SANTOS, Claudio. Augusto; DELGADO MORÁN, Juan. José;. MARTINO, Luigi; GARCÍA SEGURA, Luis, A.; DIZ CASAL, Javier, & FERNÁNDEZ-RODRÍGUEZ, Juan, Carlos. (2023). Fuzzy Logic analysis for managing Uncertain Situations. *Review of Contemporary Philosophy* Vol 22 (1), 2023 pp. 6780 -6797. https://doi.org/10.52783/rcp.1132

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

RICAURTE, Paola. (2024). Las grandes compañías tecnológicas son aliadas de gobiernos autoritarios. El País. Recuperado de https://elpais.com/america/lideresas-de-latinoamerica/2024-10-16/paola-ricaurte-las-grandes-companias-tecnologicas-son-aliadas-de-gobiernos-autoritarios.html

RODRÍGUEZ GONZÁLEZ, Víctor., PAYÁ, SANTOS, Claudio, Augusto., & PEÑA HERRERA, Bernardo. (2023). Estudio criminológico del ciberdelincuente y sus víctimas. *Cuadernos de RES PUBLICA en Derecho y criminología,* (1) 95-107. https://doi.org/10.46661/respublica.8072.

SANZ GONZÁLEZ, Roger, LUQUE JUÁREZ, José María, MARTINO, Luigi, LIZ RIVAS, Lenny, DELGADO MORÁN, Juan José, & PAYÁ SANTOS, Claudio Augusto. (2024) Artificial Intelligence Applications for Criminology and Police Sciences. *International Journal of Humanities and Social Science*. Vol. 14, No. 2, pp. 139-148. https://doi.org/10.15640/jehd.v14n2a14

SCANTAMBURLO, Teresa, Andrew CHARLESWORTH, & Nello CRISTIANINI, (2019). Machine Decisions and Human Consequences, in Karen Yeung, and Martin Lodge (eds), Algorithmic Regulation. Oxford Academic. https://doi.org/10.48550/arXiv.1811.06747

THAMER NAJM, Abdullah Abbas., HAMEED, Raed., AKRAM KADHIM, Ali., & HASHIM QASIM, Namer. (2024). Artificial intelligence and criminal liability: exploring the legal implications of aIenabled crimes. *ENCUENTROS. Revista de Ciencias Humanas, Teoría Social Y Pensamiento Crítico.,* 22 140-159. https://doi.org/10.5281/zenodo.13386675

WACHTER, Sandra., MITTELSTADT, Brent., & RUSSELL, Chris. (2018) "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", Harvard Journal of Law and Technology. 31 (2) 841-887. https://doi.org/10.2139/ssrn.3063289

WACHTER, Sandra., MITTELSTADT, Brent., & RUSSELL, Chris. (2021). Why fairness cannot be automated: Bridging the gap between AI and human rights law. Computer Law & Security Review, 43(4), 34-55. https://doi.org/10.1016/j.clsr.2021.105567

XAVIER JANUÁRIO, Tulio. Felippe. (2023). Inteligencia artificial y responsabilidad penal de personas jurídicas: un análisis de sus aspectos materiales y procesales: Un análisis de sus aspectos materiales y procesales . *Estudios Penales Y Criminológicos,* 44(Ext.), 1-39. https://doi.org/10.15304/epc.44.8902