



Influencia de la Inteligencia Artificial en el Derecho Penal: Una revisión sistemática de la literatura

Influence of Artificial Intelligence on Criminal Law: A Systematic Review of the Literature

Influência da Inteligência Artificial no Direito Penal: Uma Revisão Sistemática da Literatura

América Beatriz Benítez-Abarca ^I
americabenitezabarca@outlook.es
<https://orcid.org/0009-0008-9898-8323>

Silvia Patricia Chafla-Moina ^{II}
sp.chaflam@uea.edu.ec
<https://orcid.org/0009-0001-2412-6332>

María Alexandra Manosalvas-Vaca ^{III}
ma.manosalvasv@uea.edu.ec
<https://orcid.org/0009-0008-4152-0607>

Yajaira Marianela Saltos-Verdezoto ^{IV}
ysaltos@uea.edu.ec
<https://orcid.org/0009-0001-4081-9024>

Correspondencia: americabenitezabarca@outlook.es

Ciencias Sociales y Políticas
Artículo de Investigación

* **Recibido:** 13 de enero de 2025 * **Aceptado:** 19 de febrero de 2025 * **Publicado:** 31 de marzo de 2025

- I. Universidad Estatal Amazónica, Puyo, Ecuador.
- II. Universidad Estatal Amazónica, Puyo, Ecuador.
- III. Universidad Estatal Amazónica, Puyo, Ecuador.
- IV. Universidad Estatal Amazónica, Puyo, Ecuador.

Resumen

El presente trabajo muestra una revisión exhaustiva sobre el papel de la inteligencia artificial (IA) en el ámbito del derecho penal. Recoge para ello sus aplicaciones y ventajas, pero también los problemas éticos y legales que sus aplicaciones pueden dar lugar. En este sentido, la IA ha demostrado tener efectos positivos sobre la precisión y la eficiencia en los procesos judiciales, especialmente en áreas como la predicción de reincidencia o el análisis conductual. Por el contrario, la adopción de los sistemas de IA como herramientas en el derecho penal puede suponer también importantes riesgos, como la opacidad de los algoritmos, sesgos y problemas en el mantenimiento de la transparencia de las decisiones automatizadas. Para llevar a cabo esta revisión, se ha seguido el modelo PRISMA, interrogando estudios recientes, entre 2022 y 2025, obtenidos de las bases de datos Scopus y Web of Science. Los resultados extraídos de esta investigación se agrupan en seis categorías principales que son los tipos de IA aplicados en el derecho penal, implicaciones éticas, aspectos procesales y ejemplos de uso dados. Este análisis deja en evidencia la necesidad de regular la IA en el contexto de la justicia, y enfatiza también la importancia de la supervisión humana. Las conclusiones del trabajo muestran la necesidad de promover políticas que favorezcan la transparencia de los algoritmos, así como el respeto de los derechos fundamentales. Finalmente, también sugiere continuar futuras líneas de investigación, en dirección a evaluar empíricamente los sesgos algorítmicos y a generar algoritmos más explicables, con objeto de lograr una adopción ética y eficaz de la IA en el sistema judicial penal.

Palabras Clave: Inteligencia Artificial; Derecho Penal.

Abstract

This paper presents a comprehensive review of the role of artificial intelligence (AI) in criminal law. It examines its applications and advantages, as well as the ethical and legal issues its applications can raise. In this regard, AI has been shown to have positive effects on the accuracy and efficiency of judicial processes, especially in areas such as recidivism prediction and behavioral analysis. Conversely, the adoption of AI systems as tools in criminal law can also pose significant risks, such as the opacity of algorithms, biases, and problems maintaining the transparency of automated decisions. To conduct this review, the PRISMA model was followed, examining recent studies from 2022 to 2025, obtained from the Scopus and Web of Science databases. The results of this research are grouped into six main categories: the types of AI applied

in criminal law, ethical implications, procedural aspects, and examples of its use. This analysis highlights the need to regulate AI in the context of justice and also emphasizes the importance of human oversight. The paper's conclusions highlight the need to promote policies that favor the transparency of algorithms and respect for fundamental rights. Finally, it also suggests pursuing future lines of research to empirically evaluate algorithmic biases and generate more explainable algorithms, in order to achieve the ethical and effective adoption of AI in the criminal justice system.

Keywords: Artificial Intelligence; Criminal Law.

Resumo

Este artigo apresenta uma revisão abrangente do papel da inteligência artificial (IA) no âmbito do direito penal. Para isso, examina as suas aplicações e vantagens, mas também os problemas éticos e jurídicos que as suas aplicações podem suscitar. Neste sentido, a IA demonstrou ter efeitos positivos na precisão e eficiência dos processos judiciais, especialmente em áreas como a previsão da reincidência e a análise comportamental. Pelo contrário, a adoção de sistemas de IA como ferramentas no direito penal pode também representar riscos significativos, como a opacidade dos algoritmos, os enviesamentos e os problemas de manutenção da transparência das decisões automatizadas. Para conduzir esta revisão, seguimos o modelo PRISMA, examinando estudos recentes de 2022 a 2025, obtidos a partir das bases de dados Scopus e Web of Science. Os resultados extraídos desta pesquisa estão agrupados em seis categorias principais: os tipos de IA aplicados no direito penal, as implicações éticas, os aspetos processuais e os exemplos de utilização fornecidos. Esta análise destaca a necessidade de regulamentar a IA no contexto da justiça e enfatiza também a importância da supervisão humana. As conclusões do artigo realçam a necessidade de promover políticas que promovam a transparência nos algoritmos e respeitem os direitos fundamentais. Por fim, sugere ainda a procura de futuras linhas de investigação orientadas para a avaliação empírica de enviesamentos algorítmicos e para a geração de algoritmos mais explicáveis, de forma a alcançar a adoção ética e eficaz da IA no sistema de justiça criminal.

Palavras-chave: Inteligência Artificial; Direito penal.

Introducción

La Inteligencia Artificial (IA) ha tenido una profunda incidencia en distintas esferas, también en el ámbito del Derecho Penal. En la última década, en el contexto de las investigaciones penales, la evaluación de riesgos o la toma de decisiones judiciales, la aplicación de sistemas de IA ha promovido, tanto la eficacia como la exactitud, de los procedimientos judiciales, especialmente en ámbitos de relevancia como la predicción de la reincidencia, los análisis de patrones de conducta o la detección del crimen en tiempo real (Kleinberg et al., 2018). Estos avances son efectos de la utilización de técnicas recientes de Machine Learning y de Big Data, lo que permite realizar evaluaciones que en ocasiones rebasan las capacidades humanas. Cabe decir, sin embargo, que la integración de la IA en el ámbito del Derecho Penal también se ven acompañadas de desafíos éticos y legales, sobre todo en lo que tiene que ver con el respeto de los derechos fundamentales, la transparencia en la toma de decisiones o la mitigación del sesgo en los sistemas algoritmos (Angwin et al., 2016).

A pesar de los avances logrados, subsisten vacíos importantes de investigación que limitan nuestra comprensión de los efectos de la IA en el Derecho Penal. En efecto, la carencia de estandarización de la utilización de herramientas de IA y la escasez de estudios empíricos en los que se aplique en distintos ámbitos legales y judiciales refuerzan la necesidad de realizar investigaciones adicionales (Barocas & Selbst, 2016; Legg & Bell, 2020). Por ende, el sesgo algorítmico y la opacidad de las decisiones generadas en los procesos judiciales a partir de sistemas de IA se muestran como unas áreas poco investigadas, las cuales necesitan un análisis exhaustivo que garantice que estos sistemas se ajusten a los principios de justicia e imparcialidad (Dharneesh et al., 2024). Dando respuesta a aquellos vacíos que nos han llevado a producir esta investigación, se exploran los beneficios y también los posibles riesgos de la IA en el contexto penal, y se propone un marco de referencias críticas para una adopción responsable de las mismas en los sistemas judiciales.

Los propósitos de realizar este trabajo de investigación son por tanto los de llevar a cabo una revisión sistemática de la literatura, explorando el papel de la IA en el derecho penal desde una perspectiva holística. El primer objetivo del estudio consiste en poder determinar cuáles son las principales aplicaciones informáticas de la IA en los sistemas judiciales y de seguridad, analizando aquellos estudios que abordan las revisiones de la IA centrándose en los riesgos y beneficios que comportan y los impactos que le otorgan a la administración de justicia. Asimismo, se analizarán aquellas propuestas normativas que emergen de la literatura científica para poder contribuir a

reducir los sesgos algorítmicos y garantizar la definición de los principios de equidad en el marco de la administración de justicia penal.

Esta investigación presenta un marco completo y al hilo con el rol de la IA en el contexto del derecho penal, en el ámbito académico y práctico. La revisión sistemática aporta un conocimiento explícito y claro de los límites y del potencial de la IA en este ámbito, ofreciendo al legislador, así como al magistrado y a los actores del proceso jurídico, recomendaciones fundamentadas para una adecuada aplicación ética de las herramientas informáticas. Las implicaciones prácticas residen en la construcción de políticas que puedan llegar a elaborar la promoción de la transparencia algorítmica y la creación de estándares que lleguen a regular el funcionamiento de la IA dentro del marco de la justicia penal, velando por los derechos humanos y construyendo la confianza en el sistema judicial (Hildebrandt, 2020).

Metodología

Con el fin de conseguir los objetivos que se han formulado en esta investigación y dar respuesta a las preguntas planteadas, se llevó a cabo una revisión sistemática, la cual se ejecutó mediante un protocolo explícito y replicable basado en el modelo PRISMA, realizando una búsqueda que incluía seleccionar los estudios atendiendo a criterios de inclusión y exclusión previamente fijados (Gough et al., 2012)

Las bases de datos seleccionadas para llevar a cabo la búsqueda de literatura fue la Web of Science (WoS), debido a su extensísimo registro de publicaciones científicas revisadas por pares, y su interés en el ámbito de derecho penal e inteligencia artificial. La búsqueda de la literatura estuvo limitada al periodo de 2022 a 2025, con el fin de intentar dar cuenta de los estudios más actuales en relación con el tema a analizar. La búsqueda fue a partir de los términos "inteligencia artificial" OR "artificial intelligence" AND "derecho penal" OR "criminal law", aplicando operadores booleanos con el fin de tener mayor exactitud en los resultados que obteníamos. Se fijaron criterios específicos para incluir los estudios en el análisis. Se incluyeron aquellos artículos que trataban sobre la aplicación de la IA dentro del ámbito del derecho penal y que ofrecían datos empíricos o bien análisis críticos a propósito del tema tratado. Se excluyeron documentos duplicados, capítulos de libros, disertaciones y artículos de conferencias que no abordaban de forma directa la IA en el ámbito del derecho penal.

En relación al proceso de selección y al análisis de los datos, se realizó un proceso de selección por medio de dos fases. En primer lugar, se consultaron los títulos y resúmenes para decidir la pertinencia de los estudios. En segundo lugar, se leyeron los estudios, resultado de la selección de forma intensiva para confirmar la inclusión. Para organizar los resultados, se implanto una matriz de codificación que permitió clasificar los estudios en función de variables de interés como es la utilización de la IA en decisiones judiciales en el ámbito del derecho penal. La lectura de los estudios seleccionados fue analizada y sintetizada a través de un proceso cualitativo y cuantitativo (donde fue necesario). Para la organización de las categorías temáticas de los resultados se realiza un análisis descriptivo para ofrecer una idea global de la influencia de la IA en el derecho penal.

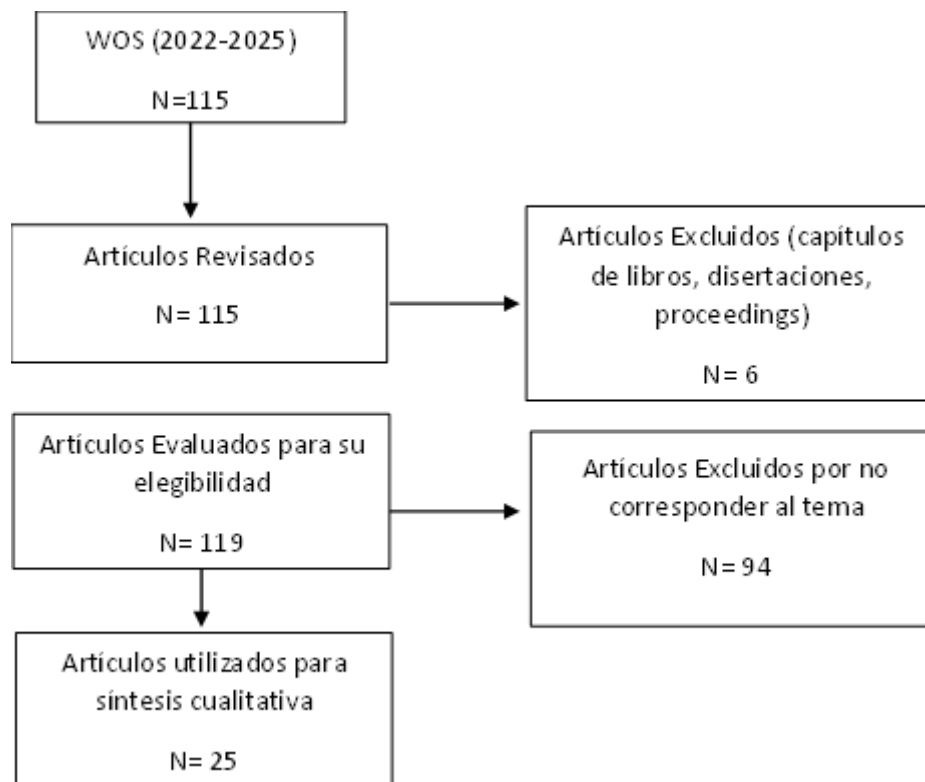


Figura 1. Diagrama de Flujo del proceso de selección del estudio

Análisis de Resultados

La revisión sistemática realizada sobre el uso de la Inteligencia Artificial (IA) en el derecho penal revela seis categorías clave de información que destacan los diversos enfoques y desafíos en la implementación de estas tecnologías en el sistema judicial. A continuación, se presentan los hallazgos en cada una de estas categorías.

Tipos de IA y sus aplicaciones en el derecho penal

La aplicación de la inteligencia artificial (IA) en el ámbito del derecho penal abarca un conjunto de tecnologías diseñadas para optimizar los sistemas de justicia penal, mejorando la eficacia y precisión en la resolución de casos. Dentro de estas tecnologías, destacan tres enfoques principales: los sistemas expertos, el aprendizaje automático y la IA para la valoración del riesgo.

En primer lugar, los sistemas expertos buscan replicar el razonamiento de un especialista en derecho penal mediante la aplicación de reglas predefinidas y conocimientos extraídos de precedentes jurisprudenciales. Estas herramientas permiten asistir a los operadores jurídicos en la toma de decisiones sobre casos de alta complejidad o en la interpretación de normativas legales (Caterini, 2022). Además, expertos en informática han desarrollado modelos capaces de proporcionar recomendaciones jurídicas fundamentadas con base en reglas específicas, facilitando así el análisis y la argumentación en el proceso penal (Dremluiga & Korobeev, 2023).

Por otro lado, el aprendizaje automático representa una de las aplicaciones más relevantes de la IA en la justicia penal. A través del procesamiento y análisis de grandes volúmenes de datos, estos algoritmos pueden identificar patrones y realizar predicciones, lo que resulta especialmente útil para el reconocimiento de tendencias delictivas y la estimación de la reincidencia criminal. Una de las ramas más avanzadas de este enfoque es el aprendizaje profundo, que emula el funcionamiento de las redes neuronales del cerebro humano, permitiendo análisis predictivos más sofisticados y precisos (Serventich, 2022).

La IA también se emplea en la valoración del riesgo, facilitando la evaluación de la probabilidad de reincidencia en un contexto judicial, así como la identificación de riesgos procesales de los acusados. Estas herramientas se basan en el análisis de datos históricos sobre conductas delictivas, factores socioeconómicos y antecedentes penales, permitiendo a jueces y fiscales realizar una evaluación informada antes de la toma de decisiones en el proceso judicial (Caterini, 2022; Serventich, 2022).

Implicaciones de la IA en el derecho penal

La incorporación de la inteligencia artificial (IA) en el sistema penal conlleva importantes implicaciones en el ámbito ético, jurisprudencial y operativo. Su aplicación en la toma de decisiones judiciales plantea debates sobre la responsabilidad penal de la IA, el impacto en los principios fundamentales del derecho penal, el rol de los jueces en la era digital y los riesgos de sesgo y discriminación.

Uno de los debates más relevantes es la posible responsabilidad penal de la IA, especialmente ante la creciente autonomía de estos sistemas en la toma de decisiones. Algunos autores argumentan que, si la IA puede actuar de manera autónoma, entonces podría considerarse que sus acciones tienen una “intención maliciosa”, lo que daría lugar a la discusión sobre la imputabilidad penal de los sistemas de IA o de quienes los programan. Esta controversia ha llevado a posturas encontradas respecto a la necesidad de regular la responsabilidad de los desarrolladores, operadores y usuarios de la IA en contextos judiciales (Mahardhika et al., 2023; Rodger et al., 2023; Zhaoxun et al., 2024).

Asimismo, el uso de la IA en el derecho penal podría afectar principios fundamentales como la legalidad, la presunción de inocencia y el derecho a la defensa. La creciente automatización en los procesos penales exige un marco regulador que garantice la protección de los derechos de las personas involucradas en estos procedimientos. Es crucial que la IA aplicada en la justicia penal sea desarrollada con mecanismos que preserven estos principios, evitando que la automatización se convierta en una barrera para el acceso a la justicia o genere fallos arbitrarios (Mahardhika et al., 2023; Serventich, 2022).

Otro aspecto de gran relevancia es el rol de la IA en la toma de decisiones judiciales. Aunque los sistemas de IA pueden servir como herramientas de apoyo para jueces y fiscales, no deben sustituir el juicio humano en la resolución de casos. La delegación excesiva de la toma de decisiones en algoritmos automatizados podría socavar la confianza pública en el sistema judicial. Para evitar estos riesgos, se requiere que los modelos de IA sean explicables y transparentes, de manera que tanto los operadores de justicia como los ciudadanos comprendan los fundamentos de sus recomendaciones o predicciones (Caterini, 2022; Ryberg, 2024; Al-Dulaimi & Mohammed, 2025). Uno de los desafíos más críticos es la posible reproducción de sesgos y discriminación por parte de los algoritmos de IA. Si bien estas tecnologías pueden utilizarse para la predicción de delitos y el análisis de patrones criminales (Kaur & Saini, 2024), también pueden perpetuar sesgos estructurales existentes en el sistema penal. Dado que los algoritmos aprenden a partir de datos históricos, es posible que reproduzcan desigualdades relacionadas con la raza, el género o el estatus socioeconómico de los acusados. La implementación de salvaguardas para mitigar estos sesgos es fundamental para garantizar la equidad en el uso de la IA en el derecho penal (Mustafa, 2024).

Se evidencia así que la integración de la inteligencia artificial en el sistema penal ofrece oportunidades significativas para mejorar la eficiencia de la justicia, pero también plantea desafíos

sustanciales en términos de responsabilidad, equidad y confianza pública. Es esencial un enfoque regulatorio riguroso que equilibre los beneficios tecnológicos con la protección de los derechos fundamentales en el ámbito penal.

Aspectos procesales de la IA en el derecho penal

La inteligencia artificial también plantea cuestiones fundamentales en relación con los procedimientos penales:

- *Admisibilidad de prueba generada por IA:* Existe una controversia respecto a la aceptabilidad de la prueba generada por IA ya que la prueba generada por IA no es suficientemente fiable y válida para los procedimientos jurisdiccionales. Las propiedades de los algoritmos, su oscuridad u opacidad, los sesgos y, por tanto, la exigencia de la admisibilidad de la prueba y su credibilidad constituyen una dificultad para la aducción de prueba generada por IA (Sayyed, 2024; Serventich, 2022).
- *Motivación de las resoluciones jurisdiccionales generadas por IA:* La incorporación de IA al procedimiento judicial que se lleva a cabo por un juez o jueza comporta la necesidad de mostrar cómo la IA ha influido en lo que dictan (Shi, 2022). La motivación de una resolución jurisdiccional a partir de IA debe ser comprensible para las partes y para la población, dado que ha de asegurarse la legitimidad del sistema judicial y el derecho de las partes a enjuiciar la decisión tomada a partir de IA (Buján, 2023; Wu et al., 2022).

Retos y oportunidades de la IA en el derecho penal

La incorporación de la inteligencia artificial (IA) en el ámbito del derecho penal ofrece oportunidades significativas para optimizar la eficiencia del sistema judicial, pero al mismo tiempo plantea desafíos que requieren una regulación clara y estrategias para garantizar su uso justo y equitativo. La IA puede mejorar la toma de decisiones, reducir la carga procesal y aumentar la precisión en los procedimientos judiciales. Sin embargo, su implementación sin un marco normativo adecuado puede generar riesgos relacionados con la ética, la privacidad y la equidad en la administración de justicia.

Uno de los aspectos más críticos en la adopción de la IA en el derecho penal es la necesidad de reglamentación y marcos éticos adecuados. A medida que estas tecnologías se integran en el sistema judicial, es fundamental contar con normativas que regulen su uso, establezcan principios éticos y garanticen que las herramientas de IA sean empleadas de manera transparente y respetuosa con los derechos fundamentales. La falta de regulación específica puede derivar en un uso

inconsistente de la IA, lo que podría dar lugar a errores en la interpretación de datos, sesgos en las decisiones automatizadas y, en casos extremos, abusos por parte de los operadores del sistema de justicia penal (DoCarmo et al., 2022; Suleimanova, 2024). Para evitar estos riesgos, es imprescindible la supervisión humana en los procesos en los que la IA desempeña un rol determinante.

Por otro lado, uno de los principales beneficios de la IA en el ámbito judicial es la reducción de la carga procesal. La automatización de tareas repetitivas, como la revisión de expedientes, la identificación de patrones en grandes volúmenes de datos y la elaboración de informes judiciales, permite a jueces y fiscales enfocarse en actividades de mayor valor agregado, como la interpretación de la ley y la deliberación en casos complejos. Esta optimización del tiempo no solo mejora la eficiencia de los procesos judiciales, sino que también contribuye a la reducción de los tiempos de espera en la resolución de casos, favoreciendo así un acceso más ágil a la justicia (Begaric et al., 2022; Granados, 2024).

Asimismo, la IA tiene el potencial de incrementar la eficiencia y precisión en la toma de decisiones judiciales. Al minimizar los errores humanos y proporcionar análisis de datos basados en evidencia, los sistemas de IA pueden ofrecer información relevante para la fundamentación de fallos y la evaluación de riesgos en el proceso penal. Sin embargo, para que la IA realmente contribuya a la justicia, es fundamental enfrentar los desafíos que representan los sesgos algorítmicos. Si los modelos de IA se entrenan con datos históricos que reflejan desigualdades estructurales, es probable que perpetúen y amplifiquen esas mismas desigualdades dentro del sistema penal (Zhang et al., 2022). Por ello, es esencial que el desarrollo de herramientas de IA en la justicia penal incorpore mecanismos de auditoría y supervisión que garanticen la equidad en su aplicación.

En conclusión, aunque la inteligencia artificial tiene el potencial de mejorar significativamente la administración de justicia penal, su implementación debe realizarse con precaución, garantizando un equilibrio entre eficiencia, transparencia y respeto a los derechos fundamentales. La regulación adecuada, el monitoreo de sesgos y la supervisión humana son elementos esenciales para asegurar que la IA se convierta en un aliado para la justicia, en lugar de un factor que contribuya a la inequidad y la deshumanización de los procesos penales.

Ejemplos concretos de IA en el derecho penal

Las aplicaciones prácticas de la inteligencia artificial en el ámbito del derecho penal son un claro ejemplo de las ventajas y de los riesgos derivados de la misma:

- *COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)*: Herramienta de estimación de riesgos que se utiliza en los Estados Unidos para pronosticar la probabilidad de que una persona vuelva a delinquir. Aunque su uso es ampliamente común, ha sido fuertemente criticada por la falta de transparencia de su algoritmo, lo que dificulta la evaluación de la imparcialidad o de su buena precisión, y por el riesgo de que puede hacer formada porque reproduce sesgos con determinadas poblaciones (Serventich, 2022).
- *Sistemas de reconocimiento facial y análisis biométrico*: Estas tecnologías se utilizan en la identificación de sospechosos en las investigaciones criminales, aunque el grado de acierto del reconocimiento facial y la utilización ética de los datos biométricos suscitan dudas, teniendo en cuenta el riesgo de violación de la privacidad y de discriminar en función de la raza y el sexo a partir de los algoritmos de reconocimiento facial (Acquaviva, 2023; Mamak, 2024).

Discusión de Resultados

La incorporación de la Inteligencia Artificial (IA) al derecho penal es un avance importante por cuanto los sistemas judiciales poseen una capacidad mayor de gestionar, analizar y predecir decisiones a partir de datos. Esta revisión refleja el papel de la IA para optimizar procesos como la evaluación de riesgos, la determinación de la reincidencia o el análisis de patrones de conducta (Kleinberg et al., 2018). A medida que se amplían las aplicaciones de la IA en el derecho penal, surgen también retos éticos, funcionales y legales, lo que tensiona principios jurídicos fundamentales como la presunción de inocencia en sentido estricto, el derecho a un debido proceso y la imparcialidad de la administración de justicia. La opacidad que poseen muchos algoritmos, como la posibilidad de sesgos en su programación o en la recogida de datos de entrenamiento, permiten aumentar las preocupaciones respecto a la transparencia y justicia de tales decisiones automatizadas (Angwin et al., 2016; Mustafa, 2024).

Uno de los aspectos que se destacan en esta revisión es la tensión que ofrece la IA frente a la eficiencia y el riesgo que conlleva. Por una parte, la automatización de las tareas del sistema penal permite a jueces y fiscales gestionar los casos de menor complejidad y así optimizar el tiempo para las decisiones de mayor calado. No obstante, esta eficiencia también viene acompañada de la posibilidad de una despersonalización de las decisiones judiciales, lo cual plantea interrogantes acerca de hasta qué punto jueces y operadores humanos deben seguir manteniendo un papel y un

control sobre las herramientas que se utilizan (Buján, 2023). El dilema de la asistencia y la delegación de decisiones a la IA es el eje transversal en este ámbito. Algunos sistemas actúan únicamente como un soporte para el juez, mientras que otros son utilizados para ofrecer recomendaciones autónomas que en algunos casos pueden condicionar de forma importante el resultado del proceso judicial.

La cuestión en torno a la responsabilidad penal de la IA también es relevante. A medida que los sistemas adquieren eficacia cada vez mayor, algunos autores postulan que sería el caso de preguntarse si la IA podría asumir una responsabilidad penal similar a la de los agentes humanos. Sin embargo, aunque la expresión "intención maliciosa" en un sistema de IA pueda parecer indefinida, la autonomía en las opciones de una toma de decisiones automatizada alimenta este debate. Por su parte, la discusión sobre de quién es la responsabilidad de errar en los casos de errores algorítmicos –si del creador del algoritmo, del usuario o del propio algoritmo– pone en evidencia la acuciante necesidad de un marco normativo que conteste a esta cuestión (Mahardhika et al., 2023; Zhaoxun et al., 2024).

Un reto ético particularmente pertinente es el posible mantenimiento o ampliación de sesgos preexistentes en los sistemas de la justicia. Los algoritmos de IA son entrenados con datos históricos que pueden resultar estar sesgados, lo que puede acontecer en decisiones que obedecen al sesgo en función de raza, género o nivel económico. Este problema se evidenció con la llegada del sistema COMPAS de evaluación del riesgo de reincidencia de los acusados en EE. UU. Por su parte, las críticas que se han suscitado sobre COMPAS –debido a su opacidad, así como al hecho de que se le acusara de discriminación hacia determinados grupos– ilustran el riesgo de que estas tecnologías mantengan desigualdades en lugar de justicia (Serventich, 2022). Por ello, los desarrolladores de IA y los legisladores deberían preparar el terreno para ver de qué forma es posible eliminar o mitigar los sesgos gracias a la preocupación de estándares de transparencia, así como de ética.

También es cierto que la admisibilidad de la evidencia producida por IA suscita difíciles cuestiones relativas a la validez y la fiabilidad de dicha evidencia en el seno de un tribunal. La opacidad algorítmica o, en otros términos, la "caja negra" de la IA complica la comprobabilidad y la discusión de las decisiones relevantes en los actos judiciales. Para poder utilizar la IA correctamente en la esfera del derecho penal, es necesario que los sistemas sea explicables y que sus procedimientos puedan revisarse. Sin embargo, el ámbito tecnológico actual no ha llegado a

despejar completamente este umbral, lo que hace que existan limitaciones a su admisión en los procedimientos judiciales formales (Sayyed, 2024; Wu et al., 2022).

De otro lado, las oportunidades de descongestión judicial y de ganar en eficiencia y en justicia que aporta la IA son irrefutables sólo si se someten a regulaciones claras que limiten estos riesgos y que permitan un uso ético de la tecnología. La IA tiene el potencial de transformar la administración de justicia penal posibilitando interiorizar la carga procesal y optimizar la utilización de recursos. Sin embargo, para que esta promesa se realice es necesaria una infraestructura regulatoria que se integre en estándares éticos y protocolos de supervisión humana que garanticen que el sistema judicial continúe con su papel de custodio de los derechos humanos y evite la delegación de decisiones en herramientas opacas o irresponsables (DoCarmo et al., 2022; Granados, 2024).

Esta revisión también pone de manifiesto la ausencia de investigación empírica sobre los efectos reales de la IA en la equidad y en la eficacia de los procedimientos judiciales. Esto es así porque aunque se han progresado en avances teóricos son escasos los estudios empíricos que midan los efectos de estas tecnologías en el día a día de los trámites judiciales. Evaluaciones longitudinales sobre la equidad, la eficacia y la percepción pública de la IA en contextos judiciales serán elementos determinantes para entender y mitigar posibles desventajas que se vayan planteando e implementar las directrices del derecho tradicional a situaciones enmarcadas por la IA.

Conclusiones

El auge de la Inteligencia Artificial en el área del derecho penal produce un desarrollo considerable, con un impacto real en la eficacia o en la calidad de los procedimientos penales. Las finalidades que se proponen con herramientas como los sistemas expertos, el machine learning o las herramientas para valorar los riesgos, han demostrado su capacidad para ayudar a la toma de decisiones, a reducir los tiempos procesales, asistir en las valoraciones del riesgo de volver a delinquir, etc. No obstante, los peligros a los que puede dar lugar la opacidad algorítmica o el sesgo inherente, pero también algunas violaciones de principios jurídicos fundamentales parecen suponer una serie de cuestiones éticas y legales que requieren poder responder urgentemente.

Esta revisión insiste en la importancia de la exigencia de poner en práctica instrumentos normativos que regulen el uso de estos trabajos tecnológicos respetando los derechos humanos sin perder la objetividad y la independencia de la actuación. En este sentido, una supervisión adecuada y el

cumplimiento de salvaguardas van a ser elementales para poder hacer un buen uso del desarrollo tecnológico sin poner en peligro la integridad del sistema judicial.

Desde un nivel teórico, esta reflexión añade literatura al reconocer la IA en el ámbito del derecho penal a partir de una visión global, de una reflexión que no haga solo hincapié en la capacidad técnica, sino que también incluya la ética y la legalidad de los usos de esta tecnología. Desde un nivel teórico, estas conclusiones vuelven a insistir en la necesidad de poder ofrecer estas estructuras teóricas que hacen cara no solo a una reflexión ética, sino a la crisis que se plantea por la implantación de la IA en derecho penal. Además, la reflexión abierta a una investigación futura que pueda poner de manifiesto de una manera más pormenorizada cuáles son las crisis que podría llegar a significar la IA, de forma especial desde la perspectiva de la opacidad de la IA, la responsabilidad penal, el sesgo algorítmico, etc. También pone de manifiesto que la integración en el estudio de la IA debe recoger diferentes enfoques y composiciones, ya que la formalización de teorías sobre la IA aplicada al derecho penal, sólo se puede construir con el apoyo de la formalización de una ciencia combinada (informática, ética etc.).

Los resultados de este trabajo de revisión a nivel práctico sugieren recomendaciones dirigidas a legisladores, jueces y profesionales del derecho. Para los legisladores, existe la necesidad de adoptar políticas públicas y de crear regulaciones que guíen el uso de la IA en procesos judiciales, a la vez que se promueve la transparencia algorítmica y se externalizan estándares de validez y fiabilidad que deben cumplir las pruebas generadas por IA. Respecto de jueces y de los operativos con el Sistema Jurídico, resultan importantes establecer prácticas que garanticen la supervisión del uso de dichas herramientas por parte de la capacidad humana y el conocimiento de las limitaciones y capacidades de la IA. En consecuencia, se debe incluir la formación práctica en el uso de estas tecnologías de modo que dicha formación permita que los profesionales del derecho sepan que uso pueden hacer de la IA en los procedimientos. Estas implicancias prácticas tienen un solo objetivo: conducir a una aplicación ética y efectiva de la IA en el derecho penal de modo que la misma contribuya al fortalecimiento del sistema judicial y no a la erosión del mismo.

Futuras Investigaciones

La incorporación de la IA en el ámbito penal demuestra la existencia de infinidad de caminos que necesitan investigación para llevar a cabo una implementación ética y conveniente. La optimización de algoritmos constituye uno de los temas de futuro. Resultan necesarios estudios que construyan y evalúen procedimientos para hacer que los algoritmos sean explicables y

comprensibles por jueces, abogados y partes intervinientes de un proceso judicial. Explorar cómo implementar tecnologías de IA que proporcionen un elevado grado de interpretabilidad así como de justificación de las decisiones, contribuirá a generar confianza del público y a promover la equidad en su aplicación.

El empírico sesgo algorítmico constituye otro de los temas relevantes. La evidencia en cuanto a la posible reproducción y extensión de la IA sobre sesgos discriminatorios en el sistema penal es todavía escasa, por cuanto que los futuros estudios empíricos pueden ser orientados a cómo diferentes algoritmos impactan a las distintas poblaciones y a cómo los datos de entrenamiento pueden ser alterados para ajustarse a éstos. La investigación también contribuiría a poder identificar y aplicar prácticas que eviten la discriminación en el sistema de justicia penal, garantizando que los principios de equidad y justicia se cumplan en todos los casos.

Referencias

1. Acquaviva, G. (2023). Crimes without Humanity? Artificial Intelligence, Meaningful Human Control, and International Criminal Law. *Journal of International Criminal Justice*, 21(5), 981-1004.
2. Al-Dulaimi, A. O. M., & Mohammed, M. A. A. W. (2025). Legal responsibility for errors caused by artificial intelligence (AI) in the public sector. *International Journal of Law and Management*.
3. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
4. Barocas, S., & Selbst, A. D. (2016). Big Data's disparate impact. *California Law Review*, 104(3), 671-732. <https://doi.org/10.15779/Z38BG31>
5. Bagaric, M., Svilar, J., Bull, M., Hunter, D., & Stobbs, N. (2022). The solution to the pervasive bias and discrimination in the criminal justice system: transparent and fair artificial intelligence. *Am. Crim. L. Rev.*, 59, 95.
6. Buján, M. V. Á. (2023). Inteligencia artificial y medidas cautelares en el proceso penal: Tutela judicial efectiva y autodeterminación informativa en potencial riesgo. *Revista española de derecho constitucional*, 43(127), 177-207.

7. Caterini, M. (2022). El sistema penal en la encrucijada ante el reto de la inteligencia artificial. *IDP: revista de Internet, derecho y política= revista d'Internet, dret i política*, (35), 5.
8. Dharneesh, B., Thenisha, S., Srithick, S. S., & Abirami, A. (2024). Why Law Firms Need to Embrace Artificial Intelligence to Transform the Indian Legal Industry. *Reshaping Intelligent Business and Industry: Convergence of AI and IoT at the Cutting Edge*, 81-95.
9. DoCarmo, T., Rea, S., Conaway, E., Emery, J., & Raval, N. (2021). The law in computation: What machine learning, artificial intelligence, and big data mean for law and society scholarship. *Law & Policy*, 43(2), 170-199.
10. Dremluiga, R., & Korobeev, A.I. (2023). Criminal Infringement on Artificial Intelligence Systems: A Criminal Law Description. *Russian Journal of Criminology*.
11. Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
12. Granados, J. S. A. P. (2024). El proceso penal mediado por inteligencia artificial. *Revista Brasileira de Direito Processual Penal*, 10, e988.
13. Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs\rand methods art:10.1186/2046-4053-1-28. *Systematic Reviews*, 1–9. <https://doi.org/10.1186/2046-4053-1-28>
14. Hildebrandt, M. (2020). *Law for computer scientists and other folk*. Oxford University Press. <https://doi.org/10.1093/oso/9780198840353.001.0001>
15. Kaur, M., & Saini, M. (2024). Role of Artificial Intelligence in the crime prediction and pattern analysis studies published over the last decade: a scientometric analysis. *Artificial Intelligence Review*, 57(8), 202.
16. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2018). Inherent trade-offs in the fair determination of risk scores. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (pp. 40-47). ACM. <https://doi.org/10.1145/3219166.3219177>
17. Legg, M., & Bell, F. (2020). *Artificial intelligence and the legal profession*. Bloomsbury Publishing.
18. Mahardhika, V., Astuti, P., & Mustaffa, A. (2023). Could Artificial Intelligence be the Subject of Criminal Law?. *Yustisia*, 12(1), 1-12.

19. Mamak, K. (2024). AGI crimes? The role of criminal law in mitigating existential risks posed by artificial general intelligence. *AI & SOCIETY*, 1-11.
20. Mustafa, A. (2024). The Future of Criminal Law in the Age of Electronic Crimes and Artificial Intelligence. *Pakistan Journal of Life and Social Sciences*, 22(2), 9519-9535
21. Rodger, H., Lensen, A., & Betkier, M. (2023). Explainable artificial intelligence for assault sentence prediction in New Zealand. *Journal of the Royal Society of New Zealand*, 53(1), 133-147.
22. Ryberg, J. (2024). Criminal Sentencing and Artificial Intelligence: What is the Input Problem?. *Criminal Law and Philosophy*, 1-18.
23. Sayyed, H. (2024). Artificial intelligence and criminal liability in India: exploring legal implications and challenges. *Cogent Social Sciences*, 10(1), 2343195.
24. Serventich, C. (2022). Inteligencia artificial en el proceso penal ¿Más vale humano conocido o algoritmo por conocer?.
25. Shi, J. (2022, June). Artificial intelligence, algorithms and sentencing in Chinese criminal justice: problems and solutions. In *Criminal Law Forum* (Vol. 33, No. 2, pp. 121-148). Dordrecht: Springer Netherlands.
26. Suleimanova, S. (2024). Comparative Legal Analysis of the Role of Artificial Intelligence in Human Rights Protection: Prospects for Europe and the Middle East. *Pakistan Journal of Criminology*, 16(3).
27. Wu, J., Wang, H., Sun, N., Wang, H., & Tatarinov, D. (2022). International Criminal Law Protection of Environmental Rights and Sentencing Based on Artificial Intelligence. *Journal of Environmental and Public Health*, 2022(1), 4064135.
28. Zea, M. P. Á., & Vázquez, A. F. Z. (2024). Aplicación de la inteligencia artificial en el Derecho Penal: problemas y desafíos.
29. Zhang, Y., Zhao, Y., & Zhao, Y. (2022). The Application of Artificial Intelligence Decision-Making Algorithm in Crisis Analysis and Optimization of the International Court System. *Mobile Information Systems*, 2022(1), 8150122.
30. Zhaoxun, C., Rajamanickam, R., & Dahlan, N. K. (2024). Exploring The Legal Subjectivity of Artificial Intelligence in Incitement to Suicide. *Jurnal IUS Kajian Hukum dan Keadilan*, 12(1), 31-42.

© 2025 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).