AulaVirtual
Generando Conocimiento

ISSN: 2665-0398

INTELIGENCIA ARTIFICIAL Y PERCEPCIÓN DE CULPABILIDAD EN DELITOS INFORMÁTICOS: UNA REVISIÓN SISTEMÁTICA DE SUS IMPLICANCIAS JUDICIALES

ARTIFICIAL INTELLIGENCE AND THE PERCEPTION OF GUILT IN CYBERCRIMES: A SYSTEMATIC REVIEW OF ITS LEGAL IMPLICATIONS

Tipo de Publicación: Articulo Científico
Recibido: 21/01/2026
Aceptado: 22/02/2026
Publicado: 30/03/2026
Código Único AV: e668
Páginas: 1(538-557)
DOI: https://doi.org/10.5281/zenodo.19334270

Autor:
Liz Tereza Palacios Esteban
Abogada
Magister en Derecho Penal
https://orcid.org/0009-0007-6863-7269
E-mail: lpalaciosest@ucvvirtual.edu.pe
Afiliación: Universidad César Vallejo
País: República del Perú

Resumen

La incorporación progresiva de herramientas de inteligencia artificial (IA) en los sistemas judiciales, especialmente en el abordaje de delitos informáticos, ha generado transformaciones sustantivas en la forma en que se produce, valora y decide la evidencia digital, planteando desafíos relevantes para la percepción de culpabilidad y la legitimidad del juicio penal. En este contexto, el objetivo de este artículo es analizar cómo la implementación de herramientas de IA afecta la percepción de culpabilidad en delitos informáticos dentro de los procesos judiciales, considerando sus efectos cognitivos, probatorios, éticos y jurídicos. Para ello, se desarrolló un artículo de revisión sistemática aplicando el método PRISMA, siguiendo un enfoque metodológico riguroso que permitió identificar, seleccionar y analizar estudios recientes publicados en revistas científicas indexadas, con énfasis en investigaciones empíricas, experimentales y normativas vinculadas al uso de IA en la justicia penal. Los resultados evidencian que la IA influye significativamente en la percepción de culpabilidad mediante mecanismos como la automatización decisional, el anclaje algorítmico, la deferencia a la autoridad técnica y la presencia de sesgos derivados de los datos y del diseño de los sistemas. Asimismo, se identificó que la falta de explicabilidad y control humano incrementa el riesgo de distorsionar principios fundamentales como la presunción de inocencia. En conclusión, el estudio demuestra que la IA no constituye una herramienta neutral en el ámbito judicial, por lo que su implementación exige marcos regulatorios, explicabilidad algorítmica y supervisión humana efectiva para garantizar una justicia penal compatible con el Estado de derecho.

Palabras Clave Inteligencia artificial, percepción de culpabilidad, delitos informáticos, sesgo algorítmico, procesos judiciales

Abstract

The progressive incorporation of artificial intelligence (AI) tools into judicial systems, especially in addressing cybercrimes, has generated substantial transformations in how digital evidence is produced, assessed, and decided, posing significant challenges to the perception of guilt and the legitimacy of criminal trials. In this context, the objective of this article is to analyze how the implementation of AI tools affects the perception of guilt in cybercrimes within judicial processes, considering their cognitive, evidentiary, ethical, and legal effects. To this end, a systematic review article was developed applying the PRISMA method, following a rigorous methodological approach that allowed for the identification, selection, and analysis of recent studies published in indexed scientific journals, with an emphasis on empirical, experimental, and normative research related to the use of AI in criminal justice. The results show that AI significantly influences the perception of guilt through mechanisms such as automated decision-making, algorithmic anchoring, deference to technical authority, and the presence of biases stemming from data and system design. Furthermore, the study identified that the lack of explainability and human oversight increases the risk of distorting fundamental principles such as the presumption of innocence. In conclusion, the study demonstrates that AI is not a neutral tool in the judicial sphere, and therefore its implementation requires regulatory frameworks, algorithmic explainability, and effective human supervision to guarantee criminal justice compatible with the rule of law.

Keywords Artificial intelligence, perception of guilt, cybercrime, algorithmic bias, judicial processes

Introducción

La intersección entre la Inteligencia Artificial (IA) y el derecho ha suscitado un creciente interés académico, especialmente en el ámbito de los delitos informáticos. A medida que las tecnologías digitales avanzan y se integran progresivamente en los procesos judiciales, emergen interrogantes fundamentales sobre el impacto de estas herramientas en la percepción de culpabilidad dentro de los escenarios jurídicos. La IA, particularmente mediante sus aplicaciones en el análisis de datos y en los sistemas de apoyo a la toma de decisiones automatizadas, ha sido concebida para optimizar la eficiencia del sistema de justicia; sin embargo, su incidencia en la subjetividad asociada a la atribución de culpabilidad y en la equidad de los juicios ha generado un debate académico cada vez más intenso (Kirsanova et al., 2021).

Uno de los desafíos más relevantes radica en comprender de qué manera la implementación de la IA puede modificar la percepción de culpabilidad en los delitos cibernéticos. Esta cuestión adquiere especial relevancia en un contexto en el que la evidencia digital desempeña un rol central en los procesos judiciales y en el que los sistemas basados en IA pueden influir en la valoración de dicha evidencia (Hani et al., 2024). Diversos estudios han señalado que, conforme se incorporan herramientas de IA en los procesos de decisión legal, estas pueden incidir de manera significativa en la forma en que se interpreta y determina la culpabilidad, lo que plantea importantes preocupaciones de carácter ético y jurídico (Zeadally et al., 2020).

Asimismo, resulta imprescindible reconocer que la aplicación de la IA en el ámbito judicial no se encuentra exenta de riesgos. El desarrollo de algoritmos con capacidad para influir en decisiones relacionadas con la culpabilidad plantea interrogantes sustanciales sobre la transparencia, la trazabilidad y la responsabilidad institucional (Kerimkhulle et al., 2023). Los sistemas de justicia deben garantizar no solo la eficiencia procesal, sino también la justicia material y la equidad, lo que exige un análisis crítico de los efectos que la automatización puede tener sobre los principios jurídicos fundamentales. Este escenario se ve agravado por la creciente complejidad de los delitos informáticos, cuya naturaleza digital dificulta su rastreo, investigación y procesamiento. En consecuencia, resulta fundamental examinar de qué manera estas tecnologías no solo asisten a los operadores y las operadoras del sistema judicial, sino también cómo podrían, de forma no intencional, afectar la equidad en la atribución de culpabilidad.

En los últimos años, han surgido diversas investigaciones orientadas a analizar la relación entre la IA y la percepción de culpabilidad en los delitos informáticos dentro de los procesos judiciales. A continuación, se destacan algunos

ISSN: 2665-0398

estudios relevantes que han contribuido de manera significativa al desarrollo del conocimiento en esta área.

En primer lugar, el estudio de Shaligar et al. (2024) examina las aplicaciones de la IA en la generación, gestión y utilización de documentos digitales y evidencia electrónica en litigios civiles y penales. Los autores señalan que estas herramientas pueden fortalecer la transparencia y la eficacia en la presentación probatoria, lo que podría incidir en la percepción de culpabilidad al aportar evidencias más consistentes y fiables en sede judicial.

De igual manera, Pristanskov et al., (2023) analizan el rol de la IA como un enfoque innovador para la aplicación de conocimientos especializados en la investigación y resolución de delitos cibernéticos. En su estudio, destacan que la adopción de algoritmos basados en IA no solo puede acelerar los procesos investigativos, sino también transformar la manera en que se construyen y perciben la culpa y la responsabilidad penal, aportando una perspectiva clave para comprender su impacto en la justicia penal contemporánea.

Por su parte, Franguloiu (2024) propone una serie de principios orientadores para la utilización de la IA en el poder judicial, a partir de las consideraciones éticas y jurídicas contenidas en la Carta Europea de Ética. Este trabajo resulta especialmente relevante en relación con la percepción de culpabilidad, al subrayar la necesidad de que el uso de la IA en las decisiones judiciales se sustente en principios éticos sólidos y en el respeto de los valores fundamentales del derecho.

Finalmente, el estudio de Sessa et al., (2024) analiza las aplicaciones actuales de la IA en el ámbito de la genética forense, ofreciendo ejemplos concretos de cómo estas tecnologías han transformado la resolución de casos penales. Si bien su enfoque se centra en el ámbito forense, sus hallazgos tienen implicancias relevantes para la percepción de culpabilidad, en la medida en que la precisión y eficacia de los sistemas de IA pueden influir en la vinculación entre personas imputadas y hechos delictivos.

En conjunto, estas investigaciones evidencian avances significativos en la comprensión de cómo la IA está reconfigurando el análisis de la culpabilidad en los delitos informáticos, proporcionando un marco contextual indispensable para el desarrollo del presente artículo de revisión sistemática. A medida que el campo continúa expandiéndose, surgen nuevos interrogantes y desafíos vinculados a los límites éticos y jurídicos del uso de la IA en el sistema de justicia penal, consolidándose como un ámbito de estudio prioritario y de alta actualidad en la criminología contemporánea.

La creciente implementación de herramientas de IA en el ámbito judicial, particularmente en el contexto de los delitos informáticos, ha intensificado los debates en torno a sus

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

repercusiones sobre la percepción de culpabilidad. Si bien existe un volumen considerable de literatura relacionada con la ciberseguridad y la aplicación de la IA, se identifican vacíos relevantes que justifican la realización de un análisis sistemático en este campo.

Un primer vacío se vincula con la escasez de estudios que examinen de manera específica cómo las herramientas de IA pueden modificar las percepciones de culpabilidad. La mayoría de las investigaciones se ha concentrado en las capacidades técnicas de estos sistemas para el procesamiento de datos, mientras que son limitados los trabajos que analizan su influencia en la interpretación de la evidencia y, en consecuencia, en la determinación de la culpabilidad dentro del proceso judicial (Hahn et al., 2022). Una evaluación integral de este impacto resulta indispensable para comprender las implicancias de la automatización en los juicios penales y para prevenir decisiones sesgadas derivadas del uso acrítico de la tecnología.

Un segundo vacío corresponde a la insuficiente problematización de las implicancias éticas y normativas asociadas al uso de la IA en el sistema judicial. La incorporación de algoritmos puede introducir sesgos que afecten la imparcialidad judicial, generando desigualdades en la percepción de culpabilidad (Shakeel et al., 2025). Aunque el sesgo algorítmico ha sido ampliamente estudiado en otros ámbitos, su incidencia específica en la justicia penal continúa siendo un campo poco desarrollado.

El tercer vacío se relaciona con la necesidad de adoptar un enfoque sistemático que articule la investigación empírica con la teoría jurídica. Si bien algunas investigaciones sugieren que la IA puede transformar tanto la recopilación y el análisis de datos como los procesos de atribución de culpabilidad, aún se carece de un marco teórico integrador que permita orientar de manera coherente la aplicación de estas tecnologías en los procesos judiciales (Akinrinola et al., 2024).

Estos vacíos fundamentan el objetivo está orientado a analizar cómo la implementación de herramientas de IA afecta la percepción de culpabilidad en los delitos informáticos dentro de los procesos judiciales.

Metodología

La revisión sistemática se desarrolló mediante el método PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Para ello, se utilizó una fórmula booleana. La fórmula es la siguiente: ("artificial intelligence" OR AI) AND ("perception of culpability" OR "guilt perception") AND ("cybercrime" OR "cyber offenses") AND ("judicial processes" OR "legal processes").

Esta estrategia de búsqueda se aplicará en la base de datos Scopus, que es reconocida por su vasta colección de literatura científica y su cobertura de estudios relevantes. Para guiar la revisión, se

ISSN: 2665-0398

formulan cinco preguntas de investigación que orientarán el análisis sistemático:

1. ¿Qué impacto tiene la implementación de herramientas de IA en la percepción de culpabilidad en procesos judiciales de delitos informáticos?
2. ¿Cómo se presentan los sesgos algorítmicos en la evaluación de culpabilidad en el contexto de la IA?
3. ¿Cuáles son las implicaciones éticas y legales del uso de IA en juicios relacionados con delitos cibernéticos?
4. ¿De qué manera influyen las percepciones sociales sobre la IA en la interpretación de culpabilidad en el ámbito judicial?
5. ¿Qué estrategias se pueden implementar para mitigar el sesgo y mejorar la equidad en la aplicación de herramientas de IA en los procesos judiciales?

Los criterios de inclusión de la presente revisión sistemática se establecerán de la siguiente manera:

1. Artículos científicos sometidos a revisión por pares.
2. Publicaciones que aborden de forma explícita la relación entre la IA y la percepción de culpabilidad en el contexto de los delitos informáticos.
3. Estudios que analicen la aplicación de dichas tecnologías en los procesos judiciales. Se priorizarán aquellos trabajos que presenten hallazgos empíricos o análisis críticos relacionados con la implementación y los efectos de las herramientas de IA en el ámbito jurídico.

De manera complementaria, se definirán criterios de exclusión con el propósito de asegurar la calidad metodológica y la pertinencia temática de la revisión. Estos criterios comprenderán:

1. Estudios que no guarden una relación directa con la IA y la atribución de culpabilidad.
2. Investigaciones que no aporten evidencia sustantiva o que se limiten a desarrollos puramente conceptuales sin aplicación práctica.
3. Materiales de divulgación, comentarios u opiniones que no hayan sido sometidos a un proceso de evaluación por pares.
4. Trabajos que carezcan de datos relevantes o que no se circunscriban al ámbito judicial.

La aplicación de estos criterios metodológicos proporcionará un marco sólido para alcanzar el objetivo de analizar cómo la implementación de herramientas de IA influye en la percepción de culpabilidad en los delitos informáticos dentro de los procesos judiciales. De este modo, se contribuirá a una comprensión integral de este fenómeno y de sus implicancias en el sistema de justicia penal.

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

ISSN: 2665-0398

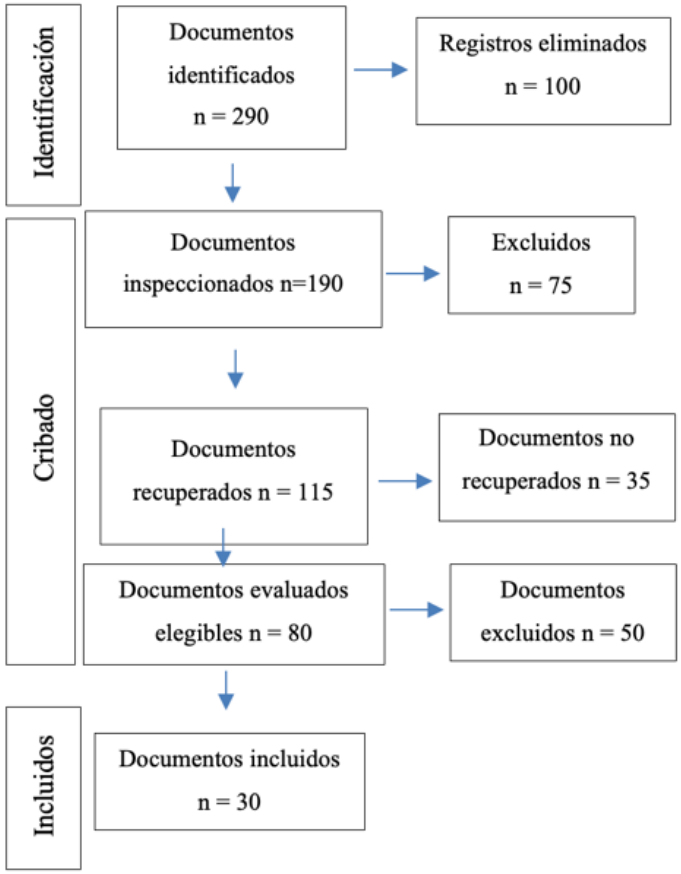

Figura 1. Identificación de estudios que utilizan el método prismático

ISSN: 2665-0398

Resultados

Autor	Herramienta IA analizada	Población / Muestra	Dimensión relevante para culpabilidad	Principales hallazgos	Contribución a la pregunta de investigación
Reedy (2023)	IA aplicada a evidencia digital forense	Casos judiciales y peritos forenses	Sesgo cognitivo y credibilidad probatoria	La IA puede reproducir sesgos humanos y generar dificultades explicativas ante el juez, afectando la valoración de la prueba digital	Evidencia que la IA influye indirectamente en la percepción de culpabilidad al reforzar la autoridad técnica de la prueba digital
Javed & Li (2025)	Sistemas de IA en adjudicación judicial	Datos judiciales (2019–2023)	Reducción / reproducción del sesgo judicial	La IA reduce sesgos en el mediano y largo plazo, pero en el corto plazo puede generar efectos ambiguos	Aporta evidencia empírica sobre cómo la IA puede modificar la percepción de culpabilidad según horizonte temporal
Fazel et al., (2022)	OxRec (algoritmo de evaluación de riesgo)	Personas procesadas penalmente	Influencia algorítmica en decisiones judiciales	Los jueces tienden a confiar en puntajes de riesgo, lo que puede afectar la severidad percibida de la conducta	Evidencia que la IA puede intensificar la percepción de peligrosidad y culpabilidad en delitos tecnológicos
Hefetz (2025)	IA en análisis forense	Expertos forenses y sistemas judiciales	Sesgo humano–IA y deferencia algorítmica	La subordinación al output algorítmico debilita el juicio crítico humano	Demuestra cómo la IA puede reforzar percepciones de culpabilidad por automatización cognitiva
Olawade et al., (2025)	IA en salud mental forense	Evaluaciones forenses	Riesgo, peligrosidad y responsabilidad penal	Los algoritmos influyen en evaluaciones que inciden directamente en decisiones judiciales	Vincula IA con construcción técnica de culpabilidad y riesgo penal
Giroux et al., (2022)	Sistemas automatizados de decisión	Consumidores (experimentos controlados)	Culpa moral y deshumanización	La interacción con IA reduce la percepción de culpa	Aporta base teórica extrapolable a jueces que interactúan con IA en procesos penales
Ghasemaghaei & Kordzadeh (2024)	Algoritmos de recomendación	Decisores humanos	Disminución de culpa por obediencia algorítmica	Las decisiones discriminatorias inducidas por IA no generan culpa percibida	Evidencia directa de desplazamiento de responsabilidad que afecta la atribución de culpabilidad
Agudo et al., (2024)	IA en procesos judiciales (human-in-the-loop)	Participantes evaluando acusados	Automatización y sesgo decisional	La IA incorrecta reduce la precisión del juicio humano	Demuestra que la IA puede distorsionar la valoración de culpabilidad cuando se introduce antes del juicio humano

Tabla 1. Impacto en la implementación de las herramientas de IA

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

Autor	Sistema o IA analizada	Tipo de sesgo identificado	Dimensión de culpabilidad	Principales hallazgos	Impacto en percepción de culpabilidad
Wu & Lin (2025)	Sistema "206" (IA judicial)	Anclaje algorítmico, evitación de responsabilidad	Evaluación judicial y fiscal	La IA refuerza decisiones previas y reduce participación del acusado	Incrementa percepción de culpabilidad por anclaje
Thamizh Mani et al., (2025)	ML en lingüística forense	Sesgo de datos, opacidad algorítmica	Atribución de autoría	ML supera a análisis humano, pero introduce sesgos no interpretables	Riesgo de sobrevaloración probatoria
Agudo et al., (2024)	IA de apoyo decisional	Automation bias	Juicio de culpabilidad	La IA errónea reduce precisión humana si se presenta antes del juicio	Distorsiona culpabilidad percibida
Chuan et al., (2024)	XAI – explanation by example	Sesgo por datasets no inclusivos	Percepción de justicia/fairness	XAI aumenta conciencia de sesgo, pero puede inducir confianza indebida	Modula percepción de culpabilidad
Barsekh-Onji et al., (2025)	IA en gobernanza pública	Sesgo estructural y de entrenamiento	Decisión administrativa	Falta de XAI afecta equidad y control humano	Impacto indirecto en culpabilidad institucional

Tabla 2. Sesgos algorítmicos en la evaluación

Autor	IA / técnica empleada	Datos / muestra	Hallazgos relevantes sobre culpabilidad / decisión judicial	Implicaciones éticas y legales identificadas
Kong et al., (2025)	Grafo heterogéneo de conocimiento jurídico; aprendizaje iterativo para predicción de decisiones penales	Dataset CAIL 2018 con más de 1.5 millones de casos penales	La IA mejora la predicción de resultados judiciales y la correspondencia entre hechos y normas, lo que puede influir indirectamente en la anticipación de culpabilidad	Riesgos asociados a la automatización del razonamiento judicial y necesidad de explicabilidad para evitar decisiones opacas que afecten el debido proceso
Peng & Lei (2024)	Modelo BERT aplicado a textos judiciales para predicción de cargos y penas	218,120 sentencias penales; submuestras específicas para duración de pena	Alta precisión en la predicción de cargos y sentencias, lo que puede reforzar percepciones previas de culpabilidad antes del juicio	Potencial afectación a la presunción de inocencia si las predicciones algorítmicas se utilizan como referencia decisoria
Han et al., (2024)	Sistema LegalAsst con representaciones estructuradas y árboles de decisión explicables	Casos judiciales estructurados para asistencia a jueces	La trazabilidad del razonamiento algorítmico mejora la comprensión del proceso de determinación de culpabilidad	Refuerza principios de transparencia, explicabilidad y control humano como garantías esenciales del debido proceso
Herrera-Tapias & Hernández Guzmán (2025)	IA generativa (ChatGPT) aplicada como apoyo en la motivación judicial	Análisis de sentencias constitucionales y uso real de IA por jueces	El uso no controlado de IA puede introducir errores que afecten la valoración de culpabilidad	Se enfatiza la responsabilidad exclusiva del juez humano, la motivación suficiente de las resoluciones y la protección del debido proceso

ISSN: 2665-0398

Autor	IA / técnica empleada	Datos / muestra	Hallazgos relevantes sobre culpabilidad / decisión judicial	Implicaciones éticas y legales identificadas
Hefetz (2023)	Algoritmos de aprendizaje automático en ciencias forenses	Evidencia forense analizada mediante sistemas automatizados	La confianza excesiva en resultados algorítmicos puede condicionar la percepción de culpabilidad	Riesgo de violación del derecho de defensa, impugnación probatoria y admisibilidad de evidencia algorítmica
Wang & Ma (2022)	Support Vector Machine y Random Forest para predicción delictiva	Casos delictivos y encuesta a 238 participantes	La predicción de conductas delictivas puede influir en valoraciones anticipadas de culpabilidad	Problemas de estigmatización, sesgos y legitimidad normativa en el uso de IA predictiva
Naik et al., (2022)	Modelos generales de IA y aprendizaje automático	Revisión conceptual de aplicaciones de IA	Identifica riesgos transversales en decisiones automatizadas que afectan derechos fundamentales	Problemas de privacidad, sesgo algorítmico, ciberseguridad y ausencia de marcos regulatorios claros

Tabla 3. Implicancias éticas y legales del uso de la IA

Autor	Tipo de IA analizada	Ámbito judicial / decisional	Dimensión de percepción social	Hallazgos relevantes sobre culpabilidad	Implicaciones éticas y legales
Engle et al., (2025)	Tecnologías digitales asociadas al cibercrimen	Evaluación social de delitos	Percepción de gravedad del delito	Los delitos informáticos son percibidos como menos graves que los delitos físicos, lo que atenúa la atribución de culpabilidad	Riesgo de subvaloración del daño y de respuestas penales desproporcionadas
Herrera-Tapias & Hernández Guzmán (2025)	IA generativa (GPT, LLMs)	Decisión judicial constitucional	Confianza y legitimidad institucional	La percepción de culpabilidad solo se considera válida cuando la IA cumple un rol auxiliar y no decisorio	Salvaguarda del debido proceso y exigencia de transparencia algorítmica
Contini et al., (2024)	Sistemas predictivos e IA generativa	Deliberación penal	Emoción, empatía y cognición judicial	La IA reduce la complejidad emotivo-cognitiva necesaria para interpretar la culpabilidad	Riesgo de deshumanización del juicio penal
Javed & Li (2025)	IA analítica y predictiva	Adjudicación judicial	Percepción de sesgo y justicia	La IA puede reducir sesgos estructurales a largo plazo, aunque genera desconfianza inicial en la valoración de culpabilidad	Necesidad de regulación para prevenir sesgos algorítmicos
Agudo et al., (2024)	Sistemas automatizados con *human-in-the-loop*	Juicios experimentales sobre culpabilidad	Automatización y obediencia cognitiva	El apoyo algorítmico erróneo incrementa errores en la atribución de culpabilidad	Riesgo de automatización acrítica del juicio humano

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

Autor	Tipo de IA analizada	Ámbito judicial / decisional	Dimensión de percepción social	Hallazgos relevantes sobre culpabilidad	Implicaciones éticas y legales
Ghasemaghaei & Kordzadeh (2024)	Algoritmos de recomendación	Toma de decisiones normativas	Confianza y desplazamiento de responsabilidad	Las decisiones injustas inducidas por IA no generan culpa percibida en los decisores	Dilución de la responsabilidad ética y jurídica
Miazek & Bocian (2025)	IA decisional	Juicio moral y de equidad	Sesgo egocéntrico y moralidad	La IA es percibida como más justa que los humanos cuando no existe interés personal	Distorsión de la percepción de imparcialidad algorítmica
Noriega (2020)	IA aplicada a interrogatorios	Investigación penal	Confianza, cooperación y sesgo	La IA puede reducir confesiones falsas al minimizar sesgos humanos	Riesgos éticos en la obtención de autoincriminaciones

Tabla 4. Percepciones sociales sobre la IA

Autor	Tipo de IA / Herramienta	Contexto de aplicación judicial	Tipo de sesgo identificado	Estrategias de mitigación del sesgo	Aportes a la equidad y percepción de culpabilidad
Borba et al., (2024)	Sistemas algorítmicos de toma de decisiones automatizadas	Uso de IA en decisiones judiciales, administrativas y cuasi-judiciales con impacto en derechos	Sesgo discriminatorio derivado de datos históricos y opacidad algorítmica	Regulación legal, evaluaciones de impacto algorítmico, mejora de calidad y diversidad de datos, control institucional	Refuerza la legitimidad de las decisiones judiciales asistidas por IA al reducir arbitrariedad y discriminación
Sovrano et al., (2025)	LLMs (ChatGPT) y herramientas de compliance (DoXpert)	Apoyo a decisiones jurídicas bajo marcos regulatorios (AI Act)	Sesgo por alucinaciones, opacidad y falta de trazabilidad	Supervisión humana obligatoria, documentación técnica estructurada, control ex ante y ex post	Reduce el riesgo de atribuciones erróneas de culpabilidad al exigir control humano efectivo
Hastings Blow et al., (2025)	Modelos generativos por difusión (Tab-DDPM)	Sistemas predictivos utilizados en justicia penal (ej. evaluación de riesgo tipo COMPAS)	Sesgo estadístico por datos desbalanceados y sobre-representación de grupos	Generación de datos sintéticos, balanceo de muestras, métricas de fairness	Disminuye desigualdades en predicciones que influyen en la percepción de culpabilidad
Herrera-Tapias & Hernández Guzmán (2025)	LLMs (GPTs)	Uso de IA como apoyo en decisiones judiciales constitucionales	Sesgo por errores semánticos y alucinaciones jurídicas	Uso auxiliar de IA, motivación judicial humana, control argumentativo	Protege el debido proceso y evita que la culpabilidad sea determinada por errores algorítmicos
Yuan et al., (2026)	Framework multi-agente con grafos legales (MAGLJP)	Predicción de sentencias con múltiples imputados	Sesgo por simplificación excesiva y falta de contextualización jurídica	Integración de conocimiento jurídico, razonamiento explicable, grafos causales	Mejora la individualización de la culpabilidad y reduce automatismos injustos
Cavus et al., (2025)	Deep Learning + XAI (RCN)	Predicción de reincidencia en justicia penal	Sesgo por desbalance de clases y opacidad del modelo	SMOTE, clustering, SHAP, explicabilidad del modelo	Fortalece la percepción de justicia al transparentar factores

ISSN: 2665-0398

Autor	Tipo de IA / Herramienta	Contexto de aplicación judicial	Tipo de sesgo identificado	Estrategias de mitigación del sesgo	Aportes a la equidad y percepción de culpabilidad
					que influyen en decisiones penales
Cirillo et al., (2020)	IA predictiva, NLP, Big Data y XAI	Aplicación indirecta al ámbito judicial en decisiones automatizadas de alto impacto	Sesgo estructural, histórico, de representación y algorítmico	Fairness algorítmica, XAI, datasets balanceados, gestión responsable de variables sensibles	Aporta marco conceptual para evitar discriminación y reforzar la legitimidad de decisiones judiciales asistidas por IA
Ferrara et al., (2024)	Machine Learning fairness-aware (MLOps)	Desarrollo de sistemas de IA utilizados en contextos judiciales y cuasi-judiciales	Sesgo por diseño, prioridades organizacionales y falta de integración de equidad	Fairness en todo el ciclo MLOps, métricas de equidad, testing y auditorías algorítmicas	Mejora la confianza pública y reduce el riesgo de culpabilidad injusta derivada de sistemas mal diseñados

Tabla 5. Estrategias de implementación para mitigar el sesgo

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

Discusión de resultados

Los resultados de la presente revisión sistemática evidencian que la implementación de herramientas de IA (IA) en los procesos judiciales vinculados con delitos informáticos incide de manera directa e indirecta en la percepción de culpabilidad. Esta incidencia se manifiesta principalmente a través de mecanismos como la autoridad técnica, el anclaje algorítmico, el desplazamiento de la responsabilidad decisional y la automatización cognitiva. Tales hallazgos responden de manera consistente al objetivo del estudio, al demostrar que la IA no opera como un instrumento neutral, sino como un agente cognitivo que reconfigura la valoración de la prueba y el juicio de culpabilidad dentro del proceso penal.

Uno de los principales resultados identifica que la aplicación de la IA en el análisis de evidencia digital y forense tiende a reforzar la credibilidad probatoria, lo que incrementa la percepción de culpabilidad de la persona imputada. Este hallazgo converge con lo señalado por Reedy (2023) y Hefetz (2025), quienes sostienen que la sofisticación técnica de los sistemas algorítmicos genera una deferencia excesiva por parte de juezas, jueces, peritas y peritos, debilitando el escrutinio crítico de la prueba. De manera concordante, Fazel et al., (2022) evidencian que los puntajes de riesgo algorítmico influyen en la severidad percibida de la conducta, intensificando la atribución de peligrosidad y culpabilidad.

Asimismo, los resultados muestran que los sesgos algorítmicos, particularmente el *automation bias* y el anclaje decisional, distorsionan la evaluación judicial de la culpabilidad cuando la IA se introduce en etapas previas al juicio humano. Este resultado es consistente con Agudo et al., (2024), quienes demuestran experimentalmente que los errores algorítmicos reducen la precisión del juicio humano, y con Wu & Lin (2025), quienes documentan cómo la utilización de IA en el ámbito judicial chino refuerza decisiones previas y limita la participación efectiva de las personas acusadas. No obstante, Chuan et al., (2024) introducen una divergencia parcial al evidenciar que los sistemas de IA explicable (*Explainable Artificial Intelligence, XAI*) pueden incrementar la conciencia sobre el sesgo, aunque dicha explicabilidad no elimina por completo la confianza indebida en el sistema.

En relación con la percepción social y moral de la culpabilidad, los resultados indican un desplazamiento de la responsabilidad desde la persona decisora humana hacia el sistema algorítmico. Este hallazgo coincide con lo expuesto por Giroux et al., (2022) y Ghasemaghaei & Kordzadeh (2024), quienes evidencian que las decisiones mediadas por IA generan una menor atribución de culpa a quienes toman decisiones humanas, incluso cuando los resultados presentan

ISSN: 2665-0398

efectos discriminatorios. De forma complementaria, Contini et al., (2024) advierten que la incorporación de IA puede reducir la dimensión emotivo-cognitiva del juicio penal, favoreciendo procesos de deshumanización que afectan la valoración individualizada de la culpabilidad.

Desde una perspectiva ética y jurídica, los resultados se alinean con los planteamientos de Peng & Lei (2024) y Kong et al., (2025), quienes alertan que los sistemas predictivos de imputación y sentencia pueden anticipar resultados judiciales y erosionar el principio de presunción de inocencia. En contraste, Han et al., (2024) demuestran que los sistemas explicables y centrados en la persona humana pueden mitigar parcialmente estos riesgos, lo que permite explicar las diferencias observadas entre modelos algorítmicos opacos y enfoques *human-in-the-loop*.

El estudio presenta diversas limitaciones. En primer lugar, la revisión se basa exclusivamente en literatura científica indexada, lo que podría excluir experiencias judiciales emergentes o prácticas institucionales no publicadas en revistas académicas. En segundo lugar, se observa una heterogeneidad metodológica significativa entre los estudios incluidos, que abarca diseños empíricos, experimentales y conceptuales, lo cual limita la comparabilidad directa de los resultados. En tercer lugar, la mayoría de las investigaciones analizadas se concentran en contextos europeos, norteamericanos y asiáticos, restringiendo la generalización de los hallazgos a sistemas judiciales latinoamericanos. Finalmente, la evidencia empírica directa basada en juezas y jueces en ejercicio sigue siendo limitada, predominando estudios experimentales o escenarios simulados.

A partir de los resultados y las limitaciones identificadas, se recomienda que futuras investigaciones desarrollen estudios empíricos de carácter longitudinal que permitan evaluar cómo la exposición sostenida a sistemas de IA modifica la percepción de culpabilidad en juezas, jueces, fiscales y personas defensoras. Asimismo, resulta necesario incorporar análisis comparativos entre distintos sistemas judiciales, especialmente en contextos latinoamericanos, a fin de evaluar la transferibilidad de estos hallazgos.

Se sugiere, además, profundizar en el diseño y la evaluación de modelos algorítmicos explicables y regulados, analizando su impacto real en la reducción del sesgo y en la protección efectiva de la presunción de inocencia. Finalmente, futuras revisiones deberían integrar de manera sistemática marcos normativos y constitucionales para evaluar la compatibilidad entre el uso de la IA, la atribución de culpabilidad y las garantías del debido proceso.

En conjunto, los resultados confirman que la implementación de herramientas de IA transforma la percepción de culpabilidad en los delitos informáticos, no solo a través de mejoras técnicas,

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

sino mediante cambios estructurales en la cognición judicial, la atribución de responsabilidad y la legitimidad del juicio penal. Ello refuerza la necesidad de adoptar enfoques regulados, explicables y centrados en el control humano como condición indispensable para una justicia penal compatible con los principios del Estado de derecho.

Conclusiones

Los hallazgos de la presente investigación evidencian que la incorporación de herramientas de IA en los procesos judiciales asociados con delitos informáticos genera efectos significativos en la percepción de culpabilidad. En particular, se identifica que la autoridad técnica atribuida a los sistemas algorítmicos, la automatización de la valoración probatoria y la presencia de sesgos algorítmicos influyen de manera determinante en la forma en que juezas, jueces y otras personas decisoras interpretan la evidencia digital.

Asimismo, los estudios analizados coinciden en que la deferencia hacia los resultados producidos por la IA puede intensificar la atribución de culpabilidad o, en determinados contextos, diluir la responsabilidad humana, configurando un cambio sustancial en la dinámica tradicional del juicio penal.

En relación con el objetivo de analizar cómo la implementación de herramientas de IA afecta la percepción de culpabilidad en delitos informáticos dentro de los procesos judiciales, los resultados permiten concluir que dicha implementación no es neutral.

La IA actúa como un mediador cognitivo que condiciona la interpretación de la prueba, la evaluación del riesgo y la anticipación de decisiones judiciales, lo que puede reforzar percepciones de culpabilidad incluso antes de que se realice un análisis exhaustivo e individualizado del caso. Este efecto se ve acentuado cuando los sistemas carecen de explicabilidad suficiente o cuando se incorporan en etapas tempranas del proceso decisorio, lo que puede afectar principios fundamentales como la presunción de inocencia y el debido proceso.

El presente trabajo se desarrolló bajo la modalidad de artículo de revisión sistemática, siguiendo un enfoque metodológico riguroso que permitió sintetizar y contrastar evidencia reciente proveniente de revistas científicas indexadas. Este diseño metodológico facilitó la identificación de patrones convergentes y divergentes en la literatura especializada, así como la integración de enfoques empíricos, experimentales y normativos, lo que fortalece la solidez de las conclusiones y su pertinencia para el debate académico y jurídico en torno al uso de la IA en la justicia penal.

Finalmente, los resultados subrayan la necesidad de avanzar hacia modelos de implementación de la IA que prioricen la explicabilidad, el control humano efectivo y una regulación jurídica clara y garantista. Las futuras

ISSN: 2665-0398

investigaciones deberían profundizar en estudios empíricos con juezas, jueces y demás personas operadoras del sistema de justicia, incorporar análisis comparados entre distintos sistemas jurídicos y evaluar de manera longitudinal los efectos de la exposición sostenida a herramientas algorítmicas.

Estas líneas de investigación permitirán consolidar un marco teórico y práctico que garantice que el uso de la IA en el ámbito judicial contribuya a una justicia penal más equitativa, sin distorsionar la percepción de culpabilidad ni comprometer los derechos fundamentales.

Referencias

Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. Cognitive Research: Principles and Implications, 9, 1. Documento en línea. Disponible https://doi.org/10.1186/s41235-023-00529-3

Akinrinola, O., Addy, W., Ajayi-Nifise, A., Odeyemi, O., & Falaiye, T. (2024). Application of machine learning in tax prediction: A review with practical approaches. *Global Journal of Engineering and Technology Advances*, 18(2), 102–117. Documento en línea. Disponible https://doi.org/10.30574/gjeta.2024.18.2.0028

Barsekh-Onji, A., Torres Hernandez, Z., & Cardoso Espinosa, E. O. (2025). Advancing smart public administration: Challenges and benefits of artificial intelligence. Urban Governance, 5, 279–292. Documento en línea. Disponible https://doi.org/10.1016/j.ugj.2025.06.003

Borba, R. L., de Paula Ferreira, I. E., & Bertucci Ramos, P. H. (2024). Addressing discriminatory bias in artificial intelligence systems operated by companies: An analysis of end-user perspectives. Technovation, 138, 103118. Documento en línea. Disponible https://doi.org/10.1016/j.technovation.2024.103118

Cavus, M., Benli, M. N., Altuntas, U., Sari, M., Ayan, H., & Ugurluoglu, Y. F. (2025). Transparent and bias-resilient AI framework for recidivism prediction using deep learning and clustering techniques in criminal justice. Applied Soft Computing, 176, 113160. Documento en línea. Disponible https://doi.org/10.1016/j.asoc.2025.113160

Chuan, C.-H., Sun, R., Tian, S., & Tsai, W.-H. S. (2024). Explainable artificial intelligence (XAI) for facilitating recognition of algorithmic bias: An experiment from imposed users' perspectives. Telematics and Informatics, 91, 102135. Documento en línea. Disponible https://doi.org/10.1016/j.tele.2024.102135

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Santuccione Chadha, A., & Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digital Medicine, 3(1), 81. Documento en línea. Disponible https://doi.org/10.1038/s41746-020-0288-5

Contini, F., Minissale, A., & Bergman Blix, S. (2024). Artificial intelligence and real decisions: Predictive systems and generative AI vs. emotive-cognitive legal deliberations. Frontiers in Sociology, 9, 1417766. Documento en línea. Disponible https://doi.org/10.3389/fsoc.2024.1417766

Engle, T. A., Maher, C. A., & Nedelec, J. L. (2025). Sellin and Wolfgang revisited: A partial replication and expansion considering cybercrime. *Journal of Criminal Justice*, 101, 102523. Documento en línea. Disponible https://doi.org/10.1016/j.jcrimjus.2025.102523

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)

Fazel, S., Fanshawe, T., & Sariaslan, A. (2022). Towards a more evidence-based risk assessment for people in the criminal justice system: The case of OxRec in the Netherlands. *European Journal on Criminal Policy and Research*, 28, 397–406. Documento en línea. Disponible https://doi.org/10.1007/s10610-022-09520-y

Ferrara, C., Sellitto, G., Ferrucci, F., Palomba, F., & De Lucia, A. (2024). Fairness-aware machine learning engineering: How far are we? Empirical Software Engineering, 29(1), 9. Documento en línea. Disponible https://doi.org/10.1007/s10664-023-10402-y

Franguloiu, S. (2024). Principles for the use of artificial intelligence (AI) in the judiciary as derived from the European Ethics Charter. Justice Efficiency and Limitations. Series VII – Social Sciences and Law, 39–46. Documento en línea. Disponible https://doi.org/10.31926/but.ssl.2023.16.65.3.5

Ghasemaghaei, M., & Kordzadeh, N. (2024). Understanding how algorithmic injustice leads to making discriminatory decisions: An obedience to authority perspective. Information & Management, 61, 103921. Documento en línea. Disponible https://doi.org/10.1016/j.im.2024.103921

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, 178, 1027–1041. Documento en línea. Disponible https://doi.org/10.1007/s10551-022-05056-7

Hahn, T., Ernsting, J., Winter, N. R., et al., (2022). An uncertainty-aware, shareable, and transparent neural network architecture for brain-age modeling. Science Advances, 8(1), eabg9471. Documento en línea. Disponible https://doi.org/10.1126/sciadv.abg9471

Han, W., Shen, J., Liu, Y., et al. (2024). LegalAsst: Human-centered and AI-empowered machine to enhance court productivity and legal assistance. Information Sciences, 679, 121052. Documento en línea. Disponible https://doi.org/10.1016/j.ins.2024.121052

Hani, U., Sohaib, O., Khan, M., Aleidi, A., & Islam, N. (2024). Psychological profiling of hackers via machine learning toward sustainable cybersecurity. Frontiers in Computer Science, 6, 1381351. Documento en línea. Disponible https://doi.org/10.3389/fcomp.2024.1381351

Hastings Blow, C., Qian, L., Gibson, C., Obiomon, P., & Dong, X. (2025). Data augmentation via diffusion model to enhance AI fairness. Frontiers in Artificial Intelligence, 8, 1530397. Documento en línea. Disponible https://doi.org/10.3389/frai.2025.1530397

Hefetz, I. (2023). Mapping AI-ethics' dilemmas in forensic case work: To trust AI or not? Forensic Science International, 350, 111807. Documento en línea. Disponible https://doi.org/10.1016/j.forsciint.2023.111807

Hefetz, I. (2025). Evaluating bias in forensic evidence: From expert analysis to AI-based decision tools. Forensic Science International: Synergy, 11, 100645. Documento en línea. Disponible https://doi.org/10.1016/j.fsisyn.2025.100645

Herrera-Tapias, B. A., & Hernández Guzmán, D. (2025). Legal hallucinations and the adoption of artificial intelligence in the judiciary. Procedia Computer Science, 257, 1184–1189. Documento en línea. Disponible https://doi.org/10.1016/j.procs.2025.03.158

Javed, K., & Li, J. (2025). Bias in adjudication: Investigating the impact of artificial intelligence, media, financial and legal institutions in pursuit of social justice. PLOS ONE, 20(1), e0315270. Documento en línea. Disponible https://doi.org/10.1371/journal.pone.0315270

Kerimkhulle, S., Dildebayeva, Z., Tokhmetov, A., et al. (2023). Fuzzy logic and its application in the assessment of information security risk of industrial Internet of Things. Symmetry, 15(10), 1958. Documento en línea. Disponible https://doi.org/10.3390/sym15101958

ISSN: 2665-0398

Kirsanova, N., Gogoleva, V., Zyabkina, T., & Semenova, K. (2021). The use of digital technologies in the administration of justice in the field of environmental crime. E3S Web of Conferences, 258, 05035. Documento en línea. Disponible https://doi.org/10.1051/e3sconf/202125805035

Kong, Y., Wang, Y.-G., Deng, H., Xiao, Z., & Zhang, Y. (2025). LF-HGRILF: A law-fact heterogeneous graph representation and iterative learning framework for legal judgment prediction. Knowledge-Based Systems, 327, 114083. Documento en línea. Disponible https://doi.org/10.1016/j.knosys.2025.114083

Miazek, K., & Bocian, K. (2025). When AI is fairer than humans: The role of egocentrism in moral and fairness judgments of AI and human decisions. Computers in Human Behavior Reports, 19, 100719. Documento en línea. Disponible https://doi.org/10.1016/j.chbr.2025.100719

Naik, N., Hameed, B. M. Z., Shetty, D. K., et al. (2022). Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? Frontiers in Surgery, 9, 862322. Documento en línea. Disponible https://doi.org/10.3389/fsurg.2022.862322

Noriega, M. (2020). The application of artificial intelligence in police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, cooperation, and false confessions. Futures, 117, 102510. Documento en línea. Disponible https://doi.org/10.1016/j.futures.2019.102510

Olawade, D. B., Ayoola, F. I., Ebo, T. O., et al., (2025). Artificial intelligence in forensic mental health: A review of applications and implications. *Journal of Forensic and Legal Medicine*, 113, 102895. Documento en línea. Disponible https://doi.org/10.1016/j.jflm.2025.102895

Peng, Y.-T., & Lei, C.-L. (2024). Using Bidirectional Encoder Representations from Transformers (BERT) to predict criminal charges and sentences from Taiwanese court judgments. PeerJ Computer Science, 10, e1841. Documento en línea. Disponible https://doi.org/10.7717/peerj-cs.1841

Pristanskov, V., Kharatishvili, A., & Evstratova, J. (2023). Artificial intelligence—A new form of using special knowledge in investigating and solving cybercrimes. *Russian Journal of Criminology*, 17(6), 586–596. Documento en línea. Disponible https://doi.org/10.17150/2500-4255.2023.17(6).586-596

Reedy, P. (2023). Interpol review of digital evidence for 2019–2022. Forensic Science International: Synergy, 6, 100313. Documento en línea. Disponible https://doi.org/10.1016/j.fsisyn.2022.100313

Sessa, F., Esposito, M., Cocimano, G., et al., (2024). Artificial intelligence and forensic genetics: Current applications and future perspectives. Applied Sciences, 14(5), 2113. Documento en línea. Disponible https://doi.org/10.3390/app14052113

Shakeel, A., Sultan, M., Idrees, S., et al., (2025). Evaluating the application of machine learning algorithms in predicting disease outcomes and enhancing diagnostic accuracy in healthcare systems. IJLSS, 3(5), 50–57. Documento en línea. Disponible https://doi.org/10.71000/ddt1bm68

Shaligar, S., Arefnia, T., & Mohammadian Amiri, M. (2024). Applications of artificial intelligence in the production and use of digital documents and electronic evidence as proof in civil and criminal litigation. Legal Studies in Digital Age, 3(2), 10–30. Documento en línea. Disponible https://doi.org/10.61838/kman.lsda.3.2.2

Sovrano, F., Hine, E., Anzolut, S., & Bacchelli, A. (2025). Simplifying software compliance: AI technologies in drafting technical documentation for the AI Act. Empirical Software Engineering,

30, 91. Documento en línea. Disponible https://doi.org/10.1007/s10664-025-10645-x

Thamizh Mani, R., Palimar, V., Pai, M. S., Shwetha, T. S., & Krishnan, M. N. (2025). An evolution of forensic linguistics: From manual analysis to machine learning – A narrative review. Forensic Science International: Reports, 11, 100417. Documento en línea. Disponible https://doi.org/10.1016/j.fsir.2025.100417

Wang, H., & Ma, S. (2022). Preventing crimes against public health with artificial intelligence and machine learning capabilities. Socio-Economic Planning Sciences, 80, 101043. Documento en línea. Disponible https://doi.org/10.1016/j.seps.2021.101043

Wu, W., & Lin, X. (2025). Access to technology, access to justice: China's artificial intelligence application in criminal proceedings. *International Journal of Law, Crime and Justice*, 81, 100741. Documento en línea. Disponible https://doi.org/10.1016/j.ijlcj.2025.100741

Yuan, W., Song, K., Jiang, Z., et al., (2026). A multi-agent framework with legal event logic graph for multi-defendant legal judgment prediction. Information Processing & Management, 63, 104319. Documento en línea. Disponible https://doi.org/10.1016/j.ipm.2025.104319

Zeadally, S., Adi, E., Baig, Z., & Khan, I. (2020). Harnessing artificial intelligence capabilities to improve cybersecurity. IEEE Access, 8, 23817–23837. Documento en línea. Disponible https://doi.org/10.1109/ACCESS.2020.2968045