

La Inteligencia Artificial como herramienta probatoria: retos y riesgos en el derecho penal contemporáneo

Artificial Intelligence as an Evidentiary Tool: Challenges and Risks in Contemporary Criminal Law

Mendoza-Valladolid, Natali Gabriela¹; Carrera-Pérez, Germán Eduardo².

Recibido: 01/04/2026

Aceptado: 10/04/2026

Publicado: 30/04/2026

Cita: Mendoza-Valladolid, N. G., & Carrera-Pérez, G. E. (2026). La Inteligencia Artificial como herramienta probatoria: retos y riesgos en el derecho penal contemporáneo. *Space Scientific Journal of Multidisciplinary*, 4(2), 31-47. <https://doi.org/10.63618/omd/ssjm/v4/n2/69>

Resumen

En la presente investigación se analizó la inteligencia artificial como herramienta probatoria e identificó los riesgos y retos en el sistema penal contemporáneo. La evolución constante del mundo y de la tecnología hacen que el derecho se encuentre a la par con los avances digitales, particularmente con las herramientas probatoria, donde la inteligencia artificial plantea oportunidades de eficiencia, pero también tensiones en relación con la fiabilidad, admisibilidad, valoración de la prueba y garantías procesales en el derecho penal contemporáneo. La metodología que se empleó fue de enfoque cualitativo, con un alcance analítico y descriptivo, con un diseño no experimental, para lo cual se empleó la técnica del análisis bibliográfico y doctrinal, con un método analítico, inductivo y comparado, empleando las fichas bibliográficas e inventario como instrumentos para la recolección de fuentes. Los resultados advierten que la ausencia de parámetros claros sobre la fiabilidad probatoria, la transparencia algorítmica, la trazabilidad los sesgos algorítmicos, opacidad, biometría y deepfakes, comprometen las garantías procesales en un sistema penal. Se concluye que la IA es admisible únicamente como herramienta auxiliar para la obtención probatoria, y no como valor probatorio automático con controles rigurosos y estándares reforzados de admisibilidad y valoración probatoria para que tenga eficacia probatoria y se garantice el debido proceso.

Palabras clave: inteligencia artificial; prueba penal; evidencia digital; debido proceso; garantías procesales.

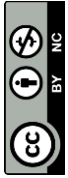
Abstract

This research analyzes artificial intelligence as an evidentiary tool and identifies the risks and challenges in the contemporary criminal justice system. The constant evolution of the world and technology means that the law must keep pace with digital advancements, particularly with evidentiary tools, where artificial intelligence presents opportunities for efficiency, but also substantial risks to procedural guarantees in contemporary criminal law. The methodology employed is qualitative, with an analytical and descriptive scope and a non-experimental design. Bibliographic and doctrinal analysis was used, employing an analytical, inductive, and comparative method, using bibliographic records and inventories as instruments for collecting sources. The results show that the evidentiary reliability of AI depends on data quality, algorithmic transparency, traceability, and the real possibility of rebuttal. Significant risks associated with bias, opacity, biometrics, and deepfakes are evident, with impacts on the presumption of innocence and equality of arms. It is concluded that AI is admissible only as an auxiliary tool for obtaining evidence, under significant human control and reinforced standards of admissibility and evidentiary assessment in order for it to have probative effectiveness.

Keywords: artificial intelligence; criminal evidence; digital evidence; due process; procedural safeguards

¹ Universidad Tecnológica Indoamerica, Facultad de Jurisprudencia y Ciencias Políticas; Ecuador, Ambato; 0009-0009-5672-2479; nmendoza11@indoamerica.edu.ec

² Universidad Tecnológica Indoamerica, Docente de Tiempo Completo de la Facultad de Jurisprudencia y Ciencias Políticas; Ecuador, Ambato; 0000-0002-0820-0600; germancarrera@uti.edu.ec



1. Introducción

En los últimos siglos, la evolución y la expansión acelerada de la tecnología y el mundo digital ha transformado diversos ámbitos de la sociedad, incluido el derecho y los sistemas de justicias en el mundo. Con frecuencia el uso de la inteligencia artificial como herramienta para la simplificación de procesos, análisis, respuestas concretas, obtención de resultados ha simplificado e impactado en los distintos sectores, especialmente en el derecho y lo relativo a la producción de pruebas electrónicas.

En el ámbito procesal y judicial, la inteligencia artificial se ha posicionado como un instrumento de suma importancia para la reconstrucción de hechos, obtener pruebas, diseñar propuestas judiciales, analizar jurisprudencia, leyes y doctrina, generar argumentos, realizar contratos con precisión, rapidez y ventaja, pese ello, el uso frecuente de esta herramienta tecnológica ha generado diversos debates y tensiones sobre los desafíos, riesgos y retos normativos y doctrinales, sobre todo en lo referente a la fiabilidad, certeza y transparencia de los algoritmos que se encuentran programados por grandes bases de datos que se alimentan con la información disponible que incorpora cada persona, lo que genera amplios cuestionamientos sobre si la inteligencia artificial se adecuado a los principios indispensables para el debido proceso, la seguridad jurídica y demás garantías procesales.

El uso de las herramientas tecnológicas como medios probatorios, no es un fenómeno reciente, sino que ha ido paulatinamente incorporado a los procesos judiciales como evidencias digitales o que tienen un origen digital, partiendo desde la aceptación de los correos electrónicos, evidencia en audio, video y otros de soporte tecnológico hasta el uso de la inteligencia artificial como herramienta probatoria en algunos sistemas procesales en diferentes países.

En la actualidad, el debate académico surge sobre la base de identificar criterios propios sobre utilidad, admisibilidad, conducencia, pertinencia y valor probatorio sobre la inteligencia artificial como herramienta probatoria en los sistemas penales contemporáneos o como uso se debe de adecuar a los sistemas procesales modernos para que las partes procesales, así como los administradores de justicia pueden generar decisiones motivadas y adecuadas al debido proceso.

En este escenario, la discusión académica se desplaza desde un plano más relativo frente a una transformación digital de los gobiernos; con el interés de vincular a la IA en sus procesos judiciales, uno de estos procesos se señala bajo el cuestionamiento del uso de la IA como una herramienta probatoria, que conlleva a estudiar indicadores que van de la mano de la evidencia algorítmica-digital, la

biometría, la exigencia de explicabilidad, los riesgos de sesgo y las garantías procesales involucradas (Pereira-Puigvert, 2024).

En la presente investigación, se abordará como diversos estudios doctrinales señalan la incorporación de la inteligencia como herramienta probatoria y cuáles son los riesgos y desafíos que se presentan en la actualidad, siendo que esta herramienta no puede observarse exclusivamente como un instrumento técnico, sino que requiere de un análisis jurídico integral, puesto que, efectos como la opacidad algorítmica, sesgos algorítmicos y los pocos estándares que existen actualmente sobre la valoración de la inteligencia artificial como herramienta probatoria, pueden llegar a dificultar el derecho a la contradicción y vulnerar principios fundamentales del debido proceso, así como el control procesal. Por ello, el campo enfatiza que la discusión sobre IA probatoria exige atender simultáneamente la fiabilidad, la trazabilidad y la posibilidad real de impugnación, porque una salida algorítmica no se legitima por su sofisticación, sino por su control procesal (García-Bernalt, 2024).

El plano literario coincide en que la inteligencia artificial se incorpora al proceso penal como un soporte técnico orientado a la obtención, depuración, correlación y presentación de información, misma que posteriormente se pretende introducir como evidencia o como un sustento pericial. Tomando en cuenta este criterio, diferentes autores explican que el valor añadido de la IA no reside únicamente en la aceleración de los tiempos de análisis, sino en su capacidad para operar sobre grandes volúmenes de datos para así revelar patrones que, sin asistencia computacional, permanecerían prácticamente inaccesibles para la investigación penal tradicional.

Desde esta mirada, el campo especializado enfatiza que la discusión sobre la IA probatoria exige atender de manera simultánea a la fiabilidad del sistema, la trazabilidad de los procesos algorítmicos y la posibilidad real de impugnación, en la medida en que una salida algorítmica no se legitima por su grado de sofisticación técnica, sino por el control procesal efectivo al que puede ser sometida (García-Bernalt, 2024).

En este plano también se consolida una agenda crítica alrededor de tres focos centrales; el primero está constituido por el uso de biometría y reconocimiento facial en tareas de identificación e investigación penal, el segundo va de concibe la implementación de sistemas de justicia predictiva junto a herramientas algorítmicas para la evaluación de riesgos, el tercer foco señala la manipulación sintética de evidencia mediante tecnologías de inteligencia artificial generativa encaminado a los deepfakes.

La hipótesis de la investigación parte del hecho de que la inteligencia artificial como herramienta probatoria no debe ser valorada bajo los mismos criterios de una prueba tradicional, toda vez que, en esta primera existen riesgos que afectan su correcta valoración y control jurídico de la inteligencia artificial como herramienta probatoria y su respeto con las garantías procesales, para lo cual, es necesario preguntar ¿Cuáles son los riesgos de la inteligencia artificial como medio de prueba y su relación con las garantías procesales? ¿se debe valorar autónomamente a la inteligencia artificial como herramienta probatoria en el proceso penal? El objetivo central de la investigación es analizar e identificar los riesgos y retos de inteligencia artificial como herramienta probatoria en el sistema penal contemporáneo, considerando los riesgos propios del uso de la inteligencia artificial y los límites que confrontan el sistema de justicia.

Riesgos derivados de la inteligencia artificial: sesgos algorítmicos, sesgos de opacidad, deepfakes

La inteligencia artificial nace como una herramienta innovadora que facilita el desarrollo de tareas que en un principio necesitan de la inteligencia humana, la cual se encuentra diseñada para que programen y alimenten sus sistemas de información, para que a través de desarrolladores establezcan puentes en función a la información disponible en sus sistemas de almacenamiento que se encargan de procesar, generar, predecir, y ejecutar mandos.

La implementación de la inteligencia artificial, a pesar de su eficiencia como herramienta para simplificar procesos cotidianos, trae consigo riesgos relacionados con los sesgos asociados a la discriminación algorítmica, sistemas biométricos, manipulación digital o deepfakes que pueden generar vulneraciones a derechos fundamentales.

Uno de los problemas en la estructura de la inteligencia artificial se deriva de la presencia de sesgos algorítmicos procedentes de los datos y diseños tecnológicos que no tienen diversidad ni representatividad por lo que tienden a beneficiar o perjudicar a determinados grupos sociales, perpetuando desigualdades sociales ya existentes. En esa línea, (Ferrante, 2021) evidencia como la Dra. Buolamwini demostró el sesgo racial presente en ciertos algoritmos de reconocimiento facial creados por Amazon, Microsoft y otros identificaban y marchaban de mejor manera con facciones de personas blancas.

Los sesgos algorítmicos son programas automatizados que predicen resultados favoreciendo a determinado grupo social en perjuicio de otros, es decir toman decisiones por identificaciones manipulables, discriminatorias y desproporcionadas de ciertos grupos sociales. Según AlgoRace (2022), un sesgo algorítmico se

produce cuando el resultado que se da un programa es diferente bajo perfiles excluyentes como género, raza, clase, religión, orientación sexual que perjudican a las minorías.

En relación a los riesgos de opacidad en la inteligencia artificial se deriva de la falta de transparencia en la programación de los sistemas de IA que no se pueden verificar o identificar que variable determina una decisión dentro del algoritmo, lo que significa incluso una dificultad para los mismos programadores, debido en que en ciertas circunstancias no son capaces de entender o predecir determinados resultados, lo cual dentro de un sistema penal y probatorio dificulta la autenticidad de la prueba.

Otro de los riesgos que nacen de la inteligencia artificial, son los deepfakes o manipulación de la información a partir de redes neuronales que crea material o información falso que parece cierto manipulando la perspectiva del que la recibe.

Este tipo de tecnología presupone un riesgo a la sociedad y las democracias, incidiendo en la toma de decisiones de las personas con información falsa o alterada, por lo que el uso indiscriminado de la inteligencia artificial sin control minucioso de los riesgos que esta conlleva supone un peligro para la sociedad.

La inteligencia artificial como herramienta probatoria en el proceso penal y las tensiones con el debido proceso

La inteligencia artificial se inserta en el proceso penal contemporáneo como una herramienta de apoyo a la actividad probatoria, no como una fuente autónoma de prueba ni como un sustituto del juicio humano. Su función principal radica en asistir a los operadores jurídicos en tareas de obtención, clasificación y análisis de información compleja, ampliando la capacidad investigativa del sistema penal sin alterar la estructura garantista que rige la valoración de los hechos.

La inteligencia artificial usada como herramienta probatoria en proceso penal se encuentra orientada en el análisis y procesamiento de la información obtenida en la investigación, es decir que esta puede ayudar en la reconstrucción de los hechos y circunstancias materias del delito, por lo tanto, la autonomía de esta, no se centra en certezas absolutas, sino que se encuentra limitada bajo el control humano, pues de lo contrario, se corre el riesgo de sesgos y erróneos análisis que pueden perjudicar una decisión judicial.

La inteligencia artificial opera como un mediador técnico del conocimiento fáctico, cuya legitimidad depende de su subordinación a las reglas procesales y no de su sofisticación tecnológica, es decir que, los medios probatorios introducidos por la

inteligencia artificial deben de ser sometidos a contradicción y evaluados conforme a los estándares tradicionales de racionalidad probatoria para que tenga eficacia jurídica.

Por lo dicho, es que la inteligencia artificial es un medio auxiliar que debe ser sometido a control humano y judicial, dado que, la autonomía en el sentido jurídico exige que los resultados que nacen de dicha herramienta tecnológica sean sometidos a verificación y contradicción, mismos que deben ser corroborados por otros medios probatorio para que se garantiza un respeto absoluto a las garantías básicas del debido proceso.

El uso de la inteligencia como medio de prueba resulta incompatible con los sistemas procesales penales, puesto que, en el razonamiento que debe realizar los jueces para valorar las pruebas para tener certeza se deben verificar los resultados con medios periféricos que corroboren la veracidad de dicha prueba, es así que, en los actuales momentos su validez se ve limitado a un medio auxiliar probatorio.

Se contempla desde una visión garantista, que la prueba penal sigue siendo, ante todo, un acto jurídico sometido a exigencias de fiabilidad, contradicción y motivación, razón por la cual la incorporación de tecnologías algorítmicas obliga a un replanteamiento crítico de la gran mayoría de los estándares clásicos de valoración racional de la prueba, con el fin de evitar que la eficiencia técnica de la herramienta termine desplazando el juicio humano provocando un debilitamiento en las bases del control judicial efectivo.

La incorporación de la inteligencia artificial en el proceso penal como herramienta probatoria ha generado tensiones en las estructuras de las garantías del debido proceso, debido los riesgos que generan los sesgos algorítmicos opacos que impactan directamente al principio de contradicción debido a que, cuando se generan evidencias desde la inteligencia artificial se ve limitado el cuestionamiento del origen de la fuente, la calidad o lógica del razonamiento que utiliza el algoritmo para generar dicha información, es así que se debilita el ejercicio de la defensa.

En el proceso penal, esta exigencia como tal no es meramente instrumental, se conecta de forma directa con el debido proceso, el derecho de defensa y con el principio de contradicción, en tanto una prueba que no puede ser explicada ni cuestionada de manera efectiva termina por comprometer la igualdad de armas entre las partes. Desde esta perspectiva, García Falconí y Barona Pazmiño (2024) sostienen que la presencia de un control humano significativo no constituye una opción técnica disponible, sino una condición jurídica indispensable para la admisibilidad y la posterior valoración de cualquier resultado algorítmico que aspire a tener relevancia probatoria.

La incorporación de sistemas de inteligencia artificial en el proceso penal introduce tensiones significativas en garantías esenciales del debido proceso como la presunción de inocencia, el derecho a la defensa, la igualdad de armas y el control judicial efectivo. La opacidad de los algoritmos especialmente en modelos predictivos y sistemas de reconocimiento facial dificulta que las partes puedan comprender, contradecir o auditar el origen de los resultados, afectando directamente el principio de contradicción y el estándar de motivación de las decisiones judiciales. Mientras que, en relación a la caja negra algorítmica, podría incidir en el derecho a la valoración de la prueba de manera racional lo que provocaría vulneraciones a la defensa si esta no con los recursos tecnológicos necesarios para impugnar la evidencia automatizada.

Cuando las evidencias o pruebas son obtenidas por un sistema de inteligencia artificial se ven afectados por la opacidad de los algoritmos, particularmente en patrones predictivos, reconocimiento facial, probabilidad de culpabilidad, afecta la presunción de inocencia, en la forma que dichos presupuestos de los sistemas de inteligencia artificial hacen que la carga de la prueba se invierta y recaiga sobre la persona procesada y no sobre la institución encargada de probar e investigar como lo es la fiscalía.

Los sesgos algorítmicos derivados de bases de datos deficientes, clasificaciones raciales o socioeconómicas y errores de entrenamiento pueden reproducir o amplificar desigualdades existentes, comprometiendo la imparcialidad judicial y la igualdad ante la ley. En conjunto, los hallazgos evidencian que la IA no solo plantea desafíos técnicos, sino que impacta directamente el núcleo garantista del proceso penal, exigiendo marcos de transparencia, auditabilidad y proporcionalidad que aseguren la plena vigencia de los derechos fundamentales.

En este contexto, las tensiones derivadas de la falta o la limitación de control judicial sobre la inteligencia artificial por la forma en que se encuentran programados los sistemas de inteligencia artificial, hacen que tanto las partes procesales como el juzgador que no necesariamente tienen conocimiento del funcionamiento de dichas herramientas tecnológicas, dificultan la valoración crítica y necesaria para una sentencia clara y certera.

Condiciones de fiabilidad, exactitud, trazabilidad, auditabilidad, margen de error de la inteligencia artificial como herramienta probatoria.

En el presupuesto de la fiabilidad la inteligencia artificial necesita que el sistema sea confiable científicamente, es decir que produzca resultados verificables, para lo cual se requiere que la información y los algoritmos sean transparentes, libres de sesgos y con el conocimiento del margen de error de los resultados.

La forma en que opera la inteligencia artificial se rige por un modelo de “caja negra”, lo que implica que, los resultados de las búsquedas o la información proporcionada por estos sistemas informáticos no brindan la suficiente seguridad en la fiabilidad y exactitud de la información, lo que implica una falta de claridad y transparencia en los procesos judiciales, puesto que erosionan la confianza de la IA como herramienta probatoria. Es así que, Abad (2025), sostiene que la falta de certeza y transparencia de cómo se generan las decisiones algorítmicas sugiere la necesidad de reformar los sistemas automatizados, los cuales deben de ser sometidos a escrutinio y contradicción, donde se asegure un debido proceso y seguridad jurídica.

La fiabilidad jurídica de la inteligencia artificial se da exclusivamente en contexto de verificación y supervisión de la información o del resultado, lo que implica que dentro del derecho penal el juzgador tenga la convicción plena de que se halla conforme con los resultados dados por la IA como herramienta probatoria y sea capaz de valorarlos y sirvan de fundamento en su decisión, en consecuencia, la fiabilidad de la inteligencia artificial implica un estricto control humano que garantice el sistema es confiable. Para Abad (2025) “La integración de la IA en los procesos de justicia y arbitraje conlleva dilemas éticos complejos relacionados con la imparcialidad y la equidad. Los algoritmos, al ser entrenados con datos históricos, pueden incorporar y perpetuar sesgos inherentes a dichos dato” (p.117).

Uno de los elementos indispensables para garantizar la eficacia de la inteligencia artificial es la trazabilidad, que implica la posibilidad de poder rastrear todo el proceso que se genera en los sistemas de inteligencia artificial, es un fundamental para que se garantice que aquello generado tenga exactitud y genere confianza al usuario, que contemplado desde la inteligencia artificial como herramienta probatoria es una forma de garantizar que se incorporado al proceso penal y que puede ser valorado de forma correcta por los jueces, es así como la trazabilidad de los sistemas de inteligencia artificial es un requisito para la confiabilidad de la misma, puesto que si se tiene un registro de donde provienen los resultados se genera una certeza clara en el proceso.

El contexto de la inteligencia artificial como herramienta probatoria, esta no debe de sustituir el trabajo humano de verificación e investigación de las fuentes generadas por dicha herramienta, debido a que pone en riesgo la eficacia probatoria que se pretende incorporar a un proceso penal, debido a que no se puede garantizar la exactitud de lo generado por una inteligencia artificial, es ahí donde radica la importancia del control, verificación y auditabilidad humana para garantizar la transparencia y como consecuencia se garantice el derecho a la contradicción que es fundamental en todo proceso judicial.

Por su parte, Garagorry (2025), manifiesta que “En una prueba, las tasas de alucinaciones de los sistemas de IA más nuevos alcanzaron el 79%. Actualmente, el modelo más confiable con una tasa de alucinaciones del 0,7% es Gemini-2.0-Flash-001 de Google”, esta información evidencia que el margen de error que tiene una herramienta de inteligencia artificial es alto, lo que permea en la fiabilidad y exactitud de la inteligencia artificial como herramienta probatoria, debido a que al ser proporcionar datos frágiles y carentes de verificación desde el origen y trazabilidad de la información vulnera la seguridad jurídica, por lo cual, para eliminar estos riesgos, es indiscutible la presencia y responsabilidad humana.

2. Materiales y Métodos

El trabajo de investigación presenta un enfoque cualitativo, posee un carácter doctrinario y jurisprudencial tendiente al análisis crítico y descriptivo orientado al estudio jurídico sobre el papel de la IA como herramienta probatoria en un proceso penal y sus estándares necesarios para que este pueda ser valorado como un estándar probatorio adecuado sin que se viole el debido proceso, el derecho a la defensa, la presunción de inocencia, es por ello que este enfoque permite comprender de forma coherente los contenidos relacionados con el tema desde lo doctrinal, jurisprudencial y legal.

Para abordar la problemática en la investigación se emplearon los métodos analítico-sintético e inductivo y comparado porque este permite desglosar los elementos doctrinarios y jurisprudenciales relacionados con el rol de la inteligencia artificial como medio probatorio dentro del derecho penal nacional como internacional, para que de esta manera sean integrados y comprender sistemáticamente los desafíos, limitaciones y retos que se presentan dentro de un proceso penal para obtener conclusiones sobre la validez, admisibilidad, fiabilidad y límites jurídicos en el derecho penal contemporáneo. El alcance de la investigación es de tipo descriptivo, porque permite identificar y describir los riesgos, incidencia, desafíos y vacíos normativos sobre el uso de la inteligencia artificial como un medio probatorio en el derecho penal y los riesgos para los estándares de admisibilidad y control judicial.

El diseño del estudio es no experimental, toda vez que, la investigación se centra en el análisis de información doctrinal y jurisprudencial, lo que guarda coherencia y relación con el carácter analítico teórico del estudio. Las técnicas de investigación que fueron empleadas en la investigación fueron el análisis dogmático jurídico y revisión bibliográfica, orientada al estudio de la literatura jurídica sobre la inteligencia artificial desde el estándar probatorio del derecho penal. Los instrumentos para la recolección de información fueron las fichas bibliográficas e inventarios sistematizados de fuentes para la selección de la información a partir de criterios de

relevancia dogmática, aporte conceptual sustantivo y disponibilidad íntegra del texto.

3. Resultados

Los resultados advierten que la fiabilidad probatoria de la inteligencia artificial depende de variables críticas como la calidad de los datos, la transparencia del algoritmo, la trazabilidad del procesamiento y la verificabilidad de los resultados. Según San Miguel (2023) en el ámbito penal, estas condiciones son esenciales para cumplir los estándares de autenticidad, integridad y pertinencia que rigen la prueba científica. La revisión revela que tecnologías como el reconocimiento facial, los modelos predictivos, el análisis automatizado de metadatos y los sistemas de machine learning presentan un potencial significativo para apoyar la identificación de patrones, la clasificación de evidencias y la reconstrucción de hechos, siempre que existan protocolos de auditoría, documentación técnica verificable y control pericial independiente. Sin embargo, la literatura muestra que la ausencia de explicabilidad especialmente en algoritmos de caja negra reduce la posibilidad de contradicción y limita su aceptación judicial.

La fiabilidad de la IA como apoyo probatorio enfrenta desafíos derivados de sesgos algorítmicos, sobreajuste de modelos, errores en bases de entrenamiento y riesgos de manipulación digital, como deepfakes o alteraciones imperceptibles en archivos electrónicos. Estas vulnerabilidades comprometen la autenticidad de la evidencia y pueden generar falsos positivos o interpretaciones distorsionadas, afectando directamente principios como presunción de inocencia, igualdad de armas y valoración racional de la prueba. Pese a ello, la revisión muestra que, cuando la IA se emplea como herramienta auxiliar y no como sustituto del criterio judicial ofrece beneficios claros en eficiencia, precisión analítica y capacidad para procesar grandes volúmenes de información. De esta manera, la literatura coincide en que la IA puede integrarse al proceso penal como prueba científica complementaria, siempre que se garantice un marco adecuado de transparencia, control técnico y supervisión humana significativa.

La doctrina especializada ha documentado que los sesgos algorítmicos, frecuentemente originados en conjuntos de datos de entrenamiento incompletos o atravesados por patrones históricos de discriminación, pueden reproducir desigualdades sociales preexistentes generando errores sistemáticos en la identificación o evaluación de las personas sometidas a investigación penal, lo que influye de manera central en el núcleo garantista de los procesos penales, llegando a afectar derechos fundamentales como la no discriminación.

A esta problemática se añade el desafío creciente que supone la manipulación sintética de evidencias como los deepfakes, fenómeno que tensiona de forma

significativa los mecanismos clásicos de autenticación, en la cadena de custodia y la valoración pericial de la prueba (Planchadell-Gargallo, 2024). Con este contexto es notable que el riesgo no se agota en la eventualidad del error técnico, sino que se proyecta en la posibilidad de que decisiones penales se sustenten en pruebas opacas, de difícil refutación o socialmente legitimadas por una apariencia de neutralidad tecnológica, lo que obliga a reforzar los estándares de admisibilidad desde una lectura estricta del debido proceso en el marco de la presunción de inocencia (Guamani-Jame et al., 2026).

Pese al avance doctrinal, los vacíos que justifican el estudio persisten; la naturaleza jurídica del resultado algorítmico, la continua oscilación entre pericia, indicio, evidencia digital autónoma o simple herramienta auxiliar. A esto se suma como problema estructural la responsabilidad y el control del riesgo cuando un sistema se integra como soporte institucional, los errores generados dejan de ser excepciones técnicas, convirtiéndose en fallos con un alto impacto penal, entonces resulta indispensable dentro del proceso que todas las partes procesales puedan auditar como un ejercicio de igualdad de armas y de contradicción las pruebas generadas por la inteligencia artificial.

Esta investigación concibe la hipótesis de que la IA tiene un potencial considerablemente alto como lo muestra la literatura actual en funciones como optimización de la gestión probatoria, sin embargo, este impacto es jurídicamente aceptable solo si se acompaña de salvaguardas verificables como la transparencia suficiente para el contradictorio, el control humano significativo y los diferentes criterios de admisibilidad que garanticen la incorporación de tasas de error, sesgos y trazabilidad. Dicho de forma directa, el sistema puede beneficiarse de herramientas que iluminen patrones, pero el proceso penal no puede renunciar a que cada parte comprenda qué se afirma, cómo se afirma y con qué grado de incertidumbre, porque allí se juega la legitimidad de la condena o la absolución

La valoración judicial de la evidencia asistida por inteligencia artificial demanda una atención especial a los riesgos probatorios que acompañan su uso, entre los que destacan los sesgos algorítmicos, los errores sistemáticos y la manipulación sintética de contenidos digitales. Estos riesgos afectan directamente la credibilidad jurídica de la prueba y pueden generar decisiones penales fundadas en resultados técnicamente sofisticados, pero jurídicamente frágiles. La proliferación de deepfakes y contenidos generados por IA compromete los mecanismos tradicionales de autenticación y obliga a reforzar los criterios de valoración judicial. La fragilidad de las pruebas digitales incorporadas a través de la IA, supone una exigencia a los administradores de justicia en donde deben de adoptar parámetros de valoración probatoria más estrictos, desde el cumplimiento de la cadena de

custodia, la integridad del archivo y la periferia con otros medios probatorios. Así, la inteligencia artificial no redefine el estándar de convicción penal, sino que impone una valoración judicial más prudente, crítica y consciente de los márgenes de incertidumbre tecnológica.

4. Discusión

La inteligencia artificial en el derecho penal genera grandes riesgos éticos y legales, que para que la aplicación de la inteligencia artificial como herramienta probatoria debe de garantizarse un uso confiable y justo, con un marco regulatorio clave que proteja el debido proceso y derechos fundamentales que aseguren el uso ético de la IA en el ámbito legal (Martínez, et al., 2025). Si bien, la inteligencia artificial en el ámbito jurídico sirve como una herramienta de apoyo para predecir los posibles escenarios legales y poder brindar una solución adecuada, no obstante, la falta de juicio moral y ético de la IA implica que funcione por patrones aprendidos, sin entender el contexto específico en cada caso, teniendo como resultados resoluciones judiciales automatizadas, relegando la valoración crítica humana que es fundamental para los sistemas de derechos.

En ese contexto, el riesgo de sesgo sistemático en sistemas de justicia debilitados a causa de la corrupción y desigualdades sociales, genera un escenario propicio para que los sesgos algorítmicos tengan consecuencias especialmente graves. (Calderón & Guaira Ramos, 2025). En Estados Unidos, los sesgos algorítmicos del sistema de sus siglas COMPAS ((Correctional Offender Management Profiling for Alternative Sanctions), que es ampliamente utilizado como una herramienta para prevenir y evaluar los riesgos de residencia de los procesados, es ampliamente criticado por la comunidad jurídica porque se ha evidenciado tendencias discriminatorias por sesgos raciales, lo que afecta particularmente a grupos racializados (Angwin et al., 2016) lo que incide de manera directa en la presunción de inocencia y la imparcialidad.

Mientras que, en Reino Unido, con su sistema de predicción de la criminalidad “PredPol”, que basado en información histórica determina los lugares donde es más alta la tasa de criminalidad, razón por la cual, ha generado críticas respecto a su uso, puesto que se centre en comunidades de bajos recursos lo que se aleja del contexto social afectando a minorías étnicas (Loaiza-Moreno, et al., 2024). La inteligencia artificial puede ser considerada como una herramienta técnica para brindar apoyo a peritos, fiscales, abogados, jueces, que procesa volúmenes grandes de información y de evidencia, no obstante, no puede ser considerado como un operador jurídico, por los riesgos que implica, la literatura contemporánea advierte que la opacidad algorítmica presente en la inteligencia artificial, dificulta la comprensión del origen de ciertas conclusiones que pueden afectar la valoración de

las pruebas y su legitimidad. El problema de estos algoritmos, es que no puede ser explicada de forma lógica, lo que provoca opacidad en su proceso de uso reduciendo la confianza en el proceso judicial (Quadra-Salcedo, 2019).

En la misma línea, García-Bernalt (2024) subraya que la opacidad de los algoritmos de “caja negra” constituye uno de los principales obstáculos para su admisibilidad, al impedir que las partes y el juez comprendan cómo se genera el resultado presentado como prueba. De forma complementaria, Medina-Martín (2024) argumenta que la evidencia algorítmica solo puede ser aceptada en el proceso penal si se somete a estándares reforzados de científicidad, equivalentes o superiores a los exigidos a la prueba pericial tradicional. En consecuencia, la admisibilidad de la IA probatoria no depende de su innovación técnica, sino de su capacidad para cumplir exigencias jurídicas verificables.

En consecuencia, la discusión subraya la necesidad de estándares de admisibilidad que incluyan tasas de error, auditorías periódicas y validación por subgrupos, tal como recomiendan los estudios que analizan reconocimiento facial y justicia predictiva (Medina-Rojas, 2025). Los resultados revelan que, a pesar de la multiplicidad de funciones de la inteligencia artificial en el ámbito del derecho penal, sin embargo, su efectividad en la práctica se ve limitada por problemas de tipo éticos y legales relacionados con posible vulneración de derechos fundamentales que, sin la supervisión humana, la inteligencia artificial como herramienta de probatoria no es confiable en el derecho penal.

De los riesgos del uso de la IA una de las áreas de posible vulneración son los derechos humanos fundamentales, incluyendo la protección de datos, privacidad y no discriminación, presente en los sesgos de datos y algoritmos, pues la IA se alimenta de datos que se introducen, y si estos datos son erróneos, incorrectos o sesgados el análisis de dicha información tendría como resultado información tergiversada o falsa (Muñoz, 2020).

Otro de los riesgos que se identificó en el uso de la inteligencia artificial como herramienta probatoria fue, la ausencia de comprensión del texto y falta de empatía, lo que sugiere que al brindar resultados automatizados por algoritmos y predicciones, no se analiza o se valora en contexto o las circunstancias de cada caso en particular, por lo cual, según Villasmil, et al., (2024) la IA carece de la capacidad para tener empatía y comprensión de los contextos que son necesarios para tomar decisiones jurídicas informadas y justas. En esa línea, los estudios revisados reafirman que la eficiencia técnica no legitima por sí sola la prueba algorítmica si no se acompaña de pericia independiente, documentación técnica y control judicial significativo, elementos que permiten preservar la motivación y la igualdad de armas (San Miguel-Caso, 2023).

La revisión destaca que la IA generativa y los deepfakes desplazan el problema hacia la autenticidad y la cadena de custodia, obligando a reforzar protocolos forenses digitales y criterios de valoración probatoria. En suma, la discusión integra la evidencia para sostener que la incorporación legítima de la IA requiere un marco garantista que armonice innovación con control humano y rendición de cuentas.

5. Conclusiones

El estudio permite concluir que las debilidades principales que se detectan en la inteligencia artificial radican principalmente en la poca fiabilidad de la información que estas generan, debido a que al no poder identificar la trazabilidad de los procesos automatizados que son características de esta herramienta, no existe certeza sobre la transparencia de las respuestas, pues además se ha detectado que el nivel de alucinación de la IA causa que se tergiverse, manipule y se obtengan información falsa como aspecto de real, puesto que los algoritmos que predicen dicha información tiene un nivel alto de margen de error, es por aquello que no puede aplicarse exclusivamente la inteligencia artificial como herramienta probatoria, cuando no existen las garantías para que produzca seguridad jurídica, sino más bien puede ser utilizada como una herramienta de apoyo para simplificar procesos siempre que la auditabilidad y verificación humana este presente.

La inteligencia artificial representa una herramienta de alto potencial para fortalecer la actividad probatoria en el proceso penal, en la medida en que facilita el análisis de grandes volúmenes de información, la identificación de patrones relevantes y la gestión más eficiente de la evidencia digital. Su incorporación no transforma la naturaleza garantista del derecho penal ni redefine por sí misma los estándares probatorios clásicos. La IA no constituye una prueba autónoma ni un mecanismo decisorio legítimo, sino un instrumento auxiliar cuya validez jurídica depende de su sometimiento a controles estrictos de fiabilidad, transparencia, trazabilidad con posibilidad efectiva de contradicción.

En este sentido, la legitimidad de la prueba asistida por IA se vincula directamente con la preservación del control humano y con la motivación racional de las decisiones judiciales, se concluye que los principales riesgos asociados al uso de IA sesgos algorítmicos, opacidad técnica, manipulación sintética de evidencias y fragilidad de la cadena de custodia digital exigen la reformulación de los criterios de admisibilidad y valoración probatoria desde una perspectiva penal garantista. El aporte central de esta investigación radica en sistematizar dichos riesgos proponiendo una aproximación teórico-metodológica que permita integrar la innovación tecnológica sin deshumanizar la justicia ni erosionar derechos fundamentales como la presunción de inocencia, la igualdad de armas y el derecho de defensa. En consecuencia, el uso legítimo de la IA en el ámbito penal requiere

marcos normativos claros, estándares periciales reforzados y una supervisión judicial activa que asegure que la eficiencia tecnológica nunca prevalezca sobre la justicia material.

CONFLICTO DE INTERESES

Indicar si existen intereses particulares por parte de los autores o de la entidad científica que pudiesen afectar directa o indirectamente a los resultados. Caso contrario de no existir conflictos ubicar “**Los autores declaran no tener ningún conflicto de intereses**”.

Referencias Bibliográficas

Abad Quevedo, C. (2025). *Inacción de la inteligencia artificial en el arbitraje en equidad: Un análisis crítico de sus limitaciones epistemológicas y prácticas en la resolución de conflictos*. Revista Ecuatoriana de Arbitraje, 15, 103–124. <https://doi.org/10.18272/rea.i15.3830>

AlgoRace. (2022). Una introducción a la IA y la discriminación algorítmica para movimientos sociales. <https://www.algorace.org/wp-content/uploads/2023/10/informe-algorace-2.pdf>

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica.

Calderón Zuleta, J. G., & Guaira Ramos, W. J. (2025). Inteligencia artificial en el procedimiento penal ecuatoriano y el análisis de la prueba en la administración de justicia. *Repositorio institucional*. <https://dspaceserver.ube.edu.ec/server/api/core/bitstreams/9a16b5d1-6df1-487f-8f15-bade2fa2979b/content>

Comisión Europea. (2020). *LIBRO BLANCO sobre la inteligencia artificial—un enfoque europeo orientado a la excelencia y la confianza*. <https://acortar.link/dYC6mW>

Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos. ¿Por qué deberían importarnos? Nueva Sociedad, 294, 27-37. <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>.

García-Bernalt, I. Y. (2024). La inteligencia artificial en el proceso penal: Eficiencia versus garantías. *IUS ET SCIENTIA*, 10(2), 80–100. <https://doi.org/10.12795/IESTSCIENTIA.2024.i02.04>

García-Falconí, R. J., & Barona-Pazmiño, K. (2024). Inteligencia artificial y proceso penal. *Revista San Gregorio*, 1(58), 101–110. <https://doi.org/10.36097/rsan.v1i58.2808>

González-García, A. (2025, marzo 6). Inteligencia artificial: Perspectiva criminológica europea y su impacto en las ciencias forenses. *10a Cumbre Mundial de Ciencias Jurídicas y Forenses*. Universidad Hispana. <https://udimundus.udima.es/handle/20.500.12226/2917>

Guamani-Jame, M., García-Segarra, H., & Herrera-Herrera, G. (2026). Responsabilidad penal de la Inteligencia Artificial: Relación con los conceptos de Dolo, la Culpa y Autoría | 593 Digital Publisher CEIT. https://www.593dp.com/index.php/593_Digital_Publisher/article/view/3717

Loaiza-Moreno, J. D., Soto-Soto, F. F., & Hoyos-Escaleras, Á. M. (2024). Revolucionando la Justicia: El impacto de la Inteligencia Artificial en el Derecho Penal. *Estudios y Perspectivas Revista Científica y Académica*, 4(3), 2155–2174. <https://doi.org/10.61384/r.c.a..v4i3.537>

Martínez, T., Llanos, R., Núñez, R., & Noboa, A. (2025). La inteligencia artificial en el derecho: transformaciones, desafíos éticos y perspectivas jurídicas. *Código Científico Revista de Investigación*, 6(1), 2044–2057. <https://doi.org/10.55813/gaea/ccri/v6/n1/1010>

Medina-Martín, E. (2024). La inteligencia artificial y su encuadre como medio de prueba. *El Criminalista Digital. Papeles de Criminología*, 12, 33–51. <https://doi.org/10.30827/cridi.31430>

Medina-Rojas, C. (2025). Robótica, Inteligencia Artificial y Derecho: Nuevas Dimensiones Jurídicas en el Siglo XXI. *Ciencia Latina Revista Científica Multidisciplinar*, 9(1), 9488–9513. https://doi.org/10.37811/cl_rcm.v9i1.16574

Muñoz Rodríguez, A. B. (2020). El impacto de la inteligencia artificial en el proceso penal = The impact of artificial intelligence on penal procedure. *Anuario de la Facultad de Derecho. Universidad de Extremadura*, 36, 695–728. <https://doi.org/10.17398/2695-7728.36.695>

Pereira-Puigvert, S. (2024). La justicia y el proceso penal en clave algorítmica. Nuevos enfoques, nuevos riesgos. *IUS ET SCIENTIA*, 10(2), 66–79. <https://doi.org/10.12795/UESTSCIENTIA.2024.i02.03>

Planchadell-Gargallo, A. (2024). Inteligencia artificial y proceso penal; especial referencia a las medidas cautelares. *Teoría & Derecho. Revista de pensamiento jurídico*, 37, 176–195. <https://doi.org/10.36151/TD.2024.110>

Quadra-Salcedo Fernández del Castillo, T. (2019). Derechos fundamentales, democracia y mercado en la edad digital. *Derecho Digital e Innovación*, (1). Madrid.

San Miguel-Caso, C. (2023). Inteligencia artificial y algoritmos: La controvertida evolución de la tutela judicial efectiva en el proceso penal. *Estudios Penales y Criminológicos*, 44(Ext.), 1–23. <https://doi.org/10.15304/epc.44.8859>

Villasmil Espinoza, J. J., Berrocal Duran, J. C., & Rodelo García, M. R. (2024). Inteligencia artificial y nuevas formas de derecho en el siglo XXI. *Revista de Ciencias Sociales*, 30(Número Especial 9), 447–458. <https://doi.org/10.31876/rcs.v30i.42324>