

Transparencia algorítmica y responsabilidad jurídica: retos de la IA en la valoración probatoria

Algorithmic transparency and legal liability: challenges of AI in evidentiary assessment

María Paola Ramos Santacruz¹ (paola0704ramos@gmail.com) (<https://orcid.org/0009-0007-6122-6184>)

Armando Rogelio Durán Ocampo² (aduran@utmachala.edu.ec) (<https://orcid.org/0000-0002-9524-0538>)

Resumen

El artículo tuvo como objetivo analizar los resultados investigativos referentes a la transparencia algorítmica y responsabilidad jurídica en la valoración de la prueba pericial asistida por IA, con el fin de proponer un estándar de admisibilidad, contradicción y valoración probatoria. Se realizó una revisión documental cualitativa, conforme a los lineamientos de PRISMA 2020, mediante la cual se identificaron y analizaron 12 estudios publicados entre 2020 y 2026. La síntesis permitió distinguir los usos de IA que están orientados a transformar, analizar o recomendar conclusiones sobre la evidencia, así como los riesgos asociados con la opacidad, el sesgo, la falta de trazabilidad, los errores de procesamiento y las asimetrías técnicas entre las partes. Los hallazgos muestran que la transparencia procesal exige, además de información sobre el sistema, documentación, explicaciones comprensibles, registros auditables y oportunidades reales de contradicción. A partir de estos elementos, se propone un modelo normativo-operativo de control probatorio para la evidencia asistida por IA, compuesto por cuatro fases: admisibilidad condicionada, transparencia contradictoria, valoración judicial motivada y atribución de responsabilidad. Se concluye que, cuando no exista posibilidad efectiva de control y contradicción, la salida algorítmica no debe constituir el fundamento único o decisivo de la decisión judicial y deberá someterse a la corroboración independiente o a las consecuencias procesales previstas por el ordenamiento aplicable.

Palabras clave: Inteligencia artificial, transparencia algorítmica, responsabilidad jurídica, valoración probatoria, prueba digital.

Abstract

The article analyzes the requirements of algorithmic transparency and legal liability applicable to artificial intelligence-assisted evidence in evidentiary activity. A qualitative documentary review was conducted, in accordance with the PRISMA 2020 guidelines, through which 12 studies published between 2020 and 2026 were identified and analyzed. The synthesis made it possible to distinguish the uses of AI aimed at

¹ Estudiante de la carrera Derecho. Universidad Técnica de Machala. Ecuador.

² Magister en Ciencias Jurídicas. Coordinador de Titulación de la carrera de Derecho. Coordinador de la Maestría en Derecho, Mención Derecho Penal y Criminología. Profesor Titular de la Universidad Técnica de Machala, Ecuador.

transforming, analyzing, or recommending conclusions about evidence, as well as the risks associated with opacity, bias, lack of traceability, processing errors, and technical asymmetries between the parties. The findings show that procedural transparency requires, in addition to information about the system, documentation, understandable explanations, auditable records, and real opportunities for adversarial challenge. Based on these elements, a normative-operational model of evidentiary control for AI-assisted evidence is proposed, consisting of four phases: conditional admissibility, adversarial transparency, reasoned judicial assessment, and attribution of liability. It is concluded that, where no effective possibility of control and adversarial challenge exists, the algorithmic output must not constitute the sole or decisive basis of the judicial decision and must be subject to independent corroboration or to the procedural consequences provided for by the applicable legal order.

Key words: Artificial intelligence, algorithmic transparency, legal accountability, evidentiary assessment, digital evidence.

Introducción

La acelerada expansión de herramientas de inteligencia artificial (IA) en la actividad probatoria ha transformado la manera en que se clasifican los documentos en la Ciencia del Derecho. Con el uso de la IA se identifican patrones, se analizan imágenes, se realizan reconocimientos biométricos, se estiman riesgos, se detecta la ocurrencia o no de algún fraude y se gestionan grandes volúmenes de evidencia digital. Esta incorporación tecnológica no es homogénea ni neutral. Pues tal y como advierten Nuredin e Inan (2024), aunque los sistemas algorítmicos se perciben como imparciales, en realidad son programados por humanos y están entrenados con datos provenientes de comunidades humanas, lo que puede reproducir sesgos e injusticias. Estudios recientes (Segovia y Flores, 2025; Mendoza y Carrera, 2026) coinciden en que la IA se inserta en el proceso judicial como un soporte técnico orientado a la obtención, depuración, correlación y presentación de información, pero su legitimidad depende de su subordinación a las reglas procesales y no de su sofisticación tecnológica.

Es posible distinguir, al menos, tres modalidades funcionales de IA en la actividad probatoria. La primera comprende la IA que produce o transforma la evidencia —transcripción automática, mejora de imagen, generación de resúmenes—, cuya admisibilidad exige verificar la fidelidad del proceso de transformación y la trazabilidad de la fuente original (Aguilar, 2024). La segunda abarca la IA que analiza evidencia existente —reconocimiento facial, detección de anomalías, clasificación documental—, donde el riesgo se desplaza hacia la opacidad de los criterios de selección y jerarquización de la información (Altamirano et al., 2026). La tercera modalidad corresponde a la IA que recomienda o predice una conclusión relevante para la decisión judicial, cuyo uso plantea la tensión más aguda con el principio del juez natural predeterminado por la ley y con la prohibición de delegación de funciones jurisdiccionales (Suárez, 2025).

La doctrina ha advertido que la incorporación de IA en la valoración probatoria no es un fenómeno futuro ni hipotético, sino una realidad en expansión que plantea profundas reflexiones epistémicas y jurídicas. En este sentido, Isaza et al. (2026) sostienen que la valoración de la prueba testifical exige un enfoque interdisciplinario que integre la epistemología jurídica, la psicología del testimonio y las garantías procesales, y advierten que la delegación de la valoración de la prueba testifical a sistemas de IA resulta epistemológica y psicológicamente problemática. Pues la IA opera con datos y correlaciones, pero carece de una comprensión semántica, de la intuición humana, la empatía y el sentido ético, que son necesarios tener en cuenta en estos procesos. Desde la perspectiva de Isaza et al. (2026), la IA puede cumplir una función auxiliar orientada a fortalecer la racionalidad del proceso, pero no puede reemplazar el juicio de ponderación, el análisis crítico y la responsabilidad ética y jurídica inherentes a la función jurisdiccional.

El problema jurídico central radica en que una salida algorítmica puede parecer objetiva por su apariencia matemática, aunque dependa de datos incompletos; tales como los criterios de diseño, los umbrales estadísticos, errores, sesgos o configuraciones que las partes no pueden conocer (Vestri, 2021). Esta apariencia de neutralidad tecnológica oculta asimetrías epistémicas que comprometen el control procesal de la prueba. Por lo que Serna (2025) advierte que la incorporación de IA en esta jurisdicción exige revisar los estándares de admisibilidad y los mecanismos de control probatorio para evitar que la opacidad algorítmica vulnere el derecho de defensa.

En este orden de ideas, Carrasco (2025) señala el modelo no sabe realmente lo que significa el texto ni tiene contexto real; lo único que hace es predecir la palabra siguiente más probable basándose en su historial de patrones, todo lo cual evidencia un déficit de sentido común y una incapacidad para realizar razonamiento abductivo. A ello se suma la tensión existente entre la transparencia algorítmica y los secretos legítimos, entre los que se pueden mencionar: la propiedad intelectual, el secreto empresarial, la ciberseguridad, la protección de datos personales y la reserva de investigaciones, que restringen el acceso al código fuente, a los datos de entrenamiento o a los registros de inferencia (Vestri, 2021).

En estos casos la consecuencia procesal es directa, si la parte afectada no puede comprender ni impugnar el resultado algorítmico, se debilitan el derecho de la defensa, el principio de contradicción, la igualdad de armas entre las partes, la motivación judicial y, en última instancia, el debido proceso (Mendoza y Carrera, 2026). 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 (Nuredin e Inan, 2024). En otras palabras, la IA no solo plantea desafíos técnicos, sino que impacta directamente en el núcleo garantista del proceso judicial, exigiendo marcos de transparencia, auditabilidad y proporcionalidad que aseguren la plena vigencia de los derechos humanos fundamentales.

Por ello, el presente artículo tiene como objetivo analizar los resultados investigativos referentes a la transparencia algorítmica y responsabilidad jurídica en la valoración de la prueba pericial asistida por IA, con el fin de proponer un estándar de admisibilidad, contradicción y valoración probatoria. Para su consecución se siguió el siguiente procedimiento: (i) se identificaron las aplicaciones de IA estudiadas en la actividad probatoria; (ii) se sistematizaron los riesgos técnicos y jurídicos asociados; (iii) se clasificaron los mecanismos de transparencia y auditabilidad existentes; (iv) se examinaron los modelos de responsabilidad aplicables; y (v) se diseñó un modelo como propuesta de control probatorio. La relevancia del estudio radica en que, como sostienen Kocagül y Kartal (2026), la arquitectura procedimental de los sistemas adversariales y mixtos fue construida sobre la premisa de que la evidencia proviene de actores humanos cuyo razonamiento puede reconstruirse y cuestionarse. La IA al ingresar como productora de evidencia, evaluadora de riesgos y configuradora de decisiones, rompe esa premisa y exige salvaguardas específicas que este trabajo busca sistematizar y proponer.

Metodología

La presente investigación se desarrolló bajo el paradigma cualitativo, siguiendo la ruta de la investigación cualitativa propuesta por Hernández y Mendoza (2023), cuyo propósito es comprender fenómenos complejos desde la perspectiva de los actores y los contextos normativos en los que se insertan. De acuerdo con estos autores, el enfoque cualitativo resulta adecuado cuando el investigador busca examinar significados, interpretaciones y procesos en profundidad, sin pretensión de generalización estadística. En coherencia con ello, se adoptó un diseño no experimental de corte transversal, dado que no se manipuló variable alguna y el análisis se concentró en la revisión documental de las fuentes normativas, doctrinales y jurisprudenciales disponibles en un momento determinado.

La investigación tiene un alcance descriptivo y analítico, orientada a identificar, sistematizar y clasificar los riesgos técnicos y jurídicos derivados del uso de IA en la actividad probatoria, así como a examinar los mecanismos de transparencia, auditabilidad y responsabilidad, aplicables.

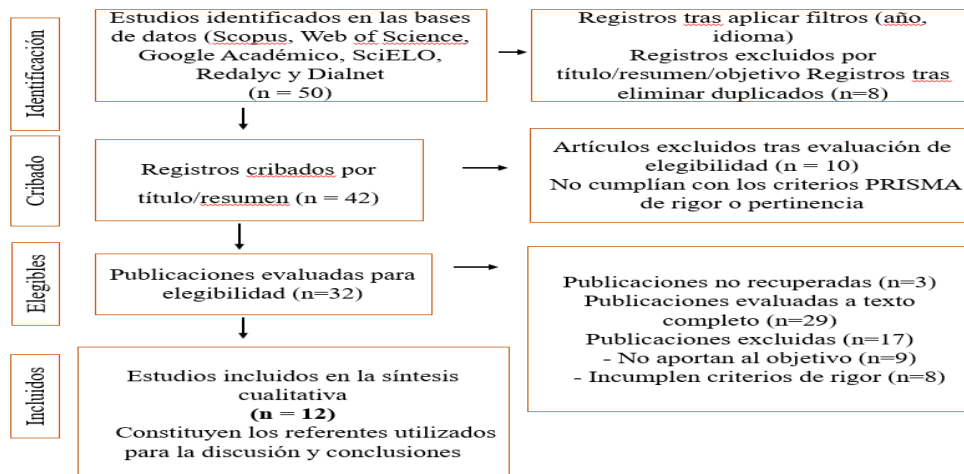
La estrategia de búsqueda y selección de fuentes se rigió por los estándares de la declaración PRISMA 2020 (Page et al., 2022), que establece una guía de 27 elementos para el reporte transparente de revisiones sistemáticas, con el fin de asegurar la reproducibilidad y minimizar los sesgos de selección. El proceso se estructuró en cuatro fases: identificación, cribado, elegibilidad e inclusión. En la fase de identificación se realizaron búsquedas sistemáticas en bases de datos académicas (Scopus, SciELO, Dialnet y Semantic Scholar), utilizando las correspondientes palabras clave de búsqueda (inteligencia artificial, justicia algorítmica, evidencia algorítmica, evaluación de evidencia, valor probatorio, debido proceso, transparencia, responsabilidad, admisibilidad).

La calidad metodológica de los estudios incluidos se evaluó mediante una ficha de análisis estructurada que registró autoría, año, tipo de publicación, metodología declarada, país de origen y aportes sustantivos a los ejes temáticos del estudio. Los hallazgos se sintetizaron mediante un análisis temático de contenido, organizado en las categorías previamente definidas. La combinación de la ruta cualitativa de Hernández y Mendoza (2023) con los estándares de reporte de PRISMA 2020 (Page et al., 2022) contribuyó a que el proceso de revisión fuera transparente, replicable y auditable, asegurando así que las conclusiones derivadas del análisis estuvieran sustentadas en una literatura pertinente y de calidad verificable, tal como exige el rigor metodológico en investigaciones jurídicas de naturaleza documental.

La representación gráfica del proceso de selección constituye el elemento central de la revisión sistemática, pues permite visualizar el tránsito desde el universo inicial de registros (50) hasta el corpus definitivo de análisis (12). El diagrama que se presenta a continuación (figura 1) documenta las cuatro fases del protocolo PRISMA 2020 — identificación, cribado, elegibilidad e inclusión—, especificando en cada etapa el número de registros y las razones de exclusión aplicadas. Su inclusión responde a la exigencia de transparencia y replicabilidad que caracteriza a este estándar metodológico, y permite analizar con precisión el itinerario seguido para la conformación de la muestra final.

Figura 1.

Flujo PRISMA (2020)



Fuente: Elaboración propia.

Dada la heterogeneidad de los estudios incluidos, no se realizó un metaanálisis; en su lugar, se aplicó una síntesis temática cualitativa organizada en cinco categorías: (i) usos de IA en el ciclo probatorio; (ii) riesgos técnicos y epistémicos; (iii) garantías procesales afectadas; (iv) transparencia, explicabilidad, auditabilidad y contradicción; y (v) modelos de responsabilidad. Esta organización permitió identificar patrones recurrentes y vacíos

en la literatura, así como contrastar los hallazgos con los estándares normativos internacionales y nacionales. La síntesis no buscó cuantificar la frecuencia de los riesgos, sino comprender su naturaleza y su incidencia en la actividad probatoria.

El cruce de las categorías permitió construir una matriz analítica que relaciona el tipo de IA, el riesgo central, la garantía afectada, la exigencia mínima y la consecuencia procesal en caso de incumplimiento. En el caso del reconocimiento facial, el riesgo central radica en los falsos positivos y el sesgo algorítmico, que afectan el derecho de defensa y la presunción de inocencia; la exigencia mínima comprende la información sobre tasas de error, la calidad de la imagen, la auditoría independiente y el peritaje contradictorio, y su incumplimiento debe conducir a que el resultado no pueda ser utilizado como prueba única o decisiva.

En la clasificación documental, el riesgo es la omisión o priorización errónea de documentos, lo que compromete la igualdad de armas y el acceso a la prueba; se exige el registro de los criterios de selección, la revisión humana y la preservación de los documentos descartados, y su incumplimiento obliga a revisar la fase de descubrimiento probatorio. En el análisis de video o audio, el riesgo es la alteración, la pérdida de contexto o la generación de alucinaciones, afectando la autenticidad e integridad de la evidencia; se requiere el archivo original, los metadatos, la identificación de la herramienta y su versión, así como un peritaje técnico, y su incumplimiento conlleva la exclusión o una reducción intensa del peso probatorio. En los sistemas de predicción o puntuación de riesgo, el riesgo central es el sesgo y la automatización indebida, que afectan la individualización, la motivación y la no discriminación; se exige una explicación comprensible, validación local y la prohibición de dependencia automática, y su incumplimiento impide que el resultado se utilice como fundamento decisivo de la resolución judicial.

Esta matriz evidencia que no todo uso de IA plantea el mismo nivel de riesgo ni exige las mismas garantías. La intensidad de las exigencias procesales debe graduarse en función de la autonomía del sistema, su incidencia en la formación de la convicción judicial y la posibilidad efectiva de contradicción. La sistematización confirma, además, que la literatura revisada coincide en la necesidad de un estándar de admisibilidad reforzado para la evidencia asistida por IA, que combine transparencia por niveles, auditoría independiente, supervisión humana efectiva y mecanismos claros de atribución de responsabilidad.

El estudio teórico se articula en torno a la tensión entre la IA como herramienta de apoyo y la IA como fuente de autoridad probatoria aparentemente objetiva. Esta tensión atraviesa los tres ejes del estudio: transparencia algorítmica, responsabilidad jurídica y valoración probatoria. La literatura revisada coincide en que la IA no sustituye la función valorativa del juez, pero sí puede influir de manera significativa en la formación de la convicción judicial si no se establecen controles procesales adecuados (Suárez, 2025; Carrasco, 2025).

Del proceso de sistematización se determinó un primer hallazgo teórico y es que la apariencia matemática de los resultados algorítmicos oculta asimetrías epistémicas. Como advierte Vestri (2021), la opacidad de los algoritmos —el fenómeno de la "caja negra"— dificulta que las partes conozcan o cuestionen los criterios utilizados para arribar a un resultado. Esta opacidad no es solo técnica, sino jurídica, pues afecta el derecho de contradicción y la igualdad de armas, pues una parte puede disponer de peritos especializados mientras la otra carece de ellos. En esta misma línea, Isaza et al. (2026) aportan una precisión epistémica fundamental, la apariencia matemática o estadística de los resultados algorítmicos generan una "ilusión de precisión" que invisibiliza los márgenes de error y las limitaciones cognitivas del sistema. Estos autores advierten que la IA opera con datos y correlaciones, pero carece de comprensión semántica, intuición, empatía y sentido ético, elementos esenciales para interpretar y ponderar pruebas como el testimonio.

Esta carencia no es un defecto menor, sino que constituye una asimetría epistémica estructural, el sistema puede producir un resultado numérico aparentemente objetivo, pero no puede explicar por qué ese resultado es correcto en términos de significado, contexto o justicia. La opacidad algorítmica, en consecuencia, no solo dificulta la contradicción procesal, sino que introduce un déficit de justificación que tensiona el deber de motivación judicial y la exigencia de una valoración racional de la prueba. La transparencia, por tanto, no puede reducirse a la publicación del código fuente, sino que debe entenderse como una verificabilidad procesal, es decir, la posibilidad efectiva de que las partes comprendan el resultado y lo impugnen (Nuredin e Inan, 2024).

Un segundo hallazgo es que la responsabilidad jurídica en el uso de IA probatoria no puede atribuirse a un único actor. La cadena de responsabilidad humana e institucional involucra al proveedor o desarrollador, a la institución que adquiere el sistema, al operador o perito que lo utiliza, a la parte que aporta la evidencia, al juez o tribunal que la valora y al Estado o poder judicial que la regula y debe supervisarla. Cada actor asume deberes específicos en función del riesgo que se introduce (Mendoza y Carrera, 2026). La responsabilidad difusa, lejos de diluir la culpa, exige mecanismos claros de atribución y rendición de cuentas. Un tercer hallazgo es que los derechos y las garantías comprometidos (debido proceso, defensa, contradicción, igualdad de armas, presunción de inocencia, tutela judicial efectiva, motivación, igualdad y no discriminación, privacidad y acceso a la justicia) configuran un núcleo garantista que la IA puede quebrantar si se utiliza sin los controles humanos y responsabilidad jurídica (Bermudez et al., 2025).

Resultados

Fundamentos teórico-jurídicos y comparados

La valoración racional de la prueba constituye el eje sobre el cual se articula la legitimidad de la decisión judicial. En los sistemas procesales contemporáneos, la sentencia debe estar motivada y sustentada en inferencias controlables, de modo que las partes puedan reconstruir el razonamiento del juzgador y, en su caso, impugnarlo.

Un resultado algorítmico, por sofisticado que parezca, no reemplaza la obligación del juez de justificar por qué considera fiable una evidencia y qué peso específico le otorga en el conjunto del material probatorio.

Como sostiene Suárez (2025), la apreciación de la prueba compete al juez dentro de los límites de la prueba tasada y bajo el principio de la sana crítica, lo que implica que las deducciones lógicas deben ser razonadas y no arbitrarias. La intervención de la IA puede auxiliar en la organización y clasificación de la información, pero no puede sustituir la operación intelectual mediante la cual el juzgador determina la eficacia de los medios de prueba practicados (Altamirano et al., 2026).

La tensión entre la IA como herramienta de apoyo y la IA como fuente de autoridad probatoria aparentemente objetiva se comprende mejor si se distinguen los tipos de prueba que pueden verse involucrados. No todo uso de IA plantea el mismo nivel de riesgo ni exige las mismas garantías procesales. Es posible diferenciar, al menos, cinco categorías: (i) prueba tradicional digitalizada, que consiste en un documento o evidencia originalmente físico que ha sido escaneado o convertido a formato digital, donde la IA puede intervenir en tareas de reconocimiento óptico de caracteres o indexación; (ii) prueba digital generada en un entorno electrónico desde su origen —correos, metadatos, registros de sistemas—, cuya autenticidad e integridad dependen de la cadena de custodia digital (Aguilar, 2024).

Por otra parte, la (iii) prueba generada por IA, que comprende contenidos creados o transformados por sistemas algorítmicos, como transcripciones automáticas, mejoras de imagen o resúmenes, donde el riesgo principal radica en la fidelidad del proceso de transformación y en la verificabilidad del resultado; (iv) prueba analizada, clasificada o priorizada por IA, en la que el sistema opera sobre evidencia preexistente para identificar patrones, detectar anomalías o jerarquizar documentos, sin producir nuevo contenido probatorio, pero introduciendo criterios de selección que pueden afectar la igualdad de armas (Altamirano et al., 2026); y la (v) decisión o recomendación algorítmica usada como insumo probatorio, donde el sistema no solo analiza, sino que predice o recomienda una conclusión relevante para la decisión judicial, planteando la tensión más aguda con el principio del juez natural predeterminado por la ley (Suárez, 2025). Esta clasificación permite advertir que la intensidad de las garantías exigibles varía en función del nivel de autonomía del sistema y de su incidencia en la formación de la convicción judicial.

La literatura especializada ha sistematizado los principios éticos que deben regir la incorporación de la IA en el ámbito jurídico. Por lo que Crespo (2025) señala que la transparencia implica que los procesos y decisiones que han sido tomados por los sistemas de IA “sean comprensibles y accesibles para todos los usuarios, lo cual es crucial en el ámbito de la justicia para garantizar que las decisiones automatizadas puedan ser auditadas y entendidas por las partes involucradas” (p. 2). A ello añade que la equidad exige que los algoritmos sean diseñados y entrenados con datos que representen adecuadamente a todas las poblaciones, evitando con ello la perpetuación

de desigualdades existentes, y que se implementen mecanismos de supervisión humana para así poder garantizar que las decisiones sean justas y equitativas.

En este sentido, Crespo (2025) identifica cinco principios éticos clave, que se analizan y adaptan a esta investigación; el primero de ellos la responsabilidad: siempre debe haber una persona capacitada técnicamente que dirija y que a su vez responda por las decisiones que sean tomadas por la IA. La explicabilidad implica que las decisiones deben ser lo más claras, técnicas y el porqué de cada decisión. La privacidad y seguridad: los sistemas deben proteger la privacidad de los datos, por ser netamente sensibles y contener información que puede ser sensible a los ataques informáticos. Otro principio ético clave es la inclusión: dado que los sistemas deben ser inclusivos a todos y todas y no discriminatorios. El logro de poder implementar estos principios trae consigo, que se logre la confianza en los sistemas de IA, y se puede contribuir a que se “utilicen de manera justa y responsable en la administración de justicia” (p. 2).

En el ámbito procesal, la transparencia tiene una finalidad específica, hacer posible el contradictorio. De poco sirve que se divulgue el código si las partes no cuentan con los elementos necesarios para comprender cómo se llegó a un resultado concreto y, sobre esa base, ejercer su derecho de defensa. Como señala Vestri (2021), la transparencia algorítmica debe registrarse conforme al conocimiento omnicomprendivo del algoritmo, incluyendo sus autores, el mecanismo de decisión y los datos considerados relevantes, y sobre la no exclusividad de la decisión algorítmica cuando esta produce efectos significativos sobre el interesado. Por otra parte, la opacidad, o el fenómeno de la “caja negra”, dificulta que las partes puedan conocer o cuestionar los criterios utilizados, afectando el derecho de contradecir la valoración (Altamirano et al., 2026).

Para superar esta limitación, resulta útil proponer una transparencia por niveles, cada uno con un contenido mínimo y una función procesal diferenciada. En el nivel informativo, se exige el aviso de uso de IA, la identidad del sistema, su finalidad y el responsable de su implementación; su función es evitar el uso oculto de herramientas algorítmicas que puedan pasar desapercibidas para las partes. En el nivel documental, se requiere información sobre la versión del sistema, el proveedor, los datos de entrada, las métricas de desempeño y las limitaciones conocidas; permite a los operadores jurídicos conocer el sistema y evaluar su pertinencia.

En el nivel explicativo, se demanda una exposición de las razones comprensibles que justifican el resultado concreto, de modo que el juez pueda motivar su decisión y las partes puedan discutirla. En el nivel auditable, se exige acceso a registros de ejecución (logs), trazabilidad del procesamiento, pruebas de desempeño y posibilidad de auditoría independiente; su función es permitir la verificación técnica del sistema. En el nivel contradictorio, se garantiza el acceso efectivo de la defensa o contraparte a los elementos necesarios para impugnar el resultado algorítmico, haciendo real la igualdad de armas. Esta última dimensión es la que conecta directamente la transparencia con el debido proceso, pues sin posibilidad efectiva de contradicción, la transparencia se reduce a una declaración formal sin consecuencias procesales (Nuredin e Inan, 2024).

La pregunta por la responsabilidad en el uso de IA en la valoración probatoria no debe formularse en términos de si “la IA es responsable”, esto sería atribuirle valores humanos. La IA no es un sujeto jurídico y, por tanto, no sustituye a los sujetos jurídicamente responsables. Lo que corresponde analizar es una cadena de responsabilidad humana e institucional, en la que cada actor asume deberes específicos en función del riesgo que introduce. Como advierten Mendoza y Carrera (2026), la responsabilidad difusa en las decisiones asistidas por IA es un fenómeno que disminuye la intención de rendir cuentas, precisamente porque la participación de múltiples actores como los usuarios, o revisores, puede tender a diluir la culpabilidad.

El proveedor o desarrollador asume el riesgo de diseño defectuoso, documentación insuficiente, sesgo algorítmico o falta de pruebas de validación. Su deber comprende documentar el sistema, evaluar riesgos, registrar versiones y advertir sobre sus límites. La institución que adquiere o implementa el sistema —poder judicial, fiscalía, tribunal— asume, por tanto, el riesgo de compra irreflexiva, falta de controles internos o dependencia excesiva del proveedor; su deber incluye realizar evaluaciones de impacto, establecer cláusulas de auditabilidad en la contratación y garantizar la supervisión humana efectiva (Vestri, 2021).

El operador o perito que utiliza el sistema asume entonces, el riesgo de uso incorrecto, mala interpretación de resultados u omisión de los límites del modelo; su deber comprende mantener la competencia técnica, preservar registros de uso y elaborar informes claros y verificables. La parte que aporta la evidencia generada o analizada por IA asume el riesgo de falta de autenticidad, información incompleta o selección interesada de resultados; su deber, por tanto, es acreditar el origen, la integridad, el método utilizado y los márgenes de error (Aguilar, 2024).

El juez o tribunal asume el riesgo de delegación indebida, valoración acrítica o falta de motivación; su deber se traduce en el control de la admisibilidad, la debida explicación del peso probatorio otorgado y la revisión humana efectiva de todo resultado algorítmico (Suárez, 2025). Por consiguiente, el Estado o poder judicial asume el riesgo de ausencia de regulación, protocolos y supervisión; su deber comprende la gobernanza, auditorías, la formación de operadores y los mecanismos necesarios de reparación frente a errores judiciales derivados del uso de IA (Bermudez et al., 2025).

La incorporación de IA en la actividad probatoria compromete un conjunto de derechos y garantías que constituyen el núcleo del debido proceso. El debido proceso exige que toda decisión judicial sea adoptada conforme a los procedimientos preestablecidos y con respeto de las garantías esenciales; la opacidad algorítmica puede deteriorar este derecho al impedir que las partes comprendan el fundamento de la decisión (Nuredin e Inan, 2024). El derecho de defensa se ve afectado cuando la parte no puede acceder a los elementos necesarios para controvertir la evidencia algorítmica, ya sea porque el código es secreto, porque los datos de entrenamiento no son accesibles o porque los criterios de inferencia no están documentados. El principio de contradicción exige que las partes tengan la oportunidad de cuestionar toda prueba que se pretenda utilizar en

su contra; cuando el resultado algorítmico es incomprensible o no replicable, la contradicción se vuelve ilusoria (Altamirano et al., 2026).

La igualdad de armas se compromete porque, en la práctica, las partes no cuentan con los mismos recursos técnicos para auditar un sistema de IA; mientras una parte puede disponer de peritos especializados, la otra puede carecer de ellos, generando una asimetría epistémica que afecta la equidad procesal. La presunción de inocencia, especialmente en el proceso penal, se ve tensionada cuando los sistemas predictivos de riesgo influyen en la decisión judicial sin que el imputado pueda comprender ni cuestionar los criterios que determinaron la evaluación (Mendoza y Carrera, 2026).

El derecho a una resolución motivada exige que el juez exprese las razones lógicas y jurídicas de su decisión, lo que resulta incompatible con la aceptación acrítica de un resultado algorítmico. La igualdad y no discriminación se ve amenazada por los sesgos algorítmicos, que pueden reproducir patrones históricos de discriminación (Nuredin e Inan, 2024). La privacidad, los datos personales y la protección de datos sensibles se comprometen cuando los sistemas de IA procesan información biométrica, genética o de salud sin las debidas garantías. Finalmente, el acceso efectivo a la justicia puede verse limitado si la complejidad técnica de los sistemas disuade a las partes de impugnar las decisiones basadas en IA, generando una brecha social, de justicia y económica entre quienes pueden costear los peritajes especializados y quienes no (Bermudez et al., 2025).

Modelo normativo-operativo de control probatorio para la evidencia asistida por IA

El modelo se articula en tres fases acumulativas de control probatorio y un componente de atribución de responsabilidad. Las fases condicionan la validez y el peso de la evidencia; el componente de responsabilidad identifica deberes institucionales. El modelo tiene un carácter sistémico, lo que hace que no se avance a la fase siguiente sin superar la anterior.

Primera fase: admisibilidad condicionada

Antes de incorporar la evidencia, se exige: identificación del sistema (nombre, versión, proveedor y finalidad); trazabilidad de la fuente (archivo original, metadatos, cadena de custodia y registro de transformaciones); documentación técnica básica (métricas, limitaciones y márgenes de error); y aviso formal a las partes sobre la intervención algorítmica. Si falta alguno de estos elementos, la evidencia no se admite; si el defecto se detecta después, se excluye, salvo subsanación que no afecte el derecho de defensa. Sin esta base, las partes ni siquiera conocen que enfrentan evidencia algorítmica, lo que vulnera la contradicción desde su nivel más elemental.

Segunda fase: transparencia contradictoria

Se exige una explicación comprensible del resultado concreto; acceso a registros de ejecución y datos de entrada; auditoría técnica por perito independiente; reproducibilidad del resultado; y derecho de la parte afectada a presentar su propio peritaje. Si la parte no puede comprender ni impugnar el resultado, la evidencia no puede ser fundamento único o decisivo y queda reducida al indicio que requiere la corroboración independiente. Sin contradicción real, la transparencia es solo una declaración formal.

Tercera fase: valoración judicial motivada

El juez debe verificar que la evidencia superó las fases anteriores; explicar expresamente el peso que le otorga dentro del conjunto probatorio; examinar críticamente el resultado, sin aceptarlo acríticamente; y contrastarlo con otras pruebas. La falta de motivación sobre el peso de la evidencia algorítmica vulnera el derecho a la resolución motivada y la tutela judicial efectiva, y la decisión es impugnabile por falta de fundamentación. La valoración de la prueba corresponde al juez bajo la sana crítica de que un resultado algorítmico, por sofisticado que parezca, no lo releva de justificar por qué lo considera fiable y qué peso le asigna (Suárez, 2025).

Componente de atribución de responsabilidad

No es una fase de validez de la prueba, sino un eje que distribuye deberes entre los actores de la cadena. El proveedor o desarrollador responde por la ejecución de un diseño defectuoso, documentación insuficiente, sesgo o falta de validación, y debe documentar el sistema, evaluar riesgos, registrar versiones y advertir sus límites; la institución que adquiere o implementa el sistema responde por compra irreflexiva, falta de controles o dependencia del proveedor, y debe realizar evaluaciones de impacto, incluir cláusulas de auditabilidad y garantizar supervisión humana. El operador o perito responde por el uso incorrecto, la mala interpretación u omisión de límites, y debe mantener competencia técnica, preservar registros y emitir informes verificables. La parte que aporta la evidencia responde por falta de autenticidad, información incompleta o selección interesada, y debe acreditar origen, integridad, método y márgenes de error. El juez responde por delegación indebida, valoración acrítica o falta de motivación, y debe controlar la admisibilidad, explicar el peso probatorio y revisar humanamente todo resultado algorítmico. El Estado o poder judicial responde por ausencia de regulación, protocolos y supervisión, y debe garantizar la gobernanza, las auditorías, formación y los mecanismos de reparación.

El incumplimiento de la primera fase impide la admisión; el de la segunda impide el uso como fundamento único o decisivo; el de la tercera vulnera el derecho a la motivación. La falta de trazabilidad o de condiciones mínimas de contradicción pueden justificar inadmisión, exclusión, nulidad, reducción de eficacia probatoria o exigencia de corroboración independiente, según la gravedad, el marco procesal y el perjuicio al derecho de defensa. El modelo se articula con los niveles de transparencia de Vestri

(2021) —informativo, documental, explicativo, auditable y contradictorio—, pero los traduce en consecuencias procesales concretas, los dos primeros corresponden a la primera fase; los tres siguientes, a la segunda; la tercera fase conecta la transparencia con el deber de motivación; y el componente de responsabilidad cierra el circuito institucional. No requiere regulación previa exhaustiva, puede aplicarse dentro del debido proceso vigente como una lista de verificación caso por caso, aunque exige la formación de operadores y los protocolos internos.

Discusión

Los resultados obtenidos confirman que la incorporación de IA en la actividad probatoria no es un fenómeno futuro ni hipotético, sino una realidad en expansión que plantea profundas reflexiones epistémicas y jurídicas. La coincidencia de los estudios en señalar la opacidad algorítmica como el riesgo más recurrente y el derecho de defensa como la garantía más afectada revela que el problema central no es la IA en sí misma, sino la falta de mecanismos procesales que permiten a las partes comprender y controlar los resultados algorítmicos. Como advierten Isaza et al. (2026), la apariencia matemática de los resultados genera una "ilusión de precisión" (p. 400) que invisibiliza los márgenes de error y las limitaciones cognitivas del sistema, lo que constituye una asimetría epistémica estructural que tensiona el deber de motivación judicial y la exigencia de una valoración racional de la prueba.

Esta conclusión es consistente con lo señalado por Serna (2025) para el proceso civil, donde la ausencia de protocolos de contradicción algorítmica puede generar una asimetría epistémica que afecta la igualdad de armas. Pues como advierten Isaza et al. (2026), la apariencia matemática de los resultados genera una "ilusión de precisión" (p. 400) que invisibiliza los márgenes de error y las limitaciones cognitivas del sistema, lo que constituye una asimetría epistémica estructural que tensiona el deber de motivación judicial y la exigencia de una valoración racional de la prueba.

La literatura revisada coincide en que la transparencia no puede reducirse a la publicación del código fuente, sino que debe entenderse como verificabilidad procesal. Esta conclusión se alinea con la propuesta de Vestri (2021) sobre la transparencia por niveles —informativa, documental, explicativa, auditable y contradictoria—, y con la caracterización de Crespo (2025) sobre la transparencia como comprensibilidad y accesibilidad de los procesos y decisiones. La novedad de este estudio radica en aplicar esa distinción al ámbito probatorio y en vincularla con las consecuencias procesales específicas que se derivan de cada nivel de incumplimiento. Mientras la transparencia informativa evita el uso oculto de IA, la transparencia contradictoria hace real la igualdad de armas; sin esta última, las demás resultan insuficientes para garantizar el debido proceso.

En cuanto a la responsabilidad jurídica, los resultados confirman que no puede atribuirse a un único actor. La cadena de responsabilidad humana e institucional que se propone en este estudio —proveedor, institución, operador, parte aportante, juez y Estado— encuentra respaldo en la literatura sobre responsabilidad distribuida y en la

crítica a la "responsabilidad difusa" que señalan Mendoza y Carrera (2026). La aportación específica de este trabajo consiste en traducir esa cadena de responsabilidad en deberes concretos y verificables para cada actor, de modo que la atribución de responsabilidad no se diluya en la complejidad técnica del sistema. Esta propuesta se alinea con la concepción relacional y procesal de la responsabilidad que desarrolla Emelu (2026), según la cual la complejidad de la IA no genera "lagunas de responsabilidad" (4), sino que exige mecanismos de trazabilidad y atribución procesal.

Los resultados también revelan una brecha entre los principios abstractos y los instrumentos implementables. Si bien la mayoría de los estudios propone soluciones como la auditoría algorítmica, la supervisión humana y el peritaje contradictorio, solo una minoría desarrolla protocolos específicos de aplicación. Esta brecha justifica la propuesta de un estándar de admisibilidad, contradicción y valoración probatoria que combine los niveles de transparencia con las exigencias mínimas por tipo de IA y las consecuencias procesales de su incumplimiento. Dicho estándar no pretende ser exhaustivo ni definitivo, sino ofrecer un marco operativo que permita a los operadores jurídicos evaluar caso por caso si la evidencia asistida por IA cumple con las garantías procesales exigibles.

Conclusiones

El análisis desarrollado permite sostener que la incorporación de la IA en la actividad probatoria no constituye un problema meramente técnico, sino una cuestión que interpela los fundamentos epistémicos del proceso judicial. La apariencia de objetividad matemática de los resultados algorítmicos encubre una asimetría estructural: el sistema puede procesar datos y arrojar correlaciones, pero carece de la capacidad de comprender el significado contextual de la evidencia, de ponderar circunstancias fácticas complejas o de justificar racionalmente por qué un resultado particular resulta correcto en términos de justicia.

Esta carencia no se resuelve con mayor sofisticación tecnológica, sino que exige reconocer que la valoración probatoria es una actividad inherentemente humana, orientada por la sana crítica, la lógica y las máximas de experiencia, y que la IA solo puede operar como un auxiliar subordinado a esas coordenadas. En consecuencia, la transparencia algorítmica debe entenderse no como la mera publicidad del código o de la arquitectura del sistema, sino como la posibilidad efectiva de que las partes comprendan, reconstruyan y contradigan el razonamiento que condujo al resultado, pues sin esa posibilidad la transparencia se reduce a una fórmula vacía que no satisface las exigencias del debido proceso.

El modelo normativo-operativo propuesto —articulado en cuatro fases acumulativas: admisibilidad condicionada, transparencia contradictoria, valoración judicial motivada y atribución de responsabilidad— ofrece una respuesta estructurada a los desafíos identificados, pero su implementación revela una tensión de fondo que el ordenamiento jurídico debe resolver. La trazabilidad, la auditabilidad y la contradicción efectiva no dependen únicamente de la buena voluntad de los operadores jurídicos, sino de

condiciones institucionales que exceden el caso concreto: la formación técnica de jueces y peritos, la disponibilidad de recursos para peritajes independientes, la existencia de protocolos de contratación pública que exijan cláusulas de auditabilidad y la regulación de la responsabilidad civil y administrativa de proveedores e instituciones.

Mientras estas condiciones no se desarrollen, el modelo corre el riesgo de convertirse en un estándar teórico inaplicable en la práctica, especialmente en sistemas judiciales con recursos limitados. La conclusión central, por tanto, no es que la IA deba excluirse de la actividad probatoria, sino que su admisibilidad y peso probatorio deben graduarse en función de la posibilidad real de control y contradicción, y que, en ausencia de esa posibilidad, la salida algorítmica no puede constituir el fundamento único o decisivo de la decisión judicial, sino que deberá someterse a corroboración independiente o a las consecuencias procesales que el ordenamiento prevea para la vulneración de garantías.

Referencias bibliográficas

- Aguilar Cabrera, D. (2024). La inteligencia artificial en la justicia: protocolos para la presentación y la valoración de prueba digital obtenida mediante IA. *Revista Oficial Del Poder Judicial*, 16(22), 475-497. <https://doi.org/10.35292/ropj.v16i22.1018>
- Altamirano Castañeda, R. L., Altamirano Sánchez, G. del R., Pérez Estrada, L. C. & Pérez Estrada, M. del C. (2026). La inteligencia artificial en la valoración de las pruebas: desafío actual en la gobernanza judicial. *Universitas Ciencia*, 14(39), 294-308. <https://doi.org/10.5281/zenodo.19578125>
- Bermudez Castro, J. C., Quiñones Li, A. E., Romero Coz, S. N., Castro Verona, F. M., Cervera Cajo, L. E. & Cabrera Chaupin, S. A. (2025). Inteligencia artificial y derechos humanos en el ordenamiento jurídico peruano: una revisión sistemática. *Revista Tribunal*, 5(13), 676-695. <https://doi.org/10.59659/revistatribunal.v5i13.295>
- Carrasco Cabezas, F. (2025). Propuesta de uso de la inteligencia artificial durante la valoración probatoria. *Laborem*, 24(31), 25-45. <https://doi.org/10.56932/laborem.24.31.1>
- Crespo Berti, L.A. (2025). El impacto de la inteligencia artificial desde las Ciencias Jurídicas. *Iustitia Socialis. Revista Arbitrada de Ciencias Jurídicas*, X(18), 1-4. <https://doi.org/10.35381/racji.v10i18.4337>

- Emelu, M. N. (2026). Formation-based AI ethics: Ethical formation, responsibility, and opportunity cost in AI ecosystems. *AI and Ethics*, 6, 1-13. <https://doi.org/10.1007/s43681-026-01165-5>
- Hernández Sampieri, R. & Mendoza Torres, C. P. (2023). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta (2.^a ed.). McGraw-Hill.
- Isaza Bedoya, J., Alzate Arenas, A., Murrillo Arroyave, S., Rubio Cardozo, M. D., Zuluaga Ramírez, S. A., Moreno Higueta, Y. A. & Castro Restrepo, S. (2026). La prueba testifical en la era de la inteligencia artificial: modelos epistémicos y desafíos para la justicia. *Revista de Derecho Procesal Contemporáneo - Semilleros*, (5), 383-403. <http://publicacionesicdp.com/index.php/revista-semilleros-icdp/article/view/695>
- Kocagül, H. & Kartal, M. (2026). AI, evidentiary authority, and the right to a fair trial in criminal proceedings. *Laws*, 15(3), 58. <https://doi.org/10.3390/laws15030058>
- 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>
- Nuredin, A. & Inan, T. C. (2024). The impact of AI-based decision-making systems on justice and equality. *International Scientific Journal Vision*, 9(2), 9-33. <https://doi.org/10.55843/ivisum24209n>
<https://visionjournal.edu.mk/social/index.php/1/article/view/161>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., Moher, D., Yepes-Nuñez, J. J., Urrutia, G., Romero-García, M., & Alonso-Fernández, S. (2022). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, 1-10. <https://doi.org/10.1186/s13643-021-01626-4>, <https://doi.org/10.1016/j.ijsu.2021.105906>
- Segovia Segovia, A. C. & Flores Quishpi, B. E. (2025). La inteligencia artificial, los derechos humanos y el sistema penal ecuatoriano: Un análisis de sus ventajas y desventajas. *Ciencia Latina Revista Científica Multidisciplinar*, 8(6), 8048-8059. https://doi.org/10.37811/cl_rcm.v8i6.15510
- Serna Acosta, A. Y. (2025). Desafíos de la inteligencia artificial en la valoración de pruebas en el proceso civil. <https://repository.ucatolicaluisamigo.edu.co/server/api/core/bitstreams/d0cb5513-7014-45f8-857a-ab1818e5bd56/content>

- Suárez Xavier, P. R. (2025). Justicia algorítmica y proceso civil: ¿puede la IA valorar la prueba en apelación? *IUS ET SCIENTIA*, 11(1), 32–53.
<https://doi.org/10.12795/IESTSCIENTIA.2025.i01.02>
- Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, 56, 368–398.
<https://dialnet.unirioja.es/servlet/articulo?codigo=7971161>

Conflicto de intereses: Los autores declaran no tener conflicto de intereses.

Contribución de los autores: Todos los autores participaron en la concepción, ejecución, análisis y redacción del estudio.

Universidad de Las Tunas en <https://ror.org/05fj29j73>