ASCE MAGAZINE ISSN: 3073–1178

Revista ASCE Magazine, Periodicidad: Trimestral, Volumen: 5, Número: 3, Año: 2026 páginas 1158-1181

Doi: https://doi.org/10.70577/asce.v5i3.1031

Recibido: 2026-06-24

Aceptado: 2026-07-04

Publicado: 2026-07-24

Divergencias técnicas, epistemológicas y jurídicas frente a la valoración automatizada de la credibilidad del testimonio mediante inteligencia artificial en el proceso penal ecuatoriano

Technical, Epistemological, and Legal Divergences in the Automated Assessment of Testimonial Credibility through Artificial Intelligence in Ecuadorian Criminal Proceedings

Autor(s)

Daniela Estefanía Zurita López [1]
Maestrante
dezurital@ube.edu.ec
https://orcid.org/0009-0000-4214-6401
Universidad Bolivariana del Ecuador
Guayaquil – Ecuador

Ariana Tatiana Viteri Galeas [2]
Maestrante
atviterig@ube.edu.ec
https://orcid.org/0009-0007-6263-8762
Universidad Bolivariana del Ecuador
Guayaquil – Ecuador

Axel Aren Zambrano Acosta [3]
Docente
aazambranoa_a@ube.edu.ec
https://orcid.org/0009-0008-0608-4627
Universidad Bolivariana del Ecuador
Guayaquil – Ecuador

Gina Roxana Arias Murillo [4]
Docente
grariasm_a@ube.edu.ec
https://orcid.org/0009-0005-1181-141X
Universidad Bolivariana del Ecuador
Guayaquil – Ecuador

Cómo citar

Zurita López , D. E., Viteri Galeas , A. T., Zambrano Acosta , A. A., & Arias Murillo , G. R. (2026). Divergencias técnicas, epistemológicas y jurídicas frente a la valoración automatizada de la credibilidad del testimonio mediante inteligencia artificial en el proceso penal ecuatoriano. *ASCE MAGAZINE*, *5*(3), 1158–1181. https://doi.org/10.70577/asce.v5i3.1031

Resumen

La incorporación de la inteligencia artificial (IA) en la justicia latinoamericana proyectó una de sus aplicaciones más controvertidas: la evaluación autónoma de la credibilidad testimonial mediante sistemas que pretenden inferir estados internos del declarante, como la sinceridad, la coherencia o la fiabilidad, a partir de marcadores verbales, vocales o conductuales, dentro del proceso penal ecuatoriano. Este artículo examinó los límites de esa pretensión y determinó en qué condiciones, si en alguna, resultó compatible con las garantías del debido proceso en Ecuador. Para ello, se realizó una revisión de alcance con orientación PRISMA-ScR, complementada con análisis dogmático-jurídico, que integró psicología forense, ciencia de la computación y derecho procesal penal a partir de 43 fuentes seleccionadas conforme a criterios explícitos de inclusión y calidad. Los resultados evidenciaron tres núcleos convergentes de restricciones: técnicas (fragilidad del ground truth, sesgos de entrenamiento, baja transferibilidad, opacidad de los modelos profundos y el problema de la tasa base), epistemológicas (la credibilidad, la veracidad, la exactitud y la fiabilidad no resultaron equivalentes entre sí, y las claves conductuales del engaño se mostraron intrínsecamente débiles) y jurídicas (incompatibilidad con la inmediación, la contradicción, la motivación y la igualdad). En las condiciones actuales de desarrollo tecnológico, el uso de sistemas de inteligencia artificial para determinar de manera independiente la credibilidad de un testimonio no resultó plenamente compatible con las garantías del debido proceso establecidas en la Constitución de la República del Ecuador. A partir de estos hallazgos, se propuso un uso responsable de la inteligencia artificial que limitó su aplicación a tareas de apoyo y complementarias, y se diseñó un protocolo de admisibilidad estructurado en tres etapas, articulable con el COIP, la LOPDP y los criterios del Sistema Interamericano de Derechos Humanos.

Palabras clave: valoración del testimonio; inteligencia artificial; identificación del engaño; garantías procesales; sesgos algorítmicos; Ecuador.

Abstract

The incorporation of artificial intelligence (AI) into Latin American justice systems produced one of its most contentious applications: the autonomous assessment of testimonial credibility through systems that seek to infer a declarant's internal states, such as sincerity, coherence, or reliability, from verbal, vocal, or behavioral markers, within Ecuadorian criminal proceedings. This article examined the limits of that aspiration and determined under what conditions, if any, it proved compatible with due-process guarantees in Ecuador. A scoping review guided by PRISMA-ScR, complemented with dogmatic legal analysis, integrated forensic psychology, computer science, and criminal procedural law across 43 sources selected under explicit inclusion and quality criteria. The results revealed three converging clusters of constraints: technical (fragile ground truth, training biases, poor transferability, opacity of deep models, and the base-rate problem), epistemological (credibility, veracity, accuracy, and reliability proved non-equivalent, and behavioral cues to deception proved intrinsically weak), and legal (incompatibility with immediacy, cross-examination, reasoned motivation, and equality). Under current conditions of technological development, the use of artificial intelligence systems to independently determine the credibility of testimony did not prove fully compatible with the due-process guarantees established in the Constitution of the Republic of Ecuador. Building on these findings, the study proposed a framework for the responsible use of artificial intelligence that limited its application to supporting and complementary tasks, and outlined a three-stage admissibility protocol designed to integrate with the COIP, the LOPDP, and the criteria of the Inter-American Human Rights System.

Keywords: assessment of testimony; artificial intelligence; deception detection; procedural guarantees; algorithmic bias; Ecuador.

Introducción

En la última década, los sistemas de justicia de América Latina incorporaron de forma progresiva herramientas de inteligencia artificial (IA) en funciones antes exclusivas de la actividad judicial. En Argentina, el sistema Prometea automatizó la elaboración de dictámenes administrativos y facilitó la consulta de jurisprudencia (Estevez et al., 2020); en Colombia, PretorIA asistió a la Corte Constitucional en la organización de acciones de tutela (Saavedra Rionda y Upegui Mejía, 2021). Estudios recientes confirmaron que la región transitó de proyectos aislados a políticas institucionales de adopción de IA judicial, con niveles de preparación y gobernanza todavía heterogéneos entre países (Pérez-Pacheco, 2025; Segura, 2023).

En Ecuador, aunque de forma menos visible, ya se emplean herramientas de apoyo para la transcripción de audiencias y el análisis de documentos en fiscalías y unidades judiciales, en línea con procesos similares en Chile, México y Brasil. Sin embargo, este avance se produjo con escaso debate público y sin un marco regulatorio específico, pese a que la doctrina comparada en derecho administrativo señaló la necesidad de extender a estos sistemas las mismas garantías que rigen a la administración pública (Boix Palop, 2020).

El objeto de este artículo se circunscribió a la evaluación autónoma de la credibilidad testimonial mediante inteligencia artificial dentro del proceso penal ecuatoriano, entendida como el uso de sistemas algorítmicos que producen, sin intervención humana determinante, una estimación sobre la sinceridad, la coherencia o la fiabilidad de un declarante a partir de indicadores verbales, vocales, faciales o conductuales. Esta delimitación exigió diferenciar un uso instrumental de la IA, limitado a transcribir, organizar datos o realizar búsquedas sin intervenir en la decisión, de un uso decisorio orientado a incidir directamente en la resolución del conflicto. La evaluación de la credibilidad se ubicó en esta segunda categoría, la más delicada: si el sistema funcionara con la fiabilidad que se le atribuye, trasladaría al mecanismo automatizado una parte central de la valoración judicial; si no, su uso podría introducir distorsiones relevantes en la administración de justicia.

El interés académico en esta temática creció de manera significativa, con modelos entrenados en bases de datos judiciales reales y sistemas multimodales que integran variables lingüísticas, vocales y faciales (Pérez-Rosas et al., 2015). Paralelamente, más de cincuenta especialistas en comunicación no verbal, psicología y derecho suscribieron una toma de posición conjunta que alertó sobre el carácter pseudocientífico que puede adquirir este tipo de análisis en contextos

forenses, por la falta de transparencia algorítmica, la presencia de sesgos estructurales y la debilidad de sus fundamentos teóricos (Denault et al., 2020).

El planteamiento central de este artículo sostuvo que esa evaluación autónoma presenta limitaciones técnicas, epistemológicas y jurídicas que, en el estado actual del conocimiento, la hacen incompatible con las garantías del debido proceso, sin que ello implique rechazar los usos auxiliares de la tecnología ni desconocer sus avances. El análisis se situó en el contexto ecuatoriano por la relevancia de su marco jurídico: la Constitución de 2008 (artículos 75 y 76) consagra garantías estrictas del debido proceso; el COIP (2014) estructura la actividad probatoria sobre la oralidad, la inmediación y la contradicción; y la LOPDP (2021) regula el tratamiento de datos y las decisiones automatizadas, en diálogo con la experiencia comparada de Argentina, Colombia, Chile, México y Brasil.

La pregunta de investigación fue: ¿en qué circunstancias, si las hay, puede considerarse compatible la evaluación de la credibilidad testimonial mediante IA con las garantías constitucionales del debido proceso y con los estándares de la valoración racional de la prueba en el proceso penal ecuatoriano? El objetivo general consistió en analizar los límites de la evaluación automatizada de credibilidad y proponer un marco de uso responsable para el contexto ecuatoriano, coherente con estándares regionales. De manera específica, el estudio buscó: (i) diferenciar los conceptos de credibilidad, veracidad, exactitud y fiabilidad; (ii) identificar las principales limitaciones técnicas de los sistemas de IA en la detección del engaño; (iii) contrastar esas limitaciones con las garantías del debido proceso en Ecuador; y (iv) proponer un protocolo de admisibilidad para su eventual uso. La contribución original radicó en articular estas tres dimensiones en un único marco operativo anclado en el derecho positivo ecuatoriano.

Marco teórico

La evaluación de la credibilidad testimonial: conceptos, métodos y estado de la evidencia

En psicología del testimonio, la credibilidad, la veracidad, la exactitud y la fiabilidad no son términos homólogos, y confundirlos conduce a errores significativos al valorar una declaración en contexto judicial (Manzanero, 2010; Vrij et al., 2022). La credibilidad describe la coherencia estructural del relato; la veracidad remite a la intención del declarante de contar los hechos tal como los percibió; la exactitud, al grado de correspondencia entre lo narrado y lo efectivamente ocurrido;

y la fiabilidad, a la consistencia de la declaración en el tiempo, sin garantizar por sí sola su corrección fáctica (DePaulo et al., 2003; Hartwig y Bond, 2011; Bogaard et al., 2022; Denault et al., 2020). Por eso es posible encontrar relatos coherentes que contienen errores, declaraciones sinceras distorsionadas por el funcionamiento normal de la memoria, o testimonios estables que en realidad fueron ensayados o influenciados por terceros (González y Manzanero, 2018; Vrij et al., 2022). Esta distinción marca los límites de cualquier evaluación: un método centrado en indicadores de fabricación no mide la exactitud de los hechos, y uno apoyado solo en la consistencia del relato no permite, por sí solo, afirmar su veracidad (Vrij et al., 2022).

En la práctica pericial iberoamericana esta diferenciación no siempre se refleja en los oficios, y jueces y fiscales suelen usar «credibilidad» como sinónimo de verdad o falsedad del testimonio. El informe pericial debe precisar qué dimensiones evaluó (estructura del relato, intención subjetiva, coherencia con otros elementos probatorios o estabilidad en el tiempo) y dejar claro que la determinación de la verdad de los hechos corresponde exclusivamente al juez o tribunal, protegiendo así la calidad metodológica de la pericia y ajustando las expectativas del sistema de justicia.

La técnica clásica en este campo es el Análisis de Contenido Basado en Criterios (CBCA), fundado en la hipótesis de Undeutsch según la cual los testimonios reales se diferencian de los inventados por su calidad narrativa, valorando la estructura lógica, los detalles específicos, la ubicación espacio-temporal y el reconocimiento de lagunas de memoria, entre otros criterios. Junto con el Reality Monitoring y el SCAN, forma parte de las herramientas verbales de evaluación de la credibilidad más estudiadas; la evidencia meta-analítica más reciente confirma una precisión moderada, con variabilidad entre evaluadores y sensibilidad a sesgos de publicación (Oberlader et al., 2021; Vrij et al., 2022), por lo que sigue siendo ampliamente utilizado en peritajes de credibilidad en Europa y en el ámbito iberoamericano (González y Manzanero, 2018). En la práctica, sin embargo, el CBCA suele interpretarse casi como un «veredicto» sobre la verdad del testimonio, cuando en realidad ofrece solo indicios que deben combinarse con los datos clínicos y el contexto del caso: si ni siquiera un instrumento validado y administrado por un especialista humano permite emitir veredictos concluyentes, la pretensión de que un algoritmo lo haga de forma autónoma exige, cuando menos, el mismo nivel de cautela.

Inteligencia artificial aplicada a inferencias sobre estados mentales

La inteligencia artificial designa sistemas diseñados para ejecutar tareas que, en condiciones ordinarias, requerirían capacidades cognitivas humanas; dentro de ella, el aprendizaje automático mejora su desempeño a partir del análisis de datos sin reglas programadas explícitamente, y el aprendizaje profundo utiliza redes neuronales de múltiples capas cuya complejidad dificulta la interpretación transparente de sus decisiones. En los últimos años se plantearon sistemas orientados a «detectar mentiras», entrenados con ejemplos de respuestas verdaderas y falsas y con modelos que combinan voz, gestos y contenido verbal (Pérez-Rosas et al., 2015; Vrij et al., 2022); aunque sus resultados iniciales parecieron prometedores, la precisión suele descender de forma importante al aplicarlos en otros idiomas, culturas o declarantes ausentes de la muestra original (Denault et al., 2020; Vrij et al., 2022), lo que invita a desconfiar de cualquier herramienta que se presente como capaz de «decir quién miente» sin una validación clara en el contexto local.

Un argumento de tipo bayesiano muestra, además, que el cribado masivo en detección del engaño produce tasas inaceptables de falsos positivos cuando la prevalencia real del engaño es baja, trasladando de hecho la carga de la prueba al sujeto erróneamente etiquetado. Sánchez-Monedero y Dencik (2022) sistematizaron tres deficiencias estructurales de estos sistemas (opacidad de los modelos profundos, riesgo sistemático de sesgo y déficit de fundamentación teórica) y señalaron que la historia del polígrafo, desacreditado pero aún presente en algunos contextos, ilustra la dificultad de retirar una tecnología una vez introducida en la práctica institucional. La irrupción de modelos lingüísticos generativos añadió una capa adicional de complejidad: Markowitz et al. (2024) mostraron que los textos engañosos producidos por humanos difieren lingüísticamente de los generados por IA, lo que plantea retos nuevos de autenticidad y atribución para el derecho probatorio.

Epistemología jurídica de la prueba: valoración racional, motivación y sana crítica

La tradición racionalista iberoamericana concibe la valoración de la prueba como una operación sujeta a control intersubjetivo y a motivación argumentativa. Autores como Ferrer Beltrán (2007), Taruffo (2002) y Gascón Abellán (2010) sostienen que un hecho se considera probado cuando aceptarlo está racionalmente justificado, en oposición a la «íntima convicción» del juez, entendida como una decisión puramente interna, no argumentada, que puede llegar a ser arbitraria. Esto exige que las decisiones se adopten a la luz del conocimiento disponible y a partir de una motivación

revisable, de modo que la forma en que se recoge y valora la prueba sea clara y controlable por las partes. Laudan (2006) añade que las herramientas novedosas deben valorarse por su capacidad para reducir errores, no por su mera novedad; y la literatura sobre ética de algoritmos muestra que la opacidad de un sistema no es solo un problema técnico, sino que determina quién responde por sus errores (Kordzadeh y Ghasemaghaei, 2022): un resultado que no puede explicarse tampoco puede controlarse racionalmente, lo que conecta esta discusión con las exigencias de motivación e inmediación del ordenamiento ecuatoriano.

Marco normativo: garantías constitucionales del Ecuador y referencias supranacionales

El marco normativo ecuatoriano se articula en tres niveles. La Constitución de 2008, en sus artículos 75 y 76, garantiza la tutela judicial efectiva, la inmediación, la presunción de inocencia, la defensa, la contradicción de la prueba y la debida motivación de las decisiones; el artículo 76, numeral 7, literal c, reconoce el derecho a ser escuchado en igualdad de condiciones, el literal h garantiza presentar y controvertir pruebas, y el literal l exige que toda resolución esté debidamente fundamentada. El COIP (2014) desarrolla estos principios sobre la oralidad, la inmediación y la contradicción.

La LOPDP (2021) resulta pertinente por tres vías concretas: un sistema que infiere sinceridad o fiabilidad a partir de rasgos vocales, faciales o lingüísticos realiza una elaboración de perfiles; los rasgos faciales y vocales usados como entrada constituyen datos biométricos sometidos a un régimen reforzado de protección; y recurrir al resultado de un sistema de IA para facilitar una decisión jurídica, por ejemplo al valorar un testimonio en un trámite penal, activa el régimen de las decisiones automatizadas y el derecho de la persona a no ser sometida a ellas sin una intervención humana significativa. Bajo esta perspectiva, la LOPDP no es una norma secundaria, sino el límite mismo que enfrenta cualquier intento de valorar la credibilidad de una declaración solo con procesos automáticos.

En el plano internacional, el Reglamento (UE) 2024/1689 clasificó los sistemas de IA según su tipología de riesgo y restringió, entre otras aplicaciones, la inferencia automatizada de emociones en contextos sensibles; la Recomendación sobre la Ética de la IA de la UNESCO (2021) trazó principios con los que los sistemas nacionales, incluido el ecuatoriano, previsiblemente deberán converger; y la experiencia comparada en administración pública aportó criterios de transparencia, explicabilidad y control aplicables también al ámbito probatorio (Cerrillo i Martínez et al., 2022).

En este contexto, la utilización de inteligencia artificial para valorar la credibilidad del testimonio debe concebirse únicamente como un recurso de apoyo al criterio profesional, y no como un mecanismo que reemplace de forma automática la apreciación humana. El perito debe detallar en sus informes la función específica de la herramienta, sus limitaciones técnicas y de interpretación, y su articulación con el resto del acervo probatorio, de modo que jueces y partes puedan valorar críticamente el peso de esa información en la decisión final.

Método

Enfoque y tipo de revisión

El enfoque de esta investigación fue cualitativo, de tipo documental e interpretativo, con una orientación jurídico-empírica que combinó dos componentes complementarios. El primero fue una revisión de alcance con orientación en las directrices PRISMA-ScR (Tricco et al., 2018), elegida por la naturaleza interdisciplinar del objeto de estudio, en el que confluyen la psicología forense, la ciencia de la computación aplicada a la inferencia de estados mentales y el derecho procesal en su dimensión constitucional; una revisión sistemática clásica habría dejado fuera buena parte del material relevante, mientras que una revisión puramente doctrinal habría silenciado el componente técnico. Se empleó la fórmula «con orientación PRISMA-ScR» porque el diseño adoptó sus principios de reporte sin satisfacer la totalidad de sus ítems formales, concebidos originalmente para literatura biomédica. El segundo componente fue un análisis dogmático-jurídico aplicado al bloque constitucional y legal ecuatoriano (Constitución, COIP y LOPDP) y a las fuentes supranacionales pertinentes, que operó como un segundo eje de trabajo cuyos resultados se integraron con los hallazgos técnicos y psicológicos mediante una matriz común de análisis por objetivo específico. Esta combinación respondió a que ninguno de los dos métodos, por separado, resultaba suficiente: la revisión de alcance sintetiza evidencia técnica y empírica dispersa, pero no resuelve cuestiones de validez normativa; el análisis dogmático-jurídico interpreta el alcance de las garantías procesales, pero no sintetiza sistemáticamente la evidencia científica que sustenta el argumento.

Bases de datos y fuentes consultadas

Se consultaron repositorios científicos indexados (Scopus, Web of Science, PsycINFO, PubMed, IEEE Xplore, ACM Digital Library, arXiv, SciELO, Redalyc, Dialnet, SSRN y DOAJ), bases oficiales y portales especializados en normativa y jurisprudencia (EUR-Lex, Registro Oficial del

Ecuador, UNESCO, Corte y Comisión Interamericana de Derechos Humanos, CEJA-JSCA), repositorios institucionales de universidades de la región (FLACSO, PUCE, UCE, USFQ, UNAM, UBA, UDP, Universidad del Rosario y Universidad Externado de Colombia), documentos de organismos internacionales (BID, OEA y CAF) y materiales de centros críticos como Dejusticia, AI Now Institute, Algorithm Watch y el Observatorio de Inteligencia Artificial Iberoamericano. Esta combinación permitió situar el análisis no solo en el plano jurídico formal, sino también dentro de la discusión actual sobre los efectos de la inteligencia artificial en los derechos fundamentales en América Latina.

Cadenas de búsqueda y descriptores

Las búsquedas se realizaron en español, portugués e inglés, con tres familias principales de descriptores: la primera combinó credibilidad e inteligencia artificial (credibility assessment, deception detection, AI, machine learning); la segunda, sesgo, opacidad y debido proceso (algorithmic bias, black-box, due process, explainability); y la tercera, el contexto regional (Latin America, Ecuador, poder judicial, justicia digital). Se añadieron cadenas auxiliares sobre CBCA, SVA, explicabilidad e interpretación judicial, ajustadas al lenguaje indexador de cada base.

Periodo y criterios de elegibilidad

El periodo de selección principal de la literatura fue 2020-2026. Esta revisión incorporó investigaciones anteriores solo cuando quedaron expresamente justificadas frente a la pregunta de investigación, por su carácter fundacional o por no contar con un equivalente posterior que las actualizara sin pérdida de precisión: Steller y Köhnken (1989) sobre el origen del CBCA; DePaulo et al. (2003) y Hartwig y Bond (2011) sobre los indicadores del engaño; Manzanero (2010) y González y Manzanero (2018) sobre el protocolo HELPT de evaluación testifical; Taruffo (2002), Ferrer Beltrán (2007), Gascón Abellán (2010) y Laudan (2006) sobre la valoración racional de la prueba; Pasquale (2015) y Burrell (2016) sobre la opacidad algorítmica; Roth (2017) sobre el «testimonio de máquina»; O'Neil (2016) sobre los efectos sociales de los sistemas algorítmicos; y Wachter et al. (2018) sobre explicaciones contrafactuales, además de Tricco et al. (2018), referencia metodológica obligada para toda revisión PRISMA-ScR. Fuera de este conjunto acotado y razonado, el resto de la literatura empírica correspondió al periodo declarado.

Las normas de inclusión comprendieron estudios empíricos con descripción explícita de la muestra y las métricas utilizadas, revisiones sistemáticas y metaanálisis en revistas indexadas, ponencias

técnicas con evaluación por pares, doctrina jurídica, normativa y jurisprudencia relevante, e informes de organismos institucionales reconocidos. Se excluyeron la literatura comercial no revisada por pares, las declaraciones empresariales no contrastadas, los informes de opinión periodística (salvo mínima contextualización) y los trabajos sin DOI activo, ISBN verificable o dirección web oficial estable, buscando que el análisis se apoyara en fuentes con un mínimo control de calidad científica y jurídica.

Proceso de selección y diagrama de flujo

El proceso de selección siguió las cuatro etapas del modelo PRISMA-ScR. En identificación, se recuperaron los registros arrojados por las cadenas de búsqueda en cada base consultada. En cribado, se eliminaron duplicados y se revisaron título y resumen conforme a los criterios de elegibilidad, descartando lo no alineado con la pregunta de investigación. En elegibilidad, se evaluaron a texto completo aproximadamente 50 registros, aplicando además el protocolo de calidad del apartado siguiente; alrededor de 7 fueron excluidos por no alcanzar el umbral mínimo exigido. En inclusión, el cuerpo final quedó integrado por 43 fuentes, documentadas en una bibliografía maestra en la que cada entrada presenta DOI activo, ISBN comprobable o URL oficial estable. La Figura 1 resume este proceso.

Figura 1. Diagrama de flujo del proceso de selección (PRISMA-ScR)

Identificación y cribado: registros localizados en las bases de datos y fuentes especializadas consultadas, sometidos a eliminación de duplicados y a cribado por título y resumen conforme a los criterios de elegibilidad descritos
Evaluación a texto completo: aproximadamente 50 registros evaluados a texto completo
Excluidos en esta etapa: aproximadamente 7 registros, por no cumplir los criterios de inclusión o el umbral de calidad exigido
Incluidos en la síntesis cualitativa: 43 fuentes

Instrumentos de recolección de datos

La recolección se apoyó en tres instrumentos: una matriz de extracción bibliográfica para cada fuente empírica o técnica (autoría, año, país, diseño, muestra, hallazgos y limitaciones declaradas); una ficha de análisis documental para las fuentes normativas y jurisprudenciales (disposición,

jerarquía normativa y relación con el debido proceso); y el protocolo de revisión de calidad descrito a continuación. Los tres instrumentos alimentaron una base de datos común organizada por los cuatro objetivos específicos del estudio.

Evaluación crítica de la calidad

Se aplicaron tres registros de evaluación según el tipo de fuente: para revisiones sistemáticas y metaanálisis, una versión adaptada de AMSTAR-2; para estudios técnicos de IA, una rúbrica basada en transparencia metodológica, validación cruzada, replicabilidad y declaración de limitaciones; y para fuentes jurídicas, un análisis doctrinal cualitativo atento al rigor argumentativo. Las fuentes que no superaron el umbral mínimo para su categoría fueron excluidas de la síntesis final.

Resultados

Los hallazgos se organizaron conforme a los cuatro objetivos específicos de la introducción. Esta sección presenta la síntesis de la evidencia para cada objetivo; su interpretación conjunta se desarrolla en la discusión.

Diferenciación conceptual entre credibilidad, veracidad, exactitud y fiabilidad

La revisión confirmó que estos cuatro conceptos no son equivalentes y que la literatura los trata como dimensiones distintas, aunque relacionadas, de la declaración testimonial (véase marco teórico). Ningún instrumento identificado, ni humano ni automatizado, logró medir simultáneamente las cuatro dimensiones con un mismo procedimiento: cualquier sistema de IA que se presente como «medidor de credibilidad» sin especificar cuál de ellas evalúa incurre en una imprecisión conceptual previa a cualquier problema técnico.

Limitaciones técnicas de los sistemas de inteligencia artificial en la detección del engaño

Los sistemas revisados se agruparon en cuatro familias arquitectónicas: análisis lingüístico mediante procesamiento del lenguaje natural, cuyo beneficio práctico fuera de condiciones experimentales resulta escaso (Vrij et al., 2022); sistemas multimodales que combinan lengua, voz, gestos y expresiones faciales, pioneramente explorados por Pérez-Rosas et al. (2015); inferencias automáticas de emociones mediante microexpresiones y señales fisiológicas, documentadas en el proyecto iBorderCtrl (Sánchez-Monedero y Dencik, 2022); y análisis del estrés vocal y polígrafos

digitales, cuya historia de límites y desacreditación advierten Suchotzki y Gamer (2024). En todos los casos, las precisiones reportadas en la literatura comercial resultaron más altas que las validadas de forma independiente, brecha que constituye, en sí misma, un problema para el derecho probatorio.

De esta evidencia se identificaron cinco limitaciones técnicas transversales:

Ground truth incierto. Las etiquetas de detección del engaño son inferencias probabilísticas, no verdades verificables: el sistema se entrena contra un referente ya incierto.

Sesgos en la formación. Los modelos reproducen o amplifican desigualdades demográficas y culturales presentes en sus datos de entrenamiento (Crawford, 2021; O'Neil, 2016).

Sobreajuste y escasa transferibilidad. Un buen desempeño en los datos de entrenamiento no se sostiene en contextos distintos, sobre todo si cambian la lengua o las referencias culturales (Sánchez-Monedero y Dencik, 2022).

Opacidad de los modelos profundos. Resulta extremadamente difícil explicar cómo un modelo avanzado llega a una salida concreta, incluso para quienes lo diseñaron (Burrell, 2016; Pasquale, 2015).

Problema de la tasa base. Cuando el engaño real es poco frecuente, incluso un sistema preciso genera muchos falsos positivos que distorsionan sus resultados prácticos (Sánchez-Monedero y Dencik, 2022).

Más allá de estos límites técnicos, la credibilidad resultó un constructo contextual complejo que depende de la memoria, el lenguaje, la emoción, la cultura y la biografía individual, por lo que reducirla a un vector numérico estático es problemático. Un sistema de IA puede además replicar o amplificar los sesgos de quien emite el juicio, apoyándose en índices que la propia psicología del testimonio ha puesto en duda (DePaulo et al., 2003; Hartwig y Bond, 2011), y cualquier sistema entrenado con decisiones judiciales previas puede perpetuar errores anteriores, comprometiendo la objetividad de decisiones futuras.

Dos casos ilustraron empíricamente estas limitaciones. iBorderCtrl mostró que el cribado masivo automatizado resulta incompatible con tasas operativas y niveles de respeto a los derechos fundamentales aceptables (Sánchez-Monedero y Dencik, 2022). COMPAS, usado para evaluar riesgo de reincidencia en Estados Unidos, evidenció que la alta complejidad técnica no garantiza precisión ni justicia material; evaluaciones posteriores confirmaron que la falta de transparencia, más que la exactitud predictiva, es la fuente principal de injusticia (Rudin et al., 2020; Wang et al.,

2023). Los sistemas comerciales de detección del engaño, por su parte, publicitan precisiones que no se sostienen al valorarse de forma independiente, lo que sugiere que su aceptación institucional responde más a la confianza en la innovación que a la evidencia disponible (Suchotzki y Gamer, 2024).

Incompatibilidad con las garantías del debido proceso en Ecuador

Las incompatibilidades jurídicas se estructuraron en torno a cuatro principios constitucionales:

Inmediación. Exige que el juzgador observe y valore directamente el testimonio; un dictamen algorítmico autónomo desplaza, en vez de complementar, ese núcleo argumentativo.

Contradicción. Las partes deben poder confrontar la prueba, pero la opacidad algorítmica restringe el acceso a sus bases; el «testimonio de máquina» dificulta su cuestionamiento (Roth, 2017).

Motivación. Exige fundamentar el vínculo entre prueba y hechos; un sistema de caja negra puede sustituir el razonamiento judicial por el asentimiento acrítico a la autoridad técnica.

Igualdad. Los sesgos algorítmicos afectan de modo desproporcionado a grupos vulnerables o minorías lingüísticas, a lo que se suma la restricción expresa de la LOPDP (2021) frente a decisiones automatizadas sin control humano.

En el ámbito regional, se identificaron desarrollos consolidados como Prometea en Argentina (Estevez et al., 2020) y PretorIA en Colombia (Saavedra Rionda y Upegui Mejía, 2021), y evidencia reciente sobre la expansión del uso de IA generativa por operadores judiciales en la región (Pérez-Pacheco, 2025; Segura, 2023). Estas herramientas, sin embargo, se concentraron en tareas rutinarias y clasificación formal de causas; ninguna abordó ni validó la credibilidad testimonial de forma autónoma. En Ecuador, la ausencia de criterios claros de admisibilidad, registros públicos de sistemas automatizados y auditorías algorítmicas institucionales reforzó la urgencia de la propuesta normativa desarrollada en el apartado siguiente.

Discusión

La interpretación conjunta de los resultados requiere situarlos frente a tres marcos teóricos que atraviesan este trabajo: la teoría racionalista de la valoración de la prueba, el marco conceptual de la psicología del testimonio y la literatura sobre opacidad y rendición de cuentas algorítmica. Estos tres marcos no operaron de forma aislada en el análisis, sino que se reforzaron mutuamente para

explicar por qué las limitaciones técnicas, epistemológicas y jurídicas convergieron en una misma conclusión.

Desde la teoría racionalista de la prueba (Ferrer Beltrán, 2007; Gascón Abellán, 2010; Laudan, 2006; Taruffo, 2002), un hecho solo puede darse por probado cuando la conclusión es susceptible de control intersubjetivo y de motivación argumentativa. Los resultados relativos a la opacidad de los modelos profundos y al problema de la tasa base mostraron, precisamente, que los sistemas de IA revisados no ofrecen ese tipo de control: sus salidas no son plenamente explicables ni, por tanto, plenamente motivables en los términos que exige esta tradición. Desde el marco conceptual de la psicología del testimonio, la evidencia sintetizada en el primer objetivo específico mostró que credibilidad, veracidad, exactitud y fiabilidad no son equivalentes; esta distinción resulta decisiva porque ningún sistema técnico identificado logró declarar, con precisión, cuál de esas dimensiones pretende medir, lo que sugiere que buena parte de la promesa comercial de estas herramientas descansa sobre una imprecisión conceptual de origen. Finalmente, la literatura sobre ética y rendición de cuentas algorítmica (Kordzadeh y Ghasemaghaei, 2022) permitió conectar ambos hallazgos con la pregunta jurídica de fondo: si un resultado no puede explicarse, tampoco puede controlarse racionalmente, y por tanto tampoco puede motivarse en el sentido exigido por el artículo 76, numeral 7, literal l, de la Constitución del Ecuador.

Con respecto al segundo objetivo específico, la revisión mostró que los problemas del ground truth incierto, los sesgos de entrenamiento, la escasa transferibilidad y la opacidad de los modelos no son fallas incidentales y corregibles con más datos o mejor ingeniería, sino limitaciones estructurales del enfoque mismo: mientras el objeto que se pretende medir, la sinceridad de una persona en un contexto específico, siga sin contar con un referente verificable de forma independiente, cualquier sistema entrenado contra ese referente heredará su incertidumbre. Esta conclusión es coherente con la evidencia comparada en América Latina: los desarrollos más consolidados de IA judicial en la región (Prometea, PretorIA) se concentraron deliberadamente en tareas administrativas y de gestión documental, evitando la valoración autónoma de la credibilidad (Pérez-Pacheco, 2025; Segura, 2023), lo que sugiere que la propia comunidad de desarrolladores e instituciones judiciales de la región identificó, aunque fuera de manera implícita, los mismos límites documentados en este artículo.

En cuanto al tercer objetivo específico, el contraste entre las limitaciones técnicas y las garantías constitucionales del debido proceso reveló una tensión que no se resuelve con ajustes técnicos

incrementales. La inmediación, la contradicción, la motivación y la igualdad no son requisitos formales que un sistema más preciso pueda satisfacer progresivamente; son garantías sustantivas orientadas a controlar el ejercicio del poder punitivo, y ese control exige explicabilidad, no solo exactitud estadística. Por ello, incluso en el escenario hipotético de que la precisión técnica de estos sistemas mejorara sustancialmente en el futuro, la incompatibilidad con el debido proceso no se resolvería automáticamente si esa mejora no viene acompañada de una transformación equivalente en materia de transparencia y control humano significativo.

La convergencia de estos tres órdenes de limitaciones (técnico, epistemológico y jurídico) condujo a la conclusión central de este artículo: en las condiciones actuales de desarrollo tecnológico, la evaluación autónoma de la credibilidad testimonial mediante inteligencia artificial no resulta compatible con las garantías del debido proceso reconocidas en el ordenamiento ecuatoriano. Esta conclusión no equivale a descartar el uso de la inteligencia artificial en el sistema de justicia penal ecuatoriano, sino a delimitar con precisión el tipo de función que puede desempeñar de forma legítima. El desarrollo operativo de esa delimitación, es decir, los principios, requisitos y el protocolo trifásico de admisibilidad que dan respuesta al cuarto objetivo específico, se presenta como una propuesta aplicada en el apartado siguiente, distinta de esta discusión interpretativa.

Propuesta de uso responsable para Ecuador

Como respuesta al cuarto objetivo de este trabajo, se desarrolló una propuesta estructurada en cinco ejes (principios rectores, requisitos técnicos, procesales e institucionales, y un protocolo operativo) orientada a que la IA se emplee solo como apoyo a la actividad judicial, sin sustituir la valoración humana, en términos compatibles con el régimen probatorio del COIP y las garantías constitucionales. Si los algoritmos pueden influir en decisiones con efectos jurídicos relevantes, es razonable exigirles estándares de control y contradicción equiparables a los de cualquier otra actividad pública (Boix Palop, 2020).

Principios rectores

Subsidiariedad. La IA opera solo como apoyo al juez; queda vedada cualquier sustitución del juicio valorativo humano.

Transparencia. El funcionamiento del sistema, sus datos de entrenamiento y sus métricas de validación deben ser accesibles a las partes y a auditorías externas.

Auditabilidad. Debe garantizarse la posibilidad de auditorías independientes sobre el código, los datos de entrada y los resultados del sistema.

No discriminación. El sistema debe acreditar documentalmente la ausencia de sesgos o impactos desproporcionados sobre grupos protegidos por la Constitución.

Supervisión humana significativa. El control humano sobre las salidas algorítmicas debe ser sustantivo y vinculante, nunca un trámite formal.

Requisitos técnicos y procesales mínimos

En el plano técnico: validación independiente en datos distintos a los de entrenamiento; métricas de error desagregadas por subgrupos demográficos relevantes en el Ecuador; documentación accesible de limitaciones y condiciones de degradación del rendimiento; preferencia por modelos interpretables (Rudin et al., 2020; Wang et al., 2023) o, en su defecto, explicaciones contrafactuales comprensibles (Wachter et al., 2018); y trazabilidad completa de cada salida.

En el plano procesal: motivación reforzada de toda resolución que incorpore insumos de IA; contradicción efectiva, con acceso íntegro de las partes a la documentación técnica; derecho a designar peritos informático y psicológico independientes para examinar los resultados del sistema; y prohibición de que cualquier resolución se fundamente de forma autónoma o determinante en la salida de un sistema de IA.

Requisitos institucionales

La propuesta incluyó formación judicial específica sobre los principios técnicos de la IA y la lectura crítica de sus salidas; un registro público obligatorio de los sistemas utilizados o en fase de prueba en la administración de justicia; auditorías algorítmicas independientes y periódicas; y un protocolo de suspensión inmediata para sistemas con fallos, sesgos detectados o errores inaceptables, evitando su uso continuado por inercia institucional.

Protocolo trifásico para Ecuador

Se propuso un protocolo de tres fases, articulado con la Constitución, el COIP y la LOPDP:

Etapa 1. Validez técnica. El sistema debe justificar ante un órgano de control el cumplimiento de los requisitos de validación, trazabilidad y ausencia de sesgos, dado que su falta figura entre los riesgos más señalados de la IA en la toma de decisiones (Sánchez-Monedero y Dencik, 2022; UNESCO, 2021).

Etapa 2. Validez procesal. En cada caso, el juez comprueba que el sistema esté validado, sus resultados sean trazables y la parte contraria haya accedido a la documentación técnica, salvaguardando la defensa frente a modelos de difícil control (Roth, 2017).

Etapa 3. Valoración judicial. La salida algorítmica se integra como elemento de apoyo en un debate contradictorio; la última palabra corresponde siempre al juzgador según la sana crítica, evitando decisiones sustentadas en sistemas opacos (Boix Palop, 2020; Reglamento (UE) 2024/1689).

Este protocolo podría implementarse mediante una reforma específica del COIP y su reglamento técnico complementario, en línea con los modelos internacionales que priorizan supervisión humana, transparencia y gestión de riesgos en el uso de IA en ámbitos sensibles (UNESCO, 2021; Reglamento (UE) 2024/1689).

Conclusiones

Primera. La evaluación autónoma de la credibilidad testimonial mediante sistemas de inteligencia artificial no resultó compatible, en el estado actual de la técnica y de la ciencia, con las garantías del debido proceso reconocidas en los artículos 75 y 76 de la Constitución de la República del Ecuador (2008), con el diseño del régimen probatorio del Código Orgánico Integral Penal (2014) ni con las restricciones expresas de la Ley Orgánica de Protección de Datos Personales (2021) respecto de las decisiones automatizadas que generan efectos jurídicos significativos.

Segunda. La incompatibilidad identificada no fue circunstancial ni pasajera, sino de naturaleza estructural, explicada por la interacción de tres tipos de limitaciones que se potencian entre sí: las limitaciones técnicas resultaron insuficientes frente a un objeto de estudio que no puede reducirse por completo a modelos matemáticos; las limitaciones epistemológicas afectaron la propia naturaleza del fenómeno que se pretende medir en psicología; y las limitaciones jurídicas respondieron a exigencias sustantivas del debido proceso y de la ética probatoria, por lo que no pueden entenderse como simples requisitos formales.

Tercera. La inteligencia artificial sí puede desempeñar funciones útiles y legítimas dentro del sistema de justicia penal ecuatoriano, como el apoyo en la transcripción de audiencias, la mejora en la búsqueda de jurisprudencia, la automatización de tareas repetitivas o el soporte auxiliar en labores periciales. Su uso, sin embargo, debe regirse estrictamente por un principio de

subsidiariedad, evitando en todo momento que estos sistemas emitan decisiones o valoraciones autónomas sobre la credibilidad de las personas.

Cuarta. El vacío regulatorio actual en el ordenamiento jurídico ecuatoriano respecto de la admisibilidad, el control y la valoración de la evidencia derivada de sistemas de inteligencia artificial en sede penal constituyó un factor de riesgo que debe subsanarse a la brevedad, mediante una reforma legislativa al COIP acompañada de su reglamentación técnica complementaria, para la cual el protocolo trifásico desarrollado en esta investigación ofrece una base operativa y jurídicamente viable.

Quinta. La experiencia comparada en América Latina mostró que los países de la región mantuvieron una postura de prudencia y contención, evitando el avance normativo o práctico hacia sistemas de evaluación autónoma de la credibilidad. Esta tendencia debe preservarse y explicitarse formalmente en las políticas públicas del Ecuador, blindando al sistema de administración de justicia frente a presiones eficientistas aisladas o frente a la importación acrítica de tecnologías extranjeras no validadas en el contexto local.

Sexta. La trayectoria histórica de cuestionamiento y posterior desacreditación del polígrafo funciona como una advertencia permanente sobre el riesgo de que tecnologías carentes de respaldo empírico unánime terminen por persistir institucionalmente debido a la inercia burocrática. En el dominio de la inteligencia artificial judicial, la regulación legal anticipada y rigurosa representa la garantía más sólida de eficiencia y de vigencia de los derechos fundamentales en el futuro.

Limitaciones del estudio y líneas futuras

Limitaciones

Este artículo tuvo naturaleza teórica e interdisciplinar, por lo que no incorporó datos primarios sobre la operación real de sistemas automatizados dentro de las salas de audiencia del sistema judicial ecuatoriano. Al tratarse de una revisión de alcance y no de una revisión sistemática exhaustiva de carácter cuantitativo, es posible que algunos desarrollos de software muy recientes o patentes comerciales cerradas no hayan quedado reflejados en el análisis. El énfasis normativo en la República del Ecuador implica que parte de las conclusiones jurídicas y de las propuestas operativas, aunque proyectables a nivel regional, requerirían un proceso de adaptación y de derecho comparado frente a los códigos procesales de otros ordenamientos. Finalmente, el examen de la detección automatizada del engaño se concentró de forma preferente en las familias arquitectónicas

dominantes en la literatura científica actual, sin agotar la totalidad de las propuestas técnicas marginales o emergentes en el campo de la ingeniería informática.

Líneas futuras

Realización de estudios empíricos de campo orientados a mapear la presencia efectiva de herramientas basadas en algoritmos predictivos dentro de la Fiscalía General del Estado y el Consejo de la Judicatura del Ecuador.

Investigaciones experimentales sobre la interacción entre operadores jurídicos y sistemas de inteligencia artificial, con atención específica al fenómeno de la deferencia algorítmica o sesgo de automatización, en la medida en que los sistemas predictivos pueden interferir en la autonomía del criterio profesional y generar sobreconfianza en las salidas automatizadas.

Análisis jurisprudenciales y doctrinales comparativos de las primeras decisiones de tribunales latinoamericanos y europeos sobre la admisibilidad de evidencia obtenida o elaborada a partir de redes neuronales profundas, que permitan observar la evolución del control judicial sobre estas tecnologías en relación con las garantías del debido proceso.

Declaraciones

Conflicto de intereses. Las autoras declaran que no existe ningún tipo de conflicto de intereses de carácter financiero, institucional o personal que haya podido afectar la objetividad, el desarrollo o los resultados de la presente investigación.

Financiación. El desarrollo de este estudio, así como el proceso de revisión y redacción del artículo, no contó con financiación externa específica ni con respaldo económico de fondos públicos o privados de desarrollo tecnológico.

Contribución de autoría (CRediT). Ambas autoras aportaron de forma equitativa al desarrollo de las distintas etapas de la investigación. Daniela Estefanía Zurita López participó en la conceptualización del estudio, el diseño metodológico de la revisión de alcance, la búsqueda en bases de datos técnicas y científicas, el análisis de las limitaciones identificadas a lo largo de la revisión de literatura, y la redacción del primer manuscrito. Ariana Tatiana Viteri Galeas participó en el análisis jurídico dogmático, la revisión del marco normativo constitucional y penal ecuatoriano, el diseño del protocolo trifásico de admisibilidad, la revisión crítica del primer manuscrito y la normalización final de citas y referencias.

Consideraciones éticas. Al tratarse de un artículo elaborado a partir de una revisión de literatura científica y de fuentes documentales de acceso público, no se requirió la evaluación de un comité de ética institucional de salud o educación.

Declaración de uso de inteligencia artificial generativa. Conforme a las recomendaciones del International Committee of Medical Journal Editors (ICMJE) y del Committee on Publication Ethics (COPE), se deja constancia de que el uso de herramientas de IA generativa se limitó a tareas de asistencia editorial: revisión de estilo, reducción de redundancias y verificación formal del formato de citación en normas APA. Todas las decisiones conceptuales, los análisis dogmático-jurídicos, las conclusiones normativas y el diseño de la propuesta para el contexto ecuatoriano son de entera autoría y responsabilidad de las investigadoras humanas. Cada fuente bibliográfica incorporada en el texto fue verificada manualmente en los repositorios indexados correspondientes, conforme al método descrito en este artículo.

Referencias Bibliográficas

Asamblea Constituyente del Ecuador. (2008). *Constitución de la República del Ecuador*. Registro Oficial No. 449, 20 de octubre de 2008. https://www.finanzas.gob.ec/constitucion-de-la-republica/

Asamblea Nacional del Ecuador. (2014). *Código Orgánico Integral Penal (COIP)*. Registro Oficial Suplemento No. 180, 10 de febrero de 2014.

Asamblea Nacional del Ecuador. (2021). *Ley Orgánica de Protección de Datos Personales (LOPDP)*. Registro Oficial Suplemento No. 459, 26 de mayo de 2021.

Boix Palop, A. (2020). Los algoritmos son reglamentos: La necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones. *Revista de Derecho Público: Teoría y Método, 1*, 223–269. https://doi.org/10.37417/RPD/vol_1_2020_33

Bogaard, G., Meijer, E. H., Vrij, A., & Nahari, G. (2022). Detecting deception using comparable truth baselines. *Psychology, Crime & Law, 29*(6), 567–583. https://doi.org/10.1080/1068316X.2022.2030334

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Cerrillo i Martínez, A., Di Lascio, F., Martín Delgado, I., & Velasco Rico, C. I. (Dirs.). (2022). *Inteligencia artificial y administraciones públicas: Una triple visión en clave comparada*. Marcial Pons.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Denault, V., Plusquellec, P., Jupe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., Sporer, S. L., Rioux-Turcotte, J., Jarry, J., Walsh, D., Otgaar, H., Viziteu, A., Talwar, V., Keatley, D. A., Blandón-Gitlin, I., Townson, C., Deslauriers-Varin, N., Lilienfeld, S. O., Patterson, M. L., … van Koppen, P. J. (2020). The analysis of nonverbal communication: The dangers of pseudoscience in security and justice contexts. *Anuario de Psicología Jurídica, 30*(1), 1–12. https://doi.org/10.5093/apj2019a9

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Estevez, E., Linares Lejarraga, S., & Fillottrani, P. (2020). *PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial.* Banco Interamericano de Desarrollo.

Ferrer Beltrán, J. (2007). *La valoración racional de la prueba.* Marcial Pons.

Gascón Abellán, M. (2010). *Los hechos en el Derecho: Bases argumentales de la prueba* (3.ª ed.). Marcial Pons.

González, J. L., & Manzanero, A. L. (2018). *Obtención y valoración del testimonio: Protocolo holístico de evaluación de la prueba testifical (HELPT).* Pirámide.

Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*(4), 643–659. https://doi.org/10.1037/a0023589

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Laudan, L. (2006). *Truth, error, and criminal law: An essay in legal epistemology.* Cambridge University Press.

Manzanero, A. L. (2010). *Memoria de testigos: Obtención y valoración de la prueba testifical.* Pirámide.

Markowitz, D. M., Hancock, J. T., & Bailenson, J. N. (2024). Linguistic markers of inherently false AI communication and intentionally false human communication: Evidence from hotel reviews. *Journal of Language and Social Psychology, 43*(1), 63–82. https://doi.org/10.1177/0261927X231200201

Oberlader, V. A., Quinten, L., Banse, R., Volbert, R., Schmidt, A. F., & Schönbrodt, F. D. (2021). Validity of content-based techniques for credibility assessment—How telling is an extended meta-analysis taking research bias into account? *Applied Cognitive Psychology, 35*(2), 393–410. https://doi.org/10.1002/acp.3776

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information.* Harvard University Press.

Pérez-Pacheco, Y. (2025). Regulación de la IA en el proceso judicial: Desafíos y oportunidades en América Latina. *Revista Especializada en Investigación Jurídica*, *9*(16), 1–25. https://doi.org/10.20983/reij.2025.1.1

Pérez-Rosas, V., Abouelenien, M., Mihalcea, R., & Burzo, M. (2015). Deception detection using real-life trial data. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, 59–66. https://doi.org/10.1145/2818346.2820758

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial). (2024). *Diario Oficial de la Unión Europea.* http://data.europa.eu/eli/reg/2024/1689/oj

Roth, A. (2017). Machine testimony. *Yale Law Journal*, *126*(7), 1972–2259.

Rudin, C., Wang, C., & Coker, B. (2020). The age of secrecy and unfairness in recidivism prediction. *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.6ed64b30

Saavedra Rionda, V. P., & Upegui Mejía, J. C. (2021). *PretorIA y la automatización del procesamiento de causas de derechos humanos.* Dejusticia.

Sánchez-Monedero, J., & Dencik, L. (2022). The politics of deceptive borders: "Biomarkers of deceit" and the case of iBorderCtrl. *Information, Communication & Society*, *25*(3), 413–430. https://doi.org/10.1080/1369118X.2020.1792530

Segura, R. E. (2023). Inteligencia artificial y administración de justicia: Desafíos derivados del contexto latinoamericano. *Revista de Bioética y Derecho, 58*, 45–72. https://doi.org/10.1344/rbd2023.58.40601

Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. En D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). Springer.

Suchotzki, K., & Gamer, M. (2024). Detecting deception with artificial intelligence: Promises and perils. *Trends in Cognitive Sciences, 28*(6), 481–483. https://doi.org/10.1016/j.tics.2024.04.002

Taruffo, M. (2002). *La prueba de los hechos* (J. Ferrer Beltrán, Trad.). Marcial Pons. (Obra original publicada en 1992).

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial.* Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Vrij, A., Granhag, P. A., Ashkenazi, T., Ganis, G., Leal, S., & Fisher, R. P. (2022). Verbal lie detection: Its past, present and future. *Brain Sciences*, *12*(12), 1644. https://doi.org/10.3390/brainsci12121644

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 841–887. https://doi.org/10.2139/ssrn.3063289

Wang, C., Han, B., Patel, B., & Rudin, C. (2023). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction. *Journal of Quantitative Criminology, 39*(2), 519–581. https://doi.org/10.1007/s10940-022-09545-w

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento:

N/A

Nota:

El artículo no es producto de una publicación anterior.