

Gobernanza de la Inteligencia Artificial Jurídica

Transparencia, explicabilidad y límites

Diego Mauricio Hernández¹
Juan Manuel Padilla²
Beliña Herrera Tapias³
Renzo Gandolfi Díaz⁴
Neuro José Zambam⁵
Pascual Ángel Gargiulo⁶
Aníbal Ríos Castrillón⁷
Ferney Rodríguez Serpa⁸

¹ Doctorado en Filosofía, Université Catholique de Louvain. Docente investigador Universidad de la Costa, Colombia. dhernand47@cuc.edu.co

² Magister en Derecho Constitucional. Docente investigador Universidad Tecnológica de Bolívar, Colombia. padillajuaan@utb.edu.co

³ Doctora en Derecho, Universidad Externado, Colombia. Decana Departamento de Derecho y Ciencia Política, Universidad de la Costa, Colombia. dhernand47@cuc.edu.co

⁴ Doctor en Derecho, Universidad de Mendoza. Docente investigador, Universidad Andrés Bello, Chile. renzo.gandolfi@unab.cl

⁵ Pós-Doutorado em Filosofia pela Universidade do Vale do Rio dos Sinos. Docente investigador, Atitus Educação, Brasil. neuro.zambam@atitus.edu.br

⁶ Médico Legista. Doctor en Medicina, Universidad Nacional de Cuyo. Docente investigador, Universidad Nacional de Cuyo, Argentina. pgargiulo@uncu.edu.ar

⁷ Docente investigador, Universidad Champagnat, Lima.

⁸ PhD. Filosofía con orientación en Ciencias Políticas. Universidad Autónoma de Nuevo León, México. Docente investigador, Universidad de la Costa, Colombia. frodrigu9@cuc.edu.co

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

Resumen

Este artículo examina si el riesgo de ritualización que la jurisprudencia constitucional colombiana advierte en materia de transparencia algorítmica, un riesgo que este artículo denomina una "nueva ordalía", categoría propia y no una expresión de la propia Corte ni de la literatura existente, es un problema exclusivo de la explicabilidad técnica o un patrón general en la gobernanza de la inteligencia artificial jurídica. A partir de un corpus combinado de 162 referencias seleccionadas, 50 de la literatura internacional sobre opacidad algorítmica y explicabilidad (XAI), y 112 de un corpus sistemático sobre razonamiento jurídico artificial construido mediante PRISMA 2020 sobre Scopus y Web of Science, de las cuales 111 quedan efectivamente citadas, se examina ese patrón en cinco dominios: la jurisprudencia colombiana misma (Sentencias T-323/2024 y T-067/2025), la explicabilidad técnica, la supervisión humana de sistemas de apoyo judicial, la adopción de inteligencia artificial generativa, y la atribución de responsabilidad y personalidad jurídica. Los hallazgos muestran que la sustitución de sustancia por ritual no es un riesgo aislado de la transparencia algorítmica. Reaparece, con la misma lógica de fondo, en la supervisión humana capturada por el sesgo de automatización, en la aceptación acrítica de resultados generativos, y en el reconocimiento formal de personalidad jurídica como sustituto de una capacidad de responsabilidad que no existe. Se concluye que la jurisprudencia colombiana ofrece, hasta ahora, el intento más explícito de bloquear ese patrón en la práctica, incluso sin nombrarlo como tal, y que su criterio de adecuación: inteligibilidad, impugnabilidad, fidelidad, puede generalizarse como estándar para evaluar cualquier mecanismo de control humano sobre sistemas de inteligencia artificial jurídica.

Palabras clave: transparencia algorítmica, explicabilidad, ordalía, opacidad epistémica, supervisión humana, sesgo de automatización, gobernanza de la inteligencia artificial.

Abstract

This article examines whether the ritualization risk that Colombian constitutional jurisprudence warns against in algorithmic transparency—a risk this article terms a "new ordeal," a category of its own rather than an expression used by the Court or by existing literature—is a problem exclusive to technical explainability or a general pattern in legal artificial intelligence governance. Drawing on a combined corpus of 162 selected references—50 from the international literature on algorithmic opacity and explainability (XAI), and 112 from a systematic corpus on artificial legal reasoning built through PRISMA 2020 across Scopus and Web of Science—, of which 111 are actually cited, the pattern is examined across five domains: Colombian jurisprudence itself, technical explainability, human oversight of judicial support systems, the adoption of generative artificial intelligence, and the attribution of liability and legal personhood. Findings show that the substitution of substance for ritual is not an isolated risk of algorithmic transparency. It reappears in human oversight captured

by automation bias, in the uncritical acceptance of generative outputs, and in the formal recognition of legal personhood as a substitute for a capacity for responsibility that does not exist. The article concludes that Colombian jurisprudence offers, so far, the most explicit attempt to block that pattern in practice, even without naming it as such, and that its adequacy criteria can be generalized as a standard for evaluating any mechanism of human control over legal artificial intelligence systems.

Keywords: algorithmic transparency, explainability, ordeal, epistemic opacity, human oversight, automation bias, artificial intelligence governance.

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

Introducción

En 2024 y 2025, la Corte Constitucional de Colombia estableció, en dos sentencias sucesivas, que la transparencia algorítmica es un derecho fundamental cuando el Estado usa inteligencia artificial para tomar decisiones que afectan a las personas. Al hacerlo, advirtió sobre un riesgo específico: buscar la transparencia de manera acrítica puede producir una actuación ritual, en la que las autoridades entregan series de código o documentación técnica, creando la apariencia de transparencia sin proporcionar explicaciones realmente comprensibles. Este artículo denomina ese riesgo una "nueva ordalía". Las ordalías medievales resolvían disputas mediante un rito performativo que transfería la responsabilidad a un agente sobrehumano, evitándose así tener que dar razones. La sustitución contemporánea que este artículo describe reemplaza al agente sobrehumano por el algoritmo, y el rito por el cumplimiento técnico.

La literatura sobre razonamiento jurídico artificial documenta, en dominios que nada tienen que ver con la explicabilidad técnica, el mismo patrón lógico. Cuando un sistema de apoyo judicial se somete a supervisión humana, esa supervisión puede degradarse en sesgo de automatización, una ratificación ritual de lo que la máquina ya decidió. Cuando un sistema de inteligencia artificial generativa produce un resultado plausible, su aceptación acrítica reproduce la misma sustitución de sustancia por apariencia. Cuando el derecho niega personalidad jurídica plena a un sistema artificial, o le atribuye responsabilidad de forma difusa entre desarrolladores y operadores, enfrenta la misma pregunta de fondo que enfrenta la explicabilidad. Si el reconocimiento formal de un mecanismo de control basta, o si ese mecanismo debe demostrar sustancia real.

La pregunta que ordena este artículo es la siguiente, ¿es la "nueva ordalía" que este artículo identifica, a partir del riesgo advertido por la jurisprudencia colombiana, un fenómeno propio de la transparencia algorítmica, o la manifestación más visible de un patrón general en la gobernanza de la inteligencia artificial jurídica? El apartado 1.1 desarrolla el concepto de ordalía nueva como categoría analítica, distinguiendo la opacidad epistémica genuina del ritual de cumplimiento que puede o no acompañarla. El apartado 1.2 describe la construcción del corpus combinado sobre el que trabaja este artículo. El apartado 2 examina el patrón en cinco dominios. La jurisprudencia colombiana, la explicabilidad técnica, la supervisión judicial, la inteligencia artificial generativa, y la responsabilidad y personalidad jurídica. El apartado 3 discute las implicaciones de encontrar, o no, el mismo patrón en los cinco dominios.

1. Marco conceptual

1.1 La nueva ordalía: de la opacidad epistémica al ritual de cumplimiento

La opacidad algorítmica no es un fenómeno único. Tiene, según una distinción ya clásica en la literatura, al menos tres causas independientes. El secreto deliberado, protegido como propiedad intelectual o secreto comercial; la brecha de competencia entre la complejidad técnica del sistema y la capacidad de su audiencia para comprenderla; y la opacidad intrínseca, inducida por la escala y la no linealidad de los modelos de aprendizaje profundo, cuya lógica interna excede la comprensión humana incluso cuando no hay secreto ni brecha de competencia de por medio (Burrell, 2016). Esta última, la opacidad epistémica en sentido estricto, describe una barrera fundamental. El camino que sigue un sistema desde la entrada hasta la salida es, literalmente, incognoscible para un observador humano, no solo difícil de explicar (Humphreys, 2009).

Esa opacidad epistémica tiene, en sí misma, un peligro legal doble. No es solo que las personas puedan ser objeto de decisiones que no pueden comprender. Es que la opacidad devora el derecho a cuestionar la evidencia, exigir razones y buscar reparación, la preocupación central que atraviesa la crítica a la "sociedad de la caja negra" (Pasquale, 2015; Ananny y Crawford, 2018). Frente a ese peligro, el derecho responde exigiendo transparencia. Pero la transparencia, entendida como mera apertura del código o de los datos, no resuelve por sí sola el problema que pretende resolver. Un sistema puede ser plenamente "abierto": código disponible, parámetros publicados, y seguir siendo epistémicamente inescrutable para cualquier persona, incluidos sus propios creadores (Castelvecchi, 2016; Lipton, 2018).

Aquí es donde entra el concepto de ordalía nueva. Las ordalías medievales eran procedimientos no testimoniales y no documentales, invocados precisamente cuando no había pruebas ordinarias disponibles. Su autoridad descansaba en un rito performativo que transfería la responsabilidad del juicio a un agente sobrehumano, no en razones. El análogo contemporáneo es preciso. Cuando un sistema jurídico exige transparencia algorítmica y la autoridad responde con un volcado de código, una matriz de coeficientes o un mapa de prominencia que nadie puede interpretar de forma significativa, el rito se cumple sin que la sustancia lo acompañe. La responsabilidad legal se desplaza de la autoridad humana a una actuación técnica de apertura, exactamente como antes se desplazaba a la voluntad divina.

La distinción operativa que este artículo utiliza para el resto del análisis es, en consecuencia, la siguiente. Un mecanismo de control sobre un sistema de inteligencia artificial, sea explicabilidad técnica, supervisión judicial, revisión humana de un resultado generativo, o atribución de responsabilidad, tiene sustancia cuando permite a la persona afectada comprender el resultado, cuestionarlo con una base real, y obtener una revisión que pueda

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

modificarlo. Tiene solo ritual cuando cumple la forma exigida por la norma sin producir ninguna de esas tres capacidades. La diferencia no siempre es visible desde fuera. Un volcado de código y una explicación genuina pueden parecer, documentalmente, lo mismo. Por eso la literatura técnica insiste en que una explicación adecuada debe ser fiel a lo que el sistema efectivamente hizo, no solo plausible para quien la recibe (Selbst y Barocas, 2018; Rudin, 2019). Una explicación plausible pero infiel no reduce la opacidad. La disfraz.

Ese mismo problema de fidelidad reaparece, con vocabulario distinto, en el resto de los dominios que este artículo examina. La supervisión humana sobre un sistema de apoyo judicial es sustancial cuando el supervisor ejerce un juicio real sobre el resultado, y es ritual cuando el sesgo de automatización lo reduce a una firma. La revisión de un resultado de inteligencia artificial generativa es sustancial cuando el usuario verifica su corrección, y es ritual cuando la plausibilidad del resultado sustituye esa verificación. El reconocimiento de responsabilidad es sustancial cuando identifica a un agente humano que efectivamente pueda responder, y es ritual cuando se diluye entre tantos actores que nadie, en la práctica, responde. La hipótesis de trabajo de este artículo es que estos cuatro fenómenos, y la explicabilidad misma, comparten una misma lógica de fondo, documentada por primera vez con precisión jurídica en la jurisprudencia colombiana, pero no exclusiva de ella.

1.2 Metodología

Este artículo trabaja sobre un corpus combinado construido en dos etapas, cada una con su propio protocolo de identificación.

La primera etapa incorpora la literatura internacional sobre opacidad algorítmica, explicabilidad y rendición de cuentas de la inteligencia artificial, reunida originalmente para un análisis doctrinal de la jurisprudencia constitucional colombiana sobre transparencia algorítmica. Esa revisión se apoyó en Scopus y Web of Science, cubriendo el periodo del 1 de enero de 2017 a septiembre de 2025, en inglés y español, combinando términos como "transparencia algorítmica", "IA explicable", explicabilidad, interpretabilidad, caja negra, contrafactual, impugnabilidad, rendición de cuentas, gobierno, sector público y debido proceso. La búsqueda inicial arrojó 140 registros de Scopus y 83 de Web of Science; tras la deduplicación por título entre bases, quedaron 173 elementos únicos, complementados con un seguimiento limitado de citas hacia atrás y hacia adelante a partir de trabajos ampliamente citados. De ese corpus, y tras eliminar duplicados internos y referencias fragmentarias que el propio documento original señala como pendientes de verificación, este artículo retoma 50 referencias.

La segunda etapa retoma el corpus sistemático construido mediante PRISMA 2020 sobre Scopus y Web of Science para una línea de trabajo independiente sobre razonamiento jurídico artificial, con 231 registros activos tras depuración de duplicados y cribado por título y

resumen. De ese corpus, este artículo selecciona 112 referencias distribuidas en cuatro dominios: explicabilidad de sistemas jurídicos (24 registros, el conjunto completo de ese núcleo temático), supervisión humana y confianza en la adjudicación algorítmica (20 registros), riesgos de fiabilidad y sobreconfianza en la inteligencia artificial generativa aplicada al derecho (24 registros), y responsabilidad y personalidad jurídica de sistemas artificiales (44 registros). La selección dentro de cada dominio privilegió los registros que tratan explícitamente supervisión, confianza, sesgo o legitimidad, sobre los que documentan capacidad técnica sin abordar esa dimensión de control.

El corpus combinado reúne, en consecuencia, 162 referencias seleccionadas: 50 de la literatura internacional sobre opacidad y explicabilidad, y 112 del corpus sistemático sobre razonamiento jurídico artificial. De ese total, 111 quedan efectivamente citadas en el desarrollo argumentativo de este artículo; el resto formó parte del acervo de trabajo sin encontrar, finalmente, un punto de integración preciso en el argumento. Ningún registro de la primera etapa incluye identificador DOI⁹ verificado en el documento de origen para todos los casos; ese dato se completa registro por registro en la fase de verificación previa a la versión final. Cuatro registros quedan marcados como fragmentarios, pendientes de confirmación bibliográfica completa.

La *Figura 1* resume el flujo combinado de identificación y selección de ambas fuentes. La *Figura 2* desglosa la distribución temporal de las referencias efectivamente citadas, por dominio.

⁹ Un DOI (Digital Object Identifier o Identificador de Objeto Digital) es un código alfanumérico único y permanente que sirve para identificar de forma unívoca un contenido electrónico, como un artículo científico, un libro o un conjunto de datos.

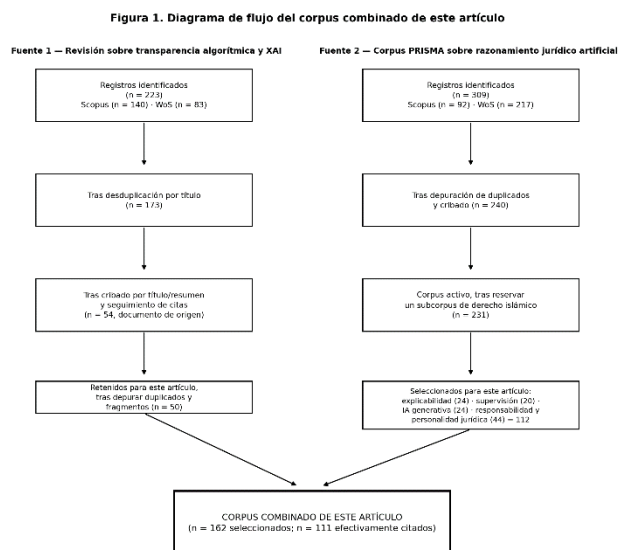


Figura 1. Elaboración propia.

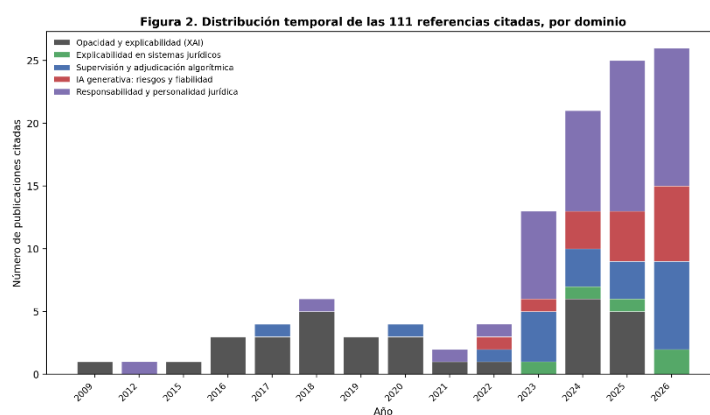


Figura 2. Elaboración propia.

2. Desarrollo

2.1 La jurisprudencia colombiana como caso: T-323/2024 y T-067/2025

La Corte Constitucional de Colombia ofrece, hasta ahora, el desarrollo jurisprudencial más explícito sobre cómo distinguir sustancia de ritual en la transparencia algorítmica. En la Sentencia T-323/2024, la Corte advierte que el uso de inteligencia artificial generativa en la adjudicación, sin salvaguardas, puede socavar el deber de motivar las decisiones, la independencia del juez natural y el derecho a la defensa; en consecuencia, la transparencia y la explicabilidad por parte de los jueces resultan indispensables para preservar el debido proceso (Corte Constitucional de Colombia, 2024). En la Sentencia T-067/2025, la Corte enmarca la transparencia algorítmica como garantía de acceso a la información pública, dirigida a que las personas comprendan cómo los sistemas automatizados del Estado procesan datos y toman decisiones que afectan derechos (Corte Constitucional de Colombia, 2025).

El detalle jurídicamente más significativo de esta jurisprudencia es su rechazo explícito a equiparar transparencia con divulgación de código fuente, más que la sola exigencia de transparencia. La Corte reconoce que publicar el código es una herramienta válida, pero advierte que no es la única, ni necesariamente la mejor en cada caso. A veces lo que se necesita es ofrecer un tipo de claridad distinto, no simplemente "abrir" el código: una explicación significativa del proceso de toma de decisiones que satisfaga la comprensión pública (Corte Constitucional de Colombia, 2025). Esa distinción es, en los términos del apartado 3.1, precisamente la línea entre ritual y sustancia. Un volcado de código satisface el ritual sin garantizar la sustancia; una explicación significativa persigue la sustancia, con o sin código de por medio.

De esa jurisprudencia se derivan trece subreglas operativas que traducen los principios constitucionales: acceso a la información, debido proceso, control judicial efectivo, en deberes concretos de diseño, divulgación y justificación. Entre ellas, tres resultan especialmente relevantes para el argumento de este artículo. La subregla de valoración probatoria establece que ordenar y valorar pruebas es una función judicial indelegable a la inteligencia artificial. La subregla de no privatización de la transparencia impide que contratos o derechos de propiedad intelectual anulen el acceso constitucional a la información. Y la subregla de acceso mediado por expertos habilita un mecanismo intermedio, revisión técnica independiente, para los casos en que la divulgación pública total no sea viable (Corte Constitucional de Colombia, 2024, 2025).

Sobre esa base, la jurisprudencia colombiana articula un criterio de adecuación de tres elementos que este artículo adopta como estándar general, no limitado a la explicabilidad

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

técnica. Una explicación es adecuada solo si es inteligible, expresada en lenguaje claro y preciso para la persona afectada; impugnabile, capaz de sostener un desafío significativo al resultado, con una vía real de reparación; y fiel, una representación precisa de lo que el sistema efectivamente hizo, no una historia plausible construida después del hecho. Los tribunales y la investigación técnica en explicabilidad coinciden en un punto crítico. La apertura sin fidelidad no produce rendición de cuentas, produce su apariencia (Selbst y Barocas, 2018).

Dos elementos adicionales de esta jurisprudencia merecen destacarse por su generalidad. Primero, la Corte vincula la transparencia algorítmica al principio democrático, entendiendo que los sistemas opacos erosionan el escrutinio público real y alimentan lo que otra literatura describe como una crisis de legitimidad "algorítmica" (Alnemr, 2024). Segundo, la Corte exige, junto a la transparencia, un funcionario humano identificable que pueda asumir la responsabilidad del resultado final, prohibiendo la aprobación automática basada únicamente en la salida de un sistema. Ese requisito no es una salvaguarda adicional entre otras. Es la condición que impide que la transparencia misma se convierta en un mecanismo de desplazamiento de responsabilidad, el riesgo exacto que el concepto de ordalía nueva nombra.

2.2 El patrón técnico: opacidad, explicabilidad y sus límites

La literatura técnica sobre inteligencia artificial explicable (XAI) documenta, con precisión creciente, por qué el criterio de fidelidad de la Corte colombiana responde a un problema real y medible, no a una exigencia retórica. El compromiso entre interpretabilidad y rendimiento predictivo es, desde hace más de una década, un lugar común de la disciplina. Los modelos simples, como la regresión lineal o los árboles de decisión pequeños, son comprensibles pero limitados; los modelos de alta capacidad, como las redes neuronales profundas, ofrecen mejor rendimiento al costo de volverse inescrutables (Lipton, 2018; Carvalho, Pereira y Cardoso, 2019). Frente a ese compromiso, una posición influyente en la literatura sostiene que la solución es evitar las cajas negras, no explicarlas: usar modelos interpretables desde el diseño en cualquier decisión de alto riesgo, en lugar de construir explicaciones posteriores para modelos que nunca debieron ser opacos (Rudin, 2019).

Las técnicas de explicación posterior más difundidas: LIME y SHAP, ilustran el problema de fidelidad de forma directa. Ambas producen aproximaciones locales. Destacan qué variables influyeron más en una predicción específica, sin reconstruir el proceso completo del modelo (Ribeiro, Singh y Guestrin, 2016; Lundberg y Lee, 2017). Esa aproximación tiene un costo. Los métodos de prominencia pueden producir mapas de calor visualmente convincentes pero semánticamente vacíos, que no superan pruebas básicas de confiabilidad (Adebayo et al., 2018), y los explicadores locales pueden ser manipulados mediante

variaciones mínimas en los datos de entrada, sin que el modelo subyacente cambie (Slack et al., 2020). Una explicación que puede ser engañada por un ajuste deliberado no es fiel en el sentido que exige la Corte colombiana. Es, en el vocabulario de este artículo, ritual con apariencia técnica.

La literatura jurídica que se ocupa específicamente de sistemas de apoyo judicial documenta el mismo problema desde el lado de la aplicación. Un modelo interpretable, aplicado a la detección de jurisprudencia sobre vivienda y desalojo dentro de la base de datos del Tribunal Europeo de Derechos Humanos, prioriza deliberadamente la trazabilidad sobre el rendimiento bruto, precisamente para que un resultado pueda auditarse (Mohammadi, Wieling y Vols, 2025). Una revisión sistemática sobre la capacidad de la inteligencia artificial para predecir decisiones judiciales confirma que buena parte de esa literatura mide precisión sin medir, con el mismo rigor, si la predicción resultante es explicable para quien debe responder por ella (Kazun, 2024). Un sistema de predicción de acusaciones y artículos legales en materia ambiental documenta ganancias de rendimiento sin resolver, de forma explícita, cómo esas predicciones se traducirían en una motivación judicial revisable (Leng et al., 2025).

Otro cuerpo de literatura examina la explicabilidad desde la cognición humana, un ángulo que la jurisprudencia colombiana también invoca al exigir que la explicación se adapte "a la audiencia". Lo que hace que una explicación sea buena no es una fórmula única. Depende de quién la recibe y en qué contexto, un hallazgo tanto de la ciencia técnica como de la ciencia cognitiva (Miller, 2019). La confianza del usuario en un sistema de asesoría basado en inteligencia artificial depende de la precisión del modelo, de sus explicaciones, y de experiencias previas de interacción que la refuerzan o la corrigen con el tiempo (Kahr et al., 2024). El proceso de traducir un cálculo estadístico en una razón utilizable por una persona exige, en consecuencia, herramientas de comunicación y no solo de matemática (Cordella y Gualdi, 2025). Investigación aplicada sobre explicabilidad en dominios administrativos: clasificación aduanera, análisis de datos tabulares, procesos de decisión crítica, confirma que ese ejercicio de traducción es factible, pero exige diseño deliberado, no un subproducto automático de la apertura técnica (O'Brien Quinn et al., 2024; Arya et al., 2025; Kalasampath et al., 2025).

Sobre el trasfondo constitucional específicamente, la literatura converge en una advertencia compartida con la jurisprudencia colombiana. La explicabilidad sin verificación de fidelidad puede generar confianza injustificada, no rendición de cuentas real (Wachter, Mittelstadt y Floridi, 2017; Goodman y Flaxman, 2017). Las explicaciones contrafactuales, qué habría tenido que cambiar en los datos para obtener un resultado distinto, ofrecen una alternativa prometedora precisamente porque no requieren abrir el modelo. Permiten a la persona afectada entender qué modificar sin exponer secretos comerciales ni exigir una comprensión

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

técnica que la mayoría de las audiencias no tiene (Wachter, Mittelstadt y Russell, 2018). Ese diseño encaja con la subregla de acceso mediado por expertos que ya documentó el apartado 2.1, y sugiere que el criterio de fidelidad no exige transparencia total del modelo, exige razones verificables sobre el caso concreto.

Finalmente, un marco normativo emergente propone estandarizar la documentación de riesgo de los sistemas de inteligencia artificial de forma legible por máquina, inspirado directamente en las categorías del Reglamento europeo de Inteligencia Artificial (Golpayegani et al., 2024). Ese esfuerzo de estandarización, leído junto con el resto de esta literatura, confirma el diagnóstico del apartado 3.1. La opacidad epistémica es un rasgo persistente de los modelos complejos, no un error corregible, y la respuesta jurídica adecuada exige que cualquier institución que use un modelo opaco demuestre, caso por caso, que puede producir razones inteligibles, impugnables y fieles, no que esa opacidad desaparezca por completo (De Bruijn, Warnier y Janssen, 2022; Larsson y Heintz, 2020; Busuioc, 2021).

2.3 La ordalía en la adjudicación algorítmica: supervisión humana y confianza

El requisito de supervisión humana que la jurisprudencia colombiana impone a los sistemas de transparencia algorítmica tiene un correlato exacto en la literatura sobre apoyo judicial. Hay que reconocer, ante todo, que el aprendizaje automático se construye sobre el desempeño humano y se compara contra él, lo que exige preguntar primero si ese desempeño humano de referencia satisface, él mismo, normas de imparcialidad y coherencia (Simon, 2026). Esa pregunta previa rara vez se formula. La literatura tiende a evaluar si la máquina alcanza el estándar humano sin verificar antes si ese estándar es, en la práctica judicial real, tan sólido como el debido proceso exige.

El mecanismo institucional propuesto con más frecuencia para resolver esta tensión es el humano en el bucle. Mantener a una persona con autoridad para revisar, corregir o revocar la decisión algorítmica. Pero la literatura reciente documenta que ese mecanismo puede convertirse, él mismo, en ritual. El argumento del humano en el bucle puede diluir en vez de reforzar la línea entre decisión humana y decisión de la máquina, incluso cuando la asistencia de inteligencia artificial es formalmente limitada (Paar, 2025). Mantener al humano en el bucle únicamente en la etapa de apelación, en vez de en cada paso del proceso, preservaría los beneficios de velocidad sin renunciar a la supervisión donde más importa, pero exige diseñar deliberadamente en qué punto exacto del proceso interviene ese humano (Cohen et al., 2023). El desafío real que expone la inteligencia artificial generativa en el razonamiento jurídico es una brecha de gobernanza del proceso, no un problema meramente tecnológico resoluble con mejores modelos (Shukla, 2026).

Los argumentos que invocan la filosofía del derecho para templar la decisión algorítmica con juicio humano suelen apelar a criterios de justicia individual que la propia automatización

pone en tensión, y examinar si esos argumentos resisten el escrutinio exige distinguir con precisión qué es lo que el juicio humano añade, más allá de su sola presencia formal (Binns, 2022). Preservar la discreción judicial en un entorno de adjudicación aumentada exige distinguir con exactitud qué preguntas siguen siendo del tipo "cómo se decide" y cuáles ya se han desplazado al tipo "cuándo delegar la decisión" (Panigrahi, Panda y Mahapatra, 2026). Un modelo híbrido de legitimidad judicial en la era algorítmica propone una teoría socio-jurídica de la adjudicación humano-máquina que reconoce fuentes plurales de legitimidad, y advierte que ninguna de esas fuentes sustituye por sí sola a las demás (Wanggai, Firdaus y Santos, 2026).

La pregunta de si la máquina puede ocupar, siquiera parcialmente, el lugar del juez tiene respuestas divergentes según el ángulo de entrada. Un capítulo dedicado explícitamente a esa pregunta concluye que la tecnología actual está lejos de sustituir el juicio judicial, aunque su influencia en la preparación de decisiones ya es significativa (Umit Atilgan, 2024). Otro trabajo, bajo el título deliberadamente irónico de la trascendencia de la ética humana, documenta que el escenario descrito hoy en las salas de audiencia, un abogado humano y un juez humano, convive ya con herramientas que participan activamente en la preparación de los argumentos que ambos presentan (Swisher, 2025). La metáfora del cuarto chino aplicada al derecho público sostiene que un sistema puede manipular símbolos jurídicos sin comprender su significado, incluso cuando produce resultados formalmente correctos, la misma distinción entre forma y sustancia que organiza este artículo (Perfetti, 2017). Un examen del núcleo no computable del derecho argumenta que buena parte de los desafíos de la inteligencia artificial jurídica surgen de rasgos persistentes del razonamiento jurídico mismo, que ninguna optimización futura eliminará, no de límites técnicos superables (Stanizzi, 2026).

La singularidad de la inteligencia artificial en el poder judicial chino se examina como una transformación institucional inminente, cuya magnitud excede la de cualquier herramienta de apoyo puntual (Ji, 2020). Una comparación entre decisión humana y algorítmica en materia penal documenta patrones de error distintos entre ambas, no una superioridad uniforme de ninguna de las dos (Karsai, 2024). Los riesgos potenciales de implementar inteligencia artificial en la toma de decisiones judiciales incluyen tanto sesgos heredados de los datos de entrenamiento como sesgos introducidos por el propio diseño del sistema de supervisión (Razmetaeva, 2024). La inteligencia artificial plantea cuestiones éticas significativas para la justicia y la rendición de cuentas del sistema judicial en su conjunto, más allá de la eficiencia de cualquier herramienta específica (Pandey, 2026). Un examen de las implicaciones legales y epistemológicas del uso de inteligencia artificial en decisiones judiciales advierte que la precisión estadística de un modelo dice poco sobre si su razonamiento puede reconstruirse y revisarse (Scamardella y Vestoso, 2023). Un enfoque deconstructivo que compara jueces humanos y artificiales evita asumir de antemano la

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

superioridad de ninguno de los dos, y encuentra en ambos las mismas vulnerabilidades al sesgo bajo presión institucional (Sağlam, 2025).

Estudios de adopción comparada confirman que este problema ya no es hipotético. Un panorama nacional e internacional documenta el amplio rango de aplicaciones con que la inteligencia artificial ya intenta atender problemas persistentes del sistema judicial, como la duración excesiva de los procesos (Lamba y Kohli, 2026). Un estudio brasileño evalúa directamente si es posible, en el estado actual de la técnica, emular tecnológicamente el razonamiento jurídico humano en la fundamentación de sentencias, y concluye que la emulación formal es más alcanzable que la emulación sustantiva (Casimiro y de Gouveia, 2023). Retomando de forma explícita la crítica de Michele Taruffo al juez y al algoritmo, un análisis reciente pregunta si la eficiencia computacional puede sustituir alguna vez la ponderación judicial, y responde que puede acelerarla sin poder sustituirla (Giabardo, 2023).

El patrón que emerge de este dominio es consistente con el de la explicabilidad técnica. La supervisión humana formalmente exigida no garantiza, por sí sola, que ese humano ejerza un juicio real. Cuando el diseño institucional no especifica en qué punto exacto interviene, con qué información y bajo qué incentivos, la supervisión corre el mismo riesgo que el volcado de código. Cumple el requisito sin producir la sustancia que el requisito busca proteger.

2.4 La ordalía en la inteligencia artificial generativa: alucinación y sobreconfianza

Si la explicabilidad enfrenta el problema de la fidelidad y la adjudicación algorítmica enfrenta el problema de la supervisión ritual, la inteligencia artificial generativa añade un tercer mecanismo de ritualización, más silencioso que los dos anteriores. La aceptación de un resultado plausible como si fuera un resultado verificado. El alcance y los límites de la inteligencia artificial en el razonamiento jurídico dependen de cuánto replica esta tecnología los procesos racionales que sostienen la interpretación del derecho, no solo de cuánto acelera tareas ya definidas (Córdoba, 2026). Un análisis crítico de los modelos de lenguaje de gran escala en el dominio legal aborda directamente los riesgos de una adopción sin control, antes de considerar cualquier forma de asistencia útil (Schilder, 2023).

La brecha entre capacidad técnica y cambio profesional real es, en sí misma, un dato revelador. La práctica legal ha cambiado relativamente poco pese a la creciente capacidad técnica de la inteligencia artificial generativa, y una explicación cognitiva de esa brecha sostiene que los profesionales luchan por comprender en profundidad sistemas cuyo funcionamiento interno no pueden auditar (Schwarcz et al., 2025). Esa resistencia admite una lectura optimista. Los abogados que integran estas herramientas en tareas acotadas, sin delegarles la ponderación final, están trazando en la práctica la misma línea que la jurisprudencia colombiana traza en la doctrina.

Los estudios técnicos sobre fiabilidad documentan fallas específicas del razonamiento generativo, no solo advertencias generales sobre su uso. La predicción de sentencias exige todavía mejoras sustantivas en razonamiento cuantitativo antes de alcanzar un desempeño constante (Han, Feng y Ng, 2026). La fiabilidad de un sistema de inteligencia artificial basado en conocimiento no puede evaluarse solo por la exactitud de la respuesta final, porque el razonamiento causal subyacente puede estar deconstruido incluso cuando la conclusión suena correcta (Park y Park, 2026). Modelos de lenguaje usados como jueces automatizados fallan al intentar aprobar la parte escrita de un examen oficial de habilitación profesional, una prueba empírica directa de sus límites (Karp, Kubaszewska y Szymanski, sin fecha). La interpretación generativa de textos legales conserva ventajas y limitaciones que la literatura examina sin resolver a favor de ninguna de las dos (Nykolyna, 2025).

El sesgo cognitivo, humano y artificial, atraviesa buena parte de este bloque. Los modelos de lenguaje de gran escala exhiben sesgos similares a los que afectan el juicio humano en la toma de decisiones legales, comprometiendo la equidad de ambos por igual (Santtila, Sun y Zappala, sin fecha). Explorar la psicología del razonamiento moral y jurídico de los modelos de lenguaje examina cuestiones éticas que se vuelven urgentes a medida que se ajustan versiones futuras de estos sistemas para tareas de alto desempeño (Almeida, Nunes y de Araújo, 2024). Un sistema de procesamiento de lenguaje sensible a la ambigüedad busca ofrecer confiabilidad en dominios de alto riesgo como el razonamiento jurídico, precisamente porque los modelos actuales todavía fallan en interpretar comunicación humana en esos contextos (Dorr, Jayaweera, Youm y Gilda, 2026).

La gobernanza y la alineación normativa conectan este dominio con el resto del artículo. Alinear la inteligencia artificial con las intenciones humanas exige que el derecho informe el diseño del sistema desde el inicio, no que lo corrija después del despliegue (Nay, 2022). Ningún banco de pruebas evalúa todavía si los modelos de lenguaje razonan correctamente sobre derechos humanos, pese a que median cada vez más decisiones sobre qué derechos humanos se realizan y cómo (Thais, Kennedy y Buchman, 2026). Los riesgos de la inteligencia artificial generativa exigen una gobernanza de múltiples herramientas que atienda violaciones de privacidad y otros riesgos sociales derivados de sistemas de uso masivo (Hu, Liu y Yu, 2024). El impacto transformador de la inteligencia artificial generativa sobre la toma de decisiones legales incluye obligaciones de divulgación que la regulación vigente todavía no resuelve con claridad (Cardoso, Chan y Pereira, 2024).

Un último grupo de estudios examina respuestas institucionales concretas a estos riesgos, con eco directo en las subreglas colombianas del apartado 4.1. Los dilemas legales y éticos que introduce la inteligencia artificial generativa en el campo educativo rompen la lógica binaria de sujeto propia de las relaciones jurídicas tradicionales (Zhang y Zeng, 2026). Diseñar barandillas de contención para sistemas de inteligencia artificial generativa en la práctica

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

legal busca atender riesgos críticos sin renunciar a los beneficios que estos sistemas ofrecen (Trivedi, 2025). El despliegue de estos sistemas en la profesión legal exige una restricción firme frente a los riesgos de alucinación y sobreconfianza, riesgos que se agravan en el dominio donde la exactitud fáctica tiene consecuencias jurídicas directas (Gültekin-Várkonyi, 2026). Facilitar decisiones administrativas más eficientes exige, al mismo tiempo, garantizar mejor acceso a explicaciones y remedios para las personas afectadas, dos objetivos que no siempre convergen en el mismo diseño de sistema (Pesch, 2025). Definir las capacidades centrales que la inteligencia artificial generativa necesitaría para operar con seguridad en el derecho es un ejercicio todavía en curso (Linna y Linna, 2026).

2.5 La ordalía institucional: responsabilidad y personalidad jurídica como sustitutos formales

El cuarto dominio de este artículo examina el riesgo de ritualización en su forma más abstracta. Cuando el derecho no puede reconstruir cómo un sistema llegó a un resultado, y tampoco puede identificar con precisión quién responde por ese resultado, la tentación institucional es resolver el vacío con una categoría formal: responsabilidad difusa, personalidad jurídica limitada, que parece cerrar la pregunta sin responderla realmente. Ese movimiento es, en sentido estricto, una ordalía institucional. Sustituye la sustancia de la rendición de cuentas por el rito de una clasificación jurídica.

La creciente autonomía de los sistemas de inteligencia artificial está remodelando supuestos jurídicos de larga data sobre la responsabilidad legal, y los marcos tradicionales de responsabilidad civil extracontractual no fueron diseñados para actores que aprenden y se adaptan con independencia progresiva (Bharti y Kaur, 2026). La atribución y la vía de responsabilidad de las entidades de inteligencia artificial exigen, en consecuencia, un rediseño de las categorías clásicas de imputación, no solo su aplicación analógica (Wang, 2026). La responsabilidad civil de asesores y abogados por daños derivados del uso de estos sistemas confirma que esa tensión ya produce litigios concretos (Szpyt, 2025), y el impacto de la inteligencia artificial sobre las reglas de responsabilidad civil se extiende a sistemas que operan fuera del control directo de un operador humano en el momento del hecho dañoso (Dajeh, 2025).

En materia penal, el desafío es todavía más agudo, porque la imputación penal exige identificar una voluntad culpable que un sistema artificial, por definición, no tiene. El derecho penal contemporáneo enfrenta un desafío significativo con la emergencia de la inteligencia artificial como entidad no humana capaz de producir resultados lesivos sin una decisión humana identificable en el momento del hecho (Adi, Darma y Kurniawan, 2024). Un estudio comparado entre distintas jurisdicciones confirma que ninguna ha resuelto de modo uniforme esa pregunta (Thuy, Nghiep y Anh, 2026), y las implicaciones de la

inteligencia artificial sobre la responsabilidad penal cuestionan directamente los fundamentos tradicionales de esa responsabilidad, contruidos sobre la culpabilidad de un agente humano (Menasria, Madjri y Beddiar, 2025). Un análisis centrado en ChatGPT examina las limitaciones de los marcos tradicionales frente a un sistema capaz de generar contenido dañoso sin instrucción específica para ese resultado (Faqr, 2025), y el alto grado de autonomía que estos sistemas alcanzarán en el futuro cercano plantea, todavía sin resolver, el problema de a quién imputar jurídicamente un resultado ilícito (Fauzan, Amarta y Fidela, sin fecha).

La aplicación concreta de inteligencia artificial dentro del proceso penal confirma que la ritualización no es una hipótesis abstracta. La legislación procesal vigente carece, en general, de definiciones jurídicas claras sobre qué constituye un sistema de inteligencia artificial, y de criterios de admisibilidad probatoria específicos para su uso (Arzumanyan, 2026). El estatuto de la inteligencia artificial dentro del proceso penal se examina desde su doble condición de instrumento probatorio y posible sujeto de imputación, sin que la doctrina procesal haya resuelto esa ambigüedad (Korotysh, 2026). La aplicación de inteligencia artificial a conceptos evaluativos dentro del proceso penal exige compatibilizar la lógica probabilística del sistema con la garantía de individualización de la pena (Barannik, Balatska y Holovko, 2026). Una integración de técnicas predictivas digitales en el sistema penal italiano documenta oportunidades de eficiencia que conviven con riesgos éticos y jurídicos todavía sin resolver (Garzo y Palumbo, 2025).

Cuando la responsabilidad individual se vuelve difícil de atribuir, la pregunta se traslada de forma natural hacia la personalidad jurídica. Si el sistema no puede ser penalmente culpable, ¿puede al menos ser civilmente responsable como sujeto propio? La literatura reciente responde con cautela creciente, no con apertura. Es más seguro otorgar personalidad jurídica a los osos que a la inteligencia artificial, sostiene un análisis que advierte contra la precipitación retórica de conceder derechos o intereses morales a estos sistemas sin resolver antes su capacidad de responsabilidad correlativa (Nelson, 2026). El reconocimiento legal de la inteligencia artificial como sujeto de derecho en Indonesia se examina explícitamente bajo el criterio de la dignidad humana, exigiendo que cualquier reconocimiento de ese tipo pase por ese filtro categórico antes que por la sola capacidad operativa del sistema (Disemadi y Sudirman, 2025). La cuestión de si las máquinas pueden ostentar derechos y responsabilidades sigue debatiéndose sin resolución uniforme, y el consenso disponible se inclina hacia fórmulas de agencia limitada antes que hacia el reconocimiento pleno (Shradda Tomar, 2026). Una propuesta de reconocimiento limitado, a medio camino entre el objeto y el sujeto de derecho, ofrece esa vía intermedia, aunque todavía no ha sido adoptada por ninguna legislación (Xudaybergenov, 2023b).

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

La analogía societaria, invocada con frecuencia en esta literatura, ilustra bien el riesgo de ritualización. La inteligencia artificial se compara a menudo con las corporaciones en los estudios jurídicos sobre su personalidad, explorando qué elementos de la ficción jurídica societaria podrían trasladarse a un agente artificial sin comprometer la coherencia del sistema (Raskulla, 2023). Pero una corporación tiene patrimonio propio, órganos de gobierno identificables y personas físicas responsables detrás de cada decisión relevante. Otorgar una etiqueta de personalidad jurídica a un sistema artificial sin exigir ese mismo sustrato institucional repetiría, en el plano de la teoría del derecho, la misma sustitución de forma por sustancia que la jurisprudencia colombiana bloquea en materia de transparencia. El estatuto legal de las entidades de inteligencia artificial (Saraswathy, 2025), la subjetividad jurídica examinada en distintos sistemas asiáticos (Cho y Tsche, 2024; Yongho, 2024; Ravizki y Lintang Yudhantaka, 2022), y el examen equilibrado de pros y contras de ese reconocimiento (Xudaybergenov, 2023a) documentan, todos, la misma conclusión provisional. Ningún sistema jurídico ha resuelto todavía cómo dotar de sustancia a una personalidad jurídica artificial sin convertirla en un mero rito de clasificación.

Un diseño de agentes de inteligencia artificial contruidos explícitamente para obedecer las leyes humanas, no solo las instrucciones de quien los despliega, ofrece una vía técnica para reducir el riesgo de responsabilidad antes de que el daño se produzca, en lugar de resolverlo después mediante ficciones de personalidad (O'Keefe, Ramakrishnan y Winter, 2025). Esa orientación: prevenir mediante diseño en vez de resolver mediante clasificación posterior, es, en los términos de este artículo, la alternativa sustantiva a la ordalía institucional. Un modelo de fiabilidad temporal en marcos de argumentación estructurada resulta relevante aquí de forma indirecta pero precisa. Cualquier estándar de responsabilidad que dependa de evaluar la razonabilidad de una inferencia automatizada necesita, primero, un modo de medir la fiabilidad de esa inferencia en el tiempo, no solo una categoría que la etiquete (Budán, Lucero, Chesnevar y Simari, 2012).

El resto de este dominio confirma la extensión del problema a ámbitos especializados. La ciberseguridad y el cibercrimen (Jaiswal y Mishra, 2024; Zverev, Bushkov y Prokopovych-Tkachenko, 2025), la responsabilidad militar por decisiones letales asistidas (Soltanzadeh, 2026; Woodcock, 2024), la propiedad intelectual y las patentes de invenciones generadas por IA (Yanisky-Ravid, 2026; Niebla Zatarain, 2018), y el derecho civil general (Kim, 2024; Hasibuan, Rawung y Wowiling, 2024; Fransisco, Hafrida y Kusniati, 2025; Oh, 2023; Meng y Cai, 2025; Tjwan, Ang y Lianto, 2026; Kudeikina y Kaija, 2024; Yan, 2023; Quattrococo, 2021; Blass, 2023; Matvieiev, 2025; Beryl Helga Fredella Hibatulloh, 2025; Monga, 2023). En ninguno de estos ámbitos la literatura documenta un caso de reconocimiento pleno de personalidad ni de imputación penal directa a un sistema artificial. Todos convergen, por caminos distintos, en la misma conclusión de fondo: la responsabilidad recae, por eliminación, en las categorías de sujeto que el derecho ya reconoce.

3. Discusión

Los cinco dominios examinados en este artículo: jurisprudencia colombiana, explicabilidad técnica, adjudicación algorítmica, inteligencia artificial generativa, y responsabilidad y personalidad jurídica, documentan, con vocabulario y evidencia distintos, el mismo patrón de riesgo. El derecho exige un mecanismo de control humano sobre un sistema que no puede supervisarse por completo. Ese mecanismo puede satisfacerse en su forma sin producir su sustancia. Y cuando eso ocurre, la responsabilidad legal se desplaza de la autoridad humana hacia una actuación técnica o institucional que ocupa el lugar de la razón sin serlo.

La respuesta a la pregunta de investigación es, en consecuencia, afirmativa. La "nueva ordalía" que este artículo identifica a partir del riesgo advertido por la jurisprudencia colombiana en materia de transparencia algorítmica no es un fenómeno exclusivo de la explicabilidad técnica. Es la manifestación más visible del patrón, y la que la jurisprudencia colombiana bloquea con mayor precisión en la práctica, aunque sin nombrarlo así; la literatura sobre razonamiento jurídico artificial documenta ese mismo patrón de forma dispersa en al menos otros tres dominios, sin la misma precisión conceptual ni el mismo nombre. Esa dispersión terminológica no es un defecto menor. Un riesgo que se nombra de formas distintas en cada dominio: sesgo de automatización en la adjudicación, alucinación aceptada acríticamente en la generación de texto, personalidad jurídica sin sustrato institucional en la responsabilidad, es más difícil de reconocer como un problema único, y por lo tanto más difícil de regular de forma coherente.

El criterio de adecuación que articula la jurisprudencia colombiana: inteligibilidad, impugnabilidad, fidelidad, ofrece, en este sentido, un aporte que excede su origen. Puede generalizarse como prueba para cualquier mecanismo de control sobre inteligencia artificial jurídica, no solo para la explicación de un modelo. Aplicado a la supervisión judicial, exige preguntar si el supervisor puede efectivamente entender, cuestionar y revisar la sugerencia algorítmica, no solo si existe formalmente la posibilidad de que lo haga. Aplicado a la inteligencia artificial generativa, exige preguntar si el usuario verifica la corrección del resultado antes de usarlo, no solo si tuvo la oportunidad de hacerlo. Aplicado a la personalidad jurídica, exige preguntar si existe un sustrato institucional capaz de responder, no solo si existe una categoría formal que lo declare.

Una limitación de este artículo merece mención explícita. El corpus combinado concentra, de forma abrumadora, jurisdicciones angloamericanas, europeas y asiáticas; solo un puñado de registros trata contextos latinoamericanos, y el propio caso central de este artículo, la jurisprudencia colombiana, llega a la literatura académica indexada de forma tardía respecto de su propia relevancia jurisprudencial. Esa asimetría sugiere que el aporte conceptual de tribunales del Sur Global a este debate puede estar sistemáticamente subrepresentado en las

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Garguilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

bases de datos que sostienen la literatura internacional, incluso cuando ese aporte, como en este caso, resulta más preciso que buena parte de la literatura que pretende sintetizar.

4. Conclusiones

La transparencia algorítmica no es, como podría parecer desde una lectura aislada de la jurisprudencia colombiana, un problema técnico contenido en la relación entre un modelo opaco y quien intenta explicarlo. Es un caso particular, mejor documentado que el resto, de una tensión que atraviesa toda la gobernanza de la inteligencia artificial jurídica. Cuando el derecho no puede verificar por completo lo que un sistema hizo, tiende a construir mecanismos de control que pueden satisfacer su propia forma sin producir la sustancia que esa forma promete.

Ese hallazgo tiene una implicación práctica directa. Evaluar si un mecanismo de control: explicabilidad, supervisión, revisión de un resultado generativo, atribución de responsabilidad, reconocimiento de personalidad jurídica, es sustantivo o ritual no exige inventar, para cada dominio, un criterio nuevo. El criterio de tres elementos que la jurisprudencia colombiana articuló para la explicabilidad técnica: inteligibilidad, impugnabilidad, fidelidad; funciona igual de bien en los otros cuatro dominios examinados en este artículo, precisamente porque el riesgo que busca prevenir depende de la relación entre una autoridad humana y un sistema que esa autoridad no puede supervisar por completo, no de la tecnología específica en juego.

Queda una pregunta que este artículo no resuelve y que merece investigación futura específica. Si el criterio colombiano puede generalizarse conceptualmente, como aquí se argumenta, ¿puede generalizarse también institucionalmente, en forma de un estándar único aplicable a explicabilidad, supervisión judicial, revisión de resultados generativos y atribución de responsabilidad, o exige cada dominio un desarrollo normativo propio que reconozca sus particularidades técnicas? La jurisprudencia colombiana ha dado, hasta ahora, la respuesta más precisa a la primera mitad de esa pregunta. La segunda mitad permanece abierta.

Referencias

- Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. *Advances in Neural Information Processing Systems*, 31.
- Adi, K., Darma, I., & Kurniawan, I. (2024). Artificial Intelligence and Criminal Liability Challenges for Contemporary Penal Law. *Pena Justisia: Media Komunikasi dan Kajian Hukum* 23(2), pp.5868-5879.
- Almeida, G., Nunes, J., & de Araújo, M. (2024). Exploring the psychology of LLMs' moral and legal reasoning. *ARTIFICIAL INTELLIGENCE*, 333.
- Alnemr, N. (2024). Democratic self-government and the algocratic shortcut: The democratic harms in algorithmic governance of society. *Contemporary Political Theory*, 23(2), 205-227.
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989.
- Arya, S., Aggarwal, S., Soni, N., Nishant, N., & Ansar, S. A. (2025). Explainable Artificial Intelligence (XAI) in critical decision-making processes. *Lecture Notes in Networks and Systems*, 1039, 445-454.
- Arzumanyan, A. (2026). The artificial intelligence use in criminal proceedings. *Полицейская и следственная деятельность (1)*, pp.263-272.
- Barannik, R., Balatska, O., & Holovko, O. (2026). ARTIFICIAL INTELLIGENCE AS A TOOL FOR APPLYING EVALUATIVE CONCEPTS IN CRIMINAL PROCEEDINGS: LEGAL AND ECONOMIC ASPECTS. *BALTIC JOURNAL OF ECONOMIC STUDIES*, 12(1), pp.125-133.
- Beryl Helga Fredella Hibatulloh (2025). Upaya Penegakan Hukum Terhadap AI (Artificial Intelligence) Sebagai Subjek Hukum Pidana dalam Perspektif Kriminologi. *TARUNALAW Journal of Law and Syariah* 3(01).
- Bharti, A. & Kaur, G. (2026). Enterprise Liability and Risk Governance in the Regulation of Artificial Intelligence. *Journal of Intellectual Property Rights*, 31(4).
- Binns, R. (2022). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *REGULATION & GOVERNANCE*, 16(1), pp.197-211.

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

- Blass, J. A. (2023). Extracting and Applying Legal Rules from Precedent Cases. *Northwestern University*.
- Budan, M. C., Lucero, M. G., Chesnevar, C. I., & Simari, G. R. (2012). Modelling time and reliability in structured argumentation frameworks. *Proceedings of the International Conference on Knowledge Representation and Reasoning*.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12.
- Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825-836.
- Cardoso, A., Chan, E., & Pereira, C. (2024). Generative Artificial Intelligence and Legal Decision-making. *GLOBAL TRADE AND CUSTOMS JOURNAL*, 19(11-12), pp.710-730.
- Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8), 832.
- Casimiro, J. S. C. & de Gouveia, L. G. (2023). JUDICIAL PROCESS AND REASONED DECISION: CURRENTLY, IN BRAZILIAN CONTEXT, IS IT POSSIBLE FOR LEGAL REASONING ARTIFICIAL INTELLIGENCE TO APPLY THE LAW JUST LIKE THE HUMAN JUDGE?. *Revista Eletronica De Direito Processual*, 24(3).
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, 538(7623), 20-23.
- Cho, J. K. & Tsche, K. J. (2024). A Study on the Legal Status and Legal Subjectivity of Artificial Intelligence (AI). *Kyung Hee Law Journal* | *경희법학* 59(3), pp.139-181.
- Cohen, I. G., Babic, B., Gerke, S., Evgeniou, T., & Wertenbroch, K. (2023). How AI can learn from the law: putting humans in the loop only on appeal. *Npj Digital Medicine*, 6(1), 160.
- Cordella, A., & Gualdi, F. (2025). Algorithmic formalization: Impacts on administrative processes. *Public Administration*, 103(2), 441-466.
- Corte Constitucional de Colombia. (2024). *Sentencia T-323 de 2024*.
- Corte Constitucional de Colombia. (2025). *Sentencia T-067 de 2025*.

- Córdoba, R. (2026). Artificial Intelligence and Legal Reasoning: An Approach to the Scope and Limits of Machine Reasoning in Law. *arXiv preprint*.
- Dajeh, B. (2025). The impact of artificial intelligence on the rules of civil liability for a machine guard. *Revista Opinião Jurídica (Fortaleza)* 23(43), pp.54-76.
- De Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, 39(2), 101666.
- Disemadi, H. & Sudirman, L. (2025). REASSESSING LEGAL RECOGNITION OF AI: HUMAN DIGNITY AND THE CHALLENGE OF AI AS A LEGAL SUBJECT IN INDONESIA. *MASALAH-MASALAH HUKUM* 54(1), pp.1-12.
- Dorr, B. J., Jayaweera, C., Youm, S., & Gilda, S. (2026). AI You Can Trust: Communication-Aware, Ambiguity-Sensitive, and Interpretable NLP for High-Stakes Domains. *Proceedings of the International Florida Artificial Intelligence Research Society Conference Flairs*, 39(1).
- Faqir, R. (2025). CHATGPT IN THE DOCK: REFLECTIONS ON THE FUTURE OF CRIMINAL LIABILITY. *ACCESS TO JUSTICE IN EASTERN EUROPE*, 8.
- Fransisco, W., Hafrida, H., & Kusniati, R. (2025). Legal responsibility for the utilization of artificial intelligence that has occurred system error from the perspective of criminal law. *Journal of Law Science* 7(1), pp.150-158.
- Garzo, G. & Palumbo, A. (2025). Legal & Ethical Implications of Predictive Digital Techniques in the Judicial Criminal Proceedings. *Isdfs 2025 13th International Symposium on Digital Forensics and Security*.
- Giabardo, C. V. (2023). ON “THE JUDGE AND THE ALGORITHM”, AGAIN. CRITICAL REFLECTIONS ON ARTIFICIAL INTELLIGENCE AND PREDICTIVE JUSTICE (STARTING FROM A CONTRIBUTION BY MICHELE TARUFFO). *Revista Italo Espanola De Derecho Procesal*, 2023(1).
- Golpayegani, D., Hupont, I., Panigutti, C., Pandit, H. J., Schade, S., O'Sullivan, D., & Lewis, D. (2024). AI cards: Towards an applied framework for machine-readable AI and risk documentation inspired by the EU AI Act. *Annual Privacy Forum*, 48-72.
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38(3), 50-57.

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Garguilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

- Gültekin-Várkonyi, G. (2026). Why Avoid Generative Legal AI Systems? Hallucination, Overreliance, and their Impact on Explainability. *Forum: Acta Juridica et Politica*, 15(2), 47-60.
- Han, S. (2025). From First Principles to Real-World Applications: Advancing LLM Reasoning From Logic to Law. *Yale University*.
- Hasibuan, R., Aurelya Jessica Rawung, & Fidel Jeremy Wowiling (2024). ARTIFICIAL INTELLIGENCE IN THE AUSPICES OF LAW: A DIVERGE PERSPECTIVE. *Mimbar Hukum - Fakultas Hukum Universitas Gadjah Mada* 36(1), pp.111-140.
- Hu, F., Liu, S., & Yu, M. (2024). Risks of Generative Artificial Intelligence and Multi-Tool Governance. *Academic Journal of Management and Social Sciences* 9(2), pp.88-93.
- Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169(3), 615-626.
- Jaiswal, A. & Mishra, P. (2024). ARTIFICIAL INTELLIGENCE (AI) AND CYBERSECURITY LAW: LEGAL ISSUES IN AI-DRIVEN CYBER DEFENSE AND OFFENSE. *ShodhKosh: Journal of Visual and Performing Arts* 5(6).
- Ji, W. (2020). The Change of Judicial Power in China in the Era of Artificial Intelligence. *ASIAN JOURNAL OF LAW AND SOCIETY*, 7(3), pp.515-530.
- Kahr, P. K., Rooks, G., Willemsen, M. C., & Snijders, C. C. P. (2024). Understanding trust and reliance development in AI advice. *ACM Transactions on Interactive Intelligent Systems*, 14(4).
- Kalasampath, K., Spoorthi, K. N., Sajeev, S., Kuppa, S. S., Ajay, K., & Maruthamuthu, A. (2025). A literature review on applications of explainable artificial intelligence (XAI). *IEEE Access*, 13, 41111-41140.
- Karsai, K. (2024). The Use of Algorithms to Support Judicial Decision-Making in Criminal Matters with a Special Focus on Trial Decisions. *Studia Iuridica Lublinensia* 33(5), pp.103-124.
- Kazun, A. P. (2024). Can artificial intelligence predict judicial decisions? A systematic review of international research. *Monitoring Obshchestvennogo Mneniya: Economic and Social Changes*, 183(5), 100.

- Kim, J. S. (2024). Issues on Generative AI (Artificial Intelligence) and Civil Law -- Focusing on the granting of legal personality or electronic personality --. *YONSEI LAW JOURNAL* | 연세법학 45, pp.1-28.
- Korotysh, E. (2026). Artificial Intelligence as a Subject and Instrument in Criminal Proceedings: Doctrinal Approaches. *The Legal World*, pp.48-51.
- Kudeikina, I. & Kaija, S. (2024). LIMITS OF THE USE OF ARTIFICIAL INTELLIGENCE IN LAW -- ETHICAL AND LEGAL ASPECTS. *Environment Technology Resources Proceedings of the International Scientific and Practical Conference 2*, pp.188-191.
- Lamba, P. & Kohli, R. (2026). AI-powered judiciary: National and international perspectives on machine learning in court administration. *Humanities and Social Sciences A Multidisciplinary Approach*.
- Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2), 1-16.
- Leng, S., Kang, X., Liang, Q., Li, X., & Fan, Y. (2025). Accusations and law articles prediction in the field of environmental protection. *Applied Sciences*, 15(1), 280.
- Linna, E. & Linna, T. (2026). Challenges for Generative AI in Legal Reasoning. *arXiv preprint*.
- Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36-43.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- Matvieiev, V. (2025). ARTIFICIAL INTELLIGENCE: SUBJECT OR OBJECT OF CIVIL LEGAL RELATIONS?. *Constitutional State (57)*, pp.173-179.
- Menasria, A., Madjri, Y., & Beddiar, M. (2025). The Implications of Artificial Intelligence on Criminal Liability in Contemporary Law. [VERIFICAR: fuente no disponible en el registro]
- Meng, Y. & Cai, J. (2025). A Novel MetaSoft Tree-Cognitive Set Model for Evaluating Criminal Litigation Efficiency under Artificial Intelligence Ecosystems. *Neutrosophic Sets and Systems*, 85.

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.
- Mohammadi, M., Wieling, M., & Vols, M. (2025). An interpretable approach to detect case law on housing and eviction issues within the HUDOC database. *Artificial Intelligence and Law*.
- Monga, G. (2023). LEGAL ACCOUNTABILITY OF AUTONOMOUS AI SYSTEMS IN CRIMINAL JUSTICE. *ShodhKosh: Journal of Visual and Performing Arts* 4(2), pp.5848-5852.
- Nay, J. (2022). Law Informs Code: A Legal Informatics Approach to Aligning Artificial Intelligence with Humans. *Arxiv* | *Total Versions*: 1.
- Nelson, J. (2026). It's Safer to Give Personhood to Bears than to Artificial Intelligence. *Arxiv* | *Total Versions*: 1.
- Niebla Zatarain, J. M. (2018). Copyright on the internet: achieving security through electronic devices an artificial intelligence approach. *The University of Edinburgh (United Kingdom)*.
- Nykolyna, K. (2025). Generative interpretation of legal texts: opportunities and limits. *Uzhhorod National University Herald Series Law* 1(87), pp.90-94.
- O'Brien Quinn, H., Sedky, M., Francis, J., & Streeton, M. (2024). Literature review of explainable tabular data analysis. *Electronics*, 13(19), 3806.
- O'Keefe, C., Ramakrishnan, K., & Winter, C. (2025). LAW-FOLLOWING AI: DESIGNING AI AGENTS TO OBEY HUMAN LAWS. *FORDHAM LAW REVIEW*, 94(1), pp.57-129.
- Oh, J. Y. (2023). Legal relations related to Artificial Intelligence. *Science, Technology and Law* | *과학기술과 법* 14(2), pp.143-166.
- Paar, E. (2025). Artificial Intelligence in Court Proceedings: Judge's Little Helper or the Beginning of AI's Hostile Takeover?. *German Law Journal*, 26(7), pp.1383-1403.
- Pandey, A. (2026). Justice, Algorithms, and Accountability: Ethical Considerations for AI in the Judiciary. *Advancing Human Dignity in Digital Workspaces Ethical AI People Analytics and Inclusive Innovation*.

- Panigrahi, S. K., Panda, S. A., & Mahapatra, N. (2026). Beyond a Reasonable Bot: Preserving Judicial Discretion in AI-Augmented Adjudication. *Artificial Intelligence and Legal Evidence the Indian Policy and Perspective*.
- Park, J. & Park, S. (2026). When correct is not enough: Deconstructing legal causal reasoning capability in models. *KNOWLEDGE-BASED SYSTEMS*, 351.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Perfetti, L. R. (2017). Beyond the chinese room. Memo for a further debate on artificial intelligence and public law. *P A Persona E Amministrazione*, 2017(1).
- Pesch, P. (2025). Potentials and Challenges of Large Language Models (LLMs) in the Context of Administrative Decision-Making. *European Journal of Risk Regulation*, 16(1), pp.76-95.
- Quattrocchio, S. (2021). For an artificial intelligence useful in criminal justice. *BIOLAW JOURNAL-RIVISTA DI BIODIRITTO*, (2), pp.387-398.
- Raskulla, S. (2023). Hybrid theory of corporate legal personhood and its application to artificial intelligence. *SN Social Sciences* 3(5).
- Ravizki, E. & Lintang Yudhantaka (2022). Artificial Intelligence Sebagai Subjek Hukum: Tinjauan Konseptual dan Tantangan Pengaturan di Indonesia. *Notaire* 5(3), pp.351-376.
- Razmetaeva, Y. (2024). Artificial Intelligence for Judicial Decision-making: Some Potential Risks. *Problems of Legality*, pp.177-191.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.
- Saraswathy, D. (2025). The Legal Status of AI Entities: Can Machines Hold Rights or Duties?. *International Journal of Judicial Science Research Studies (IJJSRS)*, pp.43-47.

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

- Sağlam, R. (2025). Equitable and Algorithmic Legal Reasoning: Deconstructive Approach to Human and Artificial Intelligence Judges. *Asian Institute of Research, Law and Human Rights Quarterly Review*, View full text.
- Scamardella, F. & Vestoso, M. (2023). Predictive models supporting judicial decision: some food for thought. *Rivista Di Filosofia Del Diritto*, 12(1), pp.135-156.
- Schilder, F. (2023). Legal Expertise Meets Artificial Intelligence: A Critical Analysis of Large Language Models as Intelligent Assistance Technology. *Ceur Workshop Proceedings*, 3423.
- Schwarcz, D., Das, D., Kang, D., & McDonnell, B. (2025). Thinking Like a Lawyer in the Age of Generative AI: Cognitive Limits on AI Adoption Among Lawyers. *Journal of Institutional and Theoretical Economics*, 182(1).
- Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, 87(3), 1085-1139.
- Shradda Tomar (2026). Legal Agency And AI: Can Machines Bear Rights And Responsibilities. *International Journal of Innovations in Science, Engineering And Management*, pp.11-20.
- Shukla, S. (2026). From Human Oversight to Cognitive Sovereignty: A Process-Based Governance Standard for AI-Assisted Legal Reasoning. *AI, Law, Politics* 2(1), pp.52-66.
- Simon, D. (2026). Can We Trust Humans with Legal Reasoning?. *Journal of Institutional and Theoretical Economics*, 182(1), pp.156-180.
- Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society*, 180-186.
- Soltanzadeh, S. (2026). Normative Reasoning and Military AI: DILEMA's Dilemmas. *Legal Ethical and Technical Dilemmas in Military Artificial Intelligence*.
- Stanizzi, S. (2026). Beyond optimization: the non-computable core of law and the limits of AI-mediated legal decision-making. *INTERNATIONAL JOURNAL OF LAW AND INFORMATION TECHNOLOGY*, 34.
- Swisher, K. (2025). The Rise of Legal Machines and the Transcendence of Human Ethics. *Arizona State University*.

- Szpyt, K. (2025). Civil Liability for Damages Caused by the Use of Artificial Intelligence Systems by Lawyers in the Provision of Legal Services. Part II. *Prawo Asekuracyjne* 2(123), pp.58-86.
- Thais, S., Kennedy, W., & Buchman, C. (2026). Toward Human Rights Benchmarking for LLMs: A Pilot Methodology. *Arxiv* | *Total Versions*: 1.
- Thuy, H., Nghiep, N., & Anh, N. (2026). CRIMINAL LIABILITY IN RELATION TO ARTIFICIAL INTELLIGENCE: A COMPARATIVE STUDY OF SELECT JURISDICTIONS AND VIETNAM. *Veredas do Direito* 23, pp.e235177-e235177.
- Tjwan, G., Ang, F., & Lianto, A. (2026). Penggunaan Artificial Intelligent Yang Melanggar Hukum. *Widya Yuridika* 9(1), pp.197-206.
- Trivedi, M. A. (2025). Guardrails Up: Designing Ethical and Regulation-Compliant Generative AI for Legal Practice. [VERIFICAR: fuente no disponible en el registro]
- Umit Atılğan, E. (2024). Will artificial intelligence sit on the judge's bench?. *Computational Intelligence and Blockchain in Complex Systems System Security and Interdisciplinary Applications*.
- 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.
- Wang, Y. (2026). Research on the Attribution and Liability Path of Artificial Intelligence Entities Infringing on Legal Interests. *Frontiers in Humanities and Social Sciences* 6(3), pp.278-288.
- Wanggai, A., Firdaus, A., & Santos, M. (2026). Hybrid Judicial Legitimacy in the Algorithmic Era: Toward a Socio-Legal Theory of Human--AI Adjudication Across Legal Traditions. *Justicia Insight* 2(2), pp.47-61.
- Woodcock, T. (2024). Human/Machine(-Learning) Interactions, Human Agency and the International Humanitarian Law Proportionality Standard. *GLOBAL SOCIETY*, 38(1), pp.100-121.
- Xudaybergenov, A. (2023a). Artificial Intelligence as a Legal Entity: A Balanced Examination of Pros and Cons. [VERIFICAR: fuente no disponible en el registro]
- Xudaybergenov, A. (2023b). Toward Legal Recognition of Artificial Intelligence Proposals for Limited Subject of Law Status. *International Journal of Law and Policy*, 1(4).

Diego Mauricio Hernández
Juan Manuel Padilla
Beliña Herrera Tapias
Renzo Gandolfi Díaz
Neuro José Zambam
Pascual Ángel Gargiuilo
Aníbal Ríos Castrillón
Ferney Rodríguez Serpa

- Yan, Q. (2023). Legal Challenges of Artificial Intelligence in the Field of Criminal Defense. *Lecture Notes in Education Psychology and Public Media* 30(1), pp.167-175.
- Yanisky-Ravid, S. (2026). Rethinking AI-generated Invention and Patent Protection. *SSRN Electronic Journal*.
- Yongho, K. (2024). A study on granting legal personality to Artificial Intelligence and responsibility for illegal acts. *Journal of the Korea Society of Computer and Information*, 29(12), 237-244.
- Zhang, R. & Zeng, J. (2026). Legal and ethical dilemmas and institutional responses to the application of generative artificial intelligence in education. *Cadernos De Dereito Actual*, (32), pp.316-335.
- Zverev, V., Bushkov, V., & Prokopovych-Tkachenko, Y. (2025). ARTIFICIAL INTELLIGENCE AND CYBERCRIME: NEW CHALLENGES AND PROSPECTS FOR LEGAL REGULATION. *Contemporary Issues in Artificial Intelligence* 1.