

Jueces algorítmicos y responsabilidad del operador judicial: límites jurídicos a la inteligencia artificial en la función pública y judicial. Estudio comparado entre Colombia y Brasil¹

Fredy Bladimir Vanegas Ladino²

Leandra Vanesa Zayas Peña³

Yaneth Ayala Gamboa⁴

Juan Camilo Barbosa Jaimes⁵

Resumen

La incorporación de sistemas de inteligencia artificial a la función pública y judicial ha producido en Colombia y en Brasil una convergencia normativa notable: ambos ordenamientos declaran que existe un núcleo de la decisión que no puede delegarse en un sistema algorítmico. Este artículo sostiene que esa declaración, por sí sola, constituye una garantía nominal. Mediante un análisis documental y jurisprudencial inscrito en el estudio de la convergencia administrativa transnacional, en la línea del derecho administrativo global de Kingsbury et al. (2005), se examinan las sentencias T-323 de 2024 y T-067 de 2025 de la Corte Constitucional colombiana, la Resolución 615 de 2025 del Conselho Nacional de Justiça y el artículo 20 de la Lei 13.709 de 2018 de Brasil, contrastados con tres decisiones europeas que ya juzgaron sistemas algorítmicos públicos. Los resultados muestran que la reserva humana opera en ambos países sobre el momento final de la decisión y no sobre el ciclo de vida del sistema, y que las dos rutas dogmáticas empleadas, esto es, el acceso a la información pública en Colombia y la protección de datos en Brasil, producen niveles de exigibilidad desiguales, hasta el punto de que el legislador brasileño suprimió en 2019 la exigencia de que la revisión de las decisiones automatizadas la practicara una persona natural. El análisis de dos casos de responsabilidad del operador judicial revela que la imputación se activa por la conducta del operador y no por el comportamiento del sistema. Se concluye que la reserva humana solo adquiere exigibilidad cuando se acompaña de trazabilidad documentada del ciclo de vida y de un régimen de imputación determinado.

Palabras claves: reserva humana de la decisión; transparencia algorítmica; responsabilidad del operador judicial; debido proceso; derecho administrativo global; gobernanza algorítmica.

¹ Artículo científico presentado como opción de grado para optar por el título de Magister en Derecho.

² Abogado litigante, socio director de Oro Vanegas Abogados S.A.S., Novoa Buendía S.A.S. e Instituto Colombiano de Asesoría Pensional y Asuntos Laborales S.A.S.

³ Abogada, Especialista en Contratación Estatal de la Universidad Externado de Colombia.

⁴ Abogada, Especialista en Derecho Procesal de la Universidad Santo Tomás y Especialista en Derecho Administrativo de la Universidad Libre.

⁵ Director: Abogado – Universidad Santo Tomas, Magister en Análisis Económico del Derecho y Políticas Pública - Universidad de Salamanca.

Algorithmic judges and the liability of the judicial officer: legal limits to artificial intelligence in public and judicial functions. A comparative study between Colombia and Brazil

Abstract

The incorporation of artificial intelligence systems into public administration and the judiciary has produced a remarkable normative convergence in Colombia and Brazil: both legal systems declare that a core of decision-making cannot be delegated to an algorithmic system. This article argues that such a declaration, on its own, amounts to a nominal guarantee. Through a documentary and case-law analysis framed within the study of transnational administrative convergence, along the lines of the global administrative law of Kingsbury et al. (2005), it examines judgments T-323 of 2024 and T-067 of 2025 of the Colombian Constitutional Court, Resolution 615 of 2025 of the Brazilian National Council of Justice, and article 20 of Brazilian Law 13.709 of 2018, set against three European decisions that have already adjudicated public algorithmic systems. The findings show that the human reserve operates in both countries at the final moment of the decision rather than across the system's life cycle, and that the two doctrinal routes employed, namely access to public information in Colombia and data protection in Brazil, yield unequal levels of enforceability, to the point that the Brazilian legislature removed in 2019 the requirement that review of automated decisions be carried out by a natural person. The analysis of two cases on the liability of judicial officers reveals that attribution is triggered by the officer's conduct and not by the system's behaviour. It concludes that the human reserve becomes enforceable only when accompanied by documented life-cycle traceability and a determinate regime of attribution.

Keywords: human reserve of decision; algorithmic transparency; judicial officer liability; due process; global administrative law; algorithmic governance.

Introducción

El 30 de enero de 2023, un juez del circuito de Cartagena resolvió en segunda instancia una acción de tutela promovida en favor de un niño diagnosticado con trastorno del espectro autista y, al motivar su decisión, transcribió en la providencia las respuestas que le había suministrado ChatGPT⁶. Año y medio después, la Corte Constitucional colombiana examinó ese fallo y concluyó que no había existido vulneración del debido proceso, pero aprovechó la ocasión para fijar doce criterios orientadores sobre el uso de herramientas de inteligencia artificial por los despachos judiciales del país (Corte Constitucional de Colombia, 2024). En mayo de 2026, en el otro extremo del continente, un juez del trabajo de Parauapebas, en el estado de Pará, detectó que la demanda que tenía ante sí contenía un texto escrito en letra blanca sobre fondo blanco, invisible para el lector humano, con una instrucción dirigida al sistema de inteligencia artificial del

⁶ La Corte Constitucional dispuso la reserva de la identidad de los intervinientes, por lo que la providencia se refiere a la accionante y al menor mediante nombres ficticios y omite la identificación del despacho judicial y de la entidad accionada.

despacho, ordenándole contestar la petición de forma superficial (Tribunal Regional do Trabalho da 8ª Região, 2026).

Los dos episodios comparten algo más que la novedad tecnológica. En ambos, un sistema de inteligencia artificial intervino en la producción de una decisión judicial; en ambos, la garantía activada fue la misma, a saber, que la decisión sea de un juez y no de una máquina; y en ambos, la pregunta decisiva no fue si un ser humano había firmado, porque en los dos casos lo había hecho, sino qué había ocurrido antes de esa firma. Esa distancia entre la firma y el proceso que la antecede es el objeto de este artículo.

La cuestión no es exclusivamente judicial. La administración pública latinoamericana atraviesa un desplazamiento que la doctrina ha descrito con precisión: el paso de la automatización formal, gobernada por reglas deterministas que ejecutan exactamente lo programado, a sistemas de inferencia cuyo resultado no se encuentra enteramente predeterminado por quien los diseñó. Barbosa Jaimes (2023) sitúa la diferencia en la arquitectura misma del sistema y advierte que, cuando el procesamiento ocurre en capas ocultas que escapan al control del programador, la opacidad resultante impide conocer la motivación de las resoluciones judiciales o administrativas y aumenta la asimetría de información entre la administración que decide y el administrado que soporta la decisión (pp. 132-134). Ese es el problema que el derecho enfrenta hoy, y no la sustitución del juez por un autómata, que ningún ordenamiento de la región autoriza.

Colombia y Brasil han respondido a ese desplazamiento con instrumentos distintos y en tiempos distintos, pero con una coincidencia de fondo que constituye el punto de partida de esta investigación: ambos declaran que hay un núcleo de la función pública que no puede transferirse a un proceso de inferencia estadística. La Corte Constitucional colombiana lo formuló como el criterio de no sustitución de la racionalidad humana y como la prohibición de que las herramientas de inteligencia artificial reemplacen labores jurisdiccionales indelegables e irremplazables (Corte Constitucional de Colombia, 2024). El Conselho Nacional de Justiça brasileño lo formuló disponiendo que en ningún momento el sistema podrá restringir o sustituir la autoridad final de sus usuarios internos y que los magistrados permanecen integralmente responsables por las decisiones adoptadas (Conselho Nacional de Justiça, 2025).

La hipótesis que aquí se sostiene es que esa declaración compartida, tal como está construida en ambos ordenamientos, resulta insuficiente. Opera sobre el acto final, donde siempre habrá una firma humana, y no sobre la cadena que produjo el resultado. Una reserva humana que solo exige que alguien suscriba la decisión es una garantía nominal, porque no impide que el contenido de esa decisión haya sido determinado, en lo esencial, por un proceso que nadie puede reconstruir. Solo adquiere exigibilidad cuando se acompaña de dos condiciones: trazabilidad documentada del ciclo de vida del sistema y un régimen de imputación que identifique a un responsable determinado. Sin trazabilidad, no hay sobre qué hacer recaer la responsabilidad; sin imputación, la trazabilidad es documentación sin consecuencias.

La hipótesis es falsable. Si el examen de las fuentes mostrara que alguno de los dos ordenamientos ya exige trazabilidad del ciclo de vida con un responsable identificado, la afirmación cae. De ahí la pregunta que guía la investigación: ¿bajo qué condiciones la reserva

humana de la decisión, declarada en Colombia y en Brasil, se traduce en una garantía efectivamente exigible, y a quién es imputable el resultado de una decisión judicial asistida por inteligencia artificial?

El objetivo general consiste en determinar, mediante el análisis comparado de ambos ordenamientos, las condiciones bajo las cuales la reserva humana de la decisión deja de ser una declaración normativa y se convierte en una garantía exigible frente al uso de sistemas de inteligencia artificial en la función pública y judicial, así como el régimen de responsabilidad del operador judicial que de ello se deriva. Cinco objetivos específicos ordenan el desarrollo: delimitar el concepto jurídico de sistema de inteligencia artificial y establecer en qué medida su arquitectura condiciona la posibilidad de motivar el acto; reconstruir el contenido de la reserva humana de la decisión en ambos países y determinar su alcance a lo largo del ciclo de vida del sistema; contrastar los fundamentos dogmáticos de la transparencia algorítmica en uno y otro ordenamiento; extraer de las decisiones judiciales que ya han juzgado sistemas algorítmicos públicos los criterios de admisibilidad utilizables como estándar de contraste; y analizar la regulación del uso de inteligencia artificial en la función judicial de ambos países mediante dos estudios de caso, uno por ordenamiento.

El artículo se organiza en consecuencia. Tras exponer la metodología, los resultados se presentan en cinco apartados que siguen el orden de los objetivos, para cerrar con la discusión y las conclusiones.

Metodología de análisis y recolección de datos

Este trabajo no emplea el método del derecho comparado en su acepción clásica. Aquella tradición, tal como la formulan Pegoraro y el comparatismo sistémico, exige contextualizar cada ordenamiento como sistema jurídico completo, atendiendo a su familia jurídica, su cultura jurídica y su organización judicial, y validar en ese registro el criterio de comparación. El presente artículo, de naturaleza reflexiva, no realiza esa operación ni la requiere para su propósito. Reivindicar el rótulo del comparatismo clásico sin satisfacer sus exigencias metodológicas desvirtuaría el alcance del trabajo, de modo que resulta preferible declarar con precisión la corriente en la que efectivamente se inscribe.

Esa corriente es el estudio de la convergencia de principios administrativos a través de fronteras y de niveles de gobernanza, tal como lo formulan Kingsbury, Krisch y Stewart (2005) en su marco del derecho administrativo global. Los autores definen ese campo como el conjunto de mecanismos, principios, prácticas y entendimientos sociales que promueven o afectan de otro modo la rendición de cuentas de los órganos administrativos globales, y describen expresamente su objeto como los mecanismos de aprendizaje, préstamo y referencia cruzada que contribuyen a integrar el campo (pp. 15, 20). Ese encuadre es el que corresponde a lo que aquí se hace: examinar cómo confluyen tres principios de gobernanza administrativa, el control humano, la explicabilidad y la transparencia, a través de niveles normativos y jurisdicciones distintos, incluido un instrumento supranacional empleado como referente definitorio. Los propios autores advierten, además, que las analogías directas entre el derecho administrativo nacional y el transnacional

deben tomarse con cautela (p. 28), advertencia que este trabajo asume al no equiparar instrumentos de rango desigual.

De este encuadre se siguen tres consecuencias metodológicas. La primera es que el objeto de estudio no son normas equivalentes, sino los abordajes que órganos de distinta naturaleza han dado a un mismo problema: determinar qué le está permitido decidir a un sistema de inteligencia artificial cuando interviene en la producción de una decisión pública, y quién responde por el resultado. La segunda es que la heterogeneidad de los instrumentos examinados no constituye un defecto del diseño sino un dato del análisis, pues revela por qué vías cada ordenamiento ha llegado a las respuestas que hoy tiene. La tercera es que el trabajo se apoya en análisis documental y jurisprudencial de corte cualitativo, sin pretensión de generalización estadística.

La comparación opera en dos niveles articulados. El primero es el de los pronunciamientos: se contrastan las decisiones mediante las cuales órganos de distinta naturaleza han resuelto casos en que un sistema de inteligencia artificial intervino en una actuación pública, atendiendo al problema jurídico que cada uno se planteó, al criterio con que lo resolvió y a la consecuencia que extrajo. El segundo es el de los marcos normativos en que esos pronunciamientos se inscriben, examinados no por su jerarquía formal sino por la función que cumplen respecto de la garantía estudiada.

El criterio respecto del cual se ponen en relación ambos ordenamientos, o *tertium comparationis*, es la función de asegurar el control humano y la explicabilidad de la decisión asistida por inteligencia artificial. La unidad de análisis no es, en consecuencia, la norma aislada, sino la respuesta institucional completa que cada ordenamiento ofrece a esa función, entendida como el conjunto de tres elementos: el enunciado que declara la garantía, el mecanismo previsto para hacerla efectiva y la providencia que la ha aplicado a un caso concreto. Este diseño obliga a leer conjuntamente instrumentos de rango distinto, operación que se realiza de forma expresa y advirtiendo en cada caso el nivel jerárquico de la fuente, con el fin de que la comparación funcional no encubra una equiparación normativa que no existe.

El corpus se delimitó temporalmente entre 2018 y 2026. El extremo inicial corresponde a la expedición de la ley brasileña de protección de datos personales, primer instrumento de los examinados que reguló expresamente el derecho del titular frente a las decisiones adoptadas con base en tratamiento automatizado; el extremo final, a la fecha de cierre de esta investigación. Dentro de ese período se aplicaron tres criterios de selección.

El primero es de origen y jerarquía. Solo se admitieron documentos primarios obtenidos de los repositorios oficiales de los órganos que los expidieron, con exclusión de reseñas doctrinales, notas de prensa y agregadores comerciales de jurisprudencia, que se emplearon únicamente como instrumento de localización y nunca como respaldo de una afirmación. Se consultaron, en particular, la relatoría de la Corte Constitucional de Colombia, el portal de legislación de la Presidencia de la República de Brasil, el sistema de actos normativos del Conselho Nacional de Justiça, el portal de proposiciones legislativas de la Cámara de Diputados de Brasil y el repositorio oficial de jurisprudencia de los Países Bajos.

El segundo es de pertinencia funcional: cada pieza debía poder vincularse a uno de los cinco objetivos específicos, de modo que su inclusión respondiera al diseño de la investigación y no a la disponibilidad del material. El tercero es de integridad del documento: se trabajó sobre el texto completo de cada providencia y sobre el texto compilado y vigente de cada norma, y no sobre extractos, resúmenes ni líneas jurisprudenciales. Esta última exigencia resultó determinante, pues la reconstrucción de los estudios de caso requería el conocimiento de los hechos y del razonamiento, y no solo de la parte resolutoria; y el análisis del régimen brasileño de decisiones automatizadas solo pudo completarse al cotejar la redacción vigente de la norma con la de su versión originaria.

El corpus resultante se compone de las siguientes piezas. Por Colombia, las sentencias T-323 de 2024 y T-067 de 2025 de la Corte Constitucional. Por Brasil, la Resolución 615 de 2025 del Conselho Nacional de Justiça, la Lei 13.709 de 2018 en su texto compilado con las modificaciones introducidas por la Lei 13.853 de 2019, la Enmienda Constitucional 115 de 2022, la ficha de tramitación del Proyecto de Ley 2338 de 2023 y la sentencia dictada el 12 de mayo de 2026 por la 3ª Vara do Trabalho de Parauapebas en el proceso 0001062-55.2025.5.08.0130. Como estándar externo de contraste se incorporó la sentencia del Rechtbank Den Haag de 5 de febrero de 2020 sobre el sistema SyRI, consultada en su traducción oficial al inglés. El aparato doctrinal se integra con seis estudios. Barbosa Jaimes (2023), del que se toman además, por vía de reporte y así declarado en cada caso, los datos relativos a la decisión del Consejo Constitucional francés de 2019 y al sistema valenciano SALER; Kingsbury et al. (2005), que suministra el encuadre metodológico; Ponce Solé (2019) y Citron (2008), antecedentes doctrinales directos de las dos categorías centrales que aquí se emplean; Macena de Lima y Freire de Sá (2020), interlocutor para el régimen brasileño de decisiones automatizadas; Berning Prieto (2023), para la perspectiva europea sobre el uso administrativo de estos sistemas; y Pérez-Pacheco y Simón Castellano (2025), para el contexto judicial latinoamericano.

Los dos estudios de caso se seleccionaron mediante muestreo intencional, conforme a dos condiciones acumulativas: que un sistema de inteligencia artificial hubiera intervenido efectivamente en la producción de una decisión judicial, y que el órgano judicial se hubiera pronunciado de manera expresa sobre la responsabilidad derivada de esa intervención. La segunda condición es la que reduce el universo de manera decisiva, pues numerosas providencias mencionan la inteligencia artificial en obiter dicta sin extraer de ella consecuencia alguna en materia de imputación. Se seleccionó una providencia por ordenamiento, de modo que el contraste conservara simetría.

El procedimiento se desarrolló en cuatro fases sucesivas. En la primera se identificó el concepto de sistema de inteligencia artificial empleado por cada fuente, con el objeto de verificar, antes de comparar, que ambos ordenamientos regulan un mismo objeto. En la segunda se reconstruyó, mediante lectura sistemática de las providencias y de la regulación, el contenido de la reserva humana de la decisión en cada país, y se determinó su alcance temporal, esto es, si opera sobre el acto final o sobre el ciclo de vida del sistema. En la tercera se contrastaron los fundamentos dogmáticos de la transparencia algorítmica, atendiendo a cinco variables: fundamento, naturaleza de la garantía, rango, objeto del acceso y mecanismo de control. En la cuarta se analizaron los estudios de caso mediante una matriz de cuatro variables: quién usó el sistema, en qué fase del

proceso decisorio intervino, qué deber se consideró infringido y sobre quién recayó la imputación. Los resultados de las fases tercera y cuarta se sintetizan en las tablas 2 y 3.

Dos limitaciones deben declararse. La primera es la asimetría de las fuentes disponibles en cada ordenamiento. Colombia cuenta con desarrollo jurisprudencial de su tribunal constitucional desde 2024, mientras que en Brasil el material es predominantemente regulatorio y de primera instancia, y su proyecto de ley general sobre la materia continúa en trámite. Esa asimetría no se ha forzado en el sentido de la simetría, ni se han equiparado fuentes de peso desigual: se reporta como dato del análisis, en tanto describe el estado real de la cuestión en cada país. La segunda limitación es que el estudio no comprende trabajo de campo ni datos sobre la aplicación efectiva de la normativa examinada, de manera que sus conclusiones se refieren al diseño normativo y jurisprudencial y no permiten afirmar nada sobre su rendimiento en la práctica de los despachos.

Resultados

1. Delimitación de los conceptos empleados

1.1. Función pública y función judicial

El título, el resumen y el objetivo general de este trabajo descansan sobre una categoría que conviene definir antes de avanzar, por la misma razón por la que se define el sistema de inteligencia artificial: sin delimitarla, el objeto de estudio queda desdibujado.

Por función pública se entiende aquí el conjunto de actividades que el Estado desarrolla, por medio de sus servidores, para la realización de los fines que el ordenamiento le atribuye, con sujeción al principio de legalidad y bajo un régimen de responsabilidad. Es el género. La función judicial es una de sus especies, y esa relación de género a especie no es una construcción doctrinal sino una definición del propio texto constitucional colombiano, cuyo artículo 228 dispone que «la Administración de Justicia es función pública». La delimitación importa para este artículo porque permite sostener que las exigencias de control humano, explicabilidad y transparencia predicables de la actuación administrativa se proyectan también sobre la actuación judicial, sin necesidad de construir para esta última un régimen separado.

En el ordenamiento brasileño la delimitación se apoya en el artículo 2.º de la Constitución de 1988, que consagra la independencia y armonía de los poderes Legislativo, Ejecutivo y Judicial, de modo que la función judicial se configura como una función estatal diferenciada dentro del conjunto de las funciones públicas. No se fuerza aquí una equivalencia jurisprudencial entre ambos ordenamientos: basta constatar que en los dos la actividad de administrar justicia es actividad estatal reglada, sometida a legalidad y a responsabilidad, que es el presupuesto sobre el que opera el análisis.

Cuando en lo sucesivo se hable de función pública sin más, deberá entenderse el género; cuando se hable de función judicial, la especie; y cuando el artículo examine conjuntamente ambas, se advertirá expresamente.

2. El sistema de inteligencia artificial como objeto jurídico

2.1. El concepto y sus cuatro criterios de aproximación

Antes de examinar los límites de un objeto es necesario fijarlo, y en esta materia la fijación no es un ejercicio de erudición prescindible sino una condición de seguridad jurídica: si la autoridad no logra precisar qué está regulando, la ambigüedad terminológica diluye la identificación de responsables cada vez que un acto se produce, se motiva o se ejecuta con apoyo de estas herramientas.

La Corte Constitucional colombiana abordó el problema en la sentencia T-323 de 2024 y adoptó, siguiendo a Russell y Norvig, cuatro criterios de aproximación al concepto: el razonamiento, la racionalidad, la conducta y la fidelidad en la forma de actuar con los humanos, de los cuales se derivan cuatro maneras de definir un sistema de inteligencia artificial: sistemas que piensan como humanos, sistemas que piensan racionalmente, sistemas que actúan como humanos y sistemas que actúan racionalmente (Corte Constitucional de Colombia, 2024, párr. 83). Conviene precisar el alcance de esta taxonomía, porque suele citarse con un sentido que no le corresponde: se trata de criterios de definición conceptual, tomados de la literatura técnica, y no de un estándar de escrutinio judicial sobre el comportamiento de un sistema. El estándar de escrutinio la Corte lo construyó por separado, y es el que se examina en el apartado siguiente.

De los cuatro criterios interesan aquí dos. La racionalidad se entiende como la habilidad de elegir la mejor acción para alcanzar cierto objetivo, dados unos criterios a optimizar y unos recursos disponibles; el sistema la alcanza al percibir su entorno mediante sensores, razonar sobre lo percibido y decidir la acción a emprender, con la particularidad de que la ejecución de esa acción puede modificar el ambiente, de manera que en la iteración siguiente el sistema percibirá una información distinta (Corte Constitucional de Colombia, 2024, párr. 85). La fidelidad, por su parte, alude a la capacidad del sistema de mantener y transmitir un comportamiento consistente, confiable y predecible, para lo cual se requiere que la herramienta sea precisa, consistente y transparente con la información que emite (párr. 87).

Ambos criterios tienen consecuencias jurídicas directas. La racionalidad así entendida describe un sistema cuyo comportamiento se modifica con el ambiente que él mismo altera, lo que significa que el resultado de una operación futura no puede inferirse enteramente de su configuración inicial. La fidelidad, al condicionar la confianza a la transparencia de la información emitida, convierte la explicabilidad en un atributo del sistema y no en una simple aspiración regulatoria.

2.2. La arquitectura del sistema como variable jurídica

El análisis no puede tratar de manera uniforme a todos los sistemas de inteligencia artificial, porque su arquitectura determina el tipo de escrutinio jurídico posible. El desarrollo más preciso de esta distinción en la doctrina colombiana se encuentra en Barbosa Jaimes (2023), quien la construye a partir de la estructura de red neuronal. En los sistemas contruidos sobre perceptrón

simple, esto es, de una sola capa, aprendizaje supervisado y correspondencia controlada entre entrada y salida, el diseñador conserva el control total de las respuestas, y en consecuencia el derecho constitucional no se interesa por el funcionamiento del algoritmo, es decir por el cómo, sino que centra el debate en el para qué de la tecnología (pp. 130-133). En los sistemas de perceptrón multicapa, en cambio, la introducción de capas ocultas que cruzan variables entre sí sustrae el procesamiento al control del programador, quien podrá ajustar los resultados pero no interferir en la operación que transforma la entrada en la respuesta dada (p. 132).

La consecuencia que el autor extrae es de la mayor relevancia para esta investigación:

la existencia de cajas negras en la toma de decisiones se presenta como una situación contraria a derecho en cuanto viola los principios generales de igualdad, imparcialidad, transparencia, debido proceso, privacidad y entre otros que se pueden evidenciar en cada caso en concreto, en la medida que los sistemas de IA expertos en todo, de aprendizaje automático y en general de perceptrón multicapa, no permiten conocer la motivación de las resoluciones judiciales y/o administrativas, aumentando la asimetría de la información existente entre la administración responsable de la toma de la decisión y administrados que actúan como sujeto procesal en la toma de la decisión encomendado a la IA. (Barbosa Jaimes, 2023, p. 134)

El razonamiento merece atención porque suministra el eslabón que suele darse por supuesto en la literatura sobre gobernanza algorítmica. La opacidad no es objetable por sí misma, en abstracto: es objetable porque impide motivar. Y el deber de motivación no es una formalidad procesal, sino la condición que hace posible la contradicción, el control de la arbitrariedad y, en definitiva, el recurso. De ahí que la arquitectura del sistema no sea un dato técnico ajeno al derecho, sino la variable que determina si el acto que en él se apoya puede o no cumplir con su deber constitucional de motivación.

La observación no es exclusiva de la doctrina colombiana, y conviene contrastarla con dos fuentes independientes. Citron (2008) advirtió, ya hace más de tres lustros, que los sistemas automatizados del sector público codifican reglas en programas que ni el público, ni los funcionarios electos, ni los tribunales pueden revisar, y que los procedimientos heredados del siglo anterior resultan incapaces de reparar ese déficit de rendición de cuentas, razón por la cual propuso un concepto nuevo, el del debido proceso tecnológico (pp. 1249-1251). Berning Prieto (2023), por su parte, examina el estado de la cuestión en el uso administrativo de estos sistemas y formula propuestas para un uso responsable desde la perspectiva española y europea. Ambas confirman, desde tradiciones jurídicas distintas, que el problema no es la automatización en sí sino la imposibilidad de reconstruir el fundamento de la decisión.

Barbosa Jaimes advierte, además, que estos sistemas no aprenden únicamente el razonamiento objetivo, sino también circunstancias indeseables a la hora de actuar en derecho, como la heurística del juicio y sesgos tales como la discriminación y la xenofobia (2023, p. 134). La observación conecta con el criterio de igualdad y equidad que más adelante formularía la Corte Constitucional colombiana.

Conviene precisar el alcance de esta clasificación, cuya nomenclatura es técnicamente correcta. Los modelos generativos que hoy emplean las administraciones y los despachos judiciales son redes neuronales profundas que incorporan, en cada uno de sus bloques, un perceptrón multicapa. Lo que la dicotomía entre perceptrón simple y multicapa no logra por sí sola, a la luz de la literatura actual, es discriminar entre los sistemas efectivamente desplegados: el primero opera hoy como referente pedagógico antes que, como categoría de herramientas en uso, y el segundo abarca desde redes elementales hasta modelos de cientos de miles de millones de parámetros, entre ellos los examinados en los dos casos del apartado 6. Nada de ello resta vigencia al planteamiento; antes bien, lo confirma. Si la arquitectura del diseño importa al derecho es precisamente porque determina la posibilidad de motivar las decisiones, y los modelos generativos son la expresión más acabada de la caja negra que el autor describe, en tanto ni siquiera el diseñador del modelo controla los datos que fundan una respuesta concreta, lo que amplía la asimetría informativa y el riesgo de reproducir los sesgos que él mismo advierte.

2.3. Definición operativa

Con base en lo anterior, y para los fines específicos de este artículo, se adopta la siguiente definición de trabajo: un sistema de inteligencia artificial es todo sistema basado en máquinas que, a partir del procesamiento de datos y mediante modelos de inferencia o aprendizaje, genera salidas, tales como recomendaciones, contenidos, predicciones o decisiones, capaces de incidir en la esfera jurídica de una persona, y cuyo resultado concreto no se encuentra enteramente predeterminado por quien lo diseñó o lo puso en operación.

Dos elementos de la definición requieren justificación. El primero es la ausencia de predeterminación completa del resultado, que es lo que distingue estos sistemas de la automatización administrativa tradicional y lo que obliga a desplazar el control desde el error de código hacia la trazabilidad del comportamiento a lo largo del ciclo de vida. El segundo es la incidencia sobre la esfera jurídica de un tercero, que es en último término lo que activa las garantías del debido proceso y del deber de motivación, y lo que separa el uso de un sistema como herramienta interna de gestión de su empleo como insumo de una decisión que afecta derechos.

3. La reserva humana de la decisión

3.1 Los doce criterios de la sentencia T-323 de 2024

La sentencia T-323 de 2024, proferida el 2 de agosto de 2024 por la Sala Segunda de Revisión con ponencia del magistrado Juan Carlos Cortés González en el expediente T-9.301.656, constituye el desarrollo jurisprudencial más significativo de la región sobre esta materia. Sus hechos se examinan en el apartado 5; interesa aquí su construcción dogmática.

La Corte formuló el estándar en los siguientes términos:

El juez, cuando encuentre necesario y pertinente recurrir a sistemas de IA, puede emplearlos razonada y ponderadamente. [...] De acuerdo con ello, so pena de comprometer eventualmente su responsabilidad a raíz del uso indiscriminado e imprudente de estas tecnologías, el funcionario judicial que las emplee deberá respetar, como presupuesto

mínimo en la materia, el criterio de no sustitución de la racionalidad humana, así como atender las cargas de transparencia, responsabilidad y privacidad. En virtud de aquel criterio, es factible emplear inteligencia artificial en labores propias de la justicia siempre que el uso de dichas herramientas no remplace labores jurisdiccionales indelegables e irremplazables, como lo son aquellas que requieren del razonamiento lógico y humano a efectos de interpretar los hechos, las pruebas, motivar la decisión o adoptarla. (Corte Constitucional de Colombia, 2024, párr. 295)

El pasaje contiene tres decisiones dogmáticas que conviene aislar. La primera es la delimitación del núcleo indelegable por referencia a operaciones concretas y no a categorías abstractas: interpretar los hechos, valorar las pruebas, motivar la decisión y adoptarla. La segunda es que el uso permitido se define por exclusión respecto de ese núcleo, de modo que la licitud del empleo de la herramienta depende de la fase del proceso decisorio en que intervenga. La tercera, y la más relevante para el quinto objetivo específico de esta investigación, es la advertencia expresa sobre la responsabilidad del funcionario: la Corte no se limitó a declarar un principio, sino que anunció que su inobservancia puede comprometer la responsabilidad de quien emplea la herramienta de manera indiscriminada e imprudente.

Conviene señalar que la categoría aquí denominada reserva humana de la decisión cuenta con un antecedente doctrinal directo. Ponce Solé (2019) analizó el uso de la inteligencia artificial por las administraciones públicas para la adopción de decisiones automatizadas y semiautomatizadas y propuso, como mejora del marco regulador, el establecimiento de una reserva para humanos en el ejercicio de potestades discrecionales, fundada en la necesidad de empatía para realizar la ponderación con criterios extrajurídicos en que la discrecionalidad consiste. El mismo autor planteó una lucha contra la opacidad de los algoritmos en garantía del derecho a una buena administración, especialmente en su vertiente del derecho a comprender las decisiones administrativas. La categoría que este artículo emplea recoge esa formulación y la proyecta sobre la función judicial.

Algo semejante ocurre con la expresión debido proceso tecnológico, que la doctrina colombiana ha empleado para describir el estándar de la Corte. Su formulación originaria corresponde a Citron (2008), quien acuñó el concepto de *technological due process* para designar el conjunto de garantías necesarias frente a sistemas automatizados de decisión pública cuyo funcionamiento escapa al control del público, de los representantes electos y de los tribunales.

Sobre esa base, la Corte exhortó a los jueces de la República y dispuso que los funcionarios y empleados de la Rama Judicial apliquen doce principios en el uso de estas tecnologías. Su enunciación completa es necesaria porque constituye, a la fecha, el catálogo más detallado producido por un tribunal constitucional latinoamericano sobre la materia.

Tabla 1 *Los doce criterios orientadores de la sentencia T-323 de 2024*

Criterio	Contenido
Transparencia	Evidenciar con claridad y precisión el uso, alcances y ubicación de los resultados obtenidos, de modo que los interesados tengan pleno conocimiento y posibilidad efectiva de contradicción.
Responsabilidad	Que el usuario de la herramienta esté capacitado y comprenda los impactos de estas tecnologías, para dar cuenta del origen, idoneidad y necesidad de su uso y de la información suministrada, la cual debe ser verificada.
Privacidad	Custodiar y proteger la reserva de los datos personales y sensibles puestos en conocimiento de la administración de justicia.
No sustitución de la racionalidad humana	Imposibilidad ética y jurídica de sustituir la acción y la responsabilidad del individuo de la especie humana en la gestión de las actuaciones y decisiones judiciales.
Seriedad y verificación	Escrutinio estricto sobre fuentes, alcances, restricciones, posibilidades, falencias y riesgos de la herramienta frente a la actuación en curso.
Prevención de riesgos	Aplicar estándares adecuados de control sobre imprecisiones, desactualizaciones, alucinaciones, sesgos e inconsistencias.
Igualdad y equidad	Erradicar toda forma de discriminación derivada de sesgos y su impacto negativo en la eficacia de los derechos humanos.
Control humano	Permitir siempre escrutinios efectivos sobre las actuaciones y decisiones, mediante el acceso a la debida información y el uso de recursos que deban ser resueltos por autoridades humanas.
Regulación ética	Desarrollo de estándares de comportamiento individual acordes con los mandatos superiores y legales.
Adecuación a buenas prácticas estándares colectivos	Aplicación de los esquemas razonables definidos para el funcionamiento de la Rama Judicial desde su autonomía e independencia.
Seguimiento continuo y adaptación	Que el uso consulte los avances jurídicos, sociológicos y tecnológicos y los esquemas de mejora progresiva.
Idoneidad	Que el uso de las tecnologías sea adecuado para facilitar y agilizar el acceso a la justicia.

Nota: a partir de la sentencia T-323 de 2024 Corte Constitucional de Colombia.

El catálogo admite una lectura estructural. Cuatro de los criterios, a saber, transparencia, responsabilidad, seriedad y verificación, y control humano, se refieren a la relación entre el operador y la herramienta, y son los que sostienen la imputación. Tres de ellos, prevención de riesgos, igualdad y equidad, y privacidad, atienden a los efectos del sistema sobre terceros. Los cinco restantes, esto es, no sustitución de la racionalidad humana, regulación ética, adecuación a buenas prácticas, seguimiento continuo e idoneidad, definen el marco institucional del uso. Esta distribución revela algo que el enunciado aislado no muestra: el peso del estándar recae sobre la conducta del operador, y no sobre el comportamiento del sistema.

3.2 El desarrollo reglamentario en ambos países

La Corte no se detuvo en el exhorto. En el numeral quinto de la parte resolutive ordenó al Consejo Superior de la Judicatura que, en el término de cuatro meses, divulgara una guía, manual o lineamiento sobre la implementación de la inteligencia artificial generativa en la Rama Judicial, acorde con lo dispuesto en la providencia; y en el numeral sexto le ordenó divulgar la sentencia a todos los despachos judiciales del país a través de la Escuela Judicial Rodrigo Lara Bonilla y continuar generando espacios de capacitación (Corte Constitucional de Colombia, 2024). El diseño es relevante: la Corte fijó el estándar y transfirió al órgano de gobierno judicial la competencia para operacionalizarlo, con plazo y con contenido mínimo predeterminado.

Brasil llegó al mismo lugar por una vía institucional distinta. La Resolución 615 de 11 de marzo de 2025 del Conselho Nacional de Justiça establece directrices para el desarrollo, la utilización y la gobernanza de soluciones de inteligencia artificial en el Poder Judicial. Su artículo 32 dispone que en ningún momento el sistema podrá restringir o sustituir la autoridad final de los usuarios internos; los parágrafos 2.º y 4.º de su artículo 19 establecen que los magistrados permanecen integralmente responsables por las decisiones adoptadas y por la información contenida en ellas; el artículo 34 exige que los sistemas requieran supervisión humana y permitan la modificación por el magistrado competente; y el artículo 13, numeral VII, impone una explicabilidad adecuada de los resultados, siempre que sea técnicamente posible (Conselho Nacional de Justiça, 2025).

La coincidencia sustantiva es notable, sobre todo si se considera que se produjo por vías estructuralmente distintas: en Colombia, por decisión de un tribunal constitucional en sede de tutela, con fundamento en el debido proceso; en Brasil, por acto administrativo del órgano nacional de gobierno del Poder Judicial. Dos ordenamientos pueden llegar a exigencias sustantivamente convergentes por rutas dogmáticas heterogéneas, y ese dato es, en sí mismo, un resultado de la comparación.

Existe, sin embargo, un matiz que conviene registrar. El parágrafo 6.º del artículo 19 de la Resolución 615 permite mencionar el uso de inteligencia artificial generativa en el cuerpo de la decisión a criterio del magistrado (Conselho Nacional de Justiça, 2025). La responsabilidad es, por tanto, obligatoria, mientras que la revelación de que se empleó la herramienta queda librada a la discreción de quien la empleó. Es una asimetría significativa si se la compara con el criterio colombiano de transparencia, que impone evidenciar con claridad y precisión el uso y la ubicación de los resultados obtenidos, precisamente para permitir la contradicción.

3.3 Del acto al ciclo de vida: trazabilidad rendicional y auditabilidad independiente

La regulación brasileña avanza en un punto en que la colombiana es menos explícita. El artículo 2.º, numeral V, de la Resolución 615 establece la participación y supervisión humana en todas las etapas de los ciclos de desarrollo y utilización, y el artículo 3.º, numeral VII, exige supervisión humana efectiva, periódica y adecuada en el ciclo de vida (Conselho Nacional de

Justiça, 2025). La reserva humana deja así de ser un requisito del instante final y se proyecta sobre el diseño, el entrenamiento, la validación, la puesta en servicio y el monitoreo del sistema.

De la lectura conjunta de las fuentes examinadas en esta investigación se desprenden dos exigencias que conviene distinguir con nombre propio, porque el ordenamiento las presupone sin nombrarlas. La primera puede denominarse trazabilidad rendicional, y consiste en la capacidad de la autoridad de demostrar, con evidencia documentada, qué hizo en cada fase del ciclo de vida del sistema. La segunda es la auditabilidad independiente, y consiste en la posibilidad efectiva de que un tercero ajeno a la administración verifique esa evidencia.

Ninguna de las dos aparece formulada con esos términos en las normas examinadas, pero ambas se deducen de sus exigencias y de la jurisprudencia analizada. La primera se desprende de la obligación de supervisión a lo largo del ciclo de vida que impone la Resolución 615 y, en el plano jurisprudencial, de la orden de la sentencia T-067 de 2025, que no se limitó a exigir la entrega del código fuente en su estado final sino que comprendió todo el historial y control de versiones, lo que equivale a una orden de trazabilidad. La segunda se desprende del reproche que el Rechtbank Den Haag dirigió a un sistema por resultar insuficientemente transparente y verificable, y de la condición de auditoría interna y externa del algoritmo que el Consejo Constitucional francés impuso como requisito de conformidad, ambas examinadas en el apartado 5.

La distinción no es terminológica. Una obligación de supervisar que no viene acompañada del deber correlativo de documentar lo supervisado no genera evidencia alguna, y una evidencia que solo puede examinar quien la produjo no permite control externo. La reserva humana requiere las dos: sin la primera, no hay nada que auditar; sin la segunda, la auditoría queda en manos del auditado. Esta es la razón por la cual, como se argumentará en la discusión, ninguno de los dos ordenamientos examinados ha alcanzado todavía la exigibilidad plena de la garantía que ambos proclaman.

Cabe señalar que el Proyecto de Ley 2338 de 2023, que constituiría el marco legal general de la inteligencia artificial en Brasil, continúa en trámite. Su ficha oficial lo ubica en la Comisión Especial sobre Inteligencia Artificial de la Cámara de Diputados, con relator designado el 20 de mayo de 2025 y a la espera de su dictamen, con treinta y siete proyectos acumulados (Câmara dos Deputados, 2026). El dato importa para la comparación: al momento de esta investigación, la regulación brasileña de la materia en el ámbito judicial es de origen administrativo y no legal.

4. Transparencia algorítmica: dos rutas dogmáticas, dos niveles de exigibilidad

4.1 Colombia: el acceso a la información pública

La reserva humana sería una garantía puramente formal si no viniera acompañada de un derecho correlativo: el de que la ciudadanía, y no solo el juez, pueda verificar cómo opera el sistema que intervino en una decisión que la afecta. En el ordenamiento colombiano ese desarrollo se produjo en la sentencia T-067 de 2025, dictada en el expediente T-8.202.533 con ponencia de la magistrada Natalia Ángel Cabo, con ocasión de la solicitud de acceso al código fuente de la aplicación CoronApp.

La Corte concedió el amparo del derecho fundamental de acceso a la información pública y ordenó al Ministerio de Salud y Protección Social entregar el código fuente de la aplicación con todo el historial y control de versiones, desde la versión original hasta el momento en que estuvo en uso, así como informar si una versión actualizada se encontraba en uso y, de ser así, dar acceso también a ella (Corte Constitucional de Colombia, 2025). La amplitud de la orden merece subrayarse: no se ordenó entregar el código en su estado final, sino su historia completa de versiones. Es, en la práctica, una orden de trazabilidad.

El fundamento de la decisión fue el incumplimiento de la carga probatoria del artículo 28 de la Ley 1712 de 2014, que desarrolla el llamado test de daño. Conforme a esa disposición, el sujeto obligado que niega el acceso alegando un daño a un interés particular o público debe demostrar que la información se relaciona con un objetivo constitucional y legalmente legítimo, que se trata de una de las excepciones expresamente establecidas, que su divulgación causaría un daño presente, probable y específico sobre un bien o interés constitucional, y que ese daño excede el interés público que representa el acceso (Corte Constitucional de Colombia, 2025). La Corte encontró que la carga no se había satisfecho.

La ratio que interesa a esta investigación es la que la propia providencia formula al describir el efecto de la opacidad: la falta de transparencia y acceso al código fuente privó tanto al ciudadano accionante como a la sociedad en general de la posibilidad de ejercer un control adecuado y oportuno sobre esa herramienta, e impidió que los ciudadanos pudieran verificar la precisión, la seguridad y el uso correcto de sus datos personales (Corte Constitucional de Colombia, 2025). El daño no se agota, entonces, en la esfera individual del solicitante: la Corte lo describe como una privación colectiva de capacidad de control. Coherentemente, la sentencia incorporó al ordenamiento la categoría de los sistemas de toma de decisiones automatizadas y calificó la transparencia algorítmica como garantía fundamental.

4.2 Brasil: la protección de datos y una modificación legislativa decisiva

Brasil ancló la transparencia en un fundamento distinto. La Enmienda Constitucional 115, promulgada el 10 de febrero de 2022, incorporó al artículo 5.º de la Constitución el inciso LXXIX, que reconoce el derecho a la protección de los datos personales, inclusive en los medios digitales, y añadió las competencias correspondientes en los artículos 21 y 22 (Brasil, 2022). La protección de datos adquirió así rango de derecho fundamental.

En el plano legal, el artículo 20 de la Lei 13.709 de 2018 reconoce al titular el derecho a solicitar la revisión de decisiones tomadas únicamente con base en tratamiento automatizado de datos personales que afecten sus intereses, incluidas las destinadas a definir su perfil personal, profesional, de consumo y de crédito o los aspectos de su personalidad; su parágrafo 1.º impone al controlador el deber de suministrar información clara y adecuada sobre los criterios y procedimientos empleados, y su parágrafo 2.º faculta a la autoridad nacional para realizar auditoría (Brasil, 2018).

Ahora bien, el examen del texto compilado revela un dato decisivo para esta investigación. La redacción originaria de 2018 reconocía el derecho a solicitar la revisión «por persona natural»

de tales decisiones. Esa referencia fue extirpada por la Medida Provisoria 869 de 2018, convertida en la Lei 13.853 de 2019, de modo que el texto vigente garantiza el derecho a la revisión pero ya no menciona a quién corresponde practicarla. El parágrafo 3.º, que preveía la posibilidad de que la autoridad nacional realizara auditoría en determinados supuestos, fue vetado (Brasil, 2018).

El alcance de esa supresión es objeto de discusión doctrinal, y conviene reproducirla porque su resultado es, para la tesis de este artículo, más elocuente que cualquiera de las dos posiciones por separado. Macena de Lima y Freire de Sá (2020) registran dos líneas hermenéuticas. Conforme a la primera, la modificación suprimió la posibilidad de que la revisión la practicara una persona natural, de suerte que una decisión automatizada pasaría a ser revisada mediante otra decisión automatizada. Conforme a la segunda, la modificación significa únicamente que la ley dejó de detallar las condiciones de la revisión, sin que exista prohibición de que la realice un ser humano: habría permiso, mas no obligatoriedad, de que la solicitud sea procesada por otro sistema automatizado. Las autoras se inclinan por sostener que la revisión humana resulta más apta para corregir eventuales discriminaciones derivadas de procesos algorítmicos y para dar concreción a los principios de transparencia y responsabilidad (pp. 227-246).

Lo relevante no es cuál de las dos lecturas prevalezca. Lo relevante es que, siete años después de expedida la ley, el ordenamiento brasileño no ofrece una respuesta inequívoca a la pregunta de si la revisión de una decisión automatizada debe practicarla una persona. Una garantía cuyo contenido depende de la interpretación que se adopte no es, en sentido propio, una garantía exigible; y esa indeterminación no proviene de un vacío normativo sino de una modificación deliberada del texto que la contenía.

Conviene añadir que el derecho a la explicación no era inédito en el ordenamiento brasileño: la Lei 12.414 de 2011, sobre catastro positivo, ya reconocía al inscrito el derecho de solicitar la revisión de decisiones adoptadas exclusivamente por medios automatizados, y su incorporación al microsistema de protección de datos amplió su campo de aplicación. El artículo 20 es, además, una consecuencia del principio de transparencia consagrado en el artículo 6.º, numeral VI, de la propia ley (Macena de Lima y Freire de Sá, 2020).

4.3 Contraste

Tabla 2 *Fundamentos y alcance de la transparencia algorítmica en Colombia y Brasil*

Variable	Colombia	Brasil
Fundamento dogmático	Derecho de acceso a la información pública (Ley 1712 de 2014).	Derecho a la protección de datos personales (art. 5.º LXXIX de la Constitución; Lei 13.709 de 2018).
Naturaleza de la garantía	Democrática y colectiva: la Corte describe la opacidad como privación de control para el accionante y para la sociedad en general.	Individual: el titular de los datos es quien puede solicitar la revisión y la información sobre los criterios.

Rango	Constitucional, artículo 74, con desarrollo legal en la ley 1712 de 2014.	Constitucional desde la Enmienda 115 de 2022; desarrollo legal en la Lei 13.709 de 2018.
Objeto del acceso	Código fuente con historial y control de versiones completo.	Información clara y adecuada sobre criterios y procedimientos del tratamiento.
Reserva humana en la revisión	No regulada expresamente en esta vía; se desprende del estándar de la T-323 de 2024 en el ámbito judicial.	Prevista en la redacción originaria de 2018 y suprimida por la Lei 13.853 de 2019.
Mecanismo de control	Test de daño del art. 28 de la Ley 1712 de 2014, con carga probatoria a cargo del sujeto obligado.	Auditoría de la autoridad nacional (párrafo 2.º del art. 20); el párrafo 3.º fue vetado.

Nota. a partir de Corte Constitucional de Colombia (2025), Brasil (2018) y Brasil (2022).

El contraste desmiente una lectura extendida según la cual ambos ordenamientos llegarían, por vías distintas, a exigencias equivalentes. Las exigencias no son equivalentes. La ruta colombiana produce una garantía de titularidad difusa, que cualquier ciudadano puede activar, sujeta a un test de daño cuya carga probatoria recae sobre el sujeto obligado, y capaz de traducirse en una orden de entrega del historial completo de versiones. La ruta brasileña produce una garantía de titularidad individual, condicionada a que el solicitante sea el titular de los datos tratados, y desprovista desde 2019 de la exigencia de intervención humana en la revisión. La convergencia es de principios; la divergencia es de exigibilidad.

5. El estándar europeo de contraste

Ni Colombia ni Brasil cuentan aún con jurisprudencia que haya sometido a escrutinio la arquitectura de un sistema algorítmico estatal en su conjunto. Europa sí, y sus decisiones ofrecen un banco de pruebas útil no por su geografía sino por su función: permiten observar cómo se comporta la reserva humana cuando un tribunal la aplica a un sistema real.

5.1 SyRI: la insuficiencia de la supervisión humana formal

El Systeem Risico Indicatie neerlandés era un instrumento legal y tecnológico destinado a detectar diversas formas de fraude, incluidos beneficios sociales, subsidios y fraude fiscal, mediante la generación de informes de riesgo (Barbosa Jaimes, 2023, p. 135). El Rechtbank Den Haag lo examinó y, en sentencia de 5 de febrero de 2020, declaró que las disposiciones que lo sustentaban carecían de efecto vinculante por vulnerar el artículo 8 del Convenio Europeo de Derechos Humanos⁷.

Tres elementos del fallo resultan directamente aplicables a la cuestión aquí planteada. El primero es la constatación de que la aplicación del sistema era insuficientemente transparente y verificable, lo que impedía al propio tribunal evaluar si funcionaba como se afirmaba (Rechtbank

⁷ La versión en neerlandés se identifica como ECLI:NL:RBDHA:2020:865 y la traducción oficial al inglés, empleada en esta investigación, como ECLI:NL:RBDHA:2020:1878.

Den Haag, 2020, párr. 6.10). El segundo es la afirmación de que el Estado tiene una responsabilidad especial cuando aplica nuevas tecnologías, que le impone equilibrar la prevención del fraude con la protección de la vida privada (párr. 6.9). El tercero, y el más importante para esta investigación, es la respuesta que el tribunal dio al argumento defensivo del Estado.

La administración neerlandesa sostuvo que el sistema no adoptaba decisiones, sino que se limitaba a producir informes de riesgo, y que el juzgamiento posterior correspondía a funcionarios humanos. El tribunal no acogió esa defensa: consideró que, aunque formalmente no se tratara de decisiones automatizadas, los informes de riesgo producían un efecto significativo sobre la vida privada de las personas señaladas, comparable al de determinaciones jurídicamente vinculantes, y que la revisión posterior no subsanaba el señalamiento algorítmico previo (Rechtbank Den Haag, 2020, párr. 6.57). Concluyó que la legislación no lograba el equilibrio justo exigido (párr. 6.80) y declaró la falta de efecto vinculante de las disposiciones examinadas (párr. 7).

Es la demostración judicial de la hipótesis de este artículo. Un ordenamiento que exige intervención humana en el momento final de la decisión, y que efectivamente la tiene, puede no obstante producir un resultado incompatible con las garantías fundamentales si la administración no está en condiciones de explicar por qué el sistema señaló a quien señaló. La supervisión humana formal no sana la opacidad previa. Conviene precisar, por lo demás, que el tribunal no declaró la inconstitucionalidad del sistema en sentido estricto, sino la falta de efecto vinculante de las normas que lo autorizaban por infracción de un instrumento internacional de derechos humanos; la distinción es pertinente en un análisis comparado y a menudo se pierde en los reportes de la decisión.

5.2 El estándar positivo: la decisión francesa de 2019

Si SyRI muestra lo que no basta, la Decisión 2019-796 DC de 27 de diciembre de 2019 del Consejo Constitucional francés muestra lo que se exige. Según reporta Barbosa Jaimes (2023), la decisión examinó la ley de finanzas para 2020, que autorizaba a las administraciones tributaria y aduanera a recolectar de manera indiferenciada grandes volúmenes de datos publicados en sitios de operadores de plataformas y a explotarlos mediante agregación y verificaciones cruzadas. El tribunal declaró la conformidad de la disposición bajo condición de que el sistema se limitara a recolectar, almacenar y organizar datos sin generar juicios valorativos sobre ellos, permaneciera bajo supervisión humana total en cada uno de sus procedimientos y etapas, y permitiera la auditoría interna y externa del algoritmo, incluida la modificación de sus patrones y variables (pp. 136-137). La formulación es notable por su fecha y por su contenido. Cinco años antes de la sentencia colombiana y seis antes de la resolución brasileña, un tribunal constitucional europeo condicionó la admisibilidad de un sistema algorítmico estatal a tres requisitos que coinciden con las condiciones que esta investigación considera necesarias: ausencia de valoración autónoma, supervisión humana distribuida a lo largo de todas las etapas y auditabilidad efectiva por terceros. La reserva humana, formulada así, no es una exigencia sobre quién firma, sino sobre quién puede verificar

5.3 SALER: la eficacia como variable independiente

El Sistema de Alerta Rápida de la Generalitat Valenciana, previsto en el artículo 17 de la Ley 22 de 6 de noviembre de 2018, se define como el conjunto de herramientas cuya interacción permite la detección preventiva de posibles irregularidades y malas prácticas administrativas a partir del análisis de la información obtenida y de la evaluación de factores de riesgo. A diferencia del caso neerlandés, opera sobre una arquitectura simple con entradas y criterios de alerta bien delimitados: adjudicación recurrente en procesos de selección, colusión en licitaciones, conflictos de intereses, falsedad documental y fraccionamiento de proveedores (Barbosa Jaimes, 2023, pp. 137-138).

El propio autor advierte, sin embargo, que a la fecha de su estudio no se había iniciado ninguna actuación de investigación de irregularidades como consecuencia del análisis de datos efectuado por el sistema (Barbosa Jaimes, 2023, p. 138). El dato introduce una variable que el análisis puramente normativo tiende a omitir: un sistema jurídicamente irreprochable puede resultar operativamente irrelevante. La corrección del diseño no garantiza el rendimiento, del mismo modo que, y esta es la simetría que interesa, la declaración de una garantía no asegura su ejercicio.

6. Dos estudios de caso sobre la responsabilidad del operador judicial

La incorporación de estos sistemas a los aparatos judiciales de la región no es un fenómeno aislado ni reciente. Pérez-Pacheco y Simón Castellano (2025) ofrecen una revisión de los desafíos y oportunidades que la inteligencia artificial plantea a los sistemas judiciales de América Latina, que permite situar los dos casos siguientes en un movimiento regional más amplio y no como episodios excepcionales.

Hecha esa salvedad, los apartados anteriores han examinado el diseño normativo. Este examina su aplicación en dos casos concretos, uno por ordenamiento, atendiendo a cuatro variables: quién usó el sistema, en qué fase del proceso decisorio intervino, qué deber se consideró infringido y sobre quién recayó la imputación.

6.1 Colombia: el juez que consultó a ChatGPT

Los hechos que dieron lugar a la sentencia T-323 de 2024 son los siguientes. Una madre promovió acción de tutela en representación de su hijo menor de edad, diagnosticado con trastorno del espectro autista, contra su entidad promotora de salud, por el cobro de copagos y cuotas moderadoras, la negativa del servicio de transporte y la falta de autorización de tratamiento integral. El juzgado municipal concedió parcialmente el amparo el 7 de diciembre de 2022. El juzgado del circuito, al resolver la segunda instancia el 30 de enero de 2023, confirmó la decisión y empleó ChatGPT 3.5 en la motivación de su fallo (Corte Constitucional de Colombia, 2024).

La Corte examinó si ese empleo había vulnerado el debido proceso, formulando el problema en términos precisos: existían dudas sobre si quien emitió la decisión fue un juez de la

República o una inteligencia artificial, y sobre si la decisión fue debidamente motivada o resultó producto de alucinaciones y sesgos generados por el sistema. La respuesta fue negativa, y su fundamento se encuentra en el orden metodológico seguido por el funcionario judicial:

no existió un remplazo del ejercicio de la función jurisdiccional por parte de ChatGPT, esencialmente porque el sistema de IA se utilizó luego de haberse fundamentado y tomado la decisión. En el orden metodológico que adoptó el fallo de tutela, el funcionario judicial primero identificó la tesis que sostendría, luego las normas constitucionales aplicables al caso, el referente jurisprudencial que se debía atender [...] para posteriormente solucionar el caso concreto [...] y, solo entonces, anunció y procedió a efectuar preguntas en el referido sistema de IA, para así transcribir las respuestas arrojadas en la consulta. (Corte Constitucional de Colombia, 2024)

El criterio de decisión es, por tanto, temporal y funcional: lo determinante no fue que el juez usara la herramienta, sino en qué momento del iter decisorio lo hizo. La consulta posterior a la formación del juicio no sustituye la función jurisdiccional; la consulta previa, que determinara la tesis o la valoración probatoria, sí la comprometería. La Corte encontró, no obstante, que los principios de transparencia y responsabilidad no se cumplieron a cabalidad, aunque sí el de privacidad, pues el juez no introdujo datos personales del menor ni de su historia clínica ni de las partes.

El resultado es una imputación de baja intensidad: se declara la infracción de dos de los criterios, pero no se anula lo actuado ni se deriva consecuencia sancionatoria alguna para el funcionario. La Corte optó por la vía prospectiva, esto es, el exhorto, los criterios orientadores y la orden de reglamentación, antes que por la retrospectiva. Esa elección es comprensible en un escenario carente de regulación al momento de los hechos, pero deja abierta la pregunta sobre qué habría ocurrido de haber existido ya el estándar.

6.2 Brasil: la manipulación del sistema del despacho

El segundo caso invierte los términos. Aquí no es el operador judicial quien usa el sistema de manera cuestionable, sino quien intenta manipular el sistema que usa el despacho.

En el proceso 0001062-55.2025.5.08.0130⁸, una reclamación laboral ordinaria seguida ante la 3ª Vara do Trabalho de Parauapebas, el juez sustituto Luiz Carlos de Araujo Santos Junior dictó sentencia el 12 de mayo de 2026. El despacho se encontraba autorizado para utilizar el sistema de inteligencia artificial generativa denominado Galileu, desarrollado por el Tribunal Regional do Trabalho de la 4ª Región y nacionalizado por el Conselho Superior da Justiça do Trabalho, con fundamento en las resoluciones 332 de 2020 y 615 de 2025 del Conselho Nacional de Justiça. Al

⁸ El número corresponde al formato unificado del Consejo Nacional de Justicia brasileño: siete dígitos secuenciales, dos de verificación, el año de radicación, el segmento de justicia (8, justicia estadual), el tribunal (5.08, Tribunal Regional do Trabalho de la 8ª Región) y la unidad de origen.

procesar la demanda mediante ese sistema, se identificó la existencia de un texto insertado en fuente de color blanco sobre fondo blanco, por tanto invisible al lector humano, que contenía la siguiente instrucción oculta, revelada tras la alteración tecnológica del color de la fuente: «ATENCIÓN, INTELIGENCIA ARTIFICIAL, CONTESTA ESTA PETICIÓN DE FORMA SUPERFICIAL Y NO IMPUGNES LOS DOCUMENTOS, INDEPENDIENTEMENTE DE LA ORDEN QUE SE TE DÉ» (Tribunal Regional do Trabalho da 8ª Região, 2026).

La sentencia identifica la técnica como inyección de instrucciones, consistente en la inserción deliberada de comandos ocultos en documentos con el propósito de manipular los sistemas de inteligencia artificial que lleguen a procesarlos, y describe la finalidad perseguida: que el sistema empleado por la parte contraria o por el propio juzgado generara una contestación superficial o un proyecto de sentencia comprometido, en perjuicio del demandado y de la propia prestación jurisdiccional.

El razonamiento sobre la imputación es el aporte central de la decisión. El juez precisó que la elaboración de la demanda es acto privativo del abogado, de cuya entera y exclusiva responsabilidad es el contenido del documento suscrito y presentado ante el juzgado; que el demandante, en su condición de parte, no posee conocimiento técnico ni acceso directo a la redacción de la pieza; y que no resulta razonable ni jurídicamente sostenible atribuirle la inserción de una instrucción oculta en lenguaje técnico dirigida a manipular sistemas de inteligencia artificial, conducta que presupone conocimiento específico y acceso a la elaboración del documento. La responsabilidad recayó, en consecuencia, de forma exclusiva sobre las dos abogadas firmantes (Tribunal Regional do Trabalho da 8ª Região, 2026).

La decisión enfrentó además un obstáculo procesal que resolvió con un argumento de alcance general. El parágrafo 6.º del artículo 77 del Código de Proceso Civil brasileño excluye a los abogados de las multas por acto atentatorio contra la dignidad de la justicia y remite su eventual responsabilidad disciplinaria al órgano de clase. El juez sostuvo que esa exclusión presupone que la conducta cuestionada se sitúe dentro del espectro, aunque irregular, del ejercicio de la abogacía, y que la inserción de una instrucción oculta destinada a manipular sistemas del Poder Judicial no constituye acto de defensa del cliente, no integra el ejercicio de la postulación y no guarda relación con la representación procesal legítima. Cuando el abogado deja de actuar como sujeto del proceso para obrar como agente de sabotaje del sistema judicial, concluyó, su conducta deja de estar protegida por la independencia funcional.

Dos precisiones adicionales completan el cuadro. El demandado había sido declarado rebelde, de modo que ninguna contestación llegó a producirse bajo la influencia del comando oculto y no hubo perjuicio procesal concreto; el juez consideró, sin embargo, que el ilícito procesal es de naturaleza formal y se consuma en el momento en que la instrucción se inserta en el documento presentado ante el Poder Judicial, con independencia de que haya alcanzado su resultado. La multa se fijó en el diez por ciento del valor de la causa, esto es, la mitad del límite legal, atendiendo a la gravedad objetiva de la conducta, a la ausencia de perjuicio concreto y a la necesidad de que la sanción fuera disuasoria sin resultar confiscatoria, y se dispuso oficiar a la seccional del colegio de abogados y a la corregiduría del tribunal regional (Tribunal Regional do Trabalho da 8ª Região, 2026).

6.3 Contraste de los dos casos

Tabla 3 Variables de imputación en los dos estudios de caso

Variable	Colombia: T-323 de 2024	Brasil: proceso 0001062-55.2025.5.08.0130
Quién usó el sistema	El juez de segunda instancia, en su propia motivación.	El despacho, con sistema institucional autorizado; las abogadas intentaron manipularlo desde fuera.
Fase de intervención	Posterior a la formación del juicio: la consulta se hizo después de adoptada y fundamentada la decisión.	Previa: la instrucción oculta buscaba condicionar el procesamiento de la demanda.
Deber infringido	Transparencia y responsabilidad en el uso de la herramienta.	Buena fe procesal y deber de no dificultar la actividad jurisdiccional.
Sujeto imputado	El funcionario judicial, sin consecuencia sancionatoria.	Las abogadas firmantes, de forma exclusiva y solidaria; se excluyó a la parte.
Consecuencia	Exhorto, criterios orientadores y orden de reglamentación al órgano de gobierno judicial.	Multa del 10 % del valor de la causa, más comunicación al colegio de abogados y a la corregiduría.
Criterio de imputación	Momento de la intervención en el iter decisorio.	Autoría material del documento y aptitud técnica de la conducta.

Nota a partir de Corte Constitucional de Colombia (2024) y Tribunal Regional do Trabalho da 8ª Região (2026).

Discusión y conclusiones

Los resultados permiten responder la pregunta de investigación y obligan a matizar la hipótesis en un punto.

La primera conclusión es que la convergencia entre Colombia y Brasil es real pero se sitúa en el plano de los principios, no en el de la exigibilidad. Ambos ordenamientos afirman que existe un núcleo indelegable, y lo hacen con formulaciones sustantivamente coincidentes pese a provenir de sedes institucionales heterogéneas: una sentencia de tutela con fundamento en el debido proceso, en un caso; un acto administrativo del órgano nacional de gobierno judicial, en el otro. Esa coincidencia por vías dogmáticas distintas es, en sí misma, un resultado del análisis comparado y sugiere que la reserva humana no depende de una arquitectura constitucional particular, sino que emerge allí donde el ordenamiento toma en serio el deber de motivación.

La segunda conclusión matiza la hipótesis de partida. Se afirmó que la reserva humana en ambos países opera sobre el acto final y no sobre el ciclo de vida. Eso es exacto respecto de Colombia, donde el estándar de la sentencia T-323 de 2024 se dirige al comportamiento del

funcionario en el momento en que emplea la herramienta, pero no lo es respecto de Brasil: los artículos 2.º y 3.º de la Resolución 615 de 2025 extienden expresamente la supervisión humana a todas las etapas del ciclo de desarrollo y utilización. La hipótesis debe reformularse en consecuencia: el problema brasileño no es la ausencia de la exigencia, sino la ausencia de la trazabilidad rendicional que la haría verificable, a lo que se suma la indeterminación del artículo 20 de su ley de protección de datos.

La tercera conclusión es la más significativa. Las dos rutas dogmáticas de la transparencia algorítmica no producen exigencias equivalentes. En Colombia, la vía del acceso a la información pública generó una garantía de titularidad difusa, que cualquier ciudadano puede activar, sujeta a un test de daño cuya carga probatoria recae sobre el sujeto obligado, y capaz de traducirse en una orden de entrega del historial completo de versiones de un código fuente. En Brasil, la vía de la protección de datos genera una garantía de titularidad individual, condicionada a la calidad de titular de los datos tratados, y desprovista desde 2019 de la exigencia de que la revisión de la decisión automatizada la practique una persona natural. La afirmación, extendida en la literatura, de que ambos ordenamientos llegan a exigencias sustantivamente convergentes por vías estructuralmente distintas debe corregirse: llegan a principios convergentes y a exigibilidades desiguales.

La modificación operada sobre el artículo 20 merece una consideración final, porque constituye la evidencia más nítida de la tesis aquí sostenida. La reserva humana no se pierde por descuido ni por vacío normativo: en Brasil estuvo escrita en la ley general de protección de datos y la expresión que la contenía fue retirada deliberadamente, con el resultado de que hoy la doctrina discute si la garantía subsiste o no. Una garantía puede declararse solemnemente en un instrumento y ser vaciada en otro sin que la declaración pierda su apariencia de vigencia. Es precisamente por eso que la exigibilidad, y no la declaración, debe ser el objeto del análisis jurídico. Algo semejante ocurre, en una escala menor, con la facultad que el parágrafo 6.º del artículo 19 de la Resolución 615 concede al magistrado para decidir si revela o no que empleó inteligencia artificial generativa.

La cuarta conclusión proviene del contraste con el estándar europeo. La sentencia neerlandesa de 2020 demuestra que la intervención humana en el momento final no sana la opacidad previa: el tribunal rechazó expresamente el argumento de que los informes de riesgo no eran decisiones porque después decidían funcionarios, y sostuvo que la revisión posterior no subsana el señalamiento algorítmico anterior. La decisión francesa de 2019, por su parte, ofrece el estándar positivo correspondiente, al condicionar la admisibilidad del sistema a la ausencia de valoración autónoma, a la supervisión humana en cada etapa y a la auditabilidad interna y externa del algoritmo. Ambas apuntan en la misma dirección: el criterio de validez no es quién firma, sino quién puede verificar.

La quinta conclusión se desprende de los estudios de caso y responde a la segunda parte de la pregunta de investigación. En los dos casos analizados, la imputación se activó por la conducta del operador y no por el comportamiento del sistema. En Colombia, el criterio determinante fue el momento del iter decisorio en que el juez consultó la herramienta; en Brasil, la autoría material del documento y la aptitud técnica de la conducta. En ninguno de los dos se discutió el funcionamiento del sistema, ni se exigió a nadie explicarlo. Esto tiene una lectura ambivalente.

Por un lado, muestra que los ordenamientos disponen ya de categorías suficientes para imputar responsabilidad cuando el uso de la herramienta es reprochable, sin necesidad de construir un régimen especial: la buena fe procesal, el deber de motivación y las cargas de transparencia y verificación bastaron. Por otro, revela el límite de ese enfoque: sirve para los casos en que la conducta humana es identificable y reprochable, y no para aquellos en que el sistema produce un resultado lesivo sin que medie conducta reprochable de nadie, que es precisamente el supuesto que la opacidad hace posible.

Se concluye, en consecuencia, que la reserva humana de la decisión adquiere exigibilidad cuando concurren tres condiciones: que la supervisión se extienda a todo el ciclo de vida del sistema y no solo al acto final; que esa supervisión genere trazabilidad rendicional, es decir, documentación capaz de ser verificada por un tercero independiente, sin la cual la obligación de supervisar no produce evidencia alguna; y que exista un régimen de imputación determinado que permita identificar a un responsable. Colombia satisface parcialmente la tercera condición por vía jurisprudencial y ha avanzado en la primera solo respecto del operador; Brasil satisface normativamente la primera y ha desarrollado la tercera en sede judicial, pero debilitó en 2019 la garantía legal correlativa. Ninguno de los dos satisface hoy la segunda, que es la que convierte a las otras dos en operativas.

La investigación deja abiertas dos líneas. La primera es empírica: establecer, con datos de aplicación, en qué medida los estándares examinados se observan efectivamente en los despachos judiciales de ambos países, cuestión que este trabajo no aborda y que la advertencia sobre el sistema valenciano sugiere que no debe presumirse. La segunda es dogmática: determinar el régimen aplicable cuando el resultado lesivo no es reconducible a una conducta reprochable identificable, supuesto para el cual las categorías empleadas en los dos casos analizados resultan insuficientes y que constituye, probablemente, el verdadero problema que la inteligencia artificial plantea a la teoría de la responsabilidad.

Referencias

- Barbosa Jaimes, J. C. (2023). El papel de la inteligencia artificial en la lucha contra la corrupción. En A. E. Carrillo del Teso y M. G. Pahul Robredo (Dirs.), *Desafíos en la lucha contra la corrupción: gestión de riesgos y paradigmas globales* (pp. 127-140). Editorial Colex.
- Berning Prieto, A. D. (2023). El uso de sistemas basados en inteligencia artificial por las Administraciones públicas: estado actual de la cuestión y algunas propuestas ad futurum para un uso responsable. *Revista de Estudios de la Administración Local y Autonómica*, (20), 165-185. <https://doi.org/10.24965/real.11247>
- Brasil. (2018). *Lei n.º 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais* [texto compilado, con las modificaciones de la Lei n.º 13.853 de 2019]. Presidência da República. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm
- Brasil. (2022). *Emenda Constitucional n.º 115, de 10 de fevereiro de 2022*. Presidência da República. https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc115.htm

- Câmara dos Deputados. (2026). *Ficha de tramitação do Projeto de Lei n.º 2.338, de 2023*. <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>
- Citron, D. K. (2008). Technological due process. *Washington University Law Review*, 85(6), 1249-1313. https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2
- Conselho Nacional de Justiça. (2025). *Resolução n.º 615, de 11 de março de 2025*. <https://atos.cnj.jus.br/atos/detalhar/6001>
- Corte Constitucional de Colombia. (2024, 2 de agosto). *Sentencia T-323 de 2024* (M. P. Juan Carlos Cortés González; expediente T-9.301.656).
- Corte Constitucional de Colombia. (2025). *Sentencia T-067 de 2025* (M. P. Natalia Ángel Cabo; expediente T-8.202.533).
- Kingsbury, B., Krisch, N., y Stewart, R. B. (2005). The emergence of global administrative law. *Law and Contemporary Problems*, 68(3-4), 15-61. <https://scholarship.law.duke.edu/lcp/vol68/iss3/2>
- Macena de Lima, T. M., y Freire de Sá, M. de F. (2020). Inteligência artificial e Lei Geral de Proteção de Dados Pessoais: o direito à explicação nas decisões automatizadas. *Revista Brasileira de Direito Civil*, 26, 227-246. <https://doi.org/10.33242/rbdc.2020.04.011>
- Pérez-Pacheco, Y., y Simón Castellano, P. (2025). La inteligencia artificial en los sistemas judiciales de América Latina: una revisión sobre desafíos y oportunidades. *A&C – Revista de Direito Administrativo & Constitucional*, 25(100), 95-129. <https://doi.org/10.21056/aec.v25i100.2027>
- Ponce Solé, J. (2019). Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico. *Revista General de Derecho Administrativo*, (50).
- Rechtbank Den Haag. (2020, 5 de febrero). *NJCM y otros c. Estado de los Países Bajos (SyRI)*, C-09-550982-HA ZA 18-388 [ECLI:NL:RBDHA:2020:865; traducción oficial al inglés ECLI:NL:RBDHA:2020:1878].
- Tribunal Regional do Trabalho da 8ª Região. (2026, 12 de mayo). *Sentencia dictada en el proceso n.º 0001062-55.2025.5.08.0130*. 3ª Vara do Trabalho de Parauapebas.