

Predecir no es juzgar: Sistemas algorítmicos de riesgo penal y presunción de inocencia en América Latina*

Predicting is not judging: Algorithmic systems of criminal risk and presumption of innocence in Latin America

Claudia Lorena Morffi Collado**

RESUMEN

En el año 2023, un juez penal de Bogotá incorporó en una decisión cautelar el resultado arrojado por PRISMA, sistema algorítmico de evaluación de riesgo de la fiscalía general de la Nación, sin que la defensa pudiera conocer las variables utilizadas ni su ponderación. Este hecho, documentado en la Sentencia T-529/2023 de la Corte Constitucional colombiana, constituye el punto de partida de este artículo. El objetivo es demostrar que la incorporación de sistemas algorítmicos de evaluación de riesgo en la justicia penal latinoamericana no ha supuesto una modernización garantista, sino la reproducción automatizada de desigualdades preexistentes. La investigación se desarrolló mediante un enfoque cualitativo, de carácter jurídico-doctrinal y comparado. A partir del examen de los casos de Argentina, Colombia, Brasil y Chile, se identifican como principal resultado las funciones atribuidas a los sistemas algorítmicos de evaluación de riesgo, sus niveles de transparencia y las tensiones que generan frente a las garantías del debido proceso.

PALABRAS CLAVE

inteligencia artificial judicial; presunción de inocencia; opacidad algorítmica; evaluación de riesgo procesal; positivismo tecnificado.

ABSTRACT

In 2023, a criminal court judge in Bogotá relied, in a precautionary ruling, on the output produced by PRISMA, a risk assessment algorithm used by the Office of the Attorney General, without the defense being able to examine either the variables considered or the weight assigned to them. This situation, recorded in Judgment T-529/2023 of the Colombian Constitutional Court, serves as the starting point for this article. The paper argues that the incorporation of algorithmic risk assessment tools into Latin American criminal justice has not led to a rights-based modernization, but rather to the automated reproduction of pre-existing inequalities. The study follows a qualitative, doctrinal, and comparative approach, grounded in the analysis of case law, statutory sources, and specialized literature. Based on a review of Argentina, Colombia, Brazil, and Chile, the findings identify the functions assigned to these systems, their levels of transparency, and the tensions they generate in relation to due process guarantees.

KEYWORDS

judicial artificial intelligence; presumption of innocence; algorithmic opacity; pretrial risk assessment; technicised positivism.

* Artículo de Investigación

** Doctora en Ciencias Jurídicas y Políticas por la Universidad Pablo de Olavide, Sevilla, España y Máster en Ciencias de la Educación Superior (lorenamorffi91@gmail.com). <https://orcid.org/0000-0002-0556-6578>.

SUMARIO

1. Introducción
2. Fundamentos y tensiones del riesgo algorítmico en el proceso penal
3. Circulación y adopción de tecnologías de riesgos en América Latina
4. Sesgos algorítmicos y reproducción de desigualdades estructurales
5. Garantías procesales bajo tensión
6. El positivismo tecnificado como marco explicativo
7. Hacia una regulación de la transparencia algorítmica judicial: referentes y garantías procesales
8. Conclusiones
9. Referencias bibliográficas

1. INTRODUCCIÓN

En el Derecho procesal penal, la restricción de la libertad antes de una condena firme constituye la intervención estatal más grave sobre la esfera individual. Su legitimidad descansa sobre una doble exigencia: que se apoye en hechos concretos y probados, y que el razonamiento que la sustenta pueda ser examinado, contradicho y recurrido. Los sistemas algorítmicos de evaluación de riesgo cuestionan esa doble exigencia en su raíz.

El episodio de Bogotá no fue una anomalía. Fue una consecuencia lógica de una tendencia regional: la adopción de herramientas de predicción actuarial en decisiones penales sin el soporte normativo ni institucional que esa adopción requiere. En Argentina, el sistema Prometea opera en la Fiscalía de la Ciudad de Buenos Aires automatizando actos procesales desde 2017 sin que exista un marco legal que defina con precisión qué puede hacer y qué no. En Brasil, el Supremo Tribunal Federal utiliza Victor para la gestión de recursos y varios tribunales han desplegado modelos predictivos a través de la plataforma Sinapses, en un ecosistema tecnológico que ninguna norma sustantiva disciplina de manera integral. En Colombia, PRISMA estima probabilidades de reincidencia que informan solicitudes de medidas cautelares. En Chile, la Defensoría Penal Pública ha documentado la incorporación de informes de riesgo no auditables en audiencias de prisión preventiva¹.

La pregunta que organiza este trabajo es: ¿en qué condiciones el uso de sistemas algorítmicos de evaluación de riesgo procesal en América Latina es compatible con la presunción de inocencia y el debido proceso, considerando la opacidad estructural de estos sistemas y el vacío regulatorio regional? La respuesta que se defiende aquí es que esa compatibilidad no existe en las condiciones actuales de implementación, y que construirla requiere algo más que regulación técnica: requiere repensar qué tipo de racionalidad puede legítimamente fundar una decisión penal.

La literatura sobre algoritmos y justicia penal se ha concentrado, con razón, en el caso estadounidense: COMPAS, State v. Loomis, el análisis de ProPublica sobre tasas de error

¹BARONA VILAR, Silvia, "Dataización de la justicia (algoritmos, inteligencia artificial y justicia, ¿el comienzo de una gran amistad?)", Revista Boliviana de Derecho, n.º 36, 2023, pp. 14-45; BARROS, Filipe et al., "Machine learning algorithms applied on classification of processes for conciliation on Brazilian labour judiciary", en Encontro Nacional de Inteligência Artificial e Computacional (ENIAC), 20.º, Belo Horizonte, Sociedade Brasileira de Computação, 2023, pp. 389-402.

diferencial por raza². Ese debate es imprescindible como referencia comparada, pero no sustituye el análisis de lo que ocurre en sistemas jurídicos con tradición continental, estructuras institucionales distintas y déficits regulatorios más pronunciados.

La contribución específica del trabajo es doble. Por un lado, ofrece el primer análisis sistemático comparado de cuatro jurisdicciones latinoamericanas (Argentina, Colombia, Brasil y Chile) que evalúa los sistemas implementados no desde su eficiencia operativa, sino desde su compatibilidad con garantías procesales concretas: motivación, contradicción y presunción de inocencia. Por otro, desarrolla el concepto de positivismo tecnificado como categoría analítica propia para describir el fenómeno por el cual la autoridad del dato estadístico desplaza la argumentación jurídica en la decisión judicial, con consecuencias que trascienden el error técnico y comprometen la legitimidad del acto jurisdiccional.

La investigación parte de un estudio de Derecho comparado con enfoque analítico-crítico. No produce datos empíricos propios: trabaja sobre fuentes documentales, jurisprudenciales y técnicas, y su contribución es interpretativa. El análisis cubre cuatro jurisdicciones: Argentina, Colombia, Brasil y Chile. Los tres primeros porque han institucionalizado sistemas con nombre propio, función declarada y documentación pública suficiente: Prometea en Argentina, PRISMA y PretorIA en Colombia, Victor y Sinapses en Brasil. Chile se incorpora porque ofrece un caso analíticamente productivo y diferente: veinte años de reforma procesal penal acusatoria en funcionamiento, un debate documentado sobre herramientas de apoyo a decisiones de prisión preventiva, y una resistencia institucional activa protagonizada por la Defensoría Penal Pública que no tiene equivalente visible en los otros casos. Esa resistencia es en sí misma un dato relevante.

México se consideró en etapas previas de investigación, pero fue excluido del análisis comparado sistemático por ausencia de documentación pública suficiente sobre sistemas algorítmicos penales concretos. Se incluye una referencia descriptiva al contexto mexicano en el apartado 4.3 por su importancia doctrinal regional, pero sin pretensión de comparabilidad sistemática con los otros cuatro casos.

Se han consultado tres tipos de fuentes. Fuentes primarias: decisiones judiciales *State v. Loomis* (Wisconsin, 2016), Sentencia T-529/2023 de la Corte Constitucional colombiana, jurisprudencia del TJUE en *SCHUFA Holding* (C-634/21, 2023) y *Dun & Bradstreet* (C-203/22, 2024), textos normativos Reglamento (UE) 2024/1689 sobre Inteligencia Artificial, RGPD y documentación técnica de los sistemas analizados cuando es públicamente accesible. Fuentes secundarias: literatura académica en derecho procesal, filosofía del derecho y ciencias de la computación³. Fuentes técnicas: el análisis de

²ANGWIN, Julia, et al., "Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks", en *ProPública*.

³BARONA VILAR, Silvia, "Algoritmización de la prueba y de la decisión judicial en el proceso penal: ¿utopía o distopía?", en Pillado González, E. (ed.), *El proceso penal ante una nueva realidad tecnológica europea*, Pamplona, Aranzadi, 2023; BELLOSO MARTÍN, Nuria, "Algoritmos predictivos al servicio de la justicia: ¿una nueva forma de minimizar el riesgo y la incertidumbre?", *Revista da Faculdade Mineira de Direito*, vol. 22, n.º 43, 2019, pp. 2-31; CORVALÁN, Juan Gustavo, "Inteligencia artificial: retos, desafíos y oportunidades. Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la justicia", *Revista de Investigações Constitucionais*, vol. 5, n.º 1, 2018, pp. 55-89; CRUZ GONZÁLEZ, Marcos, "La inteligencia artificial y los derechos humanos. Una carrera tecnológica entre Estados Unidos, China y la Unión Europea", *Revista IUS*, vol. 20, N.º 56, 2025, pp. 31-51.

ProPublica sobre COMPAS, los estudios de DRESSEL y FARID sobre precisión predictiva⁴, el análisis de CHOULDECHOVA sobre incompatibilidad matemática entre métricas de equidad, e informes institucionales de los propios sistemas⁵.

Se adopta un método funcional comparado: los sistemas se analizan en función de lo que hacen, no de cómo se denominan institucionalmente, y se evalúan según su compatibilidad con estándares procesales identificados a partir del derecho constitucional comparado y los instrumentos internacionales de derechos humanos aplicables. El análisis no es neutral respecto de sus conclusiones: parte de la premisa de que las garantías procesales son condiciones de legitimidad del sistema penal, no variables de ajuste en función de la eficiencia operativa.

Por último, el artículo no evalúa el rendimiento técnico de los sistemas analizados ni realiza análisis estadístico de sus tasas de error, tampoco examina sistemas en fase puramente experimental o sin documentación pública suficiente. Las conclusiones tienen alcance regional y no pretenden generalizarse a contextos con estructuras institucionales radicalmente distintas.

2. FUNDAMENTOS Y TENSIONES DEL RIESGO ALGORÍTMICO EN EL PROCESO PENAL

Un sistema de evaluación de riesgo algorítmico es una arquitectura computacional que transforma datos en decisiones probabilísticas. Su resultado es una puntuación o score que pretende anticipar comportamientos humanos futuros a partir de regularidades observadas en datos históricos. En el ámbito penal, esa puntuación se usa para informar decisiones sobre medidas cautelares, libertad condicional o determinación de la pena.

No se trata simplemente de un software auxiliar, su rasgo distintivo es que traduce trayectorias sociales y datos históricos en una puntuación de riesgo que pretende anticipar conductas. Para ello, el sistema procesa variables como antecedentes penales, edad, entorno socioeconómico o historial judicial, y las combina según patrones extraídos de bases de datos previas⁶. El resultado suele expresarse en una categoría (alto, medio, bajo riesgo) o en un valor numérico que se presenta como objetivable. Su lógica es actuarial, pues no examina el hecho concreto en términos de culpabilidad, sino que sitúa al individuo dentro de un grupo estadístico y le asigna el comportamiento típico de ese grupo. En ese desplazamiento (del acto al perfil) radica su principal implicación jurídica.

Además, estos sistemas no operan en el vacío, dependen de datos producidos por instituciones que ya seleccionan, clasifican y filtran la realidad (policía, fiscalía, tribunales). Por eso, más que descubrir riesgo, lo que hacen es reorganizar y proyectar hacia el futuro patrones contenidos en esos datos.

El funcionamiento del sistema opera en tres fases. Primero, la agregación: grandes volúmenes de información antecedentes, variables socioeconómicas, entorno territorial, comportamiento institucional previo se recopilan y codifican. Segundo, el entrenamiento: esos datos se procesan mediante modelos estadísticos o de aprendizaje automático

⁴DRESSEL, Julia; FARID, Hany, "The accuracy, fairness, and limits of predicting recidivism", *Science Advances*, vol. 4, n.º 1, 2018, pp. 1-56.

⁵CHOULDECHOVA, Alexandra, "Fair prediction with disparate impact: a study of bias in recidivism prediction instruments", *Big Data*, vol. 5, n.º 2, 2017, pp. 153-163.

⁶PARRA SEPÚLVEDA, Darío y CONCHA MACHUCA, Ricardo, "Inteligencia artificial y derecho. Problemas, desafíos y oportunidades", en *Revista de Derecho*, vol. 70, 2021, pp. 1-25.

que identifican correlaciones entre variables de entrada y resultados históricos, generalmente decisiones judiciales pasadas. Tercero, la predicción: el sistema aplica los patrones aprendidos al caso concreto y produce una puntuación que, lejos de ser neutral, condensa decisiones previas sobre qué medir, cómo codificarlo y qué objetivo optimizar⁷.

La distinción analíticamente central para este trabajo no es entre modelos más o menos precisos, sino entre modelos de caja negra y modelos de caja blanca. Los primeros generan resultados cuya lógica interna no puede ser reconstruida, ya sea por complejidad técnica (arquitecturas de aprendizaje profundo con millones de parámetros) o por restricciones comerciales, invocación de secreto industrial para bloquear el acceso al código. Los segundos están diseñados para ser interpretables: permiten identificar qué variables incidieron en el resultado y en qué medida.

Esta distinción no es técnica en sentido estricto: es constitucional. Un sistema de caja negra en una decisión penal crea un problema de motivación irresoluble⁸. El Reglamento (UE) 2024/1689 clasifica como sistemas de alto riesgo, entre otros, los empleados en la administración de justicia y los destinados a la predicción de reincidencia delictiva. Esa clasificación activa obligaciones reforzadas transparencia, supervisión humana, auditabilidad que en América Latina no tienen equivalente normativo.

La opacidad algorítmica no es un defecto técnico, sino una fricción directa con las garantías básicas del Estado de Derecho. Aparece cuando ni el contenido ni la lógica interna de un sistema pueden examinarse de modo que la decisión final quede desacoplada de cualquier posibilidad real de comprensión, impugnación o control.

La opacidad tiene dos fuentes distintas que a veces se confunden⁹. La primera es técnica: los modelos de aprendizaje profundo operan mediante arquitecturas de alta complejidad donde la trazabilidad de una decisión concreta se diluye en millones de parámetros ajustados iterativamente. La segunda es jurídica: las empresas desarrolladoras invocan secreto comercial para impedir el acceso a la lógica del sistema¹⁰. Cuando eso ocurre, la opacidad deja de ser una limitación técnica y se convierte en una decisión institucionalizada. Las consecuencias son cualitativamente distintas: la opacidad técnica puede mitigarse mediante herramientas de inteligencia artificial explicable (XAI); la opacidad jurídica requiere una respuesta normativa que establezca que la protección de la propiedad intelectual no puede operar como límite absoluto frente al derecho de defensa.

La falta de explicabilidad impacta en tres dimensiones procesales concretas. Primera, el derecho de defensa: sin acceso a los criterios que guiaron la evaluación, ni el juez ni las partes pueden someter el resultado a contradicción real porque no hay base sobre la que discutir su validez. Segunda, la motivación de las resoluciones: si el órgano

⁷BARONA VILAR, Silvia, "Justicia penal líquida (desde la mirada de Bauman)", *Teoría y Derecho, Revista de pensamiento crítico*, n.º 22, 2017, p. 44.

⁸CONCHA CAMACHO, William, "El derecho a la transparencia algorítmica como estándar ético en una democracia liberal", *Revista Jurídica Austral*, vol. 5, n.º 1, 2024, pp. 529-550.

⁹PASQUALE, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.

¹⁰MEDINA GUERRERO, Manuel, "El derecho a conocer los algoritmos utilizados en la toma de decisiones: aproximación desde la perspectiva del derecho fundamental a la protección de datos", *Teoría y Realidad Constitucional*, n.º 49, 2022, pp. 55-74.

judicial se limita a asumir la salida del sistema, fenómeno documentado en la psicología cognitiva como automation bias, la decisión pierde anclaje en razones verificables. Tercera, el control del sesgo: la opacidad actúa como pantalla para desigualdades presentes en los datos de entrenamiento que se reeditan a escala masiva revestidas de apariencia técnica¹¹.

El desplazamiento desde el hecho probado hacia la probabilidad estimada no es un cambio instrumental: es un cambio de paradigma. La decisión deja de apoyarse en lo que ocurrió y empieza a apoyarse en lo que el modelo anticipa que podría ocurrir. Ese cambio erosiona la presunción de inocencia con independencia de la precisión técnica del sistema, porque el problema no es si el modelo predice bien, sino si predecir es una forma legítima de fundar una decisión penal. La respuesta del Estado de Derecho a esa pregunta ha sido históricamente negativa: la peligrosidad futura no puede fundar por sí sola una restricción de libertad. Los algoritmos no cambian esa respuesta; la eluden.

3. CIRCULACIÓN Y ADOPCIÓN DE TECNOLOGÍAS DE RIESGOS EN AMÉRICA LATINA

3.1 Panorama regional: adopción sin regulación

América Latina presenta un rasgo que la distingue del modelo europeo: la tecnología avanza con rapidez en la práctica institucional mientras el correlato normativo permanece fragmentario. No existen leyes integrales que disciplinen el uso de sistemas algorítmicos en la toma de decisiones públicas. Predomina un enfoque de soft law basado en lineamientos de política pública y principios generales que no resuelven las cuestiones concretas: qué variables puede usar un sistema predictivo penal, quién lo audita, cómo puede impugnarlo la defensa, qué responsabilidad tienen los desarrolladores.

El resultado es una fase de hibridación en la que la toma de decisiones combina intervención humana y asistencia algorítmica sin condiciones claras que definan los límites de cada una¹². Ante ese vacío, algunos tribunales han comenzado a asumir un papel correctivo casuístico: limitar o suspender el uso de tecnologías cuando no se acreditan garantías suficientes. Esa respuesta es necesaria pero insuficiente: llega siempre a posteriori, caso a caso, sin efecto estructural sobre el ecosistema tecnológico.

La pregunta no es ya si la región debe adoptar estas herramientas porque ya las adoptó, sino qué condiciones mínimas deben cumplirse para que esa adopción sea compatible con las garantías que los sistemas jurídicos latinoamericanos reconocen formalmente.

3.2 Análisis comparado: Argentina, Colombia, Brasil y Chile

En Argentina, Prometea aparece configurado principalmente como una herramienta de apoyo para la automatización de tareas repetitivas y la elaboración de borradores o

¹¹RUIZ RAJO, Carlos Enrique, "Entre algoritmos y garantías: una revisión crítica sobre la inteligencia artificial en el proceso penal", *Revista de Investigación Científica Erga Omnes*, vol. 5, n.º 2, 2025, pp. 1-8; GARCÍA, Manuel Lorenzo, "La inteligencia artificial predictiva al servicio de la prevención e investigación del delito y del proceso penal", *Ciencia Policial*, vol. 183, n.º 1, 2024, pp. 91-132.

¹²DE GREGORIO, Giovanni, "From constitutional freedoms to the power of the platforms: protecting fundamental rights online in the algorithmic society", *European Journal of Legal Studies*, vol. 11, n.º 2, 2019, pp. 65-104; GIMÉNEZ, María, "Inteligencia artificial y derecho penal", *Actualidad Jurídica Iberoamericana*, n.º 10, 2019, pp. 792-843.

dictámenes a partir de precedentes, lo que justifica su clasificación como un sistema asistencial y no como un modelo de predicción penal en sentido estricto. Las fuentes institucionales y técnicas disponibles destacan su aporte a la celeridad y eficiencia del trabajo judicial y fiscal, pero no permiten reconstruir públicamente, con el mismo nivel de detalle, la lógica interna completa del sistema ni verificar la existencia de auditorías externas abiertas, razón por la cual su transparencia puede calificarse como parcial.

En Colombia, la experiencia es funcionalmente más compleja. PretorlA fue desarrollada por la Corte Constitucional para apoyar el proceso de selección y clasificación de acciones de tutela, de modo que su naturaleza es eminentemente organizativa y auxiliar, aun cuando utilice técnicas avanzadas de procesamiento de información¹³. En cambio, PRISMA se proyecta sobre un terreno más sensible, pues fue concebida para estimar riesgo de reincidencia en el contexto de solicitudes de medidas de aseguramiento. Esa diferencia explica que, dentro de un mismo país, convivan herramientas de ordenación judicial con otras de perfil claramente predictivo. No obstante, mientras PretorlA cuenta con reconocimiento institucional público sobre su finalidad y funcionamiento general, en PRISMA la información accesible externamente no permite un escrutinio suficiente de variables, ponderaciones o auditorías independientes, lo que justifica hablar de una transparencia reducida.

En Brasil, el panorama revela un mayor grado de institucionalización. Victor, en el Supremo Tribunal Federal, fue presentado como una herramienta destinada a apoyar la clasificación y triagem de recursos, especialmente en materia de repercusión general, mientras que Sinapses fue formalizada por el Conselho Nacional de Justiça como una plataforma nacional de almacenamiento, entrenamiento y compartición de modelos de inteligencia artificial para el Poder Judicial. Más recientemente, la regulación brasileña ha reforzado de manera expresa exigencias de gobernanza, auditoría, monitoreo, prevención de sesgos y transparencia, lo que sitúa al caso brasileño como uno de los más estructurados de la región. Con todo, esa mayor institucionalización no elimina la heterogeneidad: el nivel de documentación pública y de trazabilidad sigue siendo desigual según la herramienta concreta, por lo que la transparencia debe valorarse como variable y no como homogénea.

En Chile, a diferencia de los casos anteriores, no se advierte en las fuentes primarias revisadas la existencia de un sistema judicial nacional consolidado comparable a Prometea, PretorlA o Sinapses. Lo que sí aparece con claridad es un escenario de discusión doctrinal y cautela institucional, especialmente en materias sensibles como la prisión preventiva, junto con referencias aisladas a exploraciones o usos localizados de herramientas de análisis de riesgo en ámbitos del Ministerio Público. Por ello, calificar a Chile como un caso experimental o en discusión constituye una inferencia prudente a partir del estado actual de la evidencia disponible, más que una afirmación sobre un despliegue nacional consolidado.

Desde esta perspectiva, el valor del cuadro no radica solo en identificar tecnologías, sino en mostrar que el verdadero contraste comparado se produce entre modelos de apoyo administrativo, sistemas de clasificación judicial y herramientas predictivas

¹³ GUÍO, Armando, Marco ético para la inteligencia artificial en Colombia, Colombia, Consejería Presidencial para Asuntos Económicos y Transformación Digital, 2020.

con incidencia penal. Esa diferencia es decisiva, porque los riesgos jurídicos no son los mismos cuando la inteligencia artificial agiliza trámites, cuando ayuda a priorizar expedientes o cuando introduce inferencias sobre conductas futuras. En consecuencia, la comparación regional confirma que el examen crítico de estos sistemas debe centrarse en su finalidad, en la posibilidad de control humano efectivo y en las garantías de transparencia, explicabilidad y revisión.

Argentina: Prometea y el diseño asistencial

Prometea es el caso más documentado y, en cierta medida, el más cauteloso en su diseño¹⁴. Surgió con un propósito acotado, agilizar tareas repetitivas y detectar expedientes prioritarios en áreas como salud o vivienda y su evolución posterior ha incorporado funciones de búsqueda de precedentes y automatización de actos procesales. Lo relevante es la lógica bajo la cual se integra: como herramienta de asistencia, con insistencia explícita en preservar la intervención humana como instancia decisoria final. Esa cautela de diseño es real y distingue a Prometea de sistemas predictivos penales como PRISMA. Sin embargo, la ausencia de marco normativo específico significa que esa cautela depende de decisiones institucionales que pueden cambiar sin necesidad de modificar ninguna norma.

Colombia: PRISMA y el salto a la predicción penal

Colombia es el caso más problemático del análisis. PRISMA no gestiona expedientes ni clasifica tutelas: estima la probabilidad de reincidencia de una persona para informar decisiones sobre medidas de aseguramiento¹⁵. Ese es un uso cualitativamente distinto al de los otros sistemas, porque introduce la lógica actuarial en la fase más sensible del proceso penal: aquella en que se decide si una persona permanece en libertad mientras se sustancia la causa.

Las críticas documentadas son convergentes: falta de transparencia en la ponderación de variables, dependencia de datos históricos que reflejan prácticas policiales y judiciales selectivas, imposibilidad práctica de impugnación técnica por parte de la defensa. La Sentencia T-529/2023 de la Corte Constitucional comenzó a delimitar el terreno, pero lo hizo en un caso concreto y sin establecer estándares generales vinculantes para todos los usos algorítmicos en decisiones públicas.

PretorIa, por contraste, opera en la jurisdicción constitucional para gestionar el volumen masivo de acciones de tutela. Su función es organizativa, no predictiva en sentido penal. La distinción importa: las consecuencias de un error en la clasificación de una tutela y las de un error en la predicción de reincidencia no son equivalentes.

Brasil: escala sin marco

Brasil destaca por la escala de su despliegue y por la existencia de una estrategia institucional de coordinación que no tiene equivalente en la región. Victor, Sinapses y GAIA no son proyectos aislados: forman parte de un ecosistema donde distintos tribunales

¹⁴ROA AVELLA, Marcela; SANABRIA-MOYANO, Jesús; DINAS-HURTADO, Katherin, "Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos", *Revista Brasileira de Direito Processual Penal*, vol. 8, n.º 1, 2022.

¹⁵VALENCIA GALLO, Lina María, et al., "Implicaciones éticas de la utilización de la inteligencia artificial (IA) en la administración de justicia de Colombia", *Revista IUS*, vol. 20, n.º 56, 2025, pp. 47-87.

comparten y entrenan modelos, con un nivel de integración tecnológica avanzado¹⁶. Eso es exactamente lo que hace más urgente la ausencia de regulación: cuando el despliegue es masivo y coordinado, los errores sistémicos, incluyendo los sesgos, también lo son.

La amplitud del despliegue brasileño no ha ido acompañada de mecanismos equivalentes de transparencia y rendición de cuentas. La documentación técnica pública de Victor es mayor que la de otros sistemas, pero la trazabilidad de las decisiones que informan y el control judicial sobre sus resultados permanecen sin respuesta normativa clara.

Chile: resistencia institucional como dato analítico

Chile es el único caso de los cuatro en que no existe un sistema de evaluación de riesgo penal desplegado a escala nacional. Lo que existe es algo más instructivo para el análisis comparado: un debate activo sobre si debe existir, con posiciones institucionales encontradas y con resistencia documentada desde la Defensoría Penal Pública.

La Defensoría ha sostenido con consistencia que el estándar del artículo 140 del Código Procesal Penal que exige fundar la prisión preventiva en antecedentes concretos que justifiquen la necesidad de cautela es incompatible con predicciones probabilísticas basadas en perfiles. Esa posición no es retórica: ha sido sostenida en audiencias concretas y ha generado pronunciamientos de tribunales de garantía que rechazaron argumentos fiscales basados en informes de riesgo no auditables. El caso chileno muestra que la resistencia institucional es posible, funciona, y requiere capacidad técnica para materializarse. También muestra que esa resistencia no puede depender indefinidamente de la voluntad institucional de la defensa: necesita anclaje normativo.

3.3 Contexto mexicano: influencia doctrinal sin despliegue documentado

Hoy por hoy, México no dispone de una regulación jurídica integral que discipline el uso de la inteligencia artificial dentro del sistema de justicia; el país opera, más bien, en un terreno fragmentario donde predominan principios constitucionales como la igualdad, no discriminación, debido proceso y esquemas de soft law que intentan contener riesgos sin ofrecer reglas técnicas precisas.

Esta ausencia de diseño normativo detallado no es excepcional en la región: como ocurre en otros países latinoamericanos, la producción legislativa avanza con rezago frente al desarrollo tecnológico, lo que deja sin respuesta cuestiones clave como la validez, control y límites del uso de algoritmos en decisiones judiciales. En contraste, México sí ocupa un lugar protagónico en el plano académico, donde universidades y centros de investigación han construido marcos teóricos influyentes sobre argumentación jurídica, transformación digital y sesgo algorítmico, con especial atención a sus efectos en contextos sensibles como la violencia de género¹⁷.

Ante el vacío normativo interno, el referente práctico tiende a desplazarse hacia estándares internacionales, especialmente el modelo europeo de regulación de IA, que clasifica los sistemas aplicados a la justicia como de alto riesgo y exige condiciones

¹⁶SILVA VIANNA, Ludmila da, "Entre algoritmos e sentenças: o papel da inteligência artificial na administração da justiça", *Inovações jurídicas e direito contemporâneo: reflexões sobre sociedade, economia e tecnologia*. [S. l.]: Editorial Científica Digital, 2025, pp. 59-71.

¹⁷DÍEZ SPELZ, Juan Francisco, "¿Robots con derechos? La frontera entre lo humano y lo no humano. Reflexiones desde la teoría de los derechos humanos", *Revista IUS*, vol. 15, n.º 48, 2021, pp. 259-287.

estrictas de transparencia, explicabilidad y supervisión humana. Sin embargo, la falta de reglas propias no es neutra: abre espacios a la opacidad algorítmica y a la protección de secretos comerciales que pueden bloquear el escrutinio público, debilitando la posibilidad de impugnar decisiones automatizadas y, en consecuencia, erosionando garantías procesales básicas.

De ahí que la discusión doctrinal insista en la urgencia de introducir salvaguardas concretas auditorías independientes, trazabilidad de datos, derecho a explicaciones comprensibles como parte de una reformulación del debido proceso en clave tecnológica. En este escenario, México se sostiene en un equilibrio inestable: sin ley específica.

Ante el vacío normativo interno, el referente práctico se desplaza hacia estándares internacionales, especialmente el modelo europeo de regulación de IA, que clasifica los sistemas aplicados a la justicia como de alto riesgo y exige condiciones estrictas de transparencia, explicabilidad y supervisión humana. La producción académica mexicana funciona por ahora como el principal mecanismo de orientación y crítica frente a una incorporación de IA que, cuando se produzca a escala, encontrará el mismo vacío normativo que en el resto de la región.

3.4 La influencia de COMPAS: trasplante sin controles

Compas (Correctional Offender Management Profiling for Alternative Sanctions) se ha convertido en el arquetipo de los sistemas de evaluación de riesgo penal. Su centralidad no se explica por su precisión técnica sino por su validación institucional en el sistema estadounidense y por su capacidad para fijar un estándar replicable: una herramienta que traduce información biográfica y delictiva en una puntuación usable en decisiones judiciales¹⁸.

State v. Loomis (Wisconsin Supreme Court, 2016) es el punto de inflexión¹⁹. El tribunal avaló el uso de COMPAS como herramienta de apoyo con tres límites declarados: naturaleza meramente orientativa del score, obligación de advertir al juzgador sobre las limitaciones técnicas, y preservación de la autonomía judicial. Esos límites son razonables en abstracto y vacíos en la práctica si no existen mecanismos institucionales que los hagan efectivos.

Existen estudios que han documentado que COMPAS clasificaba desproporcionadamente a personas afrodescendientes como de alto riesgo en comparación con personas blancas con trayectorias similares. Los autores DRESSEL y FARID demostraron que la precisión predictiva del sistema no superaba la de personas sin formación especializada²⁰. Chouldechova mostró que las distintas métricas de equidad son matemáticamente incompatibles entre sí: un sistema no puede ser simultáneamente equitativo en todas las dimensiones relevantes²¹.

Lo que el análisis técnico de COMPAS deja al descubierto no es un error de calibración corregible²². Es el resultado lógico de entrenar un modelo con datos históricos produ-

¹⁸ANGWIN, Julia, et al., *Machine Bias...*, cit., p. 6.

¹⁹TRIBUNAL SUPREMO DE WISCONSIN. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

²⁰DRESSEL, Julia; FARID, Hany, *The accuracy, fairness, and limits of predicting...* cit., p. 45.

²¹CHOUDECHOVA, Alexandra, *Fair prediction with disparate impact: a study of bias in recidivism...*, cit., p. 153.

²²RUIZ PRADILLO, Juan Carlos, *La investigación del delito en la era digital. Los derechos fundamentales frente a las nuevas medidas tecnológicas de investigación*, Madrid, Fundación Alternativas, 2013.

cidos por un sistema de justicia penal con sesgos estructurales documentados. El sistema aprendió exactamente lo que le enseñaron.

El trasplante de esa lógica a Colombia materializado en PRISMA introduce una capa adicional de riesgo. Los sesgos estructurales de los sistemas judiciales latinoamericanos por clase, etnia, territorio son al menos tan pronunciados como los del sistema estadounidense, y los marcos de control son más débiles. Importar el modelo sin importar también los mecanismos de auditoría y control que incluso en Estados Unidos resultaron insuficientes es una decisión institucional que requiere justificación.

4. SESGOS ALGORÍTMICOS Y REPRODUCCIÓN DE DESIGUALDADES ESTRUCTURALES

4.1 Herencia estructural del sesgo: bias in, bias out

Los modelos de aprendizaje automático se entrenan con datos históricos y, en ese proceso, incorporan las regularidades que esos datos contienen: tanto decisiones correctas como prácticas problemáticas. La inteligencia artificial no actúa como corrector externo del sistema. Actúa como su heredero. Hereda sus sesgos con la misma fidelidad con que hereda sus aciertos, y los reproduce con una consistencia estadística que los hace más difíciles de cuestionar²³.

El mecanismo tiene nombre en la literatura especializada: biased training data. Cuando los datos de entrenamiento están marcados por prácticas policiales selectivas, decisiones judiciales inconsistentes, o subregistro estructural de ciertos delitos, el sistema aprende exactamente esas mismas relaciones. No distingue entre correlación estadística y justicia sustantiva. Decisiones históricamente arbitrarias o discriminatorias se convierten en insumos legítimos para predicciones futuras. La desigualdad se estabiliza bajo apariencia de coherencia estadística.

En el ámbito latinoamericano esto tiene una dimensión específica. Los sistemas de justicia penal en la región presentan sesgos documentados por clase social, etnia y territorio que son anteriores y más profundos que los que COMPAS reprodujo en el contexto estadounidense. Entrenar un modelo predictivo con datos producidos por esos sistemas no neutraliza los sesgos: los codifica. La eficiencia del algoritmo se convierte en eficiencia para reproducir la desigualdad preexistente.

4.2 Discriminación estructural por clase, etnia y territorio

El sesgo en los modelos algorítmicos no requiere incluir explícitamente variables como raza o clase para producir efectos discriminatorios. Les basta con incluir variables que se correlacionen con ellas: el entorno residencial, la trayectoria laboral, el nivel educativo, la estabilidad familiar. Esas variables funcionan como proxies. El resultado es discriminación que no se declara pero que se materializa en sus efectos. Chouldechova formalizó matemáticamente que las distintas métricas de equidad son incompatibles entre sí optimizar una empeora necesariamente otra²⁴.

El sesgo racial constituye el caso empírico más robusto en la literatura sobre evaluación algorítmica del riesgo, siendo el sistema COMPAS su exponente paradigmático:

²³FURINGO, Carla Irene, "La inteligencia artificial: ¿innovación o amenaza para las garantías procesales en la justicia penal?", *Revista de Derecho de la UNED*, n.º 36, 2026, pp. 101-132.

²⁴CHOUDECHOVA, Alexandra, *Fair prediction with disparate impact: a study of bias in recidivism...*, cit., p. 153.

evidencia independiente mostró que, ante historiales comparables, clasifica a personas afroamericanas como de alto riesgo con una tasa de falsos positivos sustancialmente superior a la de personas blancas, mientras que invierte el patrón en los falsos negativos, favoreciendo a estas últimas al infraestimar su probabilidad de reincidencia²⁵. Este comportamiento no se explica por la inclusión explícita de la variable racial formalmente ausente, sino por la explotación de variables proxy (residencia, entorno familiar, trayectorias institucionales) que permiten reconstruir correlaciones étnicas y reproducir, en clave estadística, desigualdades estructurales preexistentes.

En paralelo, los modelos incorporan un sesgo de clase al traducir indicadores de vulnerabilidad desempleo, baja escolarización, inestabilidad laboral o precariedad financiera en señales de peligrosidad o incumplimiento, de modo que la pobreza deviene un factor de riesgo autónomo: no se penaliza el hecho, sino la posición socioeconómica del sujeto, con efectos que se extienden más allá del ámbito penal hacia sistemas de scoring crediticio que restringen el acceso a bienes básicos.

A ello se suma un sesgo territorial que convierte el espacio en variable incriminatoria; el uso del código postal o de zonas catalogadas como de alta criminalidad introduce una imputación colectiva que afecta incluso a individuos sin antecedentes, y en el caso de la vigilancia predictiva genera bucles de retroalimentación: la intensificación policial en determinados barrios produce más registros y detenciones, reforzando los datos que el propio sistema interpreta como validación de su hipótesis inicial.

Este problema se agrava cuando modelos entrenados en contextos específicos se transfieren sin adaptación a otros entornos socioculturales, exportando así patrones discriminatorios desanclados de la realidad local. El resultado, en términos jurídicos, es un desplazamiento desde un derecho penal centrado en el acto hacia uno orientado al autor estadístico, donde la imputación descansa en perfiles probabilísticos más que en conductas probadas, erosionando principios nucleares como la presunción de inocencia y la igualdad ante la ley; de ahí que la respuesta doctrinal insista en auditorías independientes, trazabilidad de los modelos y mecanismos efectivos de supervisión humana capaces de intervenir sobre decisiones automatizadas con potencial discriminatorio²⁶.

La dimensión territorial añade la dinámica más problemática: los bucles de retroalimentación. La vigilancia predictiva identifica determinadas zonas como de alto riesgo a partir de datos previos; eso orienta la asignación de recursos policiales hacia esos espacios; eso aumenta la probabilidad de registrar incidentes en ellos; eso refuerza la hipótesis inicial del algoritmo. Con el tiempo, el modelo construye una imagen de la realidad que es, en buena medida, producto de su propia intervención. Eso no es predicción: es profecía autocumplida.

El problema tiene un efecto adicional que la literatura señala con menos frecuencia: desplaza el eje del juicio penal. Cuando la decisión se apoya en perfiles contruidos a partir del pasado, se diluye la centralidad del hecho concreto y reaparece, por vía técnica, una lógica próxima al derecho penal de autor: se valora al individuo por su pertenencia o su entorno más que por su conducta específica.

²⁵SILVA VIANNA, Ludmila da, Entre algoritmos e sentenças: o papel da inteligência artificial...cit., p. 63.

²⁶Saavedra Vera, et al., "La incidencia del sesgo algorítmico en la justicia predictiva del sistema judicial", Tzhoeoen, vol. 15, n.º 2, 2023, pp. 79-97; REBOLLO DELGADO, Lucrecio, Inteligencia artificial y derechos fundamentales, Madrid, Dykinson, 2023.

4.3 El círculo vicioso institucionalizador

El mecanismo completo funciona así: datos históricos con sesgos alimentan el modelo; el modelo produce predicciones que informan decisiones; esas decisiones generan nuevos datos que confirman las predicciones iniciales; esos datos se incorporan al siguiente ciclo de entrenamiento. La discriminación sistémica previa se convierte en patrón estadístico; el patrón estadístico se convierte en criterio de decisión; el criterio de decisión produce resultados que parecen confirmar que el patrón era correcto.

Lo que hace especialmente difícil interrumpir este círculo es que sus resultados tienen apariencia empírica²⁷. No son opiniones ni prejuicios: son correlaciones estadísticas respaldadas por datos reales. Una discriminación que se presenta como opinión puede impugnarse argumentativamente. Una discriminación que se presenta como estadística requiere una refutación técnica que la mayoría de los defensores no está en condiciones de producir, especialmente cuando el sistema es opaco.

5. GARANTÍAS PROCESALES BAJO TENSIÓN

5.1 El derecho a la motivación: delegación encubierta

La motivación de las resoluciones judiciales no es una formalidad externa: es el mecanismo por el cual el acto jurisdiccional se distingue del acto de poder. Una resolución motivada puede ser revisada, impugnada, corregida. Una resolución que delega su fundamento en un sistema opaco no puede ser ninguna de esas cosas.

Este deber, positivizado en preceptos como el artículo 120.3 de la Constitución española y el artículo 93, IX de la Constitución brasileña, impone al órgano jurisdiccional la carga de hacer explícita la arquitectura racional de su decisión, esto es, de exponer de manera verificable la conexión entre los hechos acreditados, las normas aplicadas y las inferencias que conducen al fallo, de modo que la resolución no se agote en una conclusión autoritativa sino que revele los criterios jurídicos que la hacen inteligible, controlable y susceptible de contradicción.

La incorporación de sistemas algorítmicos de caja negra produce lo que puede denominarse motivación delegada: la resolución menciona el resultado del sistema como si fuera un dato y construye sobre él sin poder explicar cómo se produjo. El juez no puede hacer explícito un itinerario lógico que desconoce. La consecuencia no es una motivación defectuosa: es la ausencia de motivación disfrazada de tecnología.

La doctrina ha propuesto una motivación reforzada cuando se incorporan herramientas algorítmicas: el juez debe explicitar la incidencia del resultado en su convencimiento²⁸. Esa propuesta es correcta pero insuficiente si el sistema es opaco. No se puede motivar lo que no se puede conocer. La solución no está solo en las exigencias que se imponen al juez, sino en las exigencias que se imponen al sistema antes de que el juez lo reciba. La motivación reforzada solo funciona si el sistema que informa la decisión puede explicarse.

²⁷MORÁN ESPINOSA, Alejandra, "Responsabilidad penal de la inteligencia artificial (IA). ¿La próxima frontera?", Revista IUS, vol. 15, n.º 48, 2021, pp. 290-323; MARTÍN FABA, José María, "La información sobre la lógica aplicada en el credit scoring automatizado: fácil, concisa y sin tecnicismos", Revista CESCO de Derecho de Consumo, n.º 54, 2025, pp. 15-30.

²⁸COTINO HUESO, Lorenzo, "La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial", Diario La Ley, n.º 80, 2024, pp. 5-38.

5.2 La contradicción ante una puntuación inimpugnable

El principio de contradicción exige que las partes puedan acceder, comprender y cuestionar todos los elementos que sustentan la decisión judicial. Cuando uno de esos elementos es una puntuación producida por un sistema propietario cuya lógica está protegida como secreto comercial, esa exigencia se vuelve inoperante en la práctica.

La defensa se encuentra ante una asimetría epistémica de primer orden: sabe que existe una puntuación, conoce su resultado, y no puede hacer nada más. No puede reconstruir el razonamiento que la produjo, ni identificar qué variables influyeron en ella, ni siquiera verificar si los datos de entrada eran correctos. Impugnar esa puntuación requeriría acceso a información que el secreto comercial bloquea.

State v. Loomis ilustra el problema. El tribunal admitió el uso de COMPAS sin revelar íntegramente los factores que sustentaban la puntuación. La defensa pudo conocer el resultado, pero no impugnar su fundamento. Eso no es contradicción: es acceso a la conclusión sin acceso a las premisas. En términos procesales, equivale a presentar una pericia cuyo método no puede ser examinado.

La jurisprudencia del TJUE en SCHUFA Holding (C-634/21, 2023) y Dun & Bradstreet (C-203/22, 2024) establece que el interesado tiene derecho a conocer la lógica aplicada en los sistemas de puntuación automatizada, y que ese derecho no cede frente a la protección de secretos comerciales cuando están en juego garantías fundamentales. Ese estándar no tiene equivalente jurisprudencial consolidado en América Latina.

5.3 La presunción de inocencia y el etiquetado predictivo

La presunción de inocencia impone que toda persona sea tratada como inocente hasta condena firme. Los sistemas de etiquetado predictivo de riesgo no condenan: clasifican. Esa distinción formal no resuelve el problema sustantivo. Una clasificación de alto riesgo en el contexto de una decisión cautelar produce efectos similares a los de una presunción de culpabilidad: aumenta la probabilidad de que se imponga prisión preventiva, lo que en la práctica latinoamericana puede significar meses o años de privación de libertad antes de que exista siquiera una sentencia.

El problema no es que los sistemas predigan mal, aunque en muchos casos lo hagan. El problema es que predecir no es una forma legítima de fundar una restricción de libertad en un Estado de Derecho. El proceso penal de acto exige que la decisión sobre la libertad de una persona se apoye en hechos concretos probados, no en inferencias estadísticas sobre lo que podría hacer en el futuro. Ese principio no admite excepción tecnológica.

Las variables que estos sistemas utilizan (nivel educativo, estabilidad laboral, entorno territorial) no acreditan riesgo concreto: reproducen desigualdad estructural. Una persona pobre, con empleo precario y residente en una zona estigmatizada obtendrá una puntuación de riesgo más alta que alguien con características individuales similares pero distinta posición socioeconómica. Eso no es predicción: es penalización automatizada de la vulnerabilidad.

6. EL POSITIVISMO TECNIFICADO COMO MARCO EXPLICATIVO COMO MARCO EXPLICATIVO

6.1 Genealogía del positivismo tecnificado

La literatura sobre algoritmos y derecho ha descrito con precisión los mecanismos técnicos del sesgo, la opacidad y la discriminación automatizada. Lo que ha abordado con menor profundidad es el fenómeno epistemológico que los sostiene: la sustitución de la autoridad argumentativa por la autoridad del dato. A ese fenómeno lo denominamos aquí positivismo tecnificado²⁹.

La denominación no es arbitraria. El positivismo jurídico clásico planteó que el derecho es lo que las normas dicen, con independencia de su contenido moral. Kelsen separó la validez de la justicia; Hart separó la existencia del derecho de su mérito. El positivismo tecnificado da un paso adicional: la decisión es lo que los datos indican, con independencia de la argumentación jurídica. La diferencia no es filosófica en sentido abstracto: tiene consecuencias institucionales directas.

La distinción entre esta categoría y conceptos próximos disponibles en la literatura merece precisión. La tecnocracia judicial describe una orientación institucional hacia la gestión técnica de la justicia, pero no captura el desplazamiento epistemológico específico que aquí se analiza. Delegación algorítmica describe un fenómeno de transferencia de decisiones, pero permanece en el plano descriptivo. Automatización decisional captura el aspecto procedimental pero no el ontológico: la transformación de qué cuenta como razón para decidir. El positivismo tecnificado designa precisamente ese desplazamiento: la reconfiguración de la racionalidad judicial de modo que la autoridad del dato estadístico sustituye a la justificación argumentativa como fundamento de legitimidad.

6.2 El mecanismo institucional

En el contexto judicial, el positivismo tecnificado se manifiesta cuando el juez acepta el resultado del sistema algorítmico no porque haya examinado y validado su razonamiento, sino porque confía en la autoridad del dato cuantificado. El automation bias, documentado en la psicología cognitiva y en la literatura sobre sistemas de apoyo a decisiones, describe exactamente esa dinámica: la tendencia a aceptar acríticamente los outputs de sistemas automatizados, especialmente en contextos de sobrecarga de trabajo o presión institucional. En jurisdicciones latinoamericanas con altas tasas de litigiosidad y déficits estructurales de recursos, esa presión es la norma.

El positivismo tecnificado no implica mala fe del juez. Implica algo más preocupante: una transformación silenciosa de la racionalidad judicial que no requiere ninguna decisión explícita para producirse. El juez sigue siendo formalmente el decisor; materialmente, el centro de gravedad de la decisión se ha desplazado hacia el algoritmo. La legitimidad del acto jurisdiccional descansa entonces en la autoridad del sistema, no en el razonamiento del juez. Y la autoridad del sistema descansa, a su vez, en datos que nadie examina, producidos por instituciones cuyos sesgos nadie ha corregido.

Romper esa lógica requiere algo más que regulación técnica. Requiere reafirmar que la decisión judicial es un acto de argumentación en el sentido de MacCormick, de

²⁹ESPÍN ALBA, Isabel, "Protección de datos, inteligencia artificial, score crediticio: últimos movimientos judiciales y normativos", *Revista Electrónica de Direito*, vol. 35, n.º 3, 2024, pp. 2-6; BORBÓN, Diego, "Neuroética y neuroderechos ante las propuestas de neurointervenciones coercitivas con fines de mejora moral en ámbitos penales", *Cuadernos Electrónicos de Filosofía del Derecho*, n.º 53, 2025, pp. 56-80.

Alexy, de Atienza, no de cálculo; que su legitimidad no deriva de su eficiencia sino de su justificabilidad pública; y que ningún sistema algorítmico, por preciso que sea, puede sustituir la responsabilidad del juez de razonar sobre por qué una persona merece ser privada de su libertad³⁰.

El caso *State v. Loomis* opera como punto de inflexión en la incorporación de sistemas algorítmicos al proceso penal: no inaugura solo una técnica, sino una forma de racionalidad decisional que tensiona los límites clásicos del debido proceso. En ese expediente, Eric Loomis fue condenado en Wisconsin con apoyo de un informe generado por COMPAS, herramienta que, a partir de un cuestionario extenso y datos históricos, produce una puntuación de riesgo de reincidencia; la defensa objetó su uso por la opacidad del modelo (blindado por secreto comercial), su potencial discriminatorio y su incompatibilidad con una sentencia verdaderamente individualizada.

La Wisconsin Supreme Court desestimó la impugnación, pero lo hizo bajo un esquema condicional: el output algorítmico no puede erigirse en fundamento exclusivo, debe ser tratado como insumo auxiliar, el juzgador ha de ser advertido de sus limitaciones empíricas y sesgos potenciales, y, sobre todo, debe preservar una capacidad crítica autónoma frente al resultado automatizado. Sin embargo, esta validación pragmática no neutraliza los déficits estructurales que el propio caso visibiliza: la opacidad impide un contradictorio efectivo al vedar el examen de la lógica interna del sistema; la evidencia empírica posterior, como la difundida por ProPublica, mostró patrones sistemáticos de sesgo racial en la clasificación del riesgo; y, en un plano más profundo, el modelo desplaza el eje de imputación desde el hecho probado hacia la pertenencia estadística, erosionando la individualización de la pena y aproximando el sistema a una lógica actuarial. Loomis, así, no legitima sin más la justicia algorítmica: la expone, en sus propias condiciones de posibilidad, como un dispositivo útil pero normativamente inestable, cuya adopción sin estándares de explicabilidad y control público amenaza con sustituir la deliberación judicial por una autoridad técnica difícilmente escrutable.

7. HACIA UNA REGULACIÓN DE LA TRANSPARENCIA ALGORÍTMICA JUDICIAL: REFERENTES Y GARANTÍAS PROCESALES

El Reglamento (UE) 2024/1689 ofrece la referencia normativa más desarrollada disponible. Clasifica como sistemas de alto riesgo los empleados en la administración de justicia y en la predicción de reincidencia, y les impone obligaciones reforzadas: transparencia funcional (artículo 13), supervisión humana significativa (artículo 14) y documentación técnica suficiente para auditoría³¹. La jurisprudencia del TJUE en *SCHUFA* y *Dun & Bradstreet* establece que el derecho a conocer la lógica aplicada en sistemas de puntuación automatizada no cede frente al secreto comercial cuando están en juego garantías procesales fundamentales.

³⁰ATIENZA, Manuel, *Las razones del Derecho. Teorías de la argumentación jurídica*, México, Universidad Nacional Autónoma de México, 2005.

³¹PÉREZ-UGENA, María, "Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU., China e Iberoamérica", *Anuario Iberoamericano de Justicia Constitucional*, vol. 28, n.º 1, 2024, pp. 129-156; DÍAZ SENÉS, Juan Pedro, "Panorámica internacional de la regulación de la inteligencia artificial: Reglamento europeo de inteligencia artificial (AI Act)", *Revista CESCO de Derecho de Consumo*, n.º 49, 2024, pp. 117-130.

Ese estándar europeo es útil como referencia para América Latina, pero no es directamente trasplantable. La región necesita un marco adaptado a sus condiciones institucionales: sistemas de justicia con mayor carga litigiosa, déficits estructurales de recursos, menor capacidad técnica instalada en los operadores jurídicos, y presencia dominante de proveedores privados con escasa disposición a la transparencia. Un marco regulatorio que copie las obligaciones europeas sin construir las instituciones que las hagan efectivas desplaza el problema sin resolverlo.

Independientemente del modelo regulatorio que cada jurisdicción adopte, hay tres estándares sin los cuales ningún sistema algorítmico debería utilizarse en una decisión penal.

Primero, explicabilidad. Todo sistema que informe una decisión penal debe poder explicar a la defensa, en términos comprensibles y verificables, qué variables consideró, cómo las ponderó y por qué llegó al resultado que llegó. La protección de la propiedad intelectual no puede operar como límite absoluto frente al derecho de defensa. Cuando esa información sea necesaria para impugnar una decisión, debe ser accesible. El mecanismo puede ser la mediación de una autoridad judicial o de control como el modelo europeo establece, pero la información debe llegar.

Segundo, auditabilidad. Todo sistema de evaluación de riesgo penal debe ser objeto de auditoría independiente periódica, con acceso al código fuente, a los datos de entrenamiento y a los criterios de ponderación. Esa auditoría debe incluir evaluación de sesgos y debe ser pública en sus conclusiones, aunque no necesariamente en sus detalles técnicos protegibles. Los resultados de la auditoría deben estar disponibles para la defensa en cualquier proceso en que el sistema haya informado una decisión.

Tercero, supervisión humana efectiva. El juez no puede ser un validador formal de resultados algorítmicos. Debe poder confirmar, matizar o rechazar motivadamente la recomendación del sistema, y esa capacidad requiere que el sistema sea comprensible y que el juez tenga formación suficiente para ejercer ese control de manera sustantiva. Cuando el juez incorpore una recomendación algorítmica a su decisión, debe motivar específicamente por qué la considera fiable y pertinente en el caso concreto. La referencia genérica al resultado del sistema no es motivación suficiente.

La auditoría algorítmica no es un mecanismo de mejora técnica: es una condición de legitimidad procesal. Su función en el proceso judicial es análoga a la cadena de custodia de la prueba material: garantiza que lo que llega al juez es lo que dice ser y que ha llegado de manera verificable.

La auditoría debe ser continua e integrada en todo el ciclo de vida del sistema. La auditoría *ex ante* evalúa el diseño antes del despliegue. La auditoría *ex post* periódica detecta desviaciones durante la operación. La auditoría extraordinaria, activable a instancia de parte cuando existan indicios fundados de fallos o sesgos, es la que más directamente protege el derecho de defensa.

Ninguna de estas auditorías es efectiva sin acceso a los datos de entrenamiento, al código y a los criterios de ponderación. Eso entra en tensión con los intereses comerciales de los proveedores. Esa tensión debe resolverse en favor del derecho de defensa, no del secreto empresarial, cuando los sistemas se usan en decisiones que afectan la libertad de personas. Los sistemas que no puedan ser auditados en esas condiciones no deberían utilizarse en el ámbito penal.

8. CONCLUSIONES

Primera: Lo que este trabajo ha documentado no es un problema técnico que la ingeniería pueda resolver ni un problema normativo que la regulación pueda contener por sí sola. Es un problema de racionalidad: la racionalidad actuarial y la racionalidad del proceso penal responden a lógicas diferentes, y esa diferencia no se resuelve diseñando mejores algoritmos.

Segunda: Los sistemas de evaluación de riesgo penal en América Latina comparten un conjunto de rasgos que los hacen estructuralmente incompatibles con las garantías que los propios ordenamientos de la región reconocen. Operan con opacidad suficiente para impedir la motivación real y la contradicción efectiva. Se entrenan con datos que reflejan desigualdades históricas y las reproducen con apariencia estadística de objetividad. Se despliegan en ausencia de marcos normativos que establezcan condiciones mínimas de legitimidad. Y funcionan en contextos institucionales donde la sobrecarga judicial hace probable el automation bias.

Tercera: El concepto de positivismo tecnificado describe el fenómeno resultante: la sustitución de la argumentación jurídica por la autoridad del dato, que transforma silenciosamente la racionalidad de la decisión judicial sin necesidad de ninguna decisión explícita. Ese desplazamiento se produce como resultado lógico de la interacción entre sistemas opacos, operadores sin formación técnica suficiente para cuestionarlos, y condiciones institucionales que incentivan la eficiencia por encima de la deliberación.

Cuarta: Las conclusiones del análisis comparado son coherentes con esa diagnosis. Argentina ha construido el caso más cauteloso, con un diseño asistencial que preserva formalmente la intervención humana, pero sin marco normativo que garantice que esa cautela no cambie. Colombia ha desplegado el caso más preocupante PRISMA con opacidad documentada, sin auditabilidad pública y sin respuesta regulatoria estructural tras la Sentencia T-529/2023. Brasil ha apostado por la escala sin construir los mecanismos de control equivalentes. Chile demuestra que la resistencia institucional es posible y tiene efectos reales, pero no es sustituto de regulación.

Quinta: Las propuestas de este trabajo son las condiciones mínimas sin las cuales la incorporación de sistemas algorítmicos en la justicia penal no puede considerarse compatible con los principios que esos sistemas dicen proteger. Explicabilidad, auditabilidad y supervisión humana efectiva no son aspiraciones técnicas avanzadas: son traducciones al lenguaje del diseño tecnológico de exigencias que el derecho constitucional ya establece. La cuestión de fondo no es tecnológica: es si los sistemas de justicia latinoamericanos van a utilizar la tecnología para corregir sus inercias o para darles una apariencia de objetividad que las haga más difíciles de ver y más costosas de combatir. La respuesta a esa pregunta no está en los algoritmos: está en las decisiones políticas e institucionales que determinan cómo, con qué garantías y con qué rendición de cuentas esos algoritmos se usan.

9. REFERENCIAS BIBLIOGRÁFICAS

ANGWIN, Julia, et al., "Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks", ProPública, [Consultada 23 de mayo de 2016]. [Disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>].

- ATIENZA, Manuel, *Las razones del Derecho. Teorías de la argumentación jurídica*, México, Universidad Nacional Autónoma de México, 2005.
- BARONA VILAR, Silvia, "Algoritmización de la prueba y de la decisión judicial en el proceso penal: ¿utopía o distopía?" En Pillado González, E., ed. *El proceso penal ante una nueva realidad tecnológica europea*, Pamplona, Aranzadi, 2023. [Consultada 23 de abril de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=8857602>].
- BARONA VILAR, Silvia, "Dataización de la justicia (algoritmos, inteligencia artificial y justicia, ¿el comienzo de una gran amistad?)", *Revista Boliviana de Derecho*, n.º 36, 2023, pp. 14-45. [Consultada 20 de abril de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9043836>].
- BARONA VILAR, Silvia, "Justicia penal líquida (desde la mirada de Bauman)", *Teoría y Derecho. Revista de pensamiento crítico*, n.º 22, 2017. [Consultada 20 de marzo de 2026]. [Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3510764].
- BARROS, Filipe, et al., "Machine learning algorithms applied on classification of processes for conciliation on Brazilian labour judiciary", *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 20.º, Belo Horizonte. Belo Horizonte: Sociedade Brasileira de Computação, 2023, pp. 389-402. [Consultada 19 de abril de 2026]. [Disponible en: <https://sol.sbc.org.br/index.php/eniac/article/view/25716>].
- BELLOSO MARTÍN, Nuria, "Algoritmos predictivos al servicio de la justicia: ¿una nueva forma de minimizar el riesgo y la incertidumbre?", *Revista da Faculdade Mineira de Direito*, vol. 22, n.º 43, 2019, pp. 2-31. [Consultada 2 de marzo de 2026]. [Disponible en: <https://periodicos.pucminas.br/Direito/article/view/20780>].
- BORBÓN, Diego, "Neuroética y neuroderechos ante las propuestas de neurointervenciones coercitivas con fines de mejora moral en ámbitos penales", *Cuadernos Electrónicos de Filosofía del Derecho*, n.º 53, 2025. [Consultada 20 de marzo de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=10214876>].
- CHOULDCHOVA, Alexandra, "Fair prediction with disparate impact: a study of bias in recidivism prediction instruments", *Big Data*, vol. 5, n.º 2, 2017, pp. 153-163. [Consultada 14 de febrero de 2026]. [Disponible en: <https://n9.cl/kl1ln0>].
- CONCHA CAMACHO, William, "El derecho a la transparencia algorítmica como estándar ético en una democracia liberal", *Revista Jurídica Austral*, vol. 5, n.º 1, 2024, pp. 529-550. [Consultada 25 de abril de 2026]. [Disponible en: <https://ojs.austral.edu.ar/juridicaaustral/article/view/1239>].
- CORVALÁN, Juan Gustavo, "Inteligencia artificial: retos, desafíos y oportunidades. Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la justicia", *Revista de Investigações Constitucionais*, vol. 5, n.º 1, 2018, pp. 55-89. [Consultada 15 de abril de 2026]. [Disponible en: <https://n9.cl/ua5s13>].
- COTINO HUESO, Lorenzo, "La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial", *Diario La Ley*, 2024, n.º 80. [Consultada 20 de marzo de 2026]. [Disponible en: <https://www.cotino.es/el-primer-estudio-sobre-la-primer-sentencia-del-tjue-sobre-decisiones-automatizadas/>].
- CRUZ GONZÁLEZ, Marcos, "La inteligencia artificial y los derechos humanos. Una carrera tecnológica entre Estados Unidos, China y la Unión Europea", *Revista IUS*, vol. 20, n.º 56, 2025, pp. 31-51. [Consultada 27 de marzo de 2026]. [Disponible en: <https://www.revistaius.com/index.php/ius/article/view/1059>].
- DE GREGORIO, Giovanni, "From constitutional freedoms to the power of the platforms: protecting fundamental rights online in the algorithmic society", *European Journal of Legal Studies*, vol. 11, n.º 2, 2019, pp. 65-104. [Consultada 5 de marzo de 2026]. [Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3365106].
- DÍAZ SENÉS, Juan Pedro, "Panorámica internacional de la regulación de la inteligencia artificial: Reglamento europeo de inteligencia artificial (AI Act)", *Revista CESCO de Derecho de Consumo*, n.º 49,

- 2024, pp. 117-130. [Consultada 25 de abril de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9462903>].
- DÍEZ SPELZ, Juan Francisco, "¿Robots con derechos? La frontera entre lo humano y lo no humano. Reflexiones desde la teoría de los derechos humanos", *Revista IUS*, vol. 15, n.º 48, 2021, pp. 259-287. [Consultada 25 de abril de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=8953275>].
- DRESSEL, Julia; FARID, Hany, "The accuracy, fairness, and limits of predicting recidivism", *Science Advances*, vol. 4, n.º 1, 2018, pp. 1-56. [Consultada 27 de marzo de 2026]. [Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29376122>].
- ESPÍN ALBA, Isabel, "Protección de datos, inteligencia artificial, score crediticio: últimos movimientos judiciales y normativos", *Revista Electrónica de Direito*, vol. 35, n.º 3, 2024, pp. 2-6. [Consultada 26 de abril de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9773689>].
- FURINGO, Carla Irene, "La inteligencia artificial: ¿innovación o amenaza para las garantías procesales en la justicia penal?", *Revista de Derecho de la UNED*, n.º 36, 2026, pp. 101-132. [Consultada 23 de abril de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=10611408>].
- GARCÍA, Manuel Lorenzo, "La inteligencia artificial predictiva al servicio de la prevención e investigación del delito y del proceso penal", *Ciencia Policial*, vol. 183, 2024, n.º 1, pp. 91-132. [Consultada 20 de marzo de 2026]. [Disponible en: <https://n9.cl/g2x93>].
- GIMÉNEZ, María, "Inteligencia artificial y derecho penal", *Actualidad Jurídica Iberoamericana*, n.º 10, 2019, pp. 792-843. [Consultada 6 de marzo de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6978830>].
- GUÍO, Armando, Marco ético para la inteligencia artificial en Colombia, Colombia, Consejería Presidencial para Asuntos Económicos y Transformación Digital, 2020. [Consultada 25 de marzo de 2026]. [Disponible en: <https://www.usergioarboleda.edu.co/wp-content/uploads/2021/11/Marco-etico-para-la-inteligencia-artificial-en-Colombia-Maestria-en-Inteligencia-artificial.pdf>].
- MARTÍN FABA, José María, "La información sobre la lógica aplicada en el credit scoring automatizado: fácil, concisa y sin tecnicismos", *Revista CESCO de Derecho de Consumo*, n.º 54, 2025, pp. 15-30. [Consultada 20 de marzo de 2026]. [Disponible en: <https://n9.cl/jn1917>].
- MEDINA GUERRERO, Manuel, "El derecho a conocer los algoritmos utilizados en la toma de decisiones: aproximación desde la perspectiva del derecho fundamental a la protección de datos", *Teoría y Realidad Constitucional*, n.º 49, 2022, pp. 55-74. [Consultada 20 de marzo de 2026]. [Disponible en: <https://revistas.uned.es/index.php/TRC/article/view/33847>].
- MORÁN ESPINOSA, Alejandra, "Responsabilidad penal de la inteligencia artificial (IA). ¿La próxima frontera?", *Revista IUS*, vol. 15, n.º 48, 2021, pp. 290-323. [Consultada 20 de marzo de 2026]. [Disponible en: <https://revistaius.com/index.php/ius/article/view/706>].
- PARRA SEPÚLVEDA, Darío y CONCHA MACHUCA, Ricardo, "Inteligencia artificial y derecho. Problemas, desafíos y oportunidades", *Revista de Derecho*, vol. 70, 2021, pp. 1-25. [Consultada 25 de abril de 2026]. [Disponible en: <https://revistas.javeriana.edu.co/index.php/vnijuri/article/view/35364>].
- PASQUALE, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015. [Consultada 20 de marzo de 2026]. [Disponible en: <https://n9.cl/47qzjg>].
- PÉREZ-UGENA, María, "Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU., China e Iberoamérica". *Anuario Iberoamericano de Justicia Constitucional*, vol. 28, n.º 1, 2024, pp. 129-156. [Consultada 20 de marzo de 2026]. [Disponible en: <https://recyt.fecyt.es/index.php/AJJC/article/view/108247>].
- REBOLLO DELGADO, Lucrecia, *Inteligencia artificial y derechos fundamentales*, Madrid, Dykinson, 2023.
- ROA AVELLA, Marcela; SANABRIA-MOYANO, Jesús; DINAS-HURTADO, Katherin, "Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos", *Revista Brasileira de Direito Processual Penal*, vol. 8, n.º 1, 2022. [Consultada 26 de marzo de 2026]. [Disponible en: <https://revista.ibraspp.com.br/RBDPP/article/view/615>].

- RUIZ PRADILLO, Juan Carlos, *La investigación del delito en la era digital. Los derechos fundamentales frente a las nuevas medidas tecnológicas de investigación*. Madrid, Fundación Alternativas, 2013. [Consultada 20 de marzo de 2026]. [Disponible en: <https://n9.cl/wo2vj>].
- RUIZ RAJO, Carlos Enrique, "Entre algoritmos y garantías: una revisión crítica sobre la inteligencia artificial en el proceso penal", *Revista de Investigación Científica Erga Omnes*, vol. 5, n.º 2, 2025, pp. 1-8. [Consultada 20 de marzo de 2026]. [Disponible en: <https://revistas.unh.edu.pe/index.php/rceo/article/view/763>].
- SAAVEDRA VERA, César Orlando, et al., "La incidencia del sesgo algorítmico en la justicia predictiva del sistema judicial". *Tzhoeoen*, vol. 15, n.º 2, 2023, pp. 79-97. [Consultada 25 de marzo de 2026]. [Disponible en: <https://revistas.uss.edu.pe/index.php/tzh/article/view/2592>].
- SILVA VIANNA, Ludmila da, "Entre algoritmos e sentenças: o papel da inteligência artificial na administração da justiça", *Inovações jurídicas e direito contemporâneo: reflexões sobre sociedade, economia e tecnologia*. [S. l.]: Editorial Científica Digital, 2025, pp. 59-71. [Consultada 13 de marzo de 2026]. [Disponible en: <https://periodicos.unesc.net/ojs/index.php/AnaisDirH/article/view/10501>].
- VALENCIA GALLO, Lina María et al., "Implicaciones éticas de la utilización de la inteligencia artificial (IA) en la administración de justicia de Colombia". *Revista IUS*, vol. 20, n.º 56, 2025, pp. 47-87. [Consultada 20 de marzo de 2026]. [Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=10453773>].

Fuentes jurisprudenciales

- CORTE CONSTITUCIONAL. Sentencia T-529/2023, Bogotá, Corte Constitucional, 2023. [Consultada 20 de marzo de 2026]. [Disponible en: <https://www.corteconstitucional.gov.co/relatoria/2023/t-529-23.htm>].
- TRIBUNAL DE JUSTICIA DE LA UNIÓN EUROPEA. Asunto C-203/22. Dun & Bradstreet Austria GmbH. Sentencia de 27 de febrero de 2025. [Consultada 8 de marzo de 2026]. [Disponible en: <https://n9.cl/lrt5n2>].
- TRIBUNAL DE JUSTICIA DE LA UNIÓN EUROPEA. Asunto C-634/21. SCHUFA Holding AG. Sentencia de 7 de diciembre de 2023. [Consultada 8 de marzo de 2026]. [Disponible en: <https://n9.cl/bmjomq>].
- TRIBUNAL SUPREMO DE WISCONSIN. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). [Consultada 25 de marzo de 2026]. [Disponible en: <https://courts.ca.gov/sites/default/files/courts/default/2024-12/btb24-21-3.pdf>].

Fuentes legales

- UNIÓN EUROPEA. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales. *Diario Oficial de la Unión Europea*, 2016, L 119. [Consultada 20 de abril de 2026]. [Disponible en: <https://n9.cl/is1sch>].
- UNIÓN EUROPEA. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial). *Diario Oficial de la Unión Europea*, 2024, L. [Consultada 21 de abril de 2026]. [Disponible en: <https://n9.cl/uwg0c>].