

**INTELIGENCIA ARTIFICIAL EN LA ELABORACIÓN DE SENTENCIAS PENALES  
EN COLOMBIA: TENSIONES CON EL DEBIDO PROCESO Y LOS DERECHOS  
FUNDAMENTALES**

**DIEGO FERNANDO HERNÁNDEZ DIAZ**



**FACULTAD DE DERECHO  
MAESTRÍA EN DERECHO PENAL Y DERECHO PROCESAL PENAL  
UNIVERSIDAD SANTO TOMÁS  
TUNJA  
2025**

**INTELIGENCIA ARTIFICIAL EN LA ELABORACIÓN DE SENTENCIAS PENALES  
EN COLOMBIA: TENSIONES CON EL DEBIDO PROCESO Y LOS DERECHOS  
FUNDAMENTALES**

**Por:  
DIEGO FERNANDO HERNÁNDEZ DIAZ**

**Documento final presentado como opción de grado para optar al título profesional de  
Magister en derecho penal y procesal**

**Aprobado por:  
Dr. Santiago Vásquez  
Tutor Universidad**



**FACULTAD DE DERECHO  
MAESTRÍA EN DERECHO PENAL Y DERECHO PROCESAL PENAL  
UNIVERSIDAD SANTO TOMÁS  
TUNJA  
2025**

**Nota de aceptación**

---

---

---

---

Decano Facultad Derecho

---

Tutor Universidad

---

Evaluador

Tunja, Día/ mes/ año

## DEDICATORIA

Dedico este trabajo a todas las personas que creen en la justicia como un valor humano irrenunciable, más allá de los algoritmos, las métricas y las lógicas de eficiencia, a quienes, desde el derecho, la filosofía, la tecnología y las ciencias sociales, insisten en poner la dignidad humana en el centro de toda transformación institucional.

A mis docentes, cuya rigurosidad académica y compromiso ético han trazado el camino para desarrollar una mirada crítica, reflexiva y propositiva frente a los nuevos desafíos del derecho penal contemporáneo, a mis compañeros de formación, con quienes compartí preguntas, incertidumbres y aprendizajes colectivos que nutrieron esta investigación.

A mi familia, por su paciencia, comprensión y apoyo incondicional en cada etapa del proceso, a ellos les debo no solo el tiempo para escribir, sino también la convicción de que pensar con profundidad y defender principios fundamentales sigue siendo una tarea urgente.

finalmente, dedico estas páginas a quienes, desde cualquier rincón del país, resisten a la deshumanización de la justicia y luchan, desde el aula, el estrado o la calle, por un sistema judicial garantista, accesible y verdaderamente transformador.

## AGRADECIMIENTOS

La culminación de esta monografía no habría sido posible sin el acompañamiento, la guía y el respaldo de diversas personas e instituciones que, de manera directa o indirecta, contribuyeron significativamente al desarrollo de esta investigación.

Agradezco profundamente a los docentes y asesores de la maestría, cuyas enseñanzas y observaciones rigurosas orientaron el enfoque crítico de este trabajo, sus aportes no solo fortalecieron la solidez académica del estudio, sino que también reafirmaron la importancia de defender el Estado de derecho frente a los desafíos que impone la incorporación de tecnologías como la inteligencia artificial en el ámbito judicial.

A los expertos consultados, cuyos conocimientos en derecho penal, inteligencia artificial, filosofía del derecho y derechos fundamentales permitieron enriquecer la discusión teórica y práctica planteada en estas páginas.

Agradezco también a la universidad por brindar un espacio académico plural, comprometido con el pensamiento crítico, y por fomentar investigaciones que no temen cuestionar los discursos hegemónicos en torno a la modernización judicial.

A mi familia, por ser un pilar emocional constante en este proceso, su paciencia, afecto y comprensión fueron fundamentales en los momentos de mayor exigencia y duda.

Extiendo mi gratitud a todas las personas que creen en una justicia verdaderamente humana, garante de derechos, y que no se rinde ante la promesa vacía de la automatización sin ética ni límites constitucionales.

## 1. CONTENIDO

1.	CONTENIDO .....	6
1.	INTRODUCCIÓN .....	8
1.1.	PROBLEMA DE INVESTIGACIÓN.....	12
1.2.	PLANTEAMIENTO DEL PROBLEMA .....	15
1.3.	PREGUNTA PROBLEMA.....	19
1.4.	HIPÓTESIS.....	20
1.5.	OBJETIVOS .....	21
1.5.1.	Objetivo General.....	21
1.5.2.	Objetivos Específicos.....	22
1.6.	JUSTIFICACIÓN .....	22
1.7.	METODOLOGÍA .....	24
1.7.1.	Tipo de investigación.....	24
1.7.2.	Enfoque de Investigación.....	25
1.7.3.	Método de investigación.....	27
1.8.	ACOTACIÓN MATERIAL.....	28
2.	CAPÍTULO I: FUNDAMENTOS TEÓRICO-NORMATIVOS DEL USO DE LA INTELIGENCIA ARTIFICIAL EN LA JUSTICIA PENAL COLOMBIANA: ALCANCES Y LÍMITES DEL DEBIDO PROCESO.....	30
2.1.	FUNDAMENTACIÓN TEÓRICA .....	31
2.1.1.	La relación entre inteligencia artificial y función jurisdiccional: alcances iniciales .....	32
2.1.4.	Aportes de la filosofía del derecho al debate sobre IA y jurisdicción penal.....	37
2.2.	FUNDAMENTACIÓN NORMATIVA .....	39
2.2.2.	Desarrollo legislativo en Colombia: Ley 1581 de 2012 (Habeas Data), Proyecto de Ley 154 de 2024 sobre IA y Documento CONPES 4144 de 2025 .....	44
3.	CAPÍTULO II: INTELIGENCIA ARTIFICIAL Y DEBIDO PROCESO PENAL: ANÁLISIS COMPARADO Y DIÁLOGO JURISPRUDENCIAL.....	58
3.1.	JURISPRUDENCIA COLOMBIANA SOBRE IA Y GARANTÍAS PROCESALES.....	58
3.1.1.	Sentencia T-323 de 2024 (uso de ChatGPT y debido proceso).....	60
3.1.2.	Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura.....	64
3.2.	ESTÁNDARES INTERAMERICANOS DE LA CORTE IDH Y LA CADH.....	67
3.1.3.	Principios esenciales de la justicia procesal (art. 8 CADH).....	70
3.1.4.	Casos relevantes sobre imparcialidad y transparencia.....	71
3.3.	EXPERIENCIAS EUROPEAS Y NORTEAMERICANAS.....	74
3.4.	ESTÁNDARES MÍNIMOS DE CONSTITUCIONALIDAD Y BUENAS PRÁCTICAS... ..	79
3.5.	PROPUESTA CRÍTICA PARA EL CONTEXTO COLOMBIANO .....	83
3.5.1.	Recomendaciones preliminares para el diseño de un protocolo .....	87
4.	CAPÍTULO 3: RIESGOS Y BENEFICIOS DE LA INTELIGENCIA ARTIFICIAL EN LA ELABORACIÓN DE SENTENCIAS PENALES: UNA VALORACIÓN DESDE LOS PRINCIPIOS DEL DERECHO.....	91
4.1.	RIESGOS ASOCIADOS A LA IA.....	93
4.2.	BENEFICIOS POTENCIALES DE LA IA EN LA FUNCIÓN JURISPRUDENCIAL .....	96
4.3.	VALORACIÓN CRÍTICA DESDE LOS PRINCIPIOS DEL DERECHO .....	98
5.	PROPUESTA DE PROTOCOLO PARA LA IMPLEMENTACIÓN DE INTELIGENCIA ARTIFICIAL EN LA JUSTICIA PENAL COLOMBIANA.....	103

5.1.	CRITERIOS MÍNIMOS PARA UNA IA COMPATIBLE CON EL ORDEN CONSTITUCIONAL.....	104
5.2.	PROTOCOLO PARA LA IMPLEMENTACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA ELABORACIÓN DE SENTENCIAS PENALES EN COLOMBIA .....	106
5.2.1.	Objetivo General.....	106
5.2.2.	Principios Generales .....	106
5.2.3.	Ámbito de Aplicación.....	107
5.2.4.	Controles Ex Ante.....	107
5.2.5.	Controles Ex Post.....	107
5.2.6.	Gobernanza de Datos .....	107
5.2.7.	Evaluación de Impacto Algorítmico .....	108
5.2.8.	Límites a la Automatización .....	108
5.2.9.	Formación Obligatoria.....	108
5.2.10.	Participación Democrática y Supervisión.....	109
5.2.11.	Reclamación y Reparación .....	109
5.2.12.	Cláusula Transversal.....	109
5.2.13.	Fases de desarrollo del protocolo.....	109
5.2.14.	Diseño algorítmico compatible con principios constitucionales .....	113
5.2.15.	Trazabilidad y explicabilidad del sistema (Explainable AI).....	114
5.2.16.	Entrenamiento de datos con control judicial previo .....	115
5.2.17.	Supervisión humana continua y no delegable.....	116
5.2.18.	Registro digital obligatorio de intervenciones de IA .....	117
5.2.19.	Auditoría externa e independiente del sistema .....	118
5.2.20.	Protocolo de objeción y exclusión por parte de la defensa.....	119
5.2.21.	Limitación de usos permitidos según fase procesal.....	120
5.2.22.	Inclusión de un módulo de aprendizaje supervisado .....	120
5.2.23.	Estándares de interoperabilidad con el sistema judicial .....	121
5.2.24.	Revisión judicial reforzada cuando haya intervención algorítmica.....	122
5.2.25.	Mecanismo ciudadano de vigilancia y observación tecnológica .....	123
6.	CONCLUSIONES .....	124
6.1.	PRINCIPALES HALLAZGOS DE LA INVESTIGACIÓN.....	125
6.2.	APORTES TEÓRICOS Y PRÁCTICOS .....	128
6.3.	LÍNEAS FUTURAS DE INVESTIGACIÓN.....	130
7.	BIBLIOGRAFÍA .....	134

## 1. INTRODUCCIÓN

La incorporación de la inteligencia artificial (IA) en la administración de justicia se ha convertido en uno de los debates más relevantes de los últimos años, su irrupción en distintos sistemas jurídicos del mundo ha abierto interrogantes no solo acerca de la eficiencia que promete, sino también sobre las consecuencias éticas y jurídicas que puede acarrear, en el caso colombiano, el asunto resulta especialmente delicado dentro del ámbito penal, en el que el respeto al debido proceso y a las garantías constitucionales constituye la base misma del sistema de enjuiciamiento, a partir de ello, este trabajo se orienta a examinar críticamente las tensiones que suscita la aplicación de herramientas algorítmicas en la elaboración de sentencias, preguntándose si estas tecnologías representan un paso hacia la modernización judicial o, por el contrario, un riesgo latente para los derechos fundamentales de las personas procesadas.

Diversos enfoques académicos han intentado explicar los posibles efectos de la digitalización de la justicia, Susskind (2019), por ejemplo, plantea que los tribunales del futuro serán inevitablemente más automatizados y dependientes de herramientas digitales, lo cual exige repensar el papel del juez y las dinámicas propias de la toma de decisiones, sin embargo, esa visión optimista no ha sido aceptada de forma unánime, Citron y Pasquale (2014) advierten que los sistemas algorítmicos diseñados para apoyar o reemplazar la función judicial pueden reproducir sesgos, errores y una falta de transparencia que pone en entredicho la legitimidad de las decisiones, en un sentido cercano, O'Neil (2016) demuestra cómo los denominados “algoritmos opacos” tienden a reforzar desigualdades estructurales, generando exclusiones que resultan incompatibles con los ideales de igualdad y justicia.

Estas preocupaciones han encontrado eco en la jurisprudencia constitucional colombiana, la Corte Constitucional ha insistido en que toda decisión judicial debe responder a un ejercicio

racional, argumentado y motivado, en consonancia con los principios que sustentan el Estado social de derecho, de manera expresa, el artículo 29 de la Constitución establece el debido proceso como derecho fundamental, recordando que nadie puede ser juzgado sino bajo leyes preexistentes, ante autoridad competente y con plenas garantías de defensa, contradicción y presunción de inocencia, en consecuencia, la posibilidad de automatizar sentencias penales a través de sistemas de IA debe evaluarse con extremo cuidado, pues sustituir o disminuir el papel del razonamiento humano en favor de procesos algorítmicos podría traducirse en una afectación directa de estos principios (Corte Constitucional, Sentencia C-131 de 2002).

En el escenario internacional, la Unión Europea ha marcado un precedente con la aprobación del AI Act (2021), que reconoce como de “alto riesgo” la aplicación de inteligencia artificial en la administración de justicia, esta clasificación implica límites y salvaguardas estrictas frente a la automatización de decisiones judiciales, buscando proteger la imparcialidad y la motivación de las sentencias, en el ámbito latinoamericano, aunque no existe una regulación homogénea, la Comisión Interamericana de Derechos Humanos (2021) ha formulado lineamientos que resaltan la obligación de diseñar herramientas tecnológicas compatibles con la protección de los derechos fundamentales y con la prevención de prácticas discriminatorias.

En Colombia aún no se dispone de una normativa específica que regule el uso de IA en el ámbito judicial, a pesar de ello, existen disposiciones generales que promueven la transformación digital del Estado, como la Ley 1341 de 2009 y el Decreto 620 de 2020, a ello se suma el Documento CONPES 3975 de 2019, que propone una hoja de ruta para el desarrollo de la inteligencia artificial en el país, sin embargo, estas normas resultan insuficientes para abordar los dilemas éticos, jurídicos y procesales propios del campo judicial, en contraste, la Ley Estatutaria 270 de 1996, que regula la administración de justicia, establece en su artículo 8 que las

decisiones judiciales deben ser autónomas, motivadas y dictadas por jueces investidos de jurisdicción, recordando así que cualquier intervención tecnológica debe estar subordinada al juicio humano y nunca reemplazarlo en su totalidad.

El derecho penal, por su carácter sancionador y por las consecuencias directas que produce en la libertad personal, exige un escrutinio constitucional más estricto que otras ramas del derecho, la experiencia comparada ilustra los riesgos y desafíos de introducir IA en este campo: en Estados Unidos, el uso de la herramienta COMPAS para predecir riesgos de reincidencia ha sido cuestionado por reproducir sesgos raciales y socioeconómicos; en China, la implementación de sistemas algorítmicos en procesos penales ha suscitado críticas por la ausencia de garantías procesales; mientras que en Estonia la IA se limita a tareas judiciales menores bajo estricta supervisión humana, como señala Hildebrandt (2018), estas experiencias revelan que el despliegue de inteligencia artificial en la justicia debe estar acompañado de principios éticos sólidos, salvaguardas institucionales y mecanismos de control jurisdiccional efectivo.

En el contexto colombiano, un modelo institucional responsable debería garantizar que cualquier sistema de inteligencia artificial utilizado en la elaboración de sentencias penales cuente con mecanismos de auditoría y explicaciones accesibles sobre su funcionamiento, además de asegurar en todo caso la revisión por parte de un juez humano, el principio de legalidad penal, expresado en la fórmula *nullum crimen, nulla poena sine lege*, demanda que las decisiones se sustenten en el análisis jurídico de los hechos y en la aplicación de normas vigentes, sin embargo, desde la óptica del constitucionalismo principialista, tales decisiones no pueden reducirse a un ejercicio de subsunción mecánica: involucran también la ponderación de principios fundamentales, como la dignidad humana, la igualdad o la presunción de inocencia,

esto genera tensiones con la opacidad propia de los algoritmos de aprendizaje automático, cuyo funcionamiento estadístico difícilmente puede dar cuenta de la complejidad moral y jurídica que caracteriza al derecho penal.

La jurisprudencia interamericana refuerza esta idea, la Corte Interamericana de Derechos Humanos ha subrayado que el acceso a un juez imparcial, la defensa técnica adecuada y la motivación suficiente de las decisiones son elementos esenciales del debido proceso, estos no se reducen a reglas procesales, sino que constituyen principios fundamentales que orientan la validez misma del juicio, en el caso *Barreto Leiva vs. Venezuela* (2009), la Corte señaló que la afectación de tales garantías compromete directamente la legitimidad de la justicia, recordando que la actividad jurisdiccional debe estar guiada por la protección efectiva de derechos y no por la eficiencia formal.

Bajo este panorama, la presente monografía se orienta a realizar un análisis principialista del uso de la inteligencia artificial en el ámbito penal colombiano, el propósito es examinar sus límites y posibilidades, identificando los riesgos que emergen cuando principios como la igualdad, la proporcionalidad o el derecho de defensa se ven enfrentados a lógicas algorítmicas que privilegian la predicción y la eficiencia, se trata de evaluar cómo pueden armonizarse las herramientas digitales con el respeto pleno a los derechos fundamentales, y de proponer criterios que eviten que la modernización judicial se convierta en un escenario de regresión en términos de garantías procesales.

En este sentido, el trabajo busca aportar al debate contemporáneo sobre justicia digital desde una perspectiva que reconoce el potencial de la IA, pero que exige que su uso esté siempre sometido a los principios constitucionales y a la ponderación judicial, siguiendo la tradición del análisis jurídico-crítico, se integran la revisión normativa, el estudio jurisprudencial y un marco

teórico inspirado en el constitucionalismo principialista, en línea con autores como Alexy (2007), Uprimny & Saffon (2006) y De Luis García (2023), el objetivo es esbozar un modelo de justicia que conjugue innovación tecnológica con la defensa activa de los derechos fundamentales, garantizando que la moral constitucional y la racionalidad práctica sigan siendo los referentes últimos de toda decisión penal.

### **1.1. Problema de Investigación**

En los últimos años, la inteligencia artificial (IA) ha pasado de ser un proyecto incipiente para consolidarse como una herramienta con presencia real en distintos sistemas judiciales del mundo, su incorporación ha sido presentada como una oportunidad para modernizar la justicia, acelerar los procesos y hacer más eficiente la gestión de los despachos, sin embargo, en el ámbito penal este entusiasmo se ve acompañado de un debate profundo, en el que la automatización plantea tensiones frente a los principios esenciales del derecho procesal y el riesgo de afectar garantías constitucionales fundamentales.

El caso colombiano resulta especialmente sensible, la administración de justicia penal arrastra una congestión histórica, demoras en la resolución de procesos y serias dificultades para garantizar acceso equitativo a todas las personas, en este contexto, la posibilidad de recurrir a sistemas de IA para apoyar la elaboración de sentencias aparece como una alternativa seductora frente a los problemas de eficiencia, a pesar de ello, como advierte De Luis García (2023), el debate no puede reducirse a la velocidad de la justicia, sino que debe partir del respeto a los principios fundamentales y del carácter garantista que caracteriza al derecho penal.

La especial naturaleza sancionadora del derecho penal exige un control constitucional particularmente riguroso, pues en este ámbito se encuentran en juego principios fundamentales como la legalidad, la presunción de inocencia, la imparcialidad judicial y la contradicción

procesal, todos reconocidos en el artículo 29 de la Constitución Política de Colombia, desde una perspectiva principialista, estos derechos no operan como reglas rígidas, sino como mandatos de optimización que deben ser ponderados en cada caso concreto, la introducción de algoritmos en la toma de decisiones judiciales plantea el riesgo de reducir esa complejidad a operaciones opacas y estadísticas, dificultando la transparencia, la posibilidad de auditar los razonamientos y, en consecuencia, el ejercicio de la ponderación judicial que solo puede realizar un juez humano.

La experiencia comparada confirma la relevancia de estas preocupaciones, el uso de programas como COMPAS en Estados Unidos y HART en Reino Unido ha sido cuestionado por reproducir sesgos raciales y socioeconómicos, lo que evidencia tensiones entre los principios de igualdad y de eficiencia procesal, en China, el despliegue acelerado de IA en procesos penales responde a un modelo de justicia centralizada, donde la ponderación entre seguridad y garantías individuales se inclina marcadamente hacia el primero, en contraste, Estonia ha optado por restringir la IA a funciones judiciales menores y siempre bajo supervisión directa de jueces, lo que refleja una apuesta institucional por equilibrar innovación tecnológica con el principio de control humano, como lo muestran estos ejemplos, el diseño institucional condiciona los riesgos y beneficios de la automatización (Angwin et al., 2016).

En Colombia, a pesar de los avances en política pública como el CONPES 3975 de 2019 y el Decreto 620 de 2020, no existe aún una regulación que precise los alcances del uso de la IA en materia penal, el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura establece algunas restricciones iniciales, pero no fija parámetros claros de explicabilidad, transparencia ni control humano, como advierte Pineda (2021), esta falta de regulación genera el riesgo de que se adopten sistemas sin legitimidad democrática suficiente y sin criterios claros para la ponderación

de los principios de legalidad, igualdad y debido proceso frente a la eficiencia procesal y la celeridad judicial.

La literatura especializada ha puesto de relieve la llamada opacidad algorítmica, es decir, la dificultad de comprender cómo se generan las decisiones que producen estas tecnologías, esta característica plantea tensiones con el deber constitucional de motivar las sentencias y con la exigencia de que las decisiones sean susceptibles de revisión mediante los recursos previstos en la ley, en esa medida, la introducción de IA en la justicia penal no constituye solo un reto tecnológico, sino también un desafío ético y normativo que exige definir criterios para ponderar la innovación con la deliberación judicial como eje central del proceso (Citron y Pasquale, 2014).

En línea con esta preocupación, la Corte Constitucional ha sostenido que el juez no puede trasladar a terceros la responsabilidad de decidir sobre la libertad o la culpabilidad de un individuo, en la Sentencia C-591 de 2005, reiteró que toda decisión judicial debe sustentarse en un análisis racional de las pruebas y en una argumentación jurídica sólida, tal exigencia se vería comprometida si el juez se limitara a validar de manera acrítica los resultados emitidos por un algoritmo, como plantea De Luis García (2023), el riesgo no radica en el uso de tecnología en sí mismo, sino en su potencial para erosionar el principio de imparcialidad judicial y diluir la función de motivación que le corresponde al juez.

Otro aspecto relevante se relaciona con la forma en que la IA puede reproducir patrones históricos, si el sistema judicial colombiano ha operado con sesgos hacia ciertos grupos sociales, étnicos o económicos, existe la posibilidad de que los algoritmos refuercen esas desigualdades, Binns (2018) advierte que este tipo de discriminación algorítmica puede ser más difícil de identificar y controlar que la cometida por operadores humanos, en un país marcado por desigualdades estructurales, ello plantea el desafío de diseñar marcos institucionales que

ponderen el principio de igualdad frente a la búsqueda de eficiencia en la administración de justicia.

La situación se agrava por los problemas estructurales que afectan a la justicia penal colombiana: falta de recursos, presión institucional por descongestionar los despachos y limitada formación tecnológica de jueces y funcionarios, estas condiciones pueden generar una dependencia excesiva de los sistemas de IA sin una comprensión adecuada de sus limitaciones, Hildebrandt (2018) advierte que, en tales contextos, se corre el riesgo de caer en un automatismo judicial, en el que el juez se convierte en un mero validador de resultados predefinidos, debilitando su función deliberativa y comprometiendo el principio de independencia judicial.

Desde el derecho internacional de los derechos humanos, el artículo 8 de la Convención Americana garantiza un debido proceso dirigido por jueces imparciales e independientes, cuyas decisiones sean motivadas y revisables, la Corte Interamericana, en casos como Maldonado Vargas vs. Chile (2015), ha sido clara en señalar que la eficiencia o la modernización tecnológica no pueden prevalecer sobre estas garantías mínimas, para el caso colombiano, este estándar supone que cualquier intento de digitalización de la justicia debe incorporar mecanismos que aseguren la ponderación adecuada entre innovación tecnológica y los principios fundamentales que estructuran el Estado Social de Derecho.

## **1.2. Planteamiento del problema**

En las últimas décadas, el discurso de modernización del sistema judicial ha adquirido un protagonismo creciente en América Latina, impulsado por la digitalización, la necesidad de eficiencia institucional y la influencia de modelos anglosajones de justicia digital, Colombia no ha sido ajena a este proceso, pues ha comenzado a incorporar herramientas de inteligencia

artificial (IA) en distintos niveles de la administración de justicia, desde sistemas de gestión procesal hasta proyectos piloto para la predicción de fallos y, más recientemente, iniciativas orientadas a la elaboración de sentencias, estas innovaciones, si bien prometen descongestionar los despachos, optimizar recursos y aportar celeridad, plantean un dilema de fondo: su compatibilidad con los principios del debido proceso, el control jurisdiccional y la preservación de la función deliberativa del juez (De Luis García, 2023).

La aplicación de sistemas algorítmicos en el ámbito penal resulta particularmente sensible, ya que las decisiones judiciales en esta materia inciden de manera directa en la libertad personal y otros derechos fundamentales, como señala De Luis García (2023), los riesgos de opacidad, sesgo y automatización acrítica pueden erosionar principios como el debido proceso y el derecho de defensa, la justicia penal, por su naturaleza punitiva, exige un nivel de escrutinio estricto en cuanto a la racionalidad, proporcionalidad y motivación de las decisiones, en este sentido, la incorporación de tecnologías de IA debe sustentarse en un marco normativo y ético sólido, orientado a garantizar la adecuada ponderación entre eficiencia procesal y salvaguarda de los principios constitucionales que estructuran el Estado Social de Derecho.

El problema no se reduce a una cuestión técnica, sino que también involucra dimensiones epistemológicas y jurídicas, Roa y Salazar (2022) advierten que la introducción de IA en procesos judiciales debe ser objeto de una revisión crítica, pues la supuesta objetividad tecnológica puede ocultar sesgos estructurales presentes en los datos de entrenamiento, reforzando desigualdades históricas, esta “fetichización del algoritmo” encubre relaciones de poder, exclusiones sociales y carencias institucionales que ya limitan el acceso a la justicia en contextos como el colombiano.

Preocupaciones similares han sido planteadas por O'Neil (2016), quien demuestra cómo los algoritmos de decisión tienden a replicar y amplificar exclusiones, con efectos especialmente graves para poblaciones marginadas, en el ámbito penal, esto puede traducirse en evaluaciones de riesgo de reincidencia sesgadas, recomendaciones de pena discriminatorias o valoraciones de pruebas descontextualizadas, Eubanks (2018) y Crawford (2021) coinciden en que la supuesta neutralidad matemática es ilusoria, pues todo modelo algorítmico responde a decisiones humanas, intereses institucionales y determinaciones políticas.

En Colombia, el debate normativo es aún incipiente, aunque la Corte Constitucional ha abordado aspectos relativos al habeas data, la intimidad y el uso de tecnologías estatales, como lo demuestra la Sentencia T-230 de 2021, no existe un marco regulatorio específico que delimite el uso de IA en la elaboración de sentencias penales, este vacío genera un riesgo evidente: que la innovación tecnológica avance a un ritmo mayor que las salvaguardas jurídicas necesarias para proteger los derechos fundamentales, desde la filosofía jurídica, Dworkin (1986) recuerda que la función judicial no se limita a aplicar normas de manera mecánica, sino que requiere una interpretación moral e institucional del derecho, en sintonía, Uprimny y Saffon (2006) sostienen que las decisiones judiciales deben reflejar una deliberación racional y fundada en principios, lo cual plantea un límite claro a la sustitución del juicio humano por procesos algorítmicos

En esta misma línea, De Luis García (2023) advierte que el derecho de defensa no puede garantizarse en un esquema en el que las decisiones judiciales sean delegadas parcial o totalmente a entidades no humanas, pues ello compromete tanto la legitimidad de la judicatura como el control democrático, Miró (2023) añade que la digitalización judicial, aunque necesaria, no debe convertirse en un fin en sí mismo, sino entenderse como un instrumento subordinado a límites normativos y éticos, en esa dirección, el reciente Acuerdo PCSJA24-12243 del Consejo

Superior de la Judicatura recoge esta preocupación al prohibir el uso de IA para sustituir la labor decisoria del juez y restringir su aplicación a funciones meramente auxiliares.

La falta de transparencia constituye otra preocupación central, Pasquale (2015) explica que los algoritmos operan como “cajas negras” cuyo funcionamiento interno resulta opaco, lo que contraviene la exigencia de motivación establecida en el artículo 29 de la Constitución Política, sin la posibilidad de comprender cómo se adopta una decisión o de cuestionar su razonabilidad, se vulnera el principio de contradicción y el derecho a una defensa efectiva, en consonancia, Citron y Pasquale (2014) proponen marcos regulatorios que aseguren la auditabilidad y la explicabilidad de los algoritmos, algo que en Colombia comienza a discutirse con iniciativas como el Proyecto de Ley 154 de 2024 sobre inteligencia artificial, el cual establece obligaciones de trazabilidad y transparencia para los sistemas utilizados por el Estado

A nivel internacional, los ejemplos son aleccionadores, en Estados Unidos, el software COMPAS ha sido duramente cuestionado por reproducir sesgos raciales en la determinación de riesgos de reincidencia, en Francia, el Conseil d'État (2018) impuso límites al uso de algoritmos en decisiones judiciales, sosteniendo que estas deben permanecer en manos de jueces humanos, en contraste, China ha avanzado hacia un modelo de justicia digital fuertemente centralizado, en el que la automatización tiene un papel preponderante, aunque con escasas garantías procesales, Estonia, por su parte, ha optado por restringir la IA a funciones menores bajo estricta supervisión, evidenciando que el diseño institucional determina los riesgos y alcances de estas herramientas. (Angwin et al., 2016).

El escenario colombiano, en consecuencia, se sitúa en una encrucijada, por un lado, la digitalización promete eficiencia y descongestión judicial, por otro, plantea desafíos inéditos para la preservación del debido proceso, la motivación de las sentencias y la función deliberativa

del juez natural, mientras desde un enfoque tecnocrático se defiende la IA como un instrumento neutral de apoyo, la doctrina crítica advierte que incluso una delegación parcial puede derivar en una “tercerización de la justicia” incompatible con el principio del juez natural (Prieto Sanchís, 2012; Gargarella, 2010).

Es así como la discusión no es solo institucional, sino también democrática, la ciudadanía tiene derecho a comprender cómo se adoptan las decisiones judiciales que afectan su libertad y su vida, la opacidad algorítmica puede minar la confianza pública en la justicia y erosionar su legitimidad, como advierte Morozov (2013), la fascinación con soluciones tecnológicas corre el riesgo de invisibilizar la necesidad de reformas estructurales y participativas en las instituciones, en sintonía, la Corte Constitucional colombiana, en la Sentencia T-323 de 2024, recordó que el uso de IA en funciones jurisdiccionales debe estar supeditado al control humano estricto y a la obligación de motivación judicial.

En este contexto, el presente trabajo aborda críticamente los alcances, riesgos y límites del uso de inteligencia artificial en la elaboración de sentencias penales en Colombia, más que contraponer tradición y modernización, se busca examinar cómo puede avanzar la transformación digital sin sacrificar los principios esenciales del proceso penal, el análisis se enmarca en la tradición del constitucionalismo garantista, articulando aportes de la filosofía del derecho, la ética algorítmica, la teoría crítica y la experiencia comparada, con el objetivo de contribuir a la construcción de un modelo de justicia que, sin renunciar a la innovación, asegure su carácter humano, inclusivo y democrático.

### **1.3. Pregunta problema**

La incorporación de sistemas de inteligencia artificial (IA) en la administración de justicia ha abierto un debate inaplazable sobre los límites de la automatización en un ámbito que, por su

naturaleza, exige la salvaguarda estricta de los derechos fundamentales, en materia penal, donde cobran especial relevancia las garantías previstas en el artículo 29 de la Constitución, debido proceso, contradicción, defensa y presunción de inocencia, la participación de herramientas algorítmicas en la elaboración de sentencias introduce tensiones sobre la transparencia decisoria, la motivación judicial y el papel indelegable del juez como intérprete de la ley y responsable último de la decisión.

Aunque el ordenamiento colombiano empieza a trazar algunos referentes normativos, Ley 1581 de 2012 (habeas data), Proyecto de Ley 154 de 2024 sobre IA, Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura y la estrategia CONPES 4144 de 2025, persisten vacíos y desafíos prácticos en materia de explicabilidad, trazabilidad, auditoría independiente y control humano efectivo, la jurisprudencia reciente, como la Sentencia T-323 de 2024, ha advertido que la modernización tecnológica no puede desvincular la función jurisdiccional de los principios constitucionales ni reducir la responsabilidad decisoria a “cajas negras” de difícil escrutinio público.

De esta manera, surge la siguiente pregunta problema: ¿En qué condiciones mínimas de transparencia, explicabilidad y control humano la implementación de sistemas de inteligencia artificial en la elaboración de sentencias penales en Colombia puede resultar compatible con el respeto al debido proceso y con los principios fundamentales del Estado Social de Derecho?

#### **1.4. Hipótesis**

En el escenario contemporáneo de acelerada incorporación de tecnologías emergentes en la administración de justicia, se sostiene que la implementación de sistemas de inteligencia artificial en la elaboración de sentencias penales en Colombia, cuando no se acompaña de una regulación clara, transparente y sustentada en los principios constitucionales, constituye un

riesgo cierto para las garantías procesales fundamentales, en especial para el debido proceso, si bien es cierto la IA puede ofrecer ventajas en términos de eficiencia, descongestión judicial y mayor celeridad en la resolución de procesos, su aplicación directa en funciones decisorias propias del juez penal, sin mecanismos efectivos de control humano, revisión crítica y estándares normativos rigurosos, puede producir decisiones impersonales, descontextualizadas y carentes de motivación jurídica suficiente, con la consiguiente afectación de derechos como el de defensa, la presunción de inocencia, la imparcialidad del juzgador y el acceso a una justicia sustantiva.

De igual modo, en un contexto como el colombiano, caracterizado por brechas tecnológicas, desigualdades estructurales y limitaciones en la formación técnica y ética de los operadores judiciales, la adopción no regulada de herramientas algorítmicas difícilmente asegura mejoras sustantivas en la administración de justicia; por el contrario, puede intensificar dinámicas de exclusión, reproducir sesgos preexistentes y debilitar la legitimidad democrática de las decisiones judiciales, en consecuencia, se plantea que una modernización judicial basada en inteligencia artificial solo será compatible con el orden constitucional colombiano en la medida en que se apoye en condiciones mínimas de transparencia algorítmica, mecanismos claros de rendición de cuentas y control jurisdiccional efectivo.

## **1.5. Objetivos**

### **1.5.1. Objetivo General**

Determinar críticamente la compatibilidad del uso de sistemas de inteligencia artificial en la elaboración de sentencias penales en Colombia con el debido proceso y los principios fundamentales del Estado Social de Derecho, estableciendo criterios operativos que orienten un protocolo de implementación constitucionalmente adecuado.

### **1.5.2. Objetivos Específicos**

- Evaluar de manera sistemática el marco normativo y doctrinal aplicable al uso de IA en la función jurisdiccional penal en Colombia (alcances, límites, vacíos y ambigüedades), con especial atención a la motivación judicial, la transparencia/explicabilidad y el control humano significativo.
- Comparar críticamente la jurisprudencia nacional e internacional relevante sobre IA y debido proceso en materia penal, para identificar estándares mínimos de constitucionalidad y buenas prácticas trasladables al contexto colombiano
- Valorar los riesgos y beneficios jurídicos, éticos y sociales que plantea la incorporación de inteligencia artificial en la elaboración de sentencias penales en Colombia, con el fin de establecer criterios que permitan garantizar la supremacía del debido proceso y la legitimidad de la función jurisdiccional.
- Proponer un protocolo de implementación de IA para la elaboración de sentencias penales que incorpore controles ex ante y ex post, gobernanza de datos, evaluación de impacto algorítmico, límites a la automatización y métricas de seguimiento (revocatorias, sesgo, motivación), garantizando que la IA actúe como herramienta auxiliar y no sustitutiva de la decisión judicial.

### **1.6. Justificación**

La presente investigación resulta pertinente en un momento en que la digitalización de la justicia se ha convertido en un imperativo global, en Colombia, la progresiva incorporación de herramientas de inteligencia artificial (IA) en la administración judicial plantea oportunidades innegables de modernización y eficiencia, pero también suscita preocupaciones legítimas sobre su compatibilidad con los principios constitucionales que rigen el derecho penal, la sentencia

judicial, en particular la penal, constituye el acto decisorio de mayor trascendencia dentro del Estado Social de Derecho, pues en ella se concretan derechos esenciales como la presunción de inocencia, el derecho de defensa y la libertad personal, por ello, automatizar total o parcialmente esta función demanda un escrutinio riguroso que justifica la necesidad del presente estudio.

Desde una perspectiva jurídico-constitucional, este trabajo se fundamenta en la exigencia de precisar los alcances y límites de la IA en la elaboración de sentencias penales, en coherencia con lo dispuesto en el artículo 29 de la Constitución Política y con la jurisprudencia de la Corte Constitucional (Sentencia C-591 de 2005), que reafirma el carácter indelegable de la función jurisdiccional, la motivación judicial y la imparcialidad del juez son principios que no pueden reducirse a procesos algorítmicos opacos ni a inferencias estadísticas desprovistas de control democrático, en esta línea, diversos autores han advertido que el ejercicio del poder punitivo debe desarrollarse bajo estrictos parámetros normativos y de racionalidad, lo cual resulta particularmente desafiante en un escenario de creciente automatización.

En el plano académico, la investigación aporta un enfoque crítico que busca llenar un vacío en la doctrina jurídica nacional, aunque existen avances normativos y documentos de política pública como el CONPES 3975 de 2019 o el Proyecto de Ley 154 de 2024, los estudios que analizan de manera integral los riesgos y tensiones entre la IA y el derecho penal siguen siendo limitados, la revisión sistemática de literatura y jurisprudencia nacional e internacional, articulada con debates contemporáneos sobre ética algorítmica y filosofía del derecho, permitirá ofrecer un análisis que conecte la dogmática jurídica con los desafíos tecnológicos actuales.

En el plano social y político, esta investigación cobra relevancia porque el contexto colombiano está marcado por desigualdades estructurales, sobrecarga procesal y desconfianza ciudadana hacia las instituciones judiciales, en tales condiciones, la introducción de sistemas

algorítmicos sin regulaciones robustas podría amplificar sesgos preexistentes, reproducir discriminaciones históricas y debilitar la legitimidad democrática del poder judicial, como advierten O’Neil (2016) y Pasquale (2015), la aparente neutralidad de los algoritmos puede convertirse en un riesgo mayor cuando opera en contextos desiguales, por lo tanto, resulta indispensable diseñar lineamientos que garanticen transparencia, rendición de cuentas y control humano efectivo.

La dimensión comparada refuerza la pertinencia de este trabajo, la Unión Europea, a través del AI Act, y Francia, mediante el Conseil d’État (2018), han delimitado de manera expresa el uso de la IA en funciones jurisdiccionales, reconociéndolo como un ámbito de alto riesgo que requiere supervisión humana estricta, contrastar estas experiencias con el contexto colombiano permitirá identificar buenas prácticas y advertencias relevantes para la construcción de un protocolo normativo y ético acorde con la Constitución y los derechos fundamentales, de este modo, la investigación no solo atiende una necesidad académica y jurídica, sino que también ofrece insumos para orientar políticas públicas que promuevan una justicia digital transparente, legítima y respetuosa de la dignidad humana.

## **1.7. Metodología**

### **1.7.1. Tipo de investigación**

**El presente estudio se enmarca en un enfoque descriptivo con perspectiva cualitativa,** cuyo propósito central es caracterizar el uso de la inteligencia artificial (IA) en el sistema judicial colombiano, con especial atención a su aplicación en la elaboración de sentencias penales, este enfoque metodológico resulta pertinente porque permite identificar, organizar y analizar los elementos normativos, doctrinales, jurisprudenciales y comparados que estructuran la discusión

actual sobre la automatización judicial, a fin de valorar su compatibilidad con los principios del debido proceso y los derechos fundamentales.

**El carácter descriptivo** del estudio busca ofrecer un panorama detallado y sistemático de las dimensiones jurídicas y éticas del fenómeno, en lugar de establecer relaciones causales, a su vez, **la perspectiva cualitativa** posibilita interpretar el tema de manera integral, incorporando tanto el análisis crítico de fuentes jurídicas primarias (Constitución, leyes, sentencias) como secundarias (doctrina nacional e internacional), además de experiencias comparadas que permitan comprender cómo otros sistemas judiciales han regulado o limitado el uso de IA en procesos penales.

**Metodológicamente, el trabajo se desarrolla mediante una revisión documental y un análisis jurídico-crítico, articulando la dogmática procesal penal con la teoría constitucional y la ética algorítmica, este abordaje no solo permite describir el estado actual de la discusión, sino también examinar los riesgos, tensiones y posibilidades que la inteligencia artificial plantea para la justicia penal colombiana.**

**El enfoque adoptado se proyecta hacia una dimensión propositiva, orientada a formular lineamientos normativos y éticos que puedan servir como base para un protocolo nacional de implementación de la inteligencia artificial en la administración de justicia penal, dicho protocolo deberá garantizar transparencia algorítmica, control jurisdiccional efectivo y subordinación plena de la tecnología a la Constitución y al Estado Social y Democrático de Derecho.**

### **1.7.2. Enfoque de Investigación**

El enfoque de esta investigación se fundamenta en un análisis crítico-jurídico, siguiendo la orientación metodológica planteada por **Humberto Ñaupas Gaitan** (2018), quien resalta la importancia de examinar los fenómenos sociales y jurídicos desde una perspectiva crítica que

permita identificar tanto sus implicaciones normativas como sus efectos prácticos, en este estudio, dicho enfoque resulta pertinente para evaluar la relación entre la inteligencia artificial (IA) y las garantías constitucionales en el sistema judicial colombiano, con especial atención a los principios de imparcialidad, justicia y debido proceso, de esta manera, la investigación busca comprender los riesgos y oportunidades derivados de la automatización en la elaboración de sentencias penales y analizar cómo estas innovaciones tecnológicas inciden en la protección de los derechos fundamentales de quienes participan en procesos judiciales.

De manera complementaria, se adopta un análisis comparado e interdisciplinario, también recomendado por Ñaupas (2018), que integra aportes de distintas áreas del conocimiento para abordar fenómenos complejos de manera integral, este componente metodológico permite examinar experiencias internacionales en la implementación de la IA en tribunales, identificar buenas prácticas y errores recurrentes, y valorar cuáles de estos elementos podrían adaptarse al contexto colombiano, a través de este ejercicio comparado se busca construir un marco conceptual robusto que articule dimensiones técnicas, jurídicas y éticas, y que permita anticipar tensiones entre la automatización judicial y los principios constitucionales.

La integración del análisis crítico-jurídico y del enfoque comparado interdisciplinario permite que este trabajo cumpla un doble propósito: por un lado, describir de manera rigurosa el panorama actual del uso de IA en la justicia penal colombiana, y por otro, proponer criterios normativos y éticos que orienten su implementación responsable, de este modo, la investigación no se limita a aportar al debate académico, sino que también ofrece insumos prácticos para la formulación de políticas públicas y marcos regulatorios que busquen conciliar innovación tecnológica con la protección efectiva de los derechos humanos (Ñaupas, 2018).

### **1.7.3. Método de investigación**

**El presente estudio se desarrolló bajo el método de investigación jurídica doctrinal**, entendido como aquel que se centra en el examen crítico y sistemático de normas, jurisprudencia, doctrina y literatura especializada, este método resulta pertinente al permitir identificar y evaluar el marco normativo vigente en Colombia en materia de justicia penal, así como analizar las implicaciones derivadas de la incorporación de la inteligencia artificial (IA) en la elaboración de sentencias, su propósito es valorar la compatibilidad de estas innovaciones con los principios constitucionales y las garantías fundamentales, particularmente en lo relacionado con el debido proceso y a la función indelegable del juez.

**El método doctrinal se enriquece mediante un enfoque analítico-comparado**, que posibilita contrastar la experiencia colombiana con la de otros sistemas judiciales que han implementado la IA en funciones jurisdiccionales, a través de esta perspectiva se identifican buenas prácticas, limitaciones y riesgos que permiten extraer lecciones útiles para el contexto nacional, especialmente en relación con los problemas de sesgo algorítmico, la opacidad de los sistemas y los desafíos éticos que plantea la automatización de decisiones judiciales.

**De manera complementaria, la investigación adopta un enfoque crítico**, orientado a examinar cómo la introducción de herramientas algorítmicas en la justicia penal puede tensionar los principios del Estado Social y Democrático de Derecho, esta óptica busca asegurar que cualquier propuesta normativa para el uso de la IA se fundamente en la transparencia, la rendición de cuentas y el control jurisdiccional efectivo, de modo que la innovación tecnológica no derive en la afectación de garantías esenciales como la imparcialidad, la motivación de las decisiones y la presunción de inocencia.

**En conjunto, la aplicación del método jurídico doctrinal, acompañado del análisis comparado e iluminado por una perspectiva crítica**, constituye un marco metodológico sólido

y coherente, dicho marco no solo permite describir y evaluar las condiciones actuales del uso de la inteligencia artificial en la justicia penal colombiana, sino también sustentar la propuesta de un protocolo normativo responsable que armonice la modernización tecnológica con la preservación de los principios constitucionales y la legitimidad democrática (Ñaupás, 2018).

### **1.8. Acotación material**

La presente investigación se delimita al análisis de la implementación de sistemas de inteligencia artificial (IA) en la fase resolutoria del proceso penal colombiano, específicamente en la elaboración de sentencias judiciales, este recorte responde a que la sentencia penal constituye el acto jurídico más trascendental del proceso, en el que se concreta el poder punitivo del Estado y, por ende, exige la más estricta observancia de garantías como el debido proceso, la motivación judicial, la imparcialidad del juez, el principio de contradicción y el derecho de defensa.

El estudio se centra en examinar, desde una perspectiva crítica y jurídico-constitucional, en qué medida la automatización de esta etapa resulta compatible con los principios fundamentales del Estado Social y Democrático de Derecho, y en qué condiciones mínimas de transparencia, explicabilidad y control humano sería posible concebirla como una herramienta auxiliar legítima, en consecuencia, se excluyen expresamente del análisis otras aplicaciones de IA en el ámbito judicial, tales como sistemas de gestión administrativa, software forense, herramientas de analítica jurídica o modelos de predicción de riesgo, para concentrarse en el punto neurálgico de la discusión: la función indelegable de juzgar.

Del mismo modo, la investigación no se orienta a evaluar los aspectos técnicos de la programación o funcionamiento interno de los algoritmos, sino a realizar un examen normativo, doctrinal, jurisprudencial y comparado sobre su pertinencia, riesgos y límites, la atención se

centra en los debates jurídicos y éticos que emergen de la introducción de sistemas algorítmicos en la decisión judicial, atendiendo especialmente al impacto que tienen sobre los derechos fundamentales de las personas procesadas y sobre la legitimidad democrática del poder judicial.

Si bien es cierto el eje del análisis se circunscribe al ordenamiento jurídico colombiano, se recurre también a la experiencia internacional para identificar buenas prácticas, errores y estándares mínimos de constitucionalidad, **es así como se consideran referentes como el AI Act de la Unión Europea, las restricciones del Conseil d'État en Francia y los debates en torno al software COMPAS en Estados Unidos, sin perder de vista que el marco central lo constituyen la Constitución colombiana (arts. 1, 2, 13, 29 y 228), la Ley 1581 de 2012 (habeas data), el Proyecto de Ley 154 de 2024 sobre IA, el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura, la Sentencia T-323 de 2024 de la Corte Constitucional y el Documento CONPES 4144 de 2025 sobre estrategia nacional de IA.**

De esta forma, se fija un marco preciso: no se trata de un estudio técnico de la IA en general, sino de **un análisis crítico-normativo** sobre su compatibilidad con el debido proceso en la sentencia penal, esta delimitación metodológica, coherente con los objetivos del trabajo, busca sentar bases conceptuales y normativas que permitan diseñar un protocolo de implementación constitucionalmente adecuado, **en el que la tecnología sea un instrumento auxiliar sometido a límites éticos, jurídicos y democráticos, y nunca un sustituto de la función jurisdiccional.**

## **2. CAPÍTULO I: FUNDAMENTOS TEÓRICO-NORMATIVOS DEL USO DE LA INTELIGENCIA ARTIFICIAL EN LA JUSTICIA PENAL COLOMBIANA: ALCANCES Y LÍMITES DEL DEBIDO PROCESO**

La incorporación de sistemas de inteligencia artificial (IA) en el ámbito judicial ha generado un debate sin precedentes en torno a los alcances y límites de su implementación en la justicia penal, en particular, la fase de elaboración de sentencias, considerada como el momento culminante del proceso judicial y el acto en el que se concreta el poder punitivo del Estado, exige una especial atención frente a los principios del debido proceso y las garantías constitucionales, este capítulo se propone examinar los fundamentos teóricos y normativos que enmarcan dicho fenómeno, con el fin de ofrecer una evaluación crítica sobre su compatibilidad con los valores que sustentan el Estado Social y Democrático de Derecho en Colombia.

Desde la perspectiva doctrinal, la discusión no puede reducirse a una visión tecnocrática que conciba la inteligencia artificial como un mecanismo neutral de modernización, la justicia penal es, por definición, un campo donde la imparcialidad, la motivación racional de las decisiones y la protección de los derechos fundamentales constituyen exigencias indeclinables, en este sentido, se vuelve imprescindible revisar los aportes del garantismo penal, la teoría constitucional y la ética jurídica, como marcos teóricos que permiten comprender por qué la automatización de decisiones judiciales plantea tensiones estructurales con principios esenciales como la dignidad humana, el derecho de defensa y la función indelegable del juez natural.

En paralelo, resulta necesario trazar un mapa normativo que dé cuenta tanto de los avances como de las ambigüedades que caracterizan el tratamiento jurídico del uso de IA en Colombia, el análisis abarcará la Constitución, la legislación ordinaria y estatutaria, la jurisprudencia constitucional reciente, así como las disposiciones administrativas y de política pública que orientan la transformación digital del sector judicial, a su vez, se considerarán los

estándares internacionales y las experiencias comparadas que ofrecen puntos de referencia útiles para valorar el grado de preparación del ordenamiento jurídico colombiano frente a esta innovación tecnológica, de esta manera, el capítulo sienta las bases para evaluar críticamente si la IA puede actuar como herramienta auxiliar legítima en la justicia penal o si, por el contrario, su aplicación en la elaboración de sentencias comporta riesgos de incompatibilidad con el debido proceso y los principios fundamentales.

### **2.1. Fundamentación teórica**

Las teorías acerca de la implementación de la inteligencia artificial (IA) en la justicia penal colombiana exigen partir de la comprensión de cómo esta tecnología interpela los principios esenciales del derecho procesal y constitucional, en este sentido, la IA no puede entenderse únicamente como una herramienta técnica, sino como un fenómeno que transforma la concepción misma de la función jurisdiccional, la legitimidad de la sentencia y la relación entre poder judicial y ciudadanía, la literatura especializada advierte que la digitalización de la justicia no puede reducirse a una visión instrumental, pues en el ámbito penal cualquier innovación afecta directamente derechos fundamentales como la libertad personal y el debido proceso (De Luis García, 2023; Miró, 2023).

Los debates doctrinales sobre la incorporación de sistemas algorítmicos en los tribunales suelen dividirse en dos grandes posturas, por un lado, una visión tecnocrática que resalta los beneficios en términos de eficiencia y reducción de cargas judiciales, por otro, una visión crítica que enfatiza los riesgos de deshumanización, opacidad y sesgos estructurales en las decisiones automatizadas, estas perspectivas no son necesariamente excluyentes, pero muestran la necesidad de que el análisis jurídico de la IA en la justicia penal se realice desde marcos teóricos que

integren filosofía del derecho, ética jurídica y teoría crítica, evitando reducir la discusión a un problema meramente tecnológico.

En consecuencia, la fundamentación teórica de este estudio se inscribe en el constitucionalismo contemporáneo, que entiende la administración de justicia como una función esencial del Estado sometida a parámetros de imparcialidad, motivación racional de las decisiones y respeto al juez natural, desde esta óptica, la sentencia penal no es un acto meramente técnico de aplicación de normas, sino una decisión jurídica de alto impacto político y social, que involucra la legitimidad democrática del poder punitivo. Por ello, la discusión sobre la IA en la justicia penal debe centrarse en precisar los límites y condiciones bajo los cuales su incorporación resulta compatible con los principios del Estado Social de Derecho.

### **2.1.1. La relación entre inteligencia artificial y función jurisdiccional: alcances iniciales**

La irrupción de la inteligencia artificial (IA) en el ámbito judicial ha transformado la manera en que se conciben los procesos de administración de justicia, abriendo debates que van más allá de lo meramente técnico, en la función jurisdiccional, la IA se presenta como un instrumento potencialmente útil para gestionar grandes volúmenes de información, asistir en la clasificación de casos o incluso sugerir criterios jurisprudenciales, sin embargo, cuando su uso se proyecta hacia la fase decisoria, particularmente en la elaboración de sentencias, la discusión adquiere una naturaleza radicalmente distinta, pues se tocan los límites de la potestad jurisdiccional y se pone en juego el principio de independencia judicial, como señala De Luis García (2023), no se trata únicamente de evaluar la capacidad operativa de los algoritmos, sino de analizar en qué medida estos pueden o no integrarse en un campo dominado por exigencias constitucionales y procesales.

Desde un plano conceptual, la función jurisdiccional implica más que la aplicación mecánica de normas: supone un ejercicio interpretativo, argumentativo y valorativo que conjuga hechos, derecho y principios constitucionales, la IA, por su diseño estadístico y algorítmico, carece de la dimensión hermenéutica propia de la labor judicial, lo que plantea la primera tensión esencial, mientras que los defensores de su implementación argumentan que la IA puede mejorar la eficiencia y reducir la carga de trabajo, los críticos sostienen que la delegación de tareas decisorias a sistemas automatizados puede desnaturalizar el rol del juez como garante institucional del orden constitucional y de los derechos fundamentales, en consecuencia, el debate no debe centrarse en si la IA puede producir una sentencia, sino en si esa sentencia puede considerarse jurídicamente válida y constitucionalmente legítima.

En los alcances iniciales de esta relación, resulta fundamental distinguir entre la IA como herramienta auxiliar y la IA como sustituto de la decisión jurisdiccional, en el primer caso, se reconoce la utilidad de los algoritmos para sistematizar jurisprudencia, identificar patrones o detectar contradicciones normativas, siempre bajo control humano significativo, en el segundo, se abre un terreno problemático: el reemplazo parcial o total de la motivación judicial por razonamientos algorítmicos que carecen de explicabilidad y de la prudencia que caracteriza a la decisión judicial, como advierte Miró (2023), el riesgo de “algoritmización acrítica” reside en concebir la justicia como un proceso técnico de cálculo, invisibilizando que en el derecho penal están en juego bienes supremos como la libertad y la dignidad humana.

Los alcances de esta relación en el contexto colombiano deben analizarse bajo las tensiones propias del Estado Social de Derecho, en un sistema judicial marcado por la congestión, la demora procesal y la falta de recursos, la IA aparece como una solución atractiva para agilizar trámites, no obstante, la ausencia de una regulación específica sobre su uso en

sentencias penales y las advertencias jurisprudenciales recientes, como la Sentencia T-323 de 2024 de la Corte Constitucional, que limitó el uso de ChatGPT en la judicatura, evidencian que no puede avanzarse sin un marco normativo claro, en este sentido, la incorporación de IA en la función jurisdiccional debe entenderse como un proceso condicionado, cuyos alcances iniciales están necesariamente delimitados por la supremacía del debido proceso y la preservación del juez natural como centro de la decisión judicial.

### **2.1.2. Doctrinas críticas sobre el uso de IA en la administración de justicia**

Las doctrinas críticas sobre la incorporación de inteligencia artificial en la justicia han subrayado, de manera consistente, los riesgos que esta implica para la vigencia de los derechos fundamentales y la legitimidad democrática del poder judicial, De Luis García (2023) advierte que la digitalización y algoritmización de la justicia no pueden presentarse como procesos neutrales, pues suponen una reconfiguración de la función jurisdiccional que amenaza con desplazar al juez como centro de la decisión judicial, en su análisis, la autora sostiene que la automatización de sentencias desnaturaliza la motivación judicial, transformando un ejercicio de argumentación jurídica en una operación técnica carente de deliberación, lo que compromete su validez constitucional.

En esta misma línea, Miró (2023) enfatiza que el entusiasmo por la eficiencia tecnológica ha generado una especie de “fetichismo algorítmico”, donde la promesa de objetividad y rapidez oculta riesgos de opacidad, sesgo y pérdida de control democrático, la visión crítica plantea que, lejos de reducir desigualdades, los sistemas de IA tienden a replicar estructuras de exclusión social y a reforzar dinámicas de poder que ya condicionan el acceso a la justicia en países como Colombia, esto se agrava en contextos institucionales frágiles, donde la falta de regulación

específica, como ocurre actualmente en el país, pese a avances como el Proyecto de Ley 154 de 2024, amplifica los riesgos de arbitrariedad y despersonalización en las decisiones judiciales.

La crítica también se ha centrado en el carácter opaco de los algoritmos, Pasquale (2015), con su noción de “caja negra algorítmica”, sostiene que la falta de transparencia en muchos sistemas de IA contradice directamente el principio de motivación judicial, piedra angular del debido proceso, O’Neil (2016), por su parte, ha mostrado cómo los algoritmos de decisión tienden a convertirse en “armas de destrucción matemática” al reproducir sesgos históricos que afectan con mayor severidad a las poblaciones más vulnerables, en la administración de justicia penal, esto se traduce en riesgos de discriminación en evaluaciones de reincidencia, valoración probatoria o incluso determinación de penas.

Eubanks (2018) y Citron & Pasquale (2014) destacan que la supuesta neutralidad tecnológica invisibiliza el hecho de que todo algoritmo es producto de elecciones humanas, intereses institucionales y contextos sociopolíticos, desde una perspectiva crítica, la automatización de la justicia sin transparencia y sin control humano no solo plantea riesgos para la vigencia de los derechos individuales, sino que también puede debilitar la confianza ciudadana en las instituciones judiciales, en el caso colombiano, donde la legitimidad del poder judicial se encuentra en permanente cuestionamiento, la adopción de IA sin criterios de explicabilidad, motivación suficiente y límites normativos claros podría significar un retroceso en lugar de un avance, de allí que estas doctrinas insistan en que la justicia penal, por su carácter punitivo y su relevancia social, exige un nivel reforzado de análisis y de regulación frente a cualquier innovación tecnológica.

### **2.1.3. Condicionamientos constitucionales y normativos frente a la automatización judicial: imparcialidad, motivación y juez natural**

La incorporación de inteligencia artificial en la función jurisdiccional debe analizarse a la luz de los condicionamientos constitucionales y normativos que estructuran el proceso penal, entre los que destacan la imparcialidad del juez, la motivación de las decisiones y la función indelegable del juez natural, estos elementos, reconocidos en el artículo 29 de la Constitución Política y en la jurisprudencia de la Corte Constitucional, no son simples requisitos formales, sino pilares que aseguran la coherencia y legitimidad del sistema penal, en este sentido, la automatización de sentencias sin mecanismos de control transparentes plantea el riesgo de alterar la función jurisdiccional, sustituyendo el juicio humano por operaciones algorítmicas carentes de argumentación jurídica verificable.

El principio de imparcialidad enfrenta tensiones particulares en el contexto de la automatización judicial, dado que los algoritmos pueden reproducir sesgos derivados de los datos con los que fueron entrenados, como advierte O'Neil (2016), los sistemas de decisión automatizada tienden a replicar desigualdades preexistentes, lo que en materia penal podría traducirse en resultados discriminatorios, en un país como Colombia, donde persisten profundas brechas sociales y tensiones étnicas, este riesgo adquiere una especial relevancia, no en vano, el Consejo Superior de la Judicatura, mediante el Acuerdo PCSJA24-12243, ha señalado expresamente que la IA solo puede ser usada como herramienta de apoyo, pero nunca como sustituto de la decisión jurisdiccional, justamente para preservar la imparcialidad en la administración de justicia.

La motivación de las sentencias constituye otro eje indispensable, en cuanto permite que las decisiones judiciales sean comprensibles, controlables y sujetas a contradicción, sin embargo,

la “caja negra algorítmica” descrita por Pasquale (2015) plantea un obstáculo evidente, pues en muchos sistemas resulta imposible rastrear cómo se llega a un resultado concreto, la Corte Constitucional, en la Sentencia T-323 de 2024, recordó que el uso de IA en decisiones judiciales está supeditado al control humano y al deber de motivación, dado que la opacidad algorítmica no cumple con los estándares de transparencia exigidos en el proceso penal, en consecuencia, la explicabilidad tecnológica y los mecanismos de rendición de cuentas se convierten en requisitos imprescindibles para cualquier proyecto de automatización.

El principio del juez natural constituye un límite normativo categórico, este no solo implica la existencia previa y legal del juez competente, sino también que la titularidad de la función jurisdiccional es indelegable, la sustitución del juez por un sistema automatizado resultaría incompatible con la estructura institucional prevista en la Constitución, en tanto impediría al procesado interactuar, contradecir o interpelar al decisor, en este punto, los principios de imparcialidad, motivación y juez natural se consolidan como condicionamientos indispensables frente a la automatización judicial, recordando que el proceso penal no puede reducirse a un procedimiento técnico de cálculo, sino que debe mantenerse dentro de los márgenes normativos que aseguran su legitimidad democrática.

#### **2.1.4. Aportes de la filosofía del derecho al debate sobre IA y jurisdicción penal**

Dworkin (1986) propone que el juez actúa como un intérprete moral del derecho, ponderando principios y valores en contextos de conflicto, desde esta perspectiva, la IA no puede sustituir la labor judicial, dado que carece de la capacidad de integrar criterios éticos, valorar situaciones particulares o adaptar normas a las circunstancias específicas de cada caso, en la justicia penal, donde se decide sobre la libertad y la dignidad de las personas, estas capacidades interpretativas son esenciales y no pueden ser replicadas por procesos algorítmicos.

Alexy (1993), por su parte, enfatiza que los derechos fundamentales funcionan como principios que deben optimizarse mediante juicios de proporcionalidad, este enfoque exige ponderar distintos intereses y seleccionar la alternativa que menos restrinja derechos, un proceso que requiere sensibilidad valorativa, contextualización y razonamiento moral, los sistemas automatizados, que operan bajo lógicas binarias y patrones estadísticos, presentan limitaciones estructurales para realizar estas ponderaciones, así, la IA puede asistir en la organización de información o en la detección de patrones, pero no puede reemplazar la toma de decisiones que implique equilibrio entre derechos y normas.

Hart (1961) introduce la noción de “textura abierta” del derecho, señalando que el lenguaje normativo genera márgenes de indeterminación que solo pueden resolverse mediante interpretación humana, la IA, al basarse en datos históricos, corre el riesgo de reproducir sesgos y patrones preexistentes, sin capacidad de innovar o adaptarse a nuevas situaciones, este desafío muestra que, incluso desde un enfoque positivista, la labor del juez no es sustituible, pues interpretar normas requiere creatividad, prudencia y responsabilidad frente a los efectos de las decisiones judiciales.

En conjunto, estas perspectivas coinciden en que la función judicial penal es un ejercicio de deliberación ética, interpretación normativa y juicio valorativo que no puede ser delegado a sistemas algorítmicos, la IA puede desempeñar un papel auxiliar en la sistematización de información o la identificación de patrones jurisprudenciales, pero cualquier intento de reemplazar la toma de decisiones humanas implicaría riesgos sustanciales para la coherencia del derecho, la proporcionalidad de las decisiones y la integridad del proceso judicial, la reflexión

filosófica subraya, por tanto, la necesidad de establecer límites claros y criterios operativos para la incorporación de la automatización en la justicia penal.

## **2.2. Fundamentación normativa**

El artículo 29 de la Constitución Política establece el debido proceso, que comprende la imparcialidad del juez, el derecho de defensa y la motivación de las decisiones, y constituye el marco central frente a cualquier intento de automatización judicial, ya que toda innovación tecnológica en la justicia penal debe respetar la autonomía judicial, la transparencia en la toma de decisiones y el control institucional, la jurisprudencia de la Corte Constitucional, incluyendo las Sentencias C-591 de 2005 y T-323 de 2024, ha señalado que la función jurisdiccional no puede ser desplazada por algoritmos, y que sistemas como ChatGPT solo pueden emplearse como apoyo, siempre que el juez mantenga la decisión final y se garantice claridad en la motivación.

En el ámbito legal, la Ley 1581 de 2012 sobre protección de datos personales establece principios de legalidad, finalidad, libertad, veracidad, seguridad y confidencialidad, relevantes para la aplicación de sistemas algorítmicos en la justicia penal, la información judicial es altamente sensible, por lo que el uso de IA requiere cumplir estándares estrictos de gobernanza y protección de datos, evitando riesgos sobre la intimidad y el habeas data, asimismo, el Proyecto de Ley 154 de 2024 introduce obligaciones de transparencia, explicabilidad y trazabilidad, y establece la necesidad de auditorías independientes y mecanismos de rendición de cuentas, aplicables a cualquier implementación judicial.

En el plano institucional, el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura establece límites precisos al uso de IA en la Rama Judicial, prohibiendo que estas herramientas reemplacen la labor decisoria del juez y restringiendo su uso a funciones de apoyo, como la organización documental, clasificación de expedientes o gestión de carga laboral, de

manera complementaria, el Documento CONPES 4144 de 2025, que actualiza la Estrategia Nacional de Inteligencia Artificial, reconoce el potencial de estas tecnologías para la modernización institucional, pero subraya que su implementación debe estar condicionada a la existencia de marcos regulatorios sectoriales, incluida la justicia.

En el ámbito internacional, las Directrices de la UNESCO (2021) y el AI Act de la Comisión Europea (2021) ofrecen estándares de referencia que califican el uso de IA en la función jurisdiccional como de “alto riesgo”, implicando controles reforzados, supervisión humana permanente y mecanismos de explicabilidad robustos, estas referencias internacionales evidencian que la tendencia no apunta a la sustitución del juez, sino a la incorporación regulada de la IA como herramienta auxiliar, de este modo, el marco normativo revela que la introducción de inteligencia artificial en la justicia penal colombiana requiere regulación clara, estricta y orientada al cumplimiento de principios procesales y normativos esenciales, más allá de una simple modernización tecnológica.

### **2.2.1. Marco constitucional colombiano: dignidad humana, debido proceso y función jurisdiccional**

La incorporación de la inteligencia artificial (IA) en la administración de justicia en Colombia carece, hasta el momento, de una regulación unificada y sistemática de rango legal, sin embargo, en los últimos años se han producido avances normativos y reglamentarios que, aunque fragmentarios, permiten delinear un marco jurídico emergente que condiciona y orienta su implementación, dicho marco está compuesto por normas de derecho fundamental, legislación sectorial sobre protección de datos, directrices de política pública y regulaciones internas del poder judicial, este mosaico normativo resulta clave para comprender los alcances y límites que el ordenamiento jurídico colombiano impone al uso de IA en la función jurisdiccional penal.

Uno de los pilares normativos ineludibles en esta materia es la Ley 1581 de 2012, que desarrolla el derecho fundamental al habeas data y establece principios rectores como la finalidad, libertad, veracidad, transparencia y seguridad en el tratamiento de datos personales, la implementación de sistemas de IA en el ámbito judicial implica, de manera inevitable, el tratamiento masivo de datos sensibles, muchos de ellos asociados a antecedentes penales, información procesal y datos de carácter íntimo, por lo tanto, cualquier uso de algoritmos debe sujetarse estrictamente a las disposiciones de esta ley, garantizando que el tratamiento de la información judicial respete la autorización, finalidad y proporcionalidad exigidas por el ordenamiento jurídico, como advierte la Corte Constitucional (Sentencia C-748 de 2011), el habeas data es una extensión de la dignidad humana, por lo que su vulneración a través de sistemas opacos o desproporcionados resulta inconstitucional.

En el plano específico de la Rama Judicial, un hito reciente lo constituye el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura (2024), mediante el cual se establecen lineamientos preliminares para el uso de herramientas de inteligencia artificial en las actuaciones judiciales, este acto administrativo no habilita un uso irrestricto de la IA, sino que fija límites expresos: prohíbe su utilización para tomar decisiones judiciales sin intervención humana, exige transparencia en el uso de herramientas algorítmicas, y dispone que toda integración tecnológica debe estar acompañada de un proceso de evaluación de impacto y de salvaguardas éticas y jurídicas, este acuerdo reconoce, de forma explícita, que la función de juzgar no puede ser automatizada sin menoscabar principios esenciales como el juez natural, la motivación de las decisiones y el debido proceso.

En paralelo, el Gobierno nacional ha venido desarrollando una política pública en materia de inteligencia artificial, el documento CONPES 4144 de 2025, que actualiza y complementa la

Estrategia Nacional de Inteligencia Artificial formulada en 2019 (CONPES 3975), incorpora un enfoque basado en derechos humanos y establece como uno de sus ejes la gobernanza ética de la IA en el sector público, si bien es cierto, su contenido tiene carácter programático y no normativo, el documento ofrece lineamientos relevantes para la formulación de marcos regulatorios sectoriales, entre sus recomendaciones se destacan la necesidad de garantizar la transparencia algorítmica, la explicabilidad de los sistemas, el control humano significativo y la creación de instancias de evaluación de impacto ético y jurídico, en este sentido, el CONPES 4144 funciona como un referente orientador que señala la importancia de compatibilizar la innovación tecnológica con los principios constitucionales y normativos aplicables.

De especial interés para el objeto de esta investigación es el Proyecto de Ley 154 de 2024, actualmente en trámite en el Congreso de la República, el cual busca establecer principios, definiciones y categorías de riesgo para el uso de inteligencia artificial en Colombia, este proyecto adopta un enfoque basado en el riesgo, clasificando los sistemas de IA según su nivel de impacto sobre los derechos fundamentales y disponiendo obligaciones más estrictas para aquellos considerados de “alto riesgo”, en concordancia con el AI Act de la Unión Europea, el proyecto incluye dentro de esta categoría el uso de IA en la administración de justicia, de aprobarse, esta ley obligaría a que cualquier herramienta algorítmica empleada en la elaboración de decisiones judiciales cumpla con estándares de explicabilidad, supervisión humana y auditoría independiente, lo cual representa un avance significativo hacia la regulación garantista del fenómeno.

No obstante, estos avances, el marco normativo colombiano aún presenta importantes vacíos y ambigüedades, en primer lugar, no existe una ley específica que regule el uso de inteligencia artificial en el sistema judicial penal ni que determine, de manera expresa, sus

límites funcionales, alcances éticos o condiciones técnicas mínimas, en segundo lugar, los actos administrativos vigentes (como el Acuerdo PCSJA24-12243) tienen un carácter transitorio y carecen de fuerza vinculante general para todos los operadores judiciales, en tercer lugar, no se han definido mecanismos claros de rendición de cuentas, trazabilidad de deci

siones algorítmicas ni criterios para evaluar el sesgo o la discriminación en el uso de estas tecnologías, esta falta de precisión normativa genera un escenario de inseguridad jurídica que podría afectar derechos como la defensa técnica, la igualdad procesal y el acceso a la justicia.

En este contexto, la doctrina jurídica nacional ha señalado la necesidad de construir un marco regulatorio sólido que integre el derecho constitucional, la dogmática procesal penal y la ética pública, De Luis García (2023), desde una perspectiva comparada, advierten que la falta de regulación específica puede generar “opacidad algorítmica” y “despersonalización del juicio”, situaciones que afectan la transparencia y legitimidad de la función judicial, la autora propone que cualquier norma sobre IA judicial contemple, como mínimo, mecanismos de explicabilidad, motivación clara de las decisiones y control jurisdiccional efectivo, este planteamiento coincide con la jurisprudencia de la Corte Constitucional, que ha señalado que las decisiones judiciales deben ser comprensibles, racionalmente fundamentadas y atribuibles a un sujeto responsable.

Asimismo, los estándares internacionales, como las Directrices sobre Ética de la Inteligencia Artificial de la UNESCO (2021) y el Reglamento de Inteligencia Artificial de la Unión Europea (AI Act, 2021/2024), establecen principios que Colombia está llamada a observar, incluso sin obligación jurídica directa, estas directrices recomiendan que los sistemas de IA aplicados a la justicia deben cumplir con principios como la trazabilidad, la no discriminación, la intervención humana significativa y la supervisión externa, la UNESCO, por ejemplo, subraya que la IA no debe sustituir el juicio humano en contextos donde están en juego

derechos fundamentales, incorporar estos estándares en la regulación nacional no solo fortalecería el marco jurídico interno, sino que facilitaría la cooperación internacional y el alineamiento con buenas prácticas globales.

De esta manera, aunque Colombia ha iniciado la construcción de un marco normativo para la inteligencia artificial en la justicia, este se encuentra aún en una etapa incipiente, caracterizada por normas dispersas, de baja densidad y limitada fuerza vinculante, el desafío consiste en consolidar estos lineamientos en un cuerpo normativo coherente, específico para la función jurisdiccional penal, que permita que la implementación de IA se configure como una herramienta complementaria capaz de fortalecer la transparencia, la motivación de las decisiones judiciales y la legitimidad del proceso penal, evitando que se convierta en un recurso tecnocrático que desplace la toma de decisiones humana, este reto legislativo y jurisprudencial constituye un eje central del presente trabajo, dado que condiciona la viabilidad constitucional de cualquier protocolo de implementación futura.

### **2.2.2. Desarrollo legislativo en Colombia: Ley 1581 de 2012 (Habeas Data), Proyecto de Ley 154 de 2024 sobre IA y Documento CONPES 4144 de 2025**

La evolución normativa en Colombia en torno al uso de tecnologías emergentes en la administración pública y, particularmente, en la justicia, ha sido gradual y fragmentaria, si bien es cierto, aún no existe un marco legal integral que regule expresamente la utilización de inteligencia artificial en el ámbito judicial, pueden identificarse tres hitos legislativos y de política pública que constituyen referencias obligadas: la Ley 1581 de 2012 (Habeas Data), el Proyecto de Ley 154 de 2024 sobre IA, y el Documento CONPES 4144 de 2025, cada uno de estos instrumentos, desde perspectivas distintas, aporta elementos para la comprensión de los alcances, límites y desafíos regulatorios que enfrenta Colombia en la implementación de algoritmos en la justicia penal.

**La Ley 1581 de 2012 constituye la norma de carácter general que regula la protección de datos personales en Colombia**, desarrollando el derecho fundamental al habeas data consagrado en el artículo 15 de la Constitución, esta ley impone principios como la finalidad, libertad, veracidad, transparencia, acceso y circulación restringida, proporcionalidad y seguridad, que resultan de especial relevancia frente al uso de sistemas de inteligencia artificial en la justicia, en el contexto penal, donde se manejan datos sensibles de personas procesadas, víctimas y testigos, la IA solo puede operar si respeta estrictamente las garantías de protección y confidencialidad, la jurisprudencia de la Corte Constitucional, a través de la Sentencia C-748 de 2011, ha sostenido que el habeas data es expresión de la dignidad humana y que cualquier tratamiento indebido de información sensible es inconstitucional, en consecuencia, la Ley 1581 se convierte en un límite material a la digitalización judicial y un marco de protección contra los riesgos de sesgo o filtración de información.

**Un segundo referente es el Proyecto de Ley 154 de 2024**, que constituye el primer intento legislativo de construir un marco específico de regulación para la inteligencia artificial en Colombia, inspirado en el enfoque de riesgo adoptado por el AI Act de la Unión Europea, este proyecto clasifica los sistemas de IA en diferentes niveles de riesgo (mínimo, limitado, alto e inaceptable) y establece obligaciones más estrictas para aquellos considerados de “alto riesgo”, entre estos, se incluyen expresamente los sistemas aplicados a la administración de justicia, por la incidencia directa que tienen en derechos fundamentales como la libertad personal y el acceso a la justicia, **el proyecto contempla exigencias de transparencia algorítmica, supervisión humana significativa, auditorías externas y explicabilidad de los sistemas**, buscando garantizar que la automatización no sustituya el juicio humano ni opaque la motivación judicial.

No obstante, el Proyecto de Ley 154 enfrenta desafíos relevantes, en primer lugar, su enfoque generalista podría no contemplar las particularidades de la función jurisdiccional penal, donde la exigencia de garantías es más estricta que en otros ámbitos administrativos, en segundo lugar, al encontrarse aún en trámite, su contenido está sujeto a modificaciones, lo que genera incertidumbre sobre cómo se regulará finalmente el uso de IA en la justicia, a pesar de estas limitaciones, el proyecto constituye un avance importante al reconocer la necesidad de establecer límites claros y condiciones mínimas para la aplicación de algoritmos en procesos judiciales.

**El Documento CONPES 4144 de 2025** constituye otro hito en la construcción de la política pública sobre inteligencia artificial en Colombia, a diferencia de la Ley 1581 y del Proyecto de Ley 154, no se trata de una norma de carácter vinculante, sino de un instrumento de planeación que actualiza la Estrategia Nacional de Inteligencia Artificial (CONPES 3975 de 2019), este documento incorpora un enfoque centrado en derechos humanos y ética, subrayando la importancia de la transparencia, la explicabilidad, la equidad y la supervisión humana en la adopción de IA en el sector público, en el ámbito judicial, el CONPES advierte sobre la necesidad de diseñar protocolos específicos que eviten la opacidad algorítmica y que aseguren que las decisiones automatizadas estén siempre sometidas al control humano.

El valor del CONPES 4144 radica en que proyecta una visión de Estado respecto a la incorporación de la IA en distintos sectores, incluyendo la justicia, si bien es cierto, sus lineamientos no son de obligatorio cumplimiento, sí marcan una hoja de ruta que condiciona el desarrollo legislativo y administrativo futuro, además, este documento recoge experiencias internacionales, particularmente de la Unión Europea y la UNESCO, que destacan la importancia de establecer estándares éticos universales en el uso de IA. En ese sentido, el CONPES sitúa a

Colombia dentro de una conversación global y abre la puerta a una eventual convergencia regulatoria con buenas prácticas internacionales.

Al articular estos tres referentes, la Ley 1581, el Proyecto de Ley 154 y el CONPES 4144, se evidencia un proceso de construcción normativa aún en fase embrionaria, caracterizado por avances significativos, pero también por vacíos y tensiones, mientras la Ley 1581 fija principios transversales de protección de datos, el Proyecto de Ley 154 busca establecer una regulación específica con un enfoque de riesgo, y el CONPES 4144 traza directrices éticas y de política pública, sin embargo, aún no existe una regulación sectorial que delimite expresamente cómo puede utilizarse la IA en la función jurisdiccional penal, ni cuáles son las condiciones concretas para garantizar la transparencia, la motivación judicial y la intervención decisoria del juez natural.

Es así como el desarrollo legislativo colombiano refleja una transición normativa: de un marco general de protección de datos hacia un incipiente diseño de políticas específicas para la inteligencia artificial, el desafío radica en consolidar estos avances dispersos en un cuerpo normativo integral que asegure que la IA no sustituya la racionalidad jurídica ni la independencia judicial, solo a través de una regulación robusta y garantista será posible armonizar innovación tecnológica y protección de derechos fundamentales, condición indispensable para la legitimidad de la justicia penal en un Estado Social y Democrático de Derecho.

### **2.2.3. Jurisprudencia nacional relevante: Sentencia T-323 de 2024 y Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura**

La jurisprudencia constitucional colombiana ha empezado a pronunciarse sobre el uso de inteligencia artificial en el ámbito judicial, reconociendo tanto su potencial transformador como los riesgos que implica para los derechos fundamentales, en este sentido, la Sentencia T-323 de

2024 constituye un hito, al analizar el uso de herramientas de inteligencia artificial , concretamente ChatGPT, por parte de un juez para la elaboración de una decisión judicial, la Corte Constitucional advirtió que la utilización de estas herramientas, sin un marco normativo específico ni garantías de transparencia y explicabilidad, podía afectar el debido proceso y, en particular, el derecho de defensa y la motivación de las sentencias, así, el Tribunal subrayó que la IA no puede sustituir la función argumentativa y garantista que caracteriza a la labor jurisdiccional.

Uno de los aportes centrales de la Sentencia T-323 de 2024 fue reafirmar que la sentencia judicial no es un acto meramente formal, sino un espacio de deliberación y de control constitucional en el que el juez debe exponer de manera clara las razones de su decisión, la Corte consideró que el uso acrítico de IA puede conducir a fallos carentes de fundamentación suficiente, en la medida en que los algoritmos operan sobre patrones estadísticos y no sobre criterios jurídicos ni principios constitucionales, en consecuencia, la Corte estableció que la inteligencia artificial solo puede ser utilizada como herramienta auxiliar, y nunca como sustituto de la motivación judicial ni de la función interpretativa propia del juez natural.

De manera complementaria, el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura representa el primer acto administrativo de alcance nacional que regula de forma explícita el uso de inteligencia artificial en la Rama Judicial colombiana, este acuerdo, expedido en 2024, fija límites estrictos al empleo de sistemas algorítmicos por parte de jueces y servidores judiciales, destacando que su utilización debe ser siempre secundaria y sujeta a la supervisión humana, el Consejo advirtió que los sistemas de IA no pueden tener incidencia autónoma en las decisiones de fondo, particularmente en materia penal, donde están en juego derechos fundamentales como la libertad personal y la presunción de inocencia.

El Acuerdo PCSJA24-12243 introdujo, además, parámetros de responsabilidad institucional, al exigir que cualquier uso de inteligencia artificial en la administración de justicia esté respaldado por protocolos de seguridad, registros de auditoría y mecanismos de trazabilidad que permitan verificar cómo se ha empleado la tecnología en el proceso judicial, de este modo, la norma busca garantizar que el juez no se limite a reproducir los resultados arrojados por un sistema algorítmico, sino que pueda ejercer un control crítico y motivar su decisión de acuerdo con los estándares constitucionales, en este punto, la regulación colombiana se alinea con las recomendaciones internacionales de la UNESCO (2021) y la Comisión Europea (AI Act), que enfatizan la necesidad de mantener un control humano significativo en las decisiones judiciales.

Tanto la Sentencia T-323 de 2024 como el Acuerdo PCSJA24-12243 coinciden en que el uso de IA en la justicia colombiana debe desarrollarse bajo un marco normativo que asegure la legitimidad de la función jurisdiccional, ambos instrumentos señalan que la independencia y la imparcialidad del juez no deben verse afectadas por la automatización, y que la transparencia y la explicabilidad de los sistemas son requisitos esenciales para proteger el debido proceso, al mismo tiempo, evidencian que Colombia avanza hacia un modelo regulatorio que, aunque incipiente, ya reconoce la necesidad de establecer límites claros frente a los riesgos asociados a la digitalización judicial.

Esta jurisprudencia evidencia que Colombia se encuentra en una etapa de regulación preventiva frente al uso de inteligencia artificial en la justicia penal, la Sentencia T-323 de 2024 establece criterios claros sobre los límites del uso de IA en la motivación de las decisiones judiciales, mientras que el Acuerdo PCSJA24-12243 define lineamientos administrativos destinados a asegurar el control humano y la trazabilidad de los sistemas, en conjunto, estos desarrollos constituyen un marco normativo inicial que permite orientar la implementación

tecnológica en la función jurisdiccional, sentando bases para la construcción de un protocolo regulatorio más sólido y estructurado.

#### **2.2.4. Estándares internacionales: directrices UNESCO, Comisión Europea y experiencias comparadas (Francia, UE, EE. UU.)**

La regulación internacional sobre inteligencia artificial ha adoptado un enfoque preventivo, promoviendo la utilización de la IA como herramienta complementaria en la toma de decisiones, sin reemplazar el juicio humano, organismos multilaterales como la UNESCO, en su Recomendación sobre la Ética de la Inteligencia Artificial (2021), han establecido principios de transparencia, responsabilidad y explicabilidad como condiciones mínimas para la implementación de sistemas algorítmicos en la administración de justicia, aunque este instrumento no es vinculante, sirve como marco orientador para los Estados, enfatizando que la adopción tecnológica debe estar sujeta a criterios de control, claridad en la toma de decisiones y equidad en la aplicación.

Por su parte, la Comisión Europea ha sido pionera en la construcción de un marco normativo específico, el AI Act, aprobado en 2024, clasifica el uso de sistemas de IA en el ámbito judicial como de “alto riesgo”, lo que implica la imposición de requisitos estrictos en materia de control humano, supervisión judicial y mecanismos de trazabilidad, esta clasificación parte de la premisa de que la justicia es un núcleo esencial del Estado de Derecho y, en consecuencia, no puede ser entregada a procesos automatizados sin contrapesos, el modelo europeo enfatiza la necesidad de auditorías periódicas de los algoritmos, la obligación de reportar sesgos detectados y la prohibición de utilizar sistemas opacos en la toma de decisiones jurisdiccionales.

En el caso particular de Francia, el Conseil d'État (2018) limitó expresamente el uso de algoritmos en la función jurisdiccional, señalando que las herramientas de IA pueden emplearse

como apoyo, pero nunca como sustituto de la labor del juez, este precedente consolidó un estándar que refuerza la idea de “control humano significativo” y ha marcado el debate europeo sobre la necesidad de preservar la motivación judicial como acto personalísimo, vinculado a la independencia judicial y a la legitimidad democrática de las decisiones, Francia, además, prohíbe expresamente la predicción automatizada del comportamiento judicial individual, como un mecanismo para evitar que la IA condicione las valoraciones de los jueces.

En los Estados Unidos, aunque no existe un marco unificado a nivel federal, la experiencia se ha concentrado en sistemas de “sentencing guidelines” y herramientas de predicción de riesgo como COMPAS, estas tecnologías han generado fuertes controversias debido a la falta de transparencia de sus algoritmos y la constatación de sesgos raciales y socioeconómicos en sus resultados, casos emblemáticos como *State v. Loomis* (2016) pusieron de relieve los peligros de incorporar sistemas opacos en la justicia penal, donde las decisiones afectan derechos fundamentales como la libertad personal, la experiencia estadounidense funciona como advertencia: la innovación tecnológica en justicia, sin salvaguardas claras, puede derivar en graves vulneraciones de principios constitucionales básicos.

En la Unión Europea, además del AI Act, existen documentos clave como la Carta Ética Europea sobre el uso de la Inteligencia Artificial en los Sistemas Judiciales (2018), elaborada por la Comisión Europea para la Eficiencia de la Justicia (CEPEJ), este texto identifica cinco principios rectores: respeto de los derechos fundamentales, no discriminación, calidad y seguridad, transparencia y neutralidad, y control del usuario, estos principios coinciden con la preocupación central de que la IA en justicia solo puede entenderse como un complemento, nunca como una sustitución de la función decisoria, y que cualquier automatización debe quedar bajo control humano efectivo.

La convergencia entre la UNESCO y la Comisión Europea demuestra que, en el plano internacional, se está consolidando un consenso normativo: la IA en la justicia penal representa un campo de alto riesgo que requiere límites estrictos, transparencia radical y supervisión constante, tanto las directrices como las experiencias comparadas coinciden en que los sistemas opacos o ininteligibles son incompatibles con la legitimidad democrática de la función jurisdiccional, este consenso sirve de referencia obligada para el contexto colombiano, donde aún persisten vacíos normativos y tensiones institucionales en torno al uso de nuevas tecnologías en el ámbito judicial.

La experiencia internacional evidencia que la implementación de la IA en la justicia penal exige un marco normativo robusto que garantice la protección de los derechos fundamentales, Colombia, al diseñar su propio protocolo de incorporación de estas tecnologías, no puede desconocer los estándares fijados por la UNESCO, la Comisión Europea y las experiencias de países como Francia y Estados Unidos, ignorar estos antecedentes equivaldría a repetir errores ya constatados y a arriesgar la erosión de principios esenciales como el debido proceso, la imparcialidad judicial y la función indelegable del juez natural.

### **2.3. Alcances y límites del uso de IA en la justicia penal desde el debido proceso**

La incorporación de la inteligencia artificial (IA) en la justicia penal genera un debate sobre el alcance de las herramientas algorítmicas en un ámbito donde la legitimidad de las decisiones y la protección de derechos fundamentales son prioritarias, en teoría, la IA podría aportar beneficios como eficiencia, reducción de cargas procesales y estandarización de criterios, aspectos especialmente relevantes en sistemas judiciales congestionados como el colombiano, sin embargo, el ámbito penal plantea exigencias singulares: las decisiones afectan bienes jurídicos de máxima jerarquía como la libertad personal, la honra o la vida, por lo que los

alcances de la IA deben evaluarse considerando no solo sus capacidades técnicas, sino también los límites normativos que imponen el debido proceso, la imparcialidad y la función indelegable del juez.

Un aspecto crítico en esta discusión es la motivación judicial, entendida como la explicación que legitima la decisión del juez y permite a las partes ejercer control sobre el proceso, la introducción de sistemas algorítmicos tipo “caja negra”, cuyos razonamientos internos no son comprensibles, plantea un riesgo para este principio: si una sentencia se apoya en un algoritmo que arroja resultados sin explicación verificable, el derecho de defensa y la posibilidad de contradicción se ven comprometidos, por ello, la motivación judicial no puede ser reemplazada por un output algorítmico, ya que perdería su carácter de razonamiento sometido a revisión y control.

La metáfora de la “caja negra algorítmica” refleja con claridad el dilema: cuanto más sofisticados son los modelos de IA, más difícil resulta explicar cómo se obtiene un resultado, esta opacidad puede ser aceptable en otros contextos, pero en la justicia penal, donde la transparencia y la claridad de los fundamentos son esenciales, delegar parte del razonamiento judicial a un sistema no explicable puede debilitar la legitimidad de las decisiones y limitar la comprensión de las partes involucradas.

Garantizar transparencia y explicabilidad en el uso de IA es, por tanto, una condición para la validez de las decisiones judiciales, el debido proceso requiere que los fundamentos de las sentencias sean comprensibles y auditables, algo que organismos internacionales como la Comisión Europea y la UNESCO recomiendan para sistemas de IA en la justicia, la explicabilidad no se limita a revelar códigos o fórmulas matemáticas, sino que implica traducir el razonamiento algorítmico a un lenguaje comprensible para jueces, abogados y ciudadanos, es así

como se mantiene la posibilidad de crítica y supervisión, evitando que la IA genere procesos opacos o excesivamente tecnocráticos, y asegurando que la innovación tecnológica se integre de manera compatible con los marcos normativos existentes.

El tercer elemento clave en este debate es el control humano significativo, no basta con que la IA sea transparente y explicable; es indispensable que las decisiones finales recaigan siempre en un juez que asuma responsabilidad plena sobre su razonamiento, juzgar es un acto indelegable, vinculado a la legitimidad del poder judicial y a la protección de derechos fundamentales, la IA puede sugerir, orientar o incluso predecir, pero nunca reemplazar el juicio humano, que requiere deliberación, prudencia y contextualización, la automatización sin control humano transforma al juez en un mero ejecutor de instrucciones técnicas y vacía de contenido la función decisoria.

En Colombia, persisten vacíos y ambigüedades normativas que dificultan delimitar los usos legítimos de la IA en la justicia penal, aunque la Ley 1581 de 2012 sobre protección de datos y el Proyecto de Ley 154 de 2024 ofrecen lineamientos generales, no existe un marco específico que regule directamente la aplicación de estas tecnologías en la elaboración de sentencias, el Acuerdo PCSJA24-12243 establece restricciones generales, pero no proporciona un protocolo detallado para su implementación en funciones jurisdiccionales, estos vacíos generan riesgos: la introducción de IA podría hacerse de manera improvisada, sin criterios uniformes de transparencia ni mecanismos de control claros.

La falta de regulación también genera incertidumbre práctica, por ejemplo, ¿puede un juez fundamentar parcialmente una decisión en recomendaciones algorítmicas? ¿Qué grado de responsabilidad tendría ante un error o sesgo del sistema?, en Colombia, estas preguntas aún carecen de respuesta, dejando margen para soluciones dispares que podrían afectar la coherencia

y equidad en la administración de justicia, el derecho comparado evidencia la necesidad de establecer límites claros: tanto la Unión Europea como Francia consideran de alto riesgo el uso de IA en decisiones judiciales, imponiendo requisitos estrictos de supervisión humana y trazabilidad.

La introducción de IA plantea retos directos a la legitimidad democrática de la función judicial, el juez no es solo un aplicador de normas, sino un garante de la Constitución y un intérprete responsable ante la sociedad, delegar esta función en sistemas algorítmicos genera tensiones sobre la naturaleza del poder judicial, cuya legitimidad depende de la deliberación pública, la motivación racional y la rendición de cuentas, una justicia dictada por algoritmos opacos puede percibirse como técnica, distante e impersonal, debilitando la confianza de los ciudadanos en las instituciones y afectando la percepción de legitimidad del Estado.

El riesgo no es solo simbólico, si las personas sienten que las decisiones que afectan su libertad o su honra son producto de un cálculo técnico incomprensible, la justicia pierde legitimidad como espacio de resolución racional de conflictos, la democracia constitucional exige que las sentencias sean comprensibles y puedan ser cuestionadas en instancias superiores, algo imposible si el razonamiento se oculta tras modelos matemáticos opacos, de esta manera, el problema de la legitimidad no es accesorio, sino central: la justicia no solo debe ser justa, sino también percibida como tal, y en ello la transparencia y la motivación juegan un papel esencial.

Por otro lado, el discurso de la eficiencia que suele acompañar la introducción de tecnologías en la justicia no puede desplazar las exigencias de legitimidad y garantías procesales, la congestión judicial, aunque es un problema real, no justifica la sustitución del razonamiento humano por resultados algorítmicos, la eficiencia es deseable, pero nunca a costa de sacrificar la imparcialidad, el derecho de defensa o la motivación judicial, en este sentido, la IA debe

entenderse como un instrumento auxiliar y no como un atajo para resolver problemas estructurales de la justicia que, en muchos casos, responden a falta de inversión, corrupción o ineficiencia administrativa.

Los retos que plantea la IA también se relacionan con la desigualdad estructural que caracteriza al contexto colombiano, el riesgo de sesgo algorítmico no es un problema abstracto, sino una amenaza concreta en un país atravesado por brechas socioeconómicas, étnicas y regionales, un sistema entrenado con datos sesgados podría reproducir patrones de discriminación y reforzar desigualdades históricas, afectando especialmente a poblaciones vulnerables, en este escenario, el control humano significativo y la revisión crítica se vuelven indispensables para garantizar que la tecnología no se convierta en un nuevo mecanismo de exclusión.

Es así como los alcances de la inteligencia artificial en la justicia penal colombiana solo pueden comprenderse a la luz de sus límites constitucionales, la motivación judicial, la transparencia, la explicabilidad y el control humano significativo no son simples requisitos técnicos, sino garantías esenciales del debido proceso, los vacíos y ambigüedades normativas, sumados a los riesgos de opacidad y sesgo, exigen un abordaje crítico y garantista que coloque en el centro la legitimidad democrática de la función jurisdiccional de lo contrario, la promesa de modernización podría transformarse en un retroceso, debilitando los fundamentos mismos del Estado Social de Derecho.

La revisión del marco teórico-normativo deja en evidencia que el uso de la inteligencia artificial en la justicia penal colombiana se mueve en un terreno de tensiones constantes entre innovación tecnológica y límites constitucionales, si bien es cierto es cierto, es indudable que estas herramientas pueden aportar eficiencia y estandarización en un sistema judicial marcado

por la congestión, su integración en la fase de elaboración de sentencias encuentra barreras insoslayables en principios como la motivación judicial, la transparencia y el control humano significativo, la opacidad de los algoritmos y los vacíos regulatorios existentes revelan que, en el estado actual, la IA difícilmente puede considerarse un sustituto legítimo del razonamiento judicial sin poner en riesgo el debido proceso y la dignidad humana como ejes estructurantes del Estado Social de Derecho.

De igual forma, el análisis permitió identificar que los vacíos normativos y las ambigüedades del ordenamiento jurídico colombiano no solo generan inseguridad jurídica, sino que también pueden afectar la legitimidad democrática de las decisiones judiciales, en este marco, la inteligencia artificial solo puede integrarse en la justicia penal como herramienta de apoyo técnico, subordinada al juez y su decisión final, nunca como reemplazo de su función, así, se establecen criterios mínimos para avanzar hacia un protocolo de implementación que permita aprovechar los beneficios de la tecnología, como eficiencia y manejo de información, sin comprometer la independencia, imparcialidad y motivación de las sentencias, asegurando que la modernización judicial se alinee con los principios constitucionales y la función indelegable del juez.

### **3. CAPÍTULO II: INTELIGENCIA ARTIFICIAL Y DEBIDO PROCESO PENAL: ANÁLISIS COMPARADO Y DIÁLOGO JURISPRUDENCIAL**

#### **3.1. Jurisprudencia colombiana sobre IA y garantías procesales**

La jurisprudencia nacional ha sido clara en señalar que el uso de la inteligencia artificial en la justicia no es neutro, toda herramienta tecnológica incorpora lógicas, sesgos y limitaciones que pueden impactar de manera directa en la imparcialidad y en el acceso a la justicia, de allí que el debate jurídico no se limite a una discusión técnica, sino que exija un análisis profundo de compatibilidad constitucional, este punto de partida ha orientado las primeras decisiones en las que la Corte Constitucional y el Consejo Superior de la Judicatura se han pronunciado sobre el tema.

Uno de los aspectos que más se ha resaltado en el contexto colombiano es que la función de juzgar es indelegable, lo cual implica que ningún algoritmo puede reemplazar al juez como garante de derechos, esta premisa ha sido reiterada por la doctrina y la jurisprudencia, subrayando que si bien es cierto la inteligencia artificial puede desempeñar funciones auxiliares, no puede asumir un rol decisorio en la valoración probatoria ni en la elaboración de sentencias penales, el núcleo del poder jurisdiccional exige un razonamiento humano, argumentativo y transparente, que ninguna máquina puede suplir.

A la par de esta premisa, los tribunales colombianos han comenzado a identificar ciertos riesgos asociados a la automatización judicial, entre ellos se encuentran la pérdida de trazabilidad de la motivación, la opacidad en los procesos de decisión algorítmica y la ausencia de responsabilidad clara en caso de error, estos riesgos, ya advertidos por autores como Luis García

(2023) en sus análisis críticos sobre la digitalización de la justicia, han sido reconocidos en el ámbito nacional como amenazas a la legitimidad del poder judicial.

De esta manera, la jurisprudencia colombiana no ha adoptado una postura radical de rechazo a la inteligencia artificial, por el contrario, las decisiones más recientes han abierto la puerta a un uso instrumental de estas tecnologías, siempre que se encuentren bajo condiciones mínimas de transparencia, explicabilidad y supervisión humana efectiva, en otras palabras, se admite la innovación tecnológica en el ámbito judicial, pero bajo el principio rector de que esta debe estar subordinada a las garantías constitucionales.

En este sentido, resulta fundamental comprender que la intervención de la Corte Constitucional y de la Judicatura no busca frenar la modernización, sino orientarla, lo que se persigue es establecer límites claros que eviten la delegación indebida de funciones jurisdiccionales en sistemas opacos, a la vez que se garantice que la justicia digital no se traduzca en decisiones despersonalizadas ni en violaciones al derecho de defensa, la finalidad última es que la IA sirva como herramienta de apoyo, sin desplazar la centralidad del juez en el proceso penal.

Al mismo tiempo, las decisiones jurisprudenciales han evidenciado la falta de un marco regulatorio integral en Colombia, más allá de leyes generales como la Ley 1581 de 2012 sobre protección de datos o iniciativas recientes como el Proyecto de Ley 154 de 2024, lo cierto es que aún no existe una normativa específica que regule el uso de IA en la administración de justicia, este vacío ha obligado a que sean las altas cortes y órganos de gobierno judicial quienes marquen las primeras pautas, lo que refleja la importancia de la jurisprudencia como fuente creadora de derecho en esta materia.

La jurisprudencia colombiana sobre inteligencia artificial y debido proceso penal ha comenzado a establecer lineamientos iniciales que priorizan la protección del derecho al debido proceso frente a los desafíos de la digitalización, aunque estas decisiones aún son fragmentarias, existe un consenso preliminar: la IA solo puede emplearse como herramienta auxiliar, bajo supervisión humana y con mecanismos de transparencia que permitan verificar la legitimidad de las decisiones judiciales, este marco incipiente sienta las bases para la construcción de regulaciones más claras y coherentes, orientadas a compatibilizar la eficiencia tecnológica con la preservación de los derechos fundamentales y la integridad de la función jurisdiccional.

### **3.1.1. Sentencia T-323 de 2024 (uso de ChatGPT y debido proceso).**

La Sentencia T-323 de 2024 de la Corte Constitucional marcó un hito en el debate sobre el uso de herramientas de inteligencia artificial en la justicia colombiana, pues abordó de manera directa el caso de un juez que recurrió a ChatGPT como apoyo en la motivación de una decisión judicial, aunque se trató de un asunto de tutela que cuestionaba la validez de dicha actuación, la trascendencia de la sentencia radica en la postura de la Corte frente a los alcances y límites del empleo de sistemas algorítmicos en funciones jurisdiccionales.

El primer aspecto relevante que destaca la Corte es que el uso de ChatGPT, en tanto modelo de lenguaje sin certificación jurídica ni control estatal, introduce riesgos de opacidad, inexactitud y falta de trazabilidad, en su razonamiento, la Corte enfatizó que una decisión judicial no puede apoyarse en fuentes cuya confiabilidad y verificabilidad sean dudosas, pues ello compromete el derecho a la motivación judicial y el acceso a una justicia razonada, este pronunciamiento se alinea con las críticas doctrinales de De Luis García (2023), quien ha advertido sobre la “caja negra algorítmica” como amenaza al derecho de defensa y al debido proceso.

La sentencia también insistió en que la función jurisdiccional es indelegable, recordando que el artículo 228 de la Constitución establece que la administración de justicia debe ejercerse en estricto apego a la ley y a los principios de transparencia y motivación, al utilizar ChatGPT como insumo sin advertir ni justificar su alcance, el juez involucrado trasladó parte de su responsabilidad a un sistema que no ofrece garantías de veracidad ni de control humano significativo, con ello, se desconoció la centralidad del juez como garante de los derechos fundamentales.

En este punto, la Corte resaltó la importancia del principio de publicidad y control Inter partes: si el juez acude a una herramienta tecnológica para fundamentar una decisión, las partes deben tener conocimiento expreso de ello, con el fin de ejercer contradicción y debate, el hecho de que el uso de ChatGPT no fuera transparente en el proceso vulneró el derecho de defensa de la parte accionante, ya que impidió discutir la validez de los argumentos proporcionados por la herramienta.

Otro elemento fundamental abordado en la T-323 de 2024 fue el de la fiabilidad epistémica, la Corte advirtió que los modelos de lenguaje como ChatGPT son sistemas entrenados a partir de grandes volúmenes de datos, pero carecen de comprensión jurídica y de criterios normativos vinculantes, por ello, no pueden equipararse a fuentes doctrinales, jurisprudenciales o normativas, cuya autoridad se deriva de su origen institucional, esta diferenciación permitió subrayar que la utilización de IA generativa en contextos judiciales debe ser evaluada con criterios mucho más estrictos que los que rigen en otros sectores.

La Corte no se limitó a sancionar la actuación del juez, sino que formuló orientaciones generales sobre el uso de IA en la justicia, señaló que, si bien es cierto la inteligencia artificial puede tener un papel auxiliar en la gestión de expedientes, en la sistematización de

jurisprudencia o en la búsqueda de precedentes, no puede intervenir en la elaboración de sentencias sin un marco normativo y protocolos claros que regulen su aplicación, de esta manera, la sentencia se convirtió en un precedente estructural que obliga a la Rama Judicial a diseñar lineamientos de uso responsable.

Un aporte valioso de esta decisión fue su vinculación con estándares internacionales, la Corte recordó que organismos como la Comisión Europea (AI Act, 2021) y la UNESCO (Recomendación de 2021 sobre ética de la IA) han catalogado el uso de sistemas de IA en funciones jurisdiccionales como de “alto riesgo”, lo que implica exigencias reforzadas de transparencia y supervisión humana, al traer estos referentes, la Corte situó la discusión colombiana en un contexto global, mostrando que el debate no es aislado sino parte de una preocupación compartida por la comunidad internacional.

Asimismo, la T-323 de 2024 dejó en evidencia los vacíos regulatorios existentes en Colombia, pues si bien es cierto existen normas generales como la Ley 1581 de 2012 sobre protección de datos y políticas de digitalización como el CONPES 4144 de 2025, lo cierto es que no hay una regulación específica sobre el uso de IA en procesos judiciales, ante esta ausencia, la Corte adoptó un rol activo de interpretación garantista, estableciendo que, en tanto no exista ley expresa, deben primar los principios constitucionales de dignidad humana, debido proceso y juez natural.

La decisión también es ilustrativa porque articula una visión preventiva, la Corte no esperó a que se produjeran consecuencias graves para pronunciarse, sino que actuó con base en el principio de precaución, reconociendo que el uso irresponsable de IA puede generar afectaciones irreparables a los derechos fundamentales, este enfoque precautorio se alinea con lo defendido

por Pasquale (2015), quien sostiene que la opacidad algorítmica representa una amenaza estructural para la legitimidad democrática si no se controla desde su origen.

De igual manera, la sentencia destacó la necesidad de que los jueces reciban formación tanto técnica como práctica frente al uso de IA, señalando que no basta con establecer límites normativos; es indispensable que los operadores judiciales comprendan los alcances, riesgos y capacidades de estas herramientas, de modo que puedan utilizarlas como apoyo a su razonamiento jurídico y no como sustitutos automáticos de sus decisiones, esta recomendación enfatiza la importancia de dotar a los jueces de criterios claros para evaluar y supervisar los resultados de los sistemas algorítmicos en el contexto de la justicia penal.

La Sentencia T-323 de 2024 envía un mensaje claro: la modernización judicial no puede avanzar a costa de la integridad del proceso penal, en un contexto de desconfianza institucional y desigualdades estructurales, la sustitución de la función jurisdiccional por IA sería un riesgo para la validez y legitimidad de las decisiones judiciales, la Corte subrayó que cualquier innovación tecnológica debe operar dentro del marco constitucional, destacando que la justicia no puede reducirse a cálculos estadísticos, sino que requiere decisiones fundamentadas, prudentes y ajustadas al contexto de cada caso..

Es por ello que la Sentencia T-323 de 2024 constituye un precedente cardinal en la jurisprudencia colombiana, pues establece un marco mínimo para el uso de inteligencia artificial en la justicia penal, su valor radica en que no rechaza de plano la innovación, pero fija límites claros: transparencia, explicabilidad, control humano y respeto irrestricto al debido proceso, estos lineamientos configuran una base indispensable para el diseño de futuros protocolos normativos, asegurando que la IA se utilice como herramienta auxiliar y no como sustituto del juez.

### **3.1.2. Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura.**

El Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura, expedido en 2024, constituye uno de los primeros intentos normativos internos de la Rama Judicial en Colombia para establecer parámetros sobre el uso de herramientas de inteligencia artificial en los despachos judiciales, a diferencia de la Sentencia T-323 de 2024, que se pronunció frente a un caso concreto, este Acuerdo se presenta como un instrumento de carácter administrativo y preventivo, cuyo propósito principal es delimitar los alcances y límites de la IA dentro de las funciones judiciales.

Uno de los aportes más significativos de este acuerdo es que reafirma la función indelegable del juez en la elaboración de providencias judiciales, el documento es enfático al señalar que ninguna herramienta algorítmica puede reemplazar la deliberación racional, la valoración probatoria ni la decisión final del funcionario judicial, en este sentido, el Consejo establece un marco normativo claro que previene la automatización de decisiones judiciales sin supervisión humana, en coherencia con el artículo 228 de la Constitución y con la necesidad de preservar la legitimidad y la responsabilidad en el ejercicio de la función jurisdiccional.

El Acuerdo también enfatiza el principio de transparencia, estableciendo que, cuando un despacho judicial utilice alguna aplicación de IA con fines de apoyo documental, sistematización de jurisprudencia o búsqueda de información, esta circunstancia debe ser explícitamente mencionada en las providencias, esta obligación asegura que las partes tengan conocimiento de los insumos tecnológicos que influyeron en la construcción del fallo, permitiendo que puedan ejercer control y contradicción sobre la información utilizada en el proceso.

Asimismo, el Consejo introdujo una directriz sobre la responsabilidad exclusiva del juez, incluso si se emplean herramientas de IA, el juez conserva la obligación de verificar, validar y

contextualizar los resultados generados por estas, de este modo, se evita que los algoritmos se conviertan en un “atajo epistémico” para fundamentar sentencias, tal como advierte De Luis García (2023) al señalar que la motivación judicial requiere razonamiento propio y no puede basarse en enunciados impersonales producidos por sistemas de lenguaje automático.

Un elemento innovador del Acuerdo PCSJA24-12243 es que introduce la noción de uso auxiliar de la inteligencia artificial, diferenciándola del uso decisorio, mientras que lo primero se admite como apoyo en tareas logísticas o de búsqueda, lo segundo se prohíbe expresamente en la elaboración de fallos, esta distinción es clave, pues aclara que la IA puede ser una herramienta útil en la administración judicial sin llegar a desplazar al juez en su función esencial de resolver controversias conforme a derecho.

El Acuerdo también reconoce los riesgos de sesgo y opacidad algorítmica, señalando que los jueces deben ser conscientes de que las plataformas de IA no están diseñadas con criterios jurídicos, sino estadísticos, y que, en consecuencia, pueden reproducir prejuicios o errores, en este punto, el Consejo Superior de la Judicatura coincide con las advertencias de autores como O’Neil (2016) y Pasquale (2015), quienes han denunciado los peligros de delegar decisiones sensibles a sistemas opacos que no permiten trazabilidad ni control social.

Otro aspecto relevante es que el Acuerdo se expide en un contexto de creciente presión por modernizar la justicia, la Rama Judicial colombiana ha impulsado procesos de digitalización y la implementación de expedientes electrónicos, considerando la IA como un paso adicional en esta transformación, sin embargo, el Consejo señala que la modernización no puede comprometer la legalidad ni la función indelegable del juez, por lo que condiciona el uso de estas tecnologías a un marco que asegure transparencia, explicabilidad y supervisión humana en las decisiones judiciales.

En términos prácticos, el Acuerdo PCSJA24-12243 busca unificar criterios internos para evitar que los jueces actúen de manera aislada y heterogénea frente al uso de IA, la falta de reglas claras había generado situaciones como la que dio origen a la Sentencia T-323 de 2024, donde un juez utilizó ChatGPT sin transparencia ni control, comprometiendo derechos fundamentales, con esta normativa, el Consejo procura cerrar ese vacío institucional, otorgando lineamientos mínimos que deben seguirse de manera uniforme en todo el país.

El Acuerdo también tiene un componente pedagógico y preventivo, más allá de la prohibición de ciertos usos, insta a la Rama Judicial a desarrollar programas de capacitación para los jueces y funcionarios judiciales en materia de alfabetización digital y ética del uso de IA, este énfasis en la formación resulta coherente con la necesidad de que los operadores jurídicos comprendan la naturaleza de estas herramientas, sus alcances y sus riesgos, evitando un uso acrítico o excesivamente confiado.

Sin embargo, pese a sus avances, el Acuerdo evidencia algunas limitaciones, en primer lugar, carece de fuerza legislativa, pues se trata de una disposición administrativa que regula internamente la Rama Judicial, pero no sustituye la necesidad de una ley estatutaria o de una regulación más robusta sobre IA y justicia, en segundo lugar, su alcance se limita a la función judicial, dejando por fuera otros escenarios donde la IA puede incidir, como la labor de fiscales, defensores públicos o peritos.

A pesar de estas limitaciones, el Acuerdo PCSJA24-12243 constituye un primer paso relevante hacia la construcción de un marco normativo sobre el uso de IA en la justicia penal, al delimitar claramente lo permitido y lo prohibido, contribuye a preservar la legitimidad de las decisiones judiciales y envía un mensaje inequívoco: la IA no reemplaza al juez, sino que solo puede actuar como herramienta auxiliar bajo supervisión humana.

Este Acuerdo refuerza la tesis de que la modernización tecnológica debe ir acompañada de salvaguardas constitucionales, su valor no radica únicamente en establecer restricciones, sino en articular una visión de la justicia en la que la tecnología esté subordinada a los principios del Estado Social de Derecho, así, se convierte en un referente fundamental para futuras reformas legislativas y para el diseño de protocolos de implementación de IA enmarcados en los principios de transparencia, responsabilidad y respeto por los derechos fundamentales.

### **3.2. Estándares interamericanos de la Corte IDH y la CADH**

La Convención Americana sobre Derechos Humanos (CADH), en particular su artículo 8 sobre garantías judiciales, establece criterios para evaluar innovaciones tecnológicas en los sistemas judiciales, este artículo reconoce el derecho de toda persona a ser juzgada por un juez independiente e imparcial, dentro de un plazo razonable y con los procedimientos adecuados para determinar sus derechos y obligaciones, desde esta perspectiva, la incorporación de inteligencia artificial (IA) en procesos de decisión penal debe analizarse tanto según la normativa interna como los compromisos internacionales que Colombia ha asumido como Estado parte de la CADH, en este marco, cualquier delegación de funciones judiciales a sistemas algorítmicos requiere límites claros que aseguren la correcta atribución de decisiones y eviten riesgos de afectación a los derechos reconocidos por el derecho interamericano..

La Corte Interamericana de Derechos Humanos (Corte IDH) ha desarrollado una interpretación robusta del artículo 8 de la CADH, señalando que el debido proceso es un derecho de aplicación amplia y no puede restringirse a formalidades procesales, en casos como Maldonado Vargas y otros vs. Chile (2015), el tribunal destacó que las decisiones judiciales deben ser comprensibles y motivadas, ya que la motivación va en contra de la arbitrariedad, este estándar resulta central para la discusión sobre IA: si una sentencia es emitida o influenciada por

un sistema opaco, donde las razones detrás de la decisión no son claras, se estaría incurriendo en una vulneración directa al principio de motivación y, por ende, al debido proceso interamericano.

Otro aspecto relevante es la obligación de los Estados de garantizar la imparcialidad e independencia judicial, principios reiterados en fallos como *Apitz Barbera y otros vs. Venezuela* (2008), la Corte ha insistido en que la justicia no solo debe ser imparcial, sino también parecerlo, es decir, debe generar confianza ciudadana, en este sentido, si los ciudadanos perciben que la justicia penal se encuentra en manos de algoritmos que no son auditables o que replican sesgos preexistentes, la legitimidad democrática del sistema judicial se ve socavada, la Corte IDH ha recordado en múltiples ocasiones que la legitimidad del poder judicial es un componente esencial del Estado de Derecho, y la automatización sin controles efectivos podría minar esa confianza.

Asimismo, la Corte IDH ha establecido estándares claros sobre el derecho de defensa y la igualdad de armas, reconociendo que cualquier limitación a estos principios procesales puede afectar la integridad del acceso a la justicia, en un escenario donde la IA participa en la elaboración de sentencias, surge la cuestión de cómo la defensa técnica puede contradecir o cuestionar los razonamientos de un algoritmo cuya lógica interna no es accesible, la falta de transparencia y explicabilidad de los sistemas algorítmicos impacta directamente sobre el principio de contradicción y la capacidad de las partes para ejercer un control efectivo sobre las decisiones judiciales, condicionando la correcta aplicación de los derechos reconocidos en la CADH.

La jurisprudencia interamericana también ha sido enfática en la necesidad de preservar la dignidad humana como principio rector del proceso judicial, en decisiones como *Gomes Lund y otros vs. Brasil* (2010), se afirmó que el respeto a la dignidad y a la condición humana de los procesados es un límite material al ejercicio del poder punitivo, desde esta perspectiva, reducir la

decisión judicial a un cálculo estadístico o a una correlación probabilística sin atender a la singularidad del caso y de la persona procesada constituye una forma de deshumanización incompatible con el marco interamericano, la IA, sin controles claros, corre el riesgo de producir sentencias despersonalizadas y, en consecuencia, contrarias a la dignidad que debe guiar todo proceso penal.

Con relación a la motivación judicial, la Corte IDH ha señalado que esta se vincula directamente con la legitimidad de la decisión y con el control democrático de la función jurisdiccional, los ciudadanos tienen derecho a comprender las razones por las cuales se les condena o absuelve, lo que implica que el Estado debe asegurar que las decisiones judiciales sean accesibles, claras y argumentadas, si un juez se limita a reproducir recomendaciones algorítmicas sin ejercer control crítico, se compromete el principio de motivación y, con ello, la posibilidad de revisión y contradicción efectiva de la decisión, un aspecto especialmente relevante en materia penal, donde están en juego bienes jurídicos fundamentales como la libertad personal.

Los estándares interamericanos también han enfatizado que los Estados deben adoptar medidas positivas para proteger los derechos fundamentales frente a nuevos riesgos y contextos, en el marco de la digitalización de la justicia, esto implica que la IA debe funcionar exclusivamente como apoyo técnico, sin reemplazar la decisión judicial, cualquier intento de delegar de manera automática la función del juez constituiría una vulneración del principio de juez natural, tal como lo ha interpretado la Corte IDH, al comprometer la autoridad y responsabilidad indelegable del funcionario judicial.

Es necesario subrayar que, para el sistema interamericano, el debido proceso constituye un estándar principal que orienta la actuación del Estado frente a los desafíos contemporáneos, por

ello, la eficiencia o la modernización tecnológica no pueden justificar restricciones a las garantías judiciales, la Corte IDH ha señalado que cualquier innovación en la justicia debe respetar principios fundamentales como transparencia, explicabilidad y control humano, de manera que la incorporación de IA en la justicia penal en Colombia se alinee con los compromisos internacionales asumidos bajo la CADH, asegurando que los principios esenciales de la función jurisdiccional no sean sacrificados en nombre de la tecnificación.

### **3.1.3. Principios esenciales de la justicia procesal (art. 8 CADH).**

El artículo 8 de la Convención Americana sobre Derechos Humanos establece los principios fundamentales que guían la administración de justicia, reconociendo la importancia de un proceso justo y equitativo y delimitando la obligación del Estado de asegurar que las personas sean juzgadas por tribunales competentes, independientes e imparciales, en el contexto de la inteligencia artificial en la justicia penal, este marco permite analizar si los sistemas algorítmicos cumplen con los principios de independencia judicial y de igualdad procesal entre las partes, sin asumir que la IA pueda sustituir la deliberación humana.

Entre los principios esenciales se encuentra la asignación de un juez natural y la conducción del proceso bajo reglas claras y previamente establecidas, la participación de sistemas algorítmicos en la toma de decisiones judiciales plantea la necesidad de mantener la función indelegable del juez, asegurando que cualquier apoyo tecnológico respete los principios de autoridad legítima y responsabilidad individual en la decisión final.

El derecho a la defensa y a ser oído en condiciones de igualdad constituye otro principio rector, en escenarios en los que la IA interviene en la elaboración de decisiones, surge la necesidad de que sus resultados sean comprensibles y transparentes, de manera que las partes

puedan interactuar efectivamente con el proceso, contradecir argumentos y aportar evidencia, sin que la lógica interna del algoritmo genere asimetrías que afecten la igualdad procesal.

Asimismo, el artículo 8 establece principios relativos a plazos razonables, motivación de las decisiones y publicidad, la incorporación de tecnologías digitales puede contribuir a la eficiencia y la reducción de la congestión judicial, pero estos avances deben compatibilizarse con la claridad, inteligibilidad y racionalidad de las decisiones judiciales, así como con la transparencia en su comunicación, de la misma manera, la igualdad procesal requiere atención frente a posibles sesgos algorítmicos que reproduzcan desigualdades existentes; la supervisión y explicabilidad de los sistemas se convierten así en condiciones esenciales para que la IA funcione como un instrumento de apoyo que respete los principios fundamentales de la justicia procesal.

#### **3.1.4. Casos relevantes sobre imparcialidad y transparencia.**

Uno de los ejes más sensibles en el marco del debido proceso es la imparcialidad judicial, entendida no solo como la ausencia de sesgo subjetivo por parte del juez, sino también como la existencia de garantías objetivas que transmitan confianza en la neutralidad del proceso, la Corte Interamericana de Derechos Humanos ha sostenido en casos como *Apitz Barbera y otros vs. Venezuela* (2008) que la imparcialidad debe evaluarse bajo una doble dimensión: la subjetiva, que se refiere a la convicción interna del juzgador, y la objetiva, que exige que el tribunal aparezca como independiente y libre de presiones externas, cuando se introduce la inteligencia artificial en la justicia penal, esta doble dimensión se ve tensionada, pues la opacidad de los algoritmos y la posibilidad de sesgos en su entrenamiento comprometen tanto la percepción como la realidad de la imparcialidad.

El caso *Maldonado Vargas y otros vs. Chile* (2015) ofrece un precedente importante sobre la necesidad de motivar las decisiones judiciales de manera que sean comprensibles y verificables

para las partes, allí, la Corte IDH subrayó que la transparencia no puede ser un requisito formal, sino sustantivo: las razones del juez deben ser claras, suficientes y permitir el control ciudadano, si un juez adopta recomendaciones de un sistema de IA sin explicar cómo influyeron en su razonamiento, se produciría una situación contraria a este estándar, dando lugar a decisiones que, aunque rápidas, resultan carentes de legitimidad democrática.

En el ámbito europeo, el caso *López Ribalda y otros vs. España* (TEDH, 2019) es ilustrativo respecto al equilibrio entre tecnología y derechos fundamentales, si bien es cierto es cierto, se trató de vigilancia laboral mediante cámaras ocultas, el Tribunal Europeo de Derechos Humanos enfatizó que la tecnología no puede ser utilizada de manera opaca ni desproporcionada frente a los derechos de los individuos, esta lógica resulta perfectamente trasladable al debate sobre IA judicial: la transparencia y la proporcionalidad son límites infranqueables cuando está en juego la libertad personal de los procesados.

En Colombia, la jurisprudencia constitucional ha hecho énfasis en la necesidad de que la justicia no solo sea imparcial, sino que lo parezca, como se mencionó atrás la Sentencia T-323 de 2024, que examinó el uso de ChatGPT en un fallo de tutela, evidenció que la utilización acrítica de estas herramientas puede generar dudas razonables sobre la independencia del juez, aunque en ese caso se trataba de una referencia no decisiva, la Corte advirtió que la dependencia excesiva de herramientas algorítmicas sin supervisión humana puede erosionar la confianza pública en la administración de justicia.

El Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura va en la misma dirección al establecer límites al uso de inteligencia artificial en los despachos judiciales, allí se prohíbe expresamente delegar en sistemas algorítmicos funciones decisorias y se insiste en que toda utilización debe ser auxiliar y complementaria, este acuerdo, al exigir transparencia en el uso de

IA, responde a la preocupación de que la opacidad algorítmica pueda afectar la imparcialidad objetiva, ya que una parte procesal no podría conocer ni controvertir las razones efectivas detrás de la decisión.

En el plano comparado, resulta relevante el Conseil d'État de Francia (2018), que limitó la utilización de algoritmos en la función jurisdiccional, sosteniendo que la justicia no puede convertirse en un cálculo automatizado desprovisto de control humano, la decisión enfatizó que la motivación de la sentencia es inseparable de la función del juez como garante de los derechos fundamentales, este pronunciamiento coincide con lo que la doctrina crítica, como De Luis García (2023), denomina “el riesgo de la caja negra”, una metáfora para advertir cómo la opacidad tecnológica amenaza la exigencia constitucional de fundamentación de las decisiones.

Por su parte, en Estados Unidos, los litigios en torno al uso del software COMPAS para evaluar riesgo de reincidencia han sido emblemáticos, en casos como *State v. Loomis* (2016), la Corte Suprema de Wisconsin reconoció que, si bien es cierto estas herramientas pueden servir de apoyo, su utilización sin explicabilidad puede atentar contra el derecho al debido proceso, en particular, se advirtió que los procesados no tenían forma de cuestionar los criterios de riesgo que el sistema aplicaba, lo que constituía una afectación a la transparencia y a la igualdad de armas.

Estos precedentes reflejan un consenso internacional emergente: la imparcialidad y la transparencia no son negociables en un Estado de derecho, aunque la IA pueda introducir beneficios de eficiencia, si no se garantiza la trazabilidad de las decisiones y la posibilidad de control ciudadano y procesal, se estaría afectando el núcleo esencial de las garantías judiciales mínimas, la legitimidad democrática de la justicia penal depende de que las personas perciban que las decisiones provienen de jueces humanos que deliberan, motivan y asumen responsabilidad sobre lo resuelto.

Además, debe recordarse que la imparcialidad no es un valor abstracto, sino una condición material que impacta directamente en la confianza ciudadana, un tribunal que delega excesivamente en algoritmos no solo puede tomar decisiones sesgadas, sino también transmitir la idea de que la justicia está deshumanizada, como advierte Ferrajoli (2005), el poder punitivo solo es legítimo si se ejerce bajo reglas estrictas que garanticen la protección de los derechos humanos, la transparencia judicial, en este sentido, se convierte en una salvaguarda indispensable contra los excesos de la automatización.

Los casos nacionales e internacionales revisados permiten establecer una advertencia clara: cualquier implementación de IA en la justicia penal debe garantizar imparcialidad objetiva y subjetiva, motivación judicial comprensible y plena transparencia en la utilización de algoritmos, sin estas condiciones, la función jurisdiccional corre el riesgo de ser deslegitimada y de generar una vulneración directa al artículo 8 de la CADH y a los principios fundamentales del Estado Social de Derecho.

### **3.3. Experiencias europeas y norteamericanas**

La reflexión sobre el uso de la inteligencia artificial en la justicia penal no puede limitarse al ámbito colombiano, los desarrollos más avanzados y las tensiones más visibles se han dado en Europa y en Estados Unidos, escenarios que han permitido constatar los beneficios, pero también los riesgos de introducir algoritmos en decisiones que comprometen derechos fundamentales, tanto en la tradición garantista europea como en el pragmatismo norteamericano, se han abierto debates de fondo sobre la legitimidad democrática de esta modernización judicial y sobre la necesidad de establecer límites claros frente al despliegue tecnológico.

En Europa, el Tribunal Europeo de Derechos Humanos (TEDH) ha sido un actor clave al recordar que cualquier innovación que impacte derechos debe estar sujeta a los principios de

proporcionalidad, transparencia y tutela judicial efectiva, aunque el TEDH aún no ha resuelto un caso directamente vinculado al uso de IA en decisiones judiciales, sí ha sentado doctrina aplicable en supuestos análogos, en el caso *López Ribalda y otros vs. España* (2019), relativo al uso de cámaras ocultas en el trabajo, el Tribunal sostuvo que el empleo de tecnologías debe respetar siempre el equilibrio entre eficacia y derechos fundamentales, recordando que la dignidad humana no puede ser sacrificada en aras de la eficiencia, este precedente, trasladado a la justicia penal, obliga a considerar que cualquier algoritmo empleado debe ser comprensible, revisable y controlado por jueces humanos.

El TEDH también ha insistido en que el derecho a un proceso justo (art. 6 del Convenio Europeo de Derechos Humanos) incluye el acceso a las razones que justifican una sentencia, esto conecta directamente con la problemática de la “caja negra algorítmica”, que dificulta explicar cómo un sistema automatizado llega a determinada conclusión, si un juez adopta la recomendación de un algoritmo sin comprenderla ni poder explicarla, se estaría afectando la motivación judicial y el derecho de defensa, en este sentido, la jurisprudencia europea anticipa un marco garantista que puede servir de referencia para países como Colombia.

Dentro del debate europeo, el AI Act de la Unión Europea (2024) representa un referente normativo clave al clasificar el uso de inteligencia artificial en el ámbito judicial como una aplicación de “alto riesgo”, esta categorización implica la exigencia de supervisión continua, auditorías técnicas y mecanismos de trazabilidad que aseguren la rendición de cuentas, el enfoque europeo reconoce que la justicia penal es un ámbito en el que se aplican principios esenciales como la imparcialidad, la motivación de las decisiones y la función indelegable del juez; por ello, cualquier innovación tecnológica debe insertarse dentro de un esquema regulado que preserve la autoridad legítima de los tribunales y mantenga la confianza de la ciudadanía en

las instituciones judiciales, la incorporación de IA se admite únicamente como instrumento de apoyo, subordinado a la deliberación y control humano, en línea con los principios que rigen la administración de justicia.

El AI Act no se limita a exigir transparencia, sino que también introduce el principio de explicabilidad, es decir, la obligación de que los algoritmos utilizados puedan ofrecer razones claras, accesibles y entendibles para las decisiones que apoyan, con ello, se pretende evitar que la justicia se convierta en un proceso automatizado en el que las partes no comprendan cómo se llegó al fallo, este principio constituye una advertencia importante para sistemas como el colombiano, donde aún no existe una regulación integral sobre IA en la justicia y donde el riesgo de importar soluciones opacas es considerable.

Francia ha sido pionera en establecer límites al uso de algoritmos en la jurisdicción, en 2018, el Conseil d'État emitió un pronunciamiento que prohibió expresamente la utilización de datos judiciales para elaborar perfiles predictivos de jueces y tribunales, la razón de fondo fue la protección de la independencia judicial y el respeto al principio de motivación de las sentencias, este precedente francés es relevante porque evidencia que incluso en sistemas con alto desarrollo digital, la automatización encuentra límites cuando amenaza con despersonalizar la función de juzgar.

**En contraste, Estonia representa uno de los ejemplos más avanzados de justicia digital, este país ha desarrollado proyectos piloto de “jueces virtuales”** destinados a resolver litigios de menor cuantía, aunque no se trata de casos penales graves, la experiencia Estonia muestra un laboratorio interesante sobre cómo los sistemas automatizados pueden reducir cargas judiciales y aumentar la eficiencia, sin embargo, la iniciativa ha generado críticas en torno al riesgo de trivializar la función jurisdiccional, al sustituir el razonamiento jurídico por cálculos

algorítmicos, este modelo plantea un dilema que resulta muy útil para el debate colombiano: hasta dónde puede delegarse en la IA sin desnaturalizar la justicia.

Del lado norteamericano, la experiencia más polémica ha sido el uso de sistemas de predicción de reincidencia, en particular el software COMPAS, este programa, utilizado en diversos estados, asigna puntajes de riesgo a los acusados para orientar decisiones sobre libertad condicional o sentencias, el caso *State v. Loomis* (2016) puso este debate en el centro de la discusión, cuando la Corte Suprema de Wisconsin sostuvo que COMPAS podía emplearse como una herramienta auxiliar, pero nunca como fundamento exclusivo de una decisión, sin embargo, el fallo no resolvió el problema de fondo: la falta de acceso a los criterios internos del algoritmo, que impide a los procesados ejercer un control real sobre la prueba tecnológica utilizada en su contra.

El caso COMPAS evidenció, además, que los algoritmos no son neutrales, investigaciones como la de Angwin et al. (2016) demostraron que el sistema otorgaba puntajes más altos de reincidencia a personas afroamericanas en comparación con acusados blancos con antecedentes similares, este hallazgo generó un debate nacional sobre los sesgos algorítmicos y sobre cómo la digitalización de la justicia puede reproducir y amplificar desigualdades estructurales, para Colombia, este ejemplo es fundamental, pues revela que la implementación acrítica de sistemas importados podría terminar reforzando las brechas sociales y raciales que ya atraviesan nuestro sistema penal.

Una diferencia importante entre Europa y Estados Unidos radica en el enfoque regulatorio, mientras la Unión Europea ha optado por un marco normativo preventivo y uniforme, en Estados Unidos la aproximación es fragmentada y depende en gran medida de la jurisprudencia y de regulaciones estatales, esta descentralización genera vacíos que pueden ser aprovechados por

empresas privadas para introducir herramientas sin suficientes controles, de ahí que los casos como COMPAS se conviertan en alertas tempranas sobre los peligros de la autorregulación tecnológica.

Tanto en Europa como en Estados Unidos se reitera una conclusión común: la necesidad de mantener siempre un control humano significativo sobre los sistemas de inteligencia artificial aplicados a la justicia penal, en la UE, este requisito se ha positivizado en el AI Act, mientras que en Norteamérica se ha defendido a través de pronunciamientos judiciales y del debate académico, la coincidencia muestra que el riesgo principal de la automatización judicial no es la pérdida de eficiencia, sino la deshumanización del proceso penal.

Si se compara la experiencia francesa con la estadounidense, se observa una diferencia de enfoque: mientras Francia opta por la restricción estricta, Estados Unidos permite la experimentación con herramientas como COMPAS, aun a costa de riesgos importantes para la igualdad procesal, en el medio se ubica Estonia, que experimenta con jueces virtuales, pero en casos de menor gravedad, intentando equilibrar eficiencia y legitimidad, este panorama comparado demuestra que no existe una única forma de integrar la IA en la justicia, sino que cada país adapta la innovación a su tradición jurídica y a su modelo constitucional.

El aprendizaje para Colombia es evidente: no se trata de importar mecánicamente modelos extranjeros, sino de identificar buenas prácticas y errores que permitan diseñar una regulación acorde con nuestro orden constitucional, de Europa podemos tomar el enfoque garantista del AI Act y la defensa del juez humano como figura insustituible; de Estados Unidos, la advertencia sobre los riesgos de sesgo y opacidad; y de Estonia, la cautela frente a los alcances de la automatización en casos menores, sin trasladar sin más esta lógica a la justicia penal.

La comparación internacional confirma que la clave no está en negar la IA, sino en regularla y domesticarla para que se convierta en una herramienta auxiliar y no en un sustituto del juez, tanto los tribunales europeos como la experiencia norteamericana coinciden en que la justicia penal requiere de imparcialidad, motivación y transparencia, valores que solo se garantizan si el ser humano conserva el papel central en la función jurisdiccional, la lección es clara: la IA puede modernizar, pero no reemplazar; puede asistir, pero nunca decidir por sí sola.

### **3.4. Estándares mínimos de constitucionalidad y buenas prácticas**

La comparación entre experiencias europeas, norteamericanas y las incipientes discusiones en América Latina permite extraer un conjunto de criterios comunes que pueden considerarse estándares mínimos de constitucionalidad en el uso de inteligencia artificial en la justicia penal, estos criterios no surgen de un consenso formal único, sino de la convergencia progresiva entre fallos judiciales, normas internacionales y políticas públicas, que, aunque diversas, comparten la idea de que la justicia no puede perder su carácter humano frente a la automatización.

En primer lugar, aparece de manera reiterada la exigencia de control humano significativo, tanto el AI Act europeo como la jurisprudencia estadounidense (caso *State v. Loomis*) y las advertencias de la Corte Interamericana de Derechos Humanos coinciden en que los jueces deben conservar siempre la última palabra en la decisión penal, esto significa que ningún sistema algorítmico puede emitir una sentencia de manera autónoma, ni mucho menos sustituir el razonamiento judicial, el papel del algoritmo debe limitarse a una función de apoyo, nunca de reemplazo.

Un segundo criterio común es la necesidad de transparencia y explicabilidad, la justicia, para ser legítima, requiere que las partes comprendan las razones de la decisión, este principio, recogido en el art. 6 del Convenio Europeo de Derechos Humanos y en el art. 8 de la Convención

Americana, exige que todo algoritmo utilizado en procesos penales sea auditado y explicable en términos accesibles, en Estados Unidos, el debate sobre COMPAS mostró los riesgos de usar herramientas opacas; en Europa, el AI Act ha positivizado la obligación de ofrecer explicaciones comprensibles a jueces y partes.

Un tercer estándar es la protección reforzada de los derechos fundamentales, en especial el derecho de defensa, la presunción de inocencia y la imparcialidad judicial, tanto el TEDH como la Corte IDH han recordado que no basta con que un sistema funcione técnicamente bien: su validez depende de que respete los principios estructurales del debido proceso, cualquier desviación que impida al acusado contradecir, impugnar o entender el peso que un algoritmo tuvo en su sentencia, constituye una vulneración constitucional.

Asimismo, la experiencia comparada enseña la importancia de los mecanismos de rendición de cuentas y trazabilidad algorítmica, el AI Act obliga a registrar los procesos de entrenamiento y los datos empleados, de modo que se pueda reconstruir el camino que llevó a una decisión automatizada, en paralelo, la jurisprudencia norteamericana ha advertido que la imposibilidad de auditar un algoritmo equivale a negar el derecho a la defensa, por lo tanto, un estándar mínimo para Colombia debe ser la posibilidad real de auditar, revisar y cuestionar las herramientas tecnológicas empleadas en procesos judiciales.

En el ámbito de buenas prácticas, la experiencia francesa constituye un referente en la protección de la independencia judicial, al prohibir el uso de datos para perfilar jueces o predecir fallos, Francia reconoció que la motivación de la sentencia no puede ser desplazada por estadísticas ni patrones de conducta, esta medida protege la figura del juez natural como garante del derecho, y representa una práctica trasladable al contexto colombiano, donde la confianza en

los jueces se encuentra en crisis y cualquier intento de sustituirlos por algoritmos podría profundizar la deslegitimación.

Por otro lado, la experiencia de Estonia, aunque arriesgada, ofrece un aprendizaje útil, sus jueces virtuales para casos de menor cuantía muestran que la IA puede agilizar procesos menores sin comprometer garantías esenciales, siempre que exista la posibilidad de revisión por un juez humano, esta lógica podría aplicarse en Colombia a escenarios administrativos o contravencionales, liberando carga judicial en instancias menores para que los jueces penales se concentren en casos de mayor complejidad.

La experiencia estadounidense con COMPAS, pese a sus críticas, también brinda lecciones importantes, en particular, el fallo Loomis marcó un estándar: el algoritmo puede usarse como herramienta auxiliar, pero nunca como fundamento exclusivo de una sentencia, aunque el caso no resolvió la falta de explicabilidad, sí estableció una línea que Colombia podría replicar: reconocer la existencia de la IA, pero limitar su influencia a un rol no determinante.

Un aspecto central de las buenas prácticas es la evaluación de impacto algorítmico ex ante, tanto la Unión Europea como algunos estados norteamericanos han desarrollado metodologías para identificar riesgos antes de implementar un sistema de IA en la justicia, esto implica analizar posibles sesgos, impactos diferenciados y riesgos de vulneración de derechos, Colombia, que discute actualmente el Proyecto de Ley 154 de 2024 sobre IA, tiene la oportunidad de incorporar desde el inicio esta práctica, evitando reproducir los errores del modelo estadounidense.

Otra buena práctica proviene de las directrices internacionales de la UNESCO (2021), que recomiendan que el desarrollo e implementación de IA se rijan por principios de transparencia, equidad, rendición de cuentas y participación ciudadana, trasladado a l

.a justicia penal, esto significa que no basta con regular a nivel técnico: se requiere un debate público sobre los alcances de la IA, para que su legitimidad no dependa únicamente de expertos, sino también del escrutinio democrático.

La síntesis comparada permite afirmar que los estándares mínimos giran en torno a cuatro ejes: control humano, transparencia y explicabilidad, rendición de cuentas y respeto irrestricto de los derechos fundamentales, estos elementos no solo son teóricos, sino que han sido reconocidos en normativas, fallos y políticas públicas, en consecuencia, cualquier país que aspire a integrar la IA en la justicia penal debe hacerlo bajo estas condiciones mínimas si desea preservar la legitimidad de su sistema judicial.

En el caso colombiano, la adopción de estas buenas prácticas implicaría, en primer lugar, reforzar el papel del juez natural como decisor último, asegurando que la IA no sustituya su función, en segundo lugar, establecer normas que obliguen a la trazabilidad y explicabilidad de cualquier sistema empleado en procesos judiciales, en tercer lugar, generar protocolos de auditoría independientes que permitan evaluar los algoritmos antes de su implementación y durante su uso y por último reconocer el carácter participativo del debate, de modo que los ciudadanos no perciban la digitalización de la justicia como un proceso tecnocrático impuesto desde arriba.

La experiencia comparada confirma que la IA puede integrarse a la justicia penal sin vulnerar principios fundamentales, pero solo si se asumen estándares claros y se adoptan buenas prácticas de manera estricta, la ausencia de controles, como ocurrió en Estados Unidos, conduce a la opacidad y a la reproducción de desigualdades, en cambio, la prudencia regulatoria europea ofrece un modelo más acorde con un Estado Social y Democrático de Derecho como el colombiano, la tarea, entonces, no es rechazar la innovación, sino domesticarla bajo parámetros

constitucionales que aseguren que la tecnología esté siempre al servicio de la justicia y no al revés.

### **3.5. Propuesta crítica para el contexto colombiano**

El análisis comparado de los estándares internacionales en torno al uso de inteligencia artificial en la justicia penal revela la necesidad de realizar una adaptación cuidadosa al contexto colombiano, no basta con importar fórmulas regulatorias de Europa o de Norteamérica: es imprescindible tener en cuenta las condiciones estructurales de la institucionalidad judicial colombiana, marcadas por la congestión procesal, la escasez de recursos técnicos y humanos, y la persistente desconfianza ciudadana hacia las instituciones, la propuesta crítica aquí formulada parte, entonces, de reconocer que cualquier intento de modernización tecnológica debe armonizarse con estas particularidades.

En primer lugar, el criterio de control humano significativo, ampliamente aceptado en la Unión Europea y la jurisprudencia comparada, debe traducirse en Colombia en un principio de independencia judicial reforzada, esto implica que, aunque la IA pueda ser utilizada como herramienta de apoyo, el juez colombiano debe mantener no solo la última palabra, sino la obligación de motivar de manera autónoma y comprensible cada decisión, en un sistema con graves déficits de confianza, la delegación excesiva en sistemas algorítmicos podría erosionar aún más la legitimidad de la función jurisdiccional.

En segundo lugar, el estándar de transparencia y explicabilidad enfrenta retos singulares en Colombia, la debilidad en capacidades técnicas de la mayoría de despachos judiciales hace difícil la implementación de auditorías complejas de algoritmos, por ello, la propuesta crítica es que el país adopte un modelo de transparencia escalonada, en el cual se combinen auditorías técnicas externas, realizadas por universidades y centros especializados, con explicaciones accesibles

dirigidas a las partes procesales, esto garantizaría tanto la robustez técnica como la comprensibilidad ciudadana.

Un tercer aspecto central es la necesidad de mecanismos de rendición de cuentas y trazabilidad algorítmica, mientras que en Europa existen agencias con capacidad de supervisar sistemas de alto riesgo, en Colombia no contamos con una institucionalidad tan sólida, por eso, una propuesta realista es la creación de una unidad especializada en IA judicial dentro del Consejo Superior de la Judicatura, que concentre funciones de registro, auditoría y monitoreo de impacto de los algoritmos empleados en los procesos penales, de este modo, se evita la dispersión institucional y se asegura un control más directo.

Otro elemento que debe adaptarse al contexto colombiano es la práctica de evaluación de impacto algorítmico ex ante, en un país con altos niveles de desigualdad social, cualquier algoritmo que utilice datos históricos corre el riesgo de reproducir discriminaciones estructurales, la propuesta crítica es que toda herramienta de IA destinada a la justicia penal pase por una evaluación previa obligatoria que incluya el análisis de impactos diferenciales sobre poblaciones vulnerables, como minorías étnicas, mujeres y personas en situación de pobreza, garantizando que la digitalización no amplíe brechas sociales.

El reconocimiento de los estándares internacionales también debe acompañarse de una reflexión sobre los vacíos normativos internos, si bien es cierto, el Proyecto de Ley 154 de 2024 sobre inteligencia artificial y el CONPES 4144 de 2025 constituyen avances importantes, aún no existe en Colombia una regulación específica sobre el uso de IA en la administración de justicia, en consecuencia, una de las propuestas es impulsar un marco normativo especializado que establezca límites claros a la automatización en sentencias penales, integrando principios como el juez natural, la imparcialidad y la motivación judicial.

Ahora bien, la adaptación de estándares internacionales no puede soslayar las tensiones entre modernización y garantías fundamentales, en Colombia, la presión por reducir la congestión judicial puede llevar a ver la IA como un atajo hacia la eficiencia, minimizando los riesgos que supone para el debido proceso, esta tensión exige recordar que la justicia no puede concebirse solo como un servicio eficiente, sino como un proceso de legitimación democrática, donde cada decisión afecta derechos fundamentales y debe cumplir estándares de motivación y razonabilidad.

Además, la propuesta reconoce que el uso de IA en la justicia penal plantea riesgos particulares de despersonalización de la función jurisdiccional, en un país donde la relación entre juez y ciudadano ya es distante y burocratizada, la incorporación de algoritmos sin controles adecuados podría reforzar la percepción de una justicia fría, automática y ajena a las realidades sociales, por ello, se propone que cualquier herramienta de IA sea acompañada de programas de formación en ética judicial y de estrategias de acercamiento entre jueces y comunidades, evitando que la digitalización se traduzca en mayor desconexión.

también se manifiestan en el campo de la legitimidad democrática, mientras en Europa existe una cultura jurídica consolidada que facilita el control ciudadano de las decisiones, en Colombia los déficits de transparencia y participación dificultan este proceso, la propuesta, en consecuencia, es establecer mecanismos de participación ciudadana en la discusión y evaluación de la IA judicial, como consultas públicas y observatorios de justicia digital, de manera que la modernización no se convierta en una imposición tecnocrática sino en un ejercicio deliberativo.

Un aspecto adicional que debe ser considerado en Colombia es la relación entre la IA y el principio de igualdad material, dada la fuerte brecha digital en el país, la implementación de

herramientas tecnológicas puede acentuar las desigualdades entre quienes tienen acceso a información y recursos y quienes carecen de ellos, por ello, cualquier protocolo debe incluir medidas de accesibilidad, asistencia técnica y acompañamiento para que los sectores más vulnerables no queden excluidos del ejercicio de sus derechos en procesos judiciales mediados por IA.

Asimismo, la propuesta plantea la necesidad de alinear la innovación tecnológica con los principios constitucionales fundantes, la dignidad humana, reconocida por la Corte Constitucional como eje estructurante del Estado Social de Derecho, debe ser el criterio último para evaluar la pertinencia de la IA en la justicia penal, si un algoritmo, por eficiente que sea, atenta contra la dignidad de las personas procesadas, su uso debe ser limitado o prohibido, de esta manera, la modernización tecnológica no se opone a los principios, sino que se subordina a ellos.

La propuesta para Colombia consiste en un modelo de integración prudente y garantista: la IA puede servir como herramienta auxiliar para aliviar cargas judiciales, pero siempre bajo control humano, con plena transparencia, con auditorías independientes y con participación social en su evaluación, solo de este modo será posible conciliar la necesidad de modernización con la protección de las garantías fundamentales.

La tensión entre eficiencia y garantías debe resolverse recordando que la justicia penal no es un espacio de simple gestión de expedientes, sino un ámbito donde se decide la libertad y la dignidad de las personas, la modernización digital es deseable, pero nunca a costa de vaciar de contenido los principios esenciales del debido proceso, la verdadera innovación en Colombia no será aquella que automatice más rápido, sino la que logre integrar la tecnología sin renunciar a la esencia humanista del derecho penal.

### **3.5.1. Recomendaciones preliminares para el diseño de un protocolo.**

Un primer paso indispensable para la construcción de un protocolo de implementación de inteligencia artificial en la justicia penal colombiana consiste en establecer un principio de jerarquía constitucional reforzada, esto significa que cualquier herramienta algorítmica que pretenda ser utilizada en la elaboración de sentencias debe ser sometida, antes de su adopción, a un control riguroso de compatibilidad con los artículos 29 y 228 de la Constitución, así como con los estándares de la Corte Interamericana de Derechos Humanos en materia de garantías judiciales, este control no debe ser meramente declarativo, sino que ha de materializarse en evaluaciones concretas y vinculantes, capaces de limitar o frenar proyectos que no cumplan con estas condiciones mínimas.

De manera paralela, el protocolo debe prever mecanismos de transparencia y explicabilidad de doble vía, por un lado, es indispensable que el diseño técnico de los algoritmos quede sujeto a auditorías externas, realizadas por universidades públicas y centros de investigación independientes, que puedan verificar sesgos, trazabilidad y criterios de funcionamiento por otro, debe garantizarse que las partes procesales tengan acceso a explicaciones comprensibles sobre la forma en que el sistema contribuyó a la elaboración de la decisión, de este modo, se evita que la IA se convierta en una “caja negra” inmune al control ciudadano o judicial.

Una tercera recomendación apunta a la necesidad de consolidar un control humano significativo como cláusula innegociable, no se trata de que los jueces validen mecánicamente lo que diga un sistema algorítmico, sino de que ejerzan un escrutinio autónomo, crítico y con plena responsabilidad sobre la motivación de las decisiones, el protocolo debe dejar claro que la IA

tiene una función estrictamente auxiliar, orientada a mejorar la eficiencia, pero que en ningún caso reemplaza el juicio racional y garantista que caracteriza al juez natural.

Asimismo, resulta prioritario incorporar una evaluación de impacto algorítmico ex ante y ex post, antes de que un sistema de IA sea puesto en marcha, debe evaluarse cómo afectará a poblaciones vulnerables, cómo se relaciona con las brechas digitales del país y qué riesgos supone para la igualdad material, posteriormente, debe medirse de forma periódica su impacto real, incluyendo indicadores de sesgo, número de decisiones revocadas y grado de aceptación social, este ciclo de evaluación permanente constituye una garantía de que la modernización no derive en vulneraciones encubiertas de derechos fundamentales.

El protocolo también debería contemplar la creación de una unidad especializada en justicia digital dentro del Consejo Superior de la Judicatura, con funciones de supervisión, capacitación y sanción, este organismo sería el encargado de centralizar la información sobre los sistemas utilizados, aprobar los proyectos piloto y garantizar la coherencia de las prácticas judiciales en todo el territorio nacional, sin un ente con estas competencias, la regulación corre el riesgo de dispersarse y perder eficacia práctica.

Otro aspecto fundamental es garantizar la participación ciudadana y académica en el diseño y seguimiento del protocolo, la experiencia comparada muestra que cuando la sociedad civil, las facultades de derecho y los colectivos de abogados participan en la discusión, se fortalecen los niveles de legitimidad y se construyen marcos regulatorios más sólidos, en Colombia, donde la desconfianza hacia las instituciones judiciales es alta, abrir espacios de deliberación puede contribuir a que la implementación tecnológica sea percibida como un proceso inclusivo y no como una imposición tecnocrática.

De igual manera, el protocolo debe incorporar un régimen de responsabilidad claro y escalonado, esto supone identificar las responsabilidades de los programadores, los funcionarios judiciales y las entidades que autorizan el uso de IA, la ausencia de reglas claras puede generar vacíos de imputabilidad, diluyendo la responsabilidad y dejando a los ciudadanos en situación de indefensión frente a decisiones erradas, establecer un esquema de responsabilidad compartida pero jerárquicamente definido es clave para garantizar la confianza en el sistema.

La formación de los jueces constituye otra recomendación indispensable, un protocolo de implementación que no contemple programas de capacitación ética, jurídica y técnica para los operadores judiciales está condenado al fracaso, los jueces deben comprender no solo el funcionamiento básico de los sistemas, sino también los riesgos que implican, para ejercer un control crítico y no caer en una dependencia acrítica de la tecnología, esta formación debe ser continua y adaptada a las distintas jurisdicciones del país.

Asimismo, se sugiere que el protocolo establezca límites estrictos a la automatización en la fase resolutoria, el uso de IA en sentencias penales debe estar restringido a funciones de apoyo en la búsqueda de jurisprudencia, análisis estadístico o gestión de información, pero nunca en la determinación directa de culpabilidad o pena, esta delimitación protege la indelegabilidad de la función jurisdiccional y asegura que el núcleo de la decisión permanezca en manos humanas, tal como lo exige el principio del juez natural.

Por último, el diseño del protocolo debe ir acompañado de una estrategia de comunicación y pedagogía pública, en un contexto donde el escepticismo hacia las instituciones es elevado, explicar de manera sencilla por qué se utilizan estas herramientas, cómo funcionan y qué controles se aplican resulta crucial para fortalecer la legitimidad democrática del poder judicial, una justicia digital sin pedagogía corre el riesgo de convertirse en un mecanismo incomprensible

y distante, debilitando aún más la confianza ciudadana.

#### **4. CAPÍTULO 3: RIESGOS Y BENEFICIOS DE LA INTELIGENCIA ARTIFICIAL EN LA ELABORACIÓN DE SENTENCIAS PENALES: UNA VALORACIÓN DESDE LOS PRINCIPIOS DEL DERECHO**

La incorporación de la inteligencia artificial en la elaboración de sentencias penales representa un avance tecnológico con potencial para optimizar la eficiencia del sistema judicial, especialmente en contextos de alta congestión procesal, sin embargo, esta modernización no puede evaluarse únicamente desde la perspectiva de la rapidez o la automatización de tareas rutinarias, sino que debe analizarse a la luz de los principios que rigen la función jurisdiccional, la IA puede facilitar la sistematización de jurisprudencia, la clasificación de expedientes y la organización de información probatoria, contribuyendo así a una mayor celeridad procesal, pero su aplicación requiere un marco regulatorio que asegure que estas herramientas actúen como instrumentos de apoyo, y no como sustitutos del juicio humano (Ferrajoli, 2005).

En términos del principio de legalidad, la automatización enfrenta un desafío crucial: garantizar que las decisiones derivadas de sistemas algorítmicos sean comprensibles y se ajusten a normas previamente establecidas, la opacidad de muchos algoritmos podría generar resultados imprevisibles, afectando la certeza jurídica y la seguridad de los procesados, por ello, es indispensable que la IA opere bajo parámetros claros y auditablemente vinculados a la normativa vigente, asegurando que la eficiencia tecnológica no comprometa la previsibilidad y legitimidad de las decisiones judiciales (Alexy, 1993).

El principio de proporcionalidad constituye otro eje fundamental en la valoración de la IA, ya que las medidas adoptadas deben ser adecuadas, necesarias y equilibradas respecto del fin legítimo perseguido, los algoritmos, si no son calibrados correctamente, podrían producir

decisiones desproporcionadas al no ponderar de manera adecuada factores contextuales, como la gravedad del delito o las circunstancias personales del acusado, en consecuencia, la IA debe integrarse como herramienta que apoye la toma de decisiones, manteniendo siempre la capacidad del juez para ajustar las resoluciones a cada caso concreto y garantizar la equidad en la aplicación de la ley (O’Neil, 2016).

La equidad y la no discriminación representan un desafío crítico en la justicia algorítmica, los sistemas entrenados con datos históricos tienden a replicar sesgos sociales, económicos o étnicos existentes, lo que podría generar impactos desproporcionados sobre grupos vulnerables, por ello, es imprescindible establecer mecanismos de supervisión y auditoría que detecten y corrijan estas desigualdades, asegurando que la IA no amplifique las injusticias estructurales, y garantizando que todos los procesados tengan acceso a un trato justo y equitativo en consonancia con los principios de igualdad procesal (Pasquale, 2015).

El principio de responsabilidad subraya que toda decisión judicial debe ser atribuible a sujetos humanos plenamente identificables, la automatización indiscriminada de funciones judiciales podría difuminar la rendición de cuentas, generando vacíos frente a posibles errores o arbitrariedades, por ello, es fundamental mantener un control humano significativo, de manera que el juez conserve la deliberación final, asuma responsabilidad plena y pueda intervenir cuando los resultados algorítmicos sean inadecuados o injustos, preservando así la legitimidad de la función jurisdiccional (UNESCO, 2021).

Asimismo, el principio de transparencia exige que las decisiones sean comprensibles y auditables, permitiendo a las partes procesales cuestionar y contradecir razonamientos judiciales cuando sea necesario, la complejidad de los algoritmos, conocida como la “caja negra”, plantea un riesgo directo frente a este principio, ya que la falta de explicabilidad podría afectar el

derecho de defensa y limitar la capacidad de control sobre la actuación judicial, por ello, los sistemas de IA deben incorporar mecanismos que traduzcan los resultados algorítmicos a un lenguaje accesible y verificable, garantizando que la modernización tecnológica fortalezca la confianza en la justicia y no comprometa la equidad procesal.

Los beneficios potenciales de la IA, como la reducción de tiempos de respuesta, la sistematización de información y la mejora en la gestión judicial, solo pueden ser aprovechados si se integran dentro de un marco principialista que preserve los valores esenciales del derecho penal, esto implica que cada herramienta tecnológica debe ser evaluada y supervisada según criterios de legalidad, proporcionalidad, equidad, responsabilidad y transparencia, de esta manera, la IA se convierte en un apoyo valioso, capaz de optimizar la administración de justicia penal sin socavar la legitimidad de las decisiones ni la confianza de la ciudadanía en las instituciones judiciales.

#### **4.1. Riesgos asociados a la IA**

La introducción de la inteligencia artificial en la justicia penal plantea riesgos significativos relacionados con la equidad procesal, siendo uno de los más preocupantes los sesgos algorítmicos, los sistemas de IA entrenados con datos históricos pueden reproducir patrones de discriminación existentes en la sociedad, afectando desproporcionadamente a poblaciones vulnerables, esto compromete directamente el principio de igualdad procesal, ya que los algoritmos podrían favorecer o perjudicar a ciertos grupos de manera sistemática, alterando la neutralidad y equidad que deben caracterizar toda decisión judicial (O’Neil, 2016).

Estos sesgos no siempre son evidentes a simple vista, lo que incrementa la gravedad del riesgo, en muchos casos, la discriminación se produce de manera inadvertida, pero con consecuencias concretas sobre la libertad o los derechos de los procesados, por ello, cualquier

implementación de IA en la justicia penal debe incluir mecanismos de auditoría y supervisión constante, permitiendo identificar, corregir y prevenir impactos adversos sobre la igualdad entre las partes, garantizando que la automatización no perpetúe injusticias históricas (Pasquale, 2015).

Otro riesgo relevante es la opacidad de los sistemas algorítmicos, conocida como “caja negra”, la complejidad de algunos modelos de IA impide comprender cómo se generan los resultados, dificultando la evaluación crítica de sus decisiones, esta falta de explicabilidad afecta directamente la transparencia del proceso judicial y compromete la posibilidad de contradecir o revisar razonamientos, principios esenciales en cualquier procedimiento penal que aspire a la legitimidad y confiabilidad de la justicia (Alexy, 1993).

La opacidad, además, interfiere con la rendición de cuentas y el control democrático de la función judicial, si un juez se limita a reproducir resultados algorítmicos sin comprender su fundamento, la responsabilidad de la decisión se diluye, lo que podría generar impunidad ante errores o injusticias, por ello, la exigencia de explicabilidad no es un mero formalismo, sino un mecanismo que protege la integridad del proceso y garantiza que la tecnología permanezca subordinada al juicio humano (UNESCO, 2021).

La pérdida de control humano sobre la deliberación y la valoración probatoria constituye otro riesgo crítico, la IA puede sugerir conclusiones o priorizar ciertos argumentos, pero no puede sustituir la capacidad de ponderación ética y jurídica de un juez, delegar indebidamente decisiones complejas a sistemas algorítmicos reduce la autonomía judicial y convierte la deliberación en un proceso automatizado, comprometiendo la calidad del análisis y la racionalidad de la decisión final (Ferrajoli, 2005).

Este riesgo se agrava en materia penal, donde la libertad, la reputación y los derechos fundamentales están en juego, la valoración de pruebas, la interpretación de normas y la

consideración de circunstancias particulares requieren un juicio humano informado, prudente y contextualizado, la IA, sin capacidad de razonamiento moral ni apreciación de matices, no puede garantizar que estas funciones se cumplan con la misma profundidad y responsabilidad que un juez humano (Dworkin, 1986).

Los efectos de la automatización sobre la confianza ciudadana también son significativos, si las partes perciben que las decisiones judiciales dependen de algoritmos opacos o sesgados, la legitimidad del sistema se ve erosionada, la justicia penal no solo debe ser correcta en términos formales, sino también percibida como justa y confiable por la sociedad, de modo que la implementación de IA debe priorizar mecanismos de transparencia, supervisión y explicabilidad para sostener la confianza pública (Gargarella, 2010).

La percepción de arbitrariedad se intensifica cuando los ciudadanos y profesionales del derecho no pueden comprender ni cuestionar los razonamientos algorítmicos, este efecto socava la autoridad moral del tribunal y debilita la cohesión social respecto al cumplimiento de las normas, por ello, los riesgos tecnológicos no son solo técnicos, sino profundamente institucionales, afectando la legitimidad democrática de la justicia y el reconocimiento social de sus decisiones (Prieto Sanchís, 2012).

La integración de la IA en la justicia penal requiere un equilibrio cuidadoso: aprovechar los beneficios de eficiencia y sistematización sin comprometer los principios de imparcialidad, transparencia, control humano y equidad. Identificar, prevenir y mitigar los riesgos asociados es indispensable para garantizar que la tecnología funcione como un apoyo al juez, fortaleciendo la justicia en lugar de debilitarla, la adopción responsable de la IA depende, en última instancia, de marcos normativos claros, supervisión constante y un compromiso ético con los derechos de todos los procesados.

#### **4.2. Beneficios potenciales de la IA en la función jurisprudencial**

La incorporación de inteligencia artificial en la justicia penal puede contribuir significativamente a la optimización de la gestión procesal, permitiendo un manejo más eficiente de los expedientes y reduciendo la congestión judicial, esta capacidad es especialmente relevante en sistemas donde los retrasos procesales generan demoras en la tutela de derechos fundamentales, ya que la IA puede priorizar tareas, organizar cronogramas y alertar sobre plazos vencidos, facilitando que los jueces concentren su atención en los aspectos sustantivos de cada caso (European Commission, 2021).

Además, la IA ofrece herramientas para la organización documental, permitiendo clasificar, indexar y recuperar información de manera ágil y precisa, la sistematización de expedientes, precedentes y normativas puede mejorar la eficiencia de los juzgados, asegurando que los jueces tengan acceso rápido a datos relevantes sin perder tiempo en tareas administrativas, lo que fortalece la calidad de la deliberación judicial (UNESCO, 2021).

En este sentido, los sistemas algorítmicos también pueden apoyar el análisis de jurisprudencia, identificando patrones de decisiones previas y resaltando criterios relevantes para la resolución de casos similares, esta función no reemplaza la labor del juez, pero proporciona una base informativa robusta que facilita la argumentación y la consistencia en la aplicación del derecho, promoviendo decisiones más coherentes y fundamentadas (Binns, 2020).

La IA puede, asimismo, contribuir a la consistencia de la toma de decisiones al reducir variaciones indebidas entre juzgados o entre casos comparables, al estandarizar ciertos procedimientos de análisis documental y evidencia, los sistemas algorítmicos permiten que los jueces mantengan criterios homogéneos, sin sacrificar la discrecionalidad necesaria para

ponderar factores específicos de cada proceso, lo que refuerza la predictibilidad y equidad del sistema judicial (Alexy, 1993).

Otro beneficio relevante es la rapidez en el procesamiento de información compleja, la IA puede sintetizar grandes volúmenes de datos en minutos, algo que para un equipo humano podría requerir días, liberando tiempo para que los jueces se concentren en la deliberación y valoración probatoria, asegurando que la eficiencia no se traduzca en reducción de control humano ni en afectación de derechos (Ferrajoli, 2005).

La sistematización de información mediante IA también permite identificar inconsistencias, errores formales o posibles contradicciones en expedientes judiciales, este control preventivo funciona como un mecanismo de apoyo que refuerza la calidad de las decisiones, asegurando que los jueces puedan focalizar su juicio crítico en los aspectos esenciales del caso, mientras la tecnología gestiona los datos y alertas preliminares (Pasquale, 2015).

De igual forma, la IA puede servir como herramienta de gestión de carga laboral, priorizando casos por urgencia, relevancia o complejidad, esta capacidad permite una distribución más equitativa de los expedientes entre jueces y despachos, evitando saturación y garantizando tiempos adecuados para la deliberación y fundamentación de

cada decisión, lo que impacta positivamente en la eficiencia y en el respeto al principio de plazo razonable (O'Neil, 2016).

El uso de IA también ofrece oportunidades para mejorar la transparencia del proceso, en tanto permite generar registros detallados de cada recomendación o análisis que la tecnología realiza, estos registros facilitan la trazabilidad y auditables de las decisiones, contribuyendo a que los jueces mantengan el control sobre el proceso y puedan explicar de manera clara y fundamentada las decisiones adoptadas (Dworkin, 1986).

Adicionalmente, la IA puede apoyar la investigación jurídica y la formación de criterios, al proporcionar resúmenes, alertas y análisis de tendencias jurisprudenciales, esto permite a los jueces mantenerse actualizados sobre la evolución de la interpretación de normas y criterios relevantes, potenciando la deliberación informada y la consistencia de las decisiones sin delegar la función decisoria (Gargarella, 2010).

La implementación de IA en la función jurisdiccional puede ser un aliado estratégico para fortalecer la calidad del servicio judicial, siempre que se limite a roles de apoyo técnico y no sustituya la deliberación humana, cuando se utiliza correctamente, la IA permite que la justicia sea más eficiente, transparente y coherente, conservando intacta la autoridad moral y ética del juez, mientras se aprovechan los beneficios de la tecnología para mejorar la administración de justicia (Prieto Sanchís, 2012).

#### **4.3. Valoración crítica desde los principios del derecho**

La incorporación de inteligencia artificial en la justicia penal plantea la necesidad de evaluar cómo esta tecnología interactúa con principios fundamentales como la imparcialidad, la motivación de las decisiones y la función indelegable del juez, aunque la IA ofrece ventajas técnicas, su uso inadecuado puede generar tensiones con estos principios, ya que algoritmos entrenados en datos históricos pueden reproducir sesgos estructurales, afectando la igualdad procesal y comprometiendo la percepción de neutralidad que debe caracterizar al juzgador (O'Neil, 2016).

En cuanto a la imparcialidad, los sistemas algorítmicos deben diseñarse para evitar la reproducción de desigualdades sociales y prejuicios históricos, si se implementan con criterios transparentes y con auditorías periódicas, la IA puede apoyar decisiones más coherentes y homogéneas entre diferentes juzgados, reduciendo la influencia de factores subjetivos, pero esto

solo es efectivo cuando el juez mantiene el control crítico y puede corregir posibles desviaciones, preservando así la esencia de su independencia (Binns, 2020).

Respecto a la motivación de la decisión, la IA puede ser un instrumento que facilite la estructuración de argumentos y la identificación de precedentes relevantes, lo que contribuye a una fundamentación más sólida de las sentencias, sin embargo, si el juez adopta recomendaciones algorítmicas sin un análisis crítico, la motivación se vuelve superficial, limitando la comprensibilidad y la transparencia de la decisión judicial, por ello, la trazabilidad de las recomendaciones y la posibilidad de explicarlas en términos comprensibles para todas las partes es indispensable (Pasquale, 2015)

La función del juez natural constituye otro eje crítico de evaluación, ningún sistema de IA puede asumir la deliberación final ni la valoración probatoria de manera autónoma, pues la función jurisdiccional es indelegable, la tecnología debe limitarse a apoyar la recolección de información, la sistematización de jurisprudencia o la gestión documental, mientras que el juzgador sigue siendo responsable de la interpretación de normas y la ponderación de principios en cada caso (Ferrajoli, 2005).

La compatibilidad de la IA con los estándares de control humano depende de la existencia de mecanismos de supervisión efectivos, esto implica que cualquier recomendación algorítmica sea verificable, auditada y sometida a juicio crítico por el juez, solo bajo estas condiciones es posible que la IA contribuya a la eficiencia procesal sin erosionar la autoridad judicial ni la legitimidad de las decisiones (Alexy, 1993).

Además, la trazabilidad de los sistemas algorítmicos se convierte en un principio fundamental para garantizar transparencia y responsabilidad, cada paso del análisis computacional debe ser documentado y susceptible de revisión, de modo que el juez pueda

fundamentar plenamente sus decisiones y las partes puedan ejercer su derecho de contradicción, reforzando la confianza en el proceso penal y evitando que la IA se convierta en una “caja negra” opaca (Dworkin, 1986).

En escenarios de implementación óptimos, la IA puede maximizar beneficios como la reducción de congestión judicial, la organización documental y la sistematización de precedentes, sin comprometer los principios del derecho, esto requiere una planificación estratégica, capacitación de jueces y protocolos claros que definan los límites de la intervención algorítmica, priorizando siempre la deliberación humana sobre la automatización (UNESCO, 2021).

No obstante, también se deben contemplar riesgos residuales, como la posible dependencia excesiva de la tecnología, la introducción de sesgos inadvertidos y la percepción de arbitrariedad, la mitigación de estos riesgos exige un enfoque principialista, donde cada decisión asistida por IA sea supervisada, justificada y controlada de manera que los principios de imparcialidad, motivación y juez natural permanezcan intactos (Gargarella, 2010).

Es así como la implementación de la IA en la justicia penal requiere un equilibrio cuidadoso entre eficiencia y legitimidad, cuando se respeta la autoridad del juez, se asegura la explicabilidad de los sistemas y se mantiene un control humano significativo, la IA puede convertirse en una herramienta estratégica que fortalezca la justicia, promoviendo decisiones más rápidas, coherentes y fundamentadas, sin comprometer los pilares del debido proceso ni los estándares del Estado de Derecho (Prieto Sanchís, 2012).

Por ello la implementación de inteligencia artificial en la justicia penal debe enfocarse en diseñar protocolos claros que delimiten estrictamente la función de los sistemas algorítmicos como herramientas de apoyo, evitando que sustituyan la deliberación y la decisión final del juez,

los protocolos deberían establecer criterios de transparencia, explicabilidad y trazabilidad, garantizando que cada recomendación pueda ser revisada, evaluada y fundamentada por el juzgador, de modo que se preserve la integridad de los principios jurídicos fundamentales, como la imparcialidad, la motivación de la decisión y la función indelegable del juez.

Asimismo, se recomienda que los protocolos contemplen la capacitación continua de los operadores judiciales, integrando formación ética, técnica y procedimental sobre el uso de la IA, esta preparación permitirá a los jueces interpretar adecuadamente las sugerencias algorítmicas, evaluar posibles sesgos y mantener control sobre la valoración probatoria, asegurando que la tecnología complemente, y no desplace, el juicio humano, la formación especializada también contribuirá a generar confianza en el proceso penal por parte de la ciudadanía y a consolidar la legitimidad institucional de la justicia.

En términos legislativos, es imperativo establecer normas específicas que regulen el uso de IA en la función jurisdiccional penal, diferenciando claramente entre tareas de apoyo técnico y decisiones sustantivas de carácter judicial, la regulación debería incorporar requisitos mínimos de control humano, estándares de auditoría, obligaciones de transparencia y mecanismos de rendición de cuentas, de manera que se eviten prácticas arbitrarias o inconsistentes, y se preserve la autoridad del juez frente a cualquier influencia algorítmica.

Las líneas de acción deben orientarse hacia un marco normativo coherente y articulado con los compromisos internacionales en materia de derechos humanos y debido proceso, esto implica que cualquier protocolo o ley sobre IA judicial contemple la igualdad procesal, la motivación de las decisiones, el derecho de defensa y la protección frente a sesgos, asegurando que la modernización tecnológica se traduzca en eficiencia y calidad procesal, sin comprometer

la legitimidad de la función judicial ni los estándares de justicia que sustentan el Estado de Derecho.

Estos análisis sobre riesgos, beneficios y principios jurídicos muestran que la incorporación de la inteligencia artificial en la justicia penal no puede ser improvisada ni carecer de un marco normativo claro, por el contrario, la evidencia sugiere que su implementación debe estructurarse a través de protocolos precisos que delimiten funciones, establezcan controles de supervisión humana y garanticen la trazabilidad de cada recomendación algorítmica, solo de esta manera será posible equilibrar la eficiencia tecnológica con la protección de los derechos fundamentales y la autoridad del juez.

En este sentido, el siguiente capítulo presenta un protocolo de implementación de IA en la función jurisdiccional penal, diseñado para maximizar los beneficios de la tecnología sin comprometer los principios del derecho, este protocolo establece lineamientos sobre capacitación judicial, auditorías de sistemas, trazabilidad de decisiones y mecanismos de supervisión, ofreciendo un modelo operativo que traduce las recomendaciones teóricas en acciones concretas y reguladas, sirviendo como puente entre la innovación tecnológica y la preservación de la legitimidad judicial.

## **5. PROPUESTA DE PROTOCOLO PARA LA IMPLEMENTACIÓN DE INTELIGENCIA ARTIFICIAL EN LA JUSTICIA PENAL COLOMBIANA**

La incorporación de tecnologías digitales en los sistemas judiciales plantea desafíos inéditos para el derecho penal, donde están en juego bienes jurídicos fundamentales como la libertad, la dignidad y el debido proceso, en Colombia, esta realidad se acentúa por el carácter acusatorio del sistema penal y la fragilidad institucional que limita el acceso efectivo a la justicia, en este contexto, la implementación de inteligencia artificial (IA) en la elaboración de sentencias no puede considerarse un simple avance técnico, sino un proceso político-jurídico que requiere garantías reforzadas, supervisión democrática y una reflexión constitucional profunda, la IA puede ser útil como herramienta de apoyo en tareas como gestión documental, consulta de jurisprudencia o detección de patrones, pero no debe asumir funciones decisorias que comprometan la autonomía judicial ni los derechos procesales (De Luis García, 2023).

La necesidad de un protocolo para su implementación se fundamenta en el principio de precaución, dado que la automatización de decisiones penales conlleva riesgos estructurales de despersonalización de la justicia y debilitamiento del vínculo entre juez y procesado, experiencias internacionales muestran que la ausencia de marcos regulatorios robustos permite que sistemas opacos influyan en decisiones sin posibilidad de impugnación o auditoría, un riesgo que en Colombia se agrava por brechas digitales, debilidad institucional y limitada formación técnica de los operadores, por ello, cualquier propuesta de implementación debe contemplar criterios normativos, éticos y formativos, asegurando que la IA complemente pero nunca sustituya las funciones humanas esenciales del juzgador, en particular la capacidad deliberativa y la motivación de las decisiones, pilares de la justicia penal (Ferrajoli, 2001).

El protocolo propuesto se estructura sobre cinco dimensiones clave: compatibilidad constitucional, elementos estructurales del modelo de IA, fases de implementación gradual,

referencias comparadas de buenas prácticas y articulación con el bloque de constitucionalidad, se concibe como un instrumento normativo abierto, sujeto a revisión crítica y deliberación democrática, orientado a garantizar que la modernización tecnológica fortalezca la justicia sin sacrificar derechos fundamentales, lejos de idealizar o demonizar la IA, el enfoque busca ofrecer herramientas críticas y propositivas para optimizar procesos judiciales, proteger a los más vulnerables y consolidar una justicia penal transparente, humanista y garantista, que combine innovación tecnológica con respeto irrestricto al debido proceso.

### **5.1. Criterios mínimos para una IA compatible con el orden constitucional**

El proceso penal constituye uno de los espacios más sensibles del ordenamiento jurídico, dado que activa el poder punitivo del Estado y exige respeto irrestricto a las garantías fundamentales, en este contexto, la incorporación de inteligencia artificial (IA) en la elaboración de sentencias debe evaluarse bajo criterios estrictos de compatibilidad con la Constitución, no solo en términos de legalidad formal, sino desde los principios estructurales del Estado social y democrático de derecho, la legitimidad del derecho penal moderno se sustenta en su carácter garantista, donde cada regla procesal limita el poder del Estado frente al ciudadano y asegura que la tecnología no sustituya funciones humanas esenciales, como la motivación judicial y la deliberación ética (Ferrajoli, 2001).

Dentro de los criterios fundamentales se encuentran la irrenunciabilidad de la motivación judicial, la transparencia tecnológica y el control humano significativo, toda decisión penal debe estar debidamente fundamentada, comprensible y auditada, evitando la opacidad de algoritmos que impida reconstruir la cadena lógica de la decisión (Raso et al., 2018; Pasquale, 2015).

Asimismo, el juez debe intervenir activamente en la valoración jurídica y en la ponderación de derechos, sin convertirse en un mero validador de recomendaciones

automatizadas, garantizando así la deliberación ética y el cumplimiento del debido proceso (De Luis García, 2023).

Otro aspecto esencial es la no discriminación algorítmica y la responsabilidad institucional, la IA puede heredar sesgos estructurales presentes en los datos históricos, reproduciendo patrones injustos, por lo que se requieren mecanismos de detección y corrección de sesgos, además de monitoreo permanente, al mismo tiempo, la complejidad técnica no puede diluir responsabilidades: cada decisión basada en IA debe ser trazable y asignable a operadores, desarrolladores o autoridades judiciales, asegurando accountability y cumplimiento de estándares internacionales de derechos humanos (Angwin et al., 2016; Eubanks, 2018; Hirsi Jamaa vs. Italia, 2012).

Otros criterios de relevancia incluyen la proporcionalidad tecnológica, la formación ética y jurídica de los operadores, la interoperabilidad normativa y la participación ciudadana, la IA debe aplicarse solo cuando aporte valor real, reforzando la justicia sin vulnerar garantías; los jueces y defensores deben comprender su funcionamiento y límites, los sistemas deben armonizarse con la Constitución y tratados internacionales; y la ciudadanía debe participar en el diseño y control de las políticas de IA para asegurar un uso democrático y transparente.

Se requiere garantizar la revisión judicial efectiva, la implementación gradual y el respeto al principio de dignidad humana, toda decisión asistida por IA debe poder ser impugnada y evaluada críticamente, avanzando con prudencia bajo pilotos y monitoreo independiente, la IA nunca puede reemplazar la valoración subjetiva del juez, quien debe preservar su rol de garante de derechos, asegurando que la tecnología complemente la justicia sin comprometer la dignidad ni los principios fundamentales del proceso penal colombiano.

## **5.2. Protocolo para la Implementación de la Inteligencia Artificial en la Elaboración de Sentencias Penales en Colombia**

### **5.2.1. Objetivo General**

Diseñar un marco normativo, técnico y ético para la implementación de sistemas de inteligencia artificial en la elaboración de sentencias penales, asegurando que su uso funcione como herramienta auxiliar y esté alineado con los principios fundamentales del derecho, como la imparcialidad, la motivación de las decisiones y la función indelegable del juez, de manera que fortalezca la legitimidad, transparencia y eficacia del poder judicial.

### **5.2.2. Principios Generales**

El protocolo se basa en los siguientes principios fundamentales:

- **Auxiliaridad de la IA:** La IA solo apoya el proceso judicial; la decisión final corresponde siempre al juez.
- **Debido proceso y motivación judicial:** Toda recomendación automatizada debe ser explicable, auditada y motivada jurídicamente.
- **Transparencia y trazabilidad:** Los sistemas deben permitir la reconstrucción completa del razonamiento algorítmico.
- **No discriminación y equidad:** Se deben implementar mecanismos de detección y corrección de sesgos que puedan afectar derechos por género, raza, estrato o región.
- **Responsabilidad institucional:** Cada decisión asistida por IA debe contar con responsables claros: desarrolladores, operadores y jueces.
- **Protección de datos y privacidad:** Garantizar la seguridad, confidencialidad y calidad de los datos utilizados.

### **5.2.3. Ámbito de Aplicación**

- Etapas asistidas por IA: gestión documental, consulta de jurisprudencia, análisis de patrones y generación de borradores no vinculantes.
- Etapas exclusivas del juez: calificación jurídica, valoración probatoria, determinación de culpabilidad y fundamentación final de la sentencia.

### **5.2.4. Controles Ex Ante**

- Evaluación de Impacto Jurídico y Ético: Antes de implementar cualquier sistema, se realizará un análisis de compatibilidad con la Constitución, el debido proceso y los estándares internacionales de derechos humanos.
- Autorización de Uso: Los sistemas deben ser aprobados por un comité interdisciplinario conformado por jueces, expertos en derecho, técnicos en IA y representantes de la sociedad civil.
- Pruebas Piloto en Contextos No Punitivos: La IA debe ser probada en entornos simulados para evaluar riesgos de sesgo, errores y funcionamiento operativo sin afectar derechos reales.

### **5.2.5. Controles Ex Post**

- Auditorías Algorítmicas Periódicas: Revisiones programadas de resultados y funcionamiento para detectar sesgos, errores o recomendaciones inconsistentes.
- Monitoreo de Sesgos y Revocatorias: Seguimiento de indicadores sobre decisiones revocadas, patrones discriminatorios o inconsistencias en la motivación de la sentencia.
- Revisión Judicial Efectiva: Cualquier decisión asistida por IA debe ser impugnabile y revisable por las partes, con posibilidad de solicitar peritajes técnicos.

### **5.2.6. Gobernanza de Datos**

- Calidad y actualización: Los datos deben ser veraces, completos y actualizados.

- Acceso controlado: Solo personal autorizado puede manipular o consultar datos, bajo protocolos de seguridad y confidencialidad.
- Registro y trazabilidad: Todo acceso o modificación de datos debe quedar registrado para garantizar transparencia.

### **5.2.7. Evaluación de Impacto Algorítmico**

- Explicabilidad Algorítmica: Cada recomendación debe poder reconstruirse y explicarse jurídicamente.
- Métricas de Seguimiento: Evaluar consistencia, motivación, revocatorias y sesgo en las decisiones.
- Pruebas de Equidad: Comprobación de que no existen patrones discriminatorios según género, raza, estrato o territorio.

### **5.2.8. Límites a la Automatización**

- La IA no puede emitir decisiones vinculantes.
- La valoración jurídica, ponderación de derechos y determinación de culpabilidad son funciones exclusivas del juez.
- Las recomendaciones deben considerarse como apoyo, no sustitución, de la función jurisdiccional.

### **5.2.9. Formación Obligatoria**

- Capacitación ética, jurídica y técnica para jueces, fiscales y defensores.
- Alfabetización digital crítica que permita comprender límites, riesgos y responsabilidades de la IA.
- Estrategias de divulgación para partes procesales sobre funcionamiento y alcance de la IA.

### 5.2.10. Participación Democrática y Supervisión

- Creación de comités interdisciplinarios para supervisión continua.
- Escenarios de deliberación pública con sociedad civil, universidades y gremios profesionales.-
- Revisión y actualización periódica del protocolo con participación de todos los actores interesados.

### 5.2.11. Reclamación y Reparación

- Mecanismos accesibles para impugnar decisiones asistidas por IA.
- Posibilidad de anulación de sentencias basadas en información sesgada o errónea.
- Procedimientos de reparación rápida y efectiva para proteger derechos vulnerados.

### 5.2.12. Cláusula Transversal

El protocolo reconoce explícitamente que la IA es una herramienta auxiliar y no un sustituto del juicio humano, ningún sistema puede operar autónomamente ni desplazar al juez como garante de los derechos fundamentales; cualquier desviación de este principio constituye un riesgo grave para la legitimidad del poder judicial.

### 5.2.13. Fases de desarrollo del protocolo

#### 🚦 Fase 1: Evaluación Previa y Aprobación del Sistema de IA

- ❖ **Objetivo:** Determinar la viabilidad jurídica, ética y técnica del sistema de IA antes de su implementación.
- ❖ **Responsables:** Consejo Superior de la Judicatura, Comité Interdisciplinario de Evaluación Ética y Jurídica.
- ❖ **Actividades clave:**
  - ✓ Presentación técnica del sistema de IA por parte de los desarrolladores.

- ✓ Revisión del algoritmo y su compatibilidad con los principios constitucionales (legalidad, dignidad, presunción de inocencia, imparcialidad).

- ✓ Evaluación de impacto ético, jurídico y social.

- ✓ Simulaciones controladas en ambientes no judiciales.

❖ **Criterios de verificación:**

- ✓ Informe técnico-jurídico de viabilidad.

- ✓ Certificado de transparencia y trazabilidad.

- ✓ Auditoría externa independiente.

- ✓ Dictamen favorable del comité interdisciplinario.

✚ **Fase 2: Pilotaje Controlado y Formación Judicial**

- ❖ **Objetivo:** Verificar el funcionamiento del sistema en entornos reales sin comprometer garantías procesales.

- ❖ **Responsables:** Escuela Judicial Rodrigo Lara Bonilla, jueces penales voluntarios, universidades aliadas.

- ❖ **Actividades clave:**

- ✓ Capacitación obligatoria a jueces, fiscales y defensores sobre IA, ética judicial y derechos digitales.

- ✓ Pruebas piloto de IA en procesos penales simulados.

- ✓ Evaluación del comportamiento del algoritmo frente a casos con complejidades jurídicas.

- ❖ **Criterios de verificación:**

- ✓ Informes de seguimiento pedagógico.

- ✓ Registro audiovisual y documental de simulaciones.

- ✓ Evaluación de satisfacción y comprensión por parte de los operadores judiciales.
- ✓ No generación de sesgos o errores significativos.

### ✚ **Fase 3: Implementación Gradual Supervisada**

- ❖ **Objetivo:** Iniciar el uso del sistema de IA como herramienta auxiliar, bajo control judicial estricto.
- ❖ **Responsables:** Órganos jurisdiccionales penales, Defensoría del Pueblo, Procuraduría General de la Nación.
- ❖ **Actividades clave:**
  - ✓ Uso asistido de IA en la redacción de borradores de sentencias penales (no vinculantes).
  - ✓ Evaluación crítica del juez frente a cada sugerencia emitida.
  - ✓ Obligación del juez de motivar la aceptación o el rechazo del insumo automatizado.
  - ✓ Registro obligatorio en el expediente de cada intervención del sistema de IA.
- ❖ **Criterios de verificación:**
  - ✓ Seguimiento por el Observatorio de Justicia Digital.
  - ✓ Registro nacional de decisiones asistidas.
  - ✓ Reportes bimensuales de funcionamiento.
  - ✓ Retroalimentación continua por parte de los operadores judiciales.

### ✚ **Fase 4: Evaluación, Auditoría y Revisión de Impacto**

- ❖ **Objetivo:** Verificar que la aplicación del sistema de IA respeta los derechos fundamentales y fortalece la función judicial.

- ❖ **Responsables:** Corte Constitucional, Universidad Nacional, Veedurías Ciudadanas, organizaciones de derechos humanos.
- ❖ **Actividades clave:**
  - ✓ Auditoría algorítmica permanente y externa.
  - ✓ Análisis estadístico de sesgos, errores o patrones discriminatorios.
  - ✓ Consultas públicas y encuestas a usuarios del sistema judicial.
  - ✓ Evaluación del acceso, comprensión y legitimidad de las decisiones automatizadas.
- ❖ **Criterios de verificación:**
  - ✓ Informe anual de cumplimiento.
  - ✓ Base de datos de casos críticos o litigios asociados a la IA.
  - ✓ Participación ciudadana documentada.
  - ✓ Mecanismos de corrección o suspensión inmediata si se detecta afectación a derechos.

#### ✚ **Fase 5: Ajuste Normativo y Consolidación Legal**

- ❖ **Objetivo:** Convertir el protocolo en política pública judicial y fuente normativa obligatoria.
- ❖ **Responsables:** Congreso de la República, Ministerio de Justicia, Corte Suprema de Justicia.
- ❖ **Actividades clave:**
  - ✓ Proyecto de ley para regular la IA en la justicia penal.
  - ✓ Inclusión del protocolo en el Código de Procedimiento Penal.
  - ✓ Integración del protocolo en los planes de transformación digital del poder judicial

❖ **Criterios de verificación:**

- ✓ Aprobación legislativa de la ley correspondiente.
- ✓ Modificación normativa explícita.
- ✓ Incorporación en reglamentos judiciales internos.
- ✓ Observancia por todos los operadores jurídicos

✚ **Principios Transversales del Protocolo**

- ✓ Primacía del juicio humano. La IA nunca reemplaza al juez.
- ✓ Transparencia algorítmica. Todo procesamiento debe ser explicable.
- ✓ Debido proceso ampliado. Derecho a controvertir decisiones asistidas por IA.
- ✓ Pro homine. En caso de duda, prevalece la protección del procesado.
- ✓ Participación ciudadana. Observación, crítica y control democrático continuo.

**5.2.14. Diseño algorítmico compatible con principios constitucionales**

Un sistema de inteligencia artificial diseñado para intervenir en el ámbito judicial penal debe construirse con una arquitectura centrada en los principios del Estado social de derecho, tal como lo exige la Constitución Política de Colombia (artículos 1, 2 y 29), esto implica que desde su diseño no se priorice la eficiencia operativa sobre los derechos fundamentales, sino que se incorporen, como criterios de programación, valores como la dignidad humana, el derecho a la defensa, la presunción de inocencia, y la imparcialidad judicial, esta exigencia no puede ser vista como una carga, sino como una garantía institucional frente al uso de tecnologías disruptivas que podrían desbordar las competencias constitucionales de la función judicial.

Además, el diseño algorítmico debe prever explícitamente escenarios de control, exclusión y reversibilidad cuando el sistema detecte datos o patrones que puedan derivar en decisiones discriminatorias o injustas, siguiendo lo propuesto por Elena de Luis García (2023),

esto significa incorporar desde el código “umbrales de veto constitucional”, es decir, limitaciones estructurales que impidan al sistema sugerir decisiones que vulneren garantías mínimas, estos umbrales deben ser definidos por expertos en derecho penal, constitucional y ética jurídica, en lugar de desarrolladores exclusivamente técnicos, para asegurar una visión humanista y garantista en la codificación del sistema.

El protocolo de diseño debe contemplar la pluralidad social y jurídica del país, el sistema no puede asumir una visión única del derecho penal, sino que debe ser sensible a contextos socioculturales diversos, incluyendo variables como etnicidad, género, estratificación social, y condiciones de vulnerabilidad, de esta manera, se evita una estandarización nociva de la justicia, y se garantiza que el sistema actúe conforme a una racionalidad jurídica ajustada al principio pro persona, propio del bloque de constitucionalidad en Colombia.

#### **5.2.15. Trazabilidad y explicabilidad del sistema (Explainable AI)**

Toda operación realizada por el algoritmo quede registrada en un historial accesible y auditable, incluyendo las variables utilizadas, las ponderaciones aplicadas y la secuencia lógica de razonamiento seguida, esto responde al principio de publicidad y al derecho a la defensa, permitiendo que las partes procesales puedan comprender y controvertir cualquier recomendación o resultado emitido por la IA, la jurisprudencia constitucional ha establecido que todo elemento que influya en la decisión judicial debe ser susceptible de control y crítica razonada (Sentencia C-252 de 2019).

En cuanto a la explicabilidad, el sistema debe ser capaz de generar resultados que puedan traducirse a lenguaje jurídico claro, comprensible y accesible para los jueces, las partes y la ciudadanía, no basta con que la IA arroje un resultado estadístico o un valor numérico, es imprescindible que explique el razonamiento detrás de la sugerencia, incluyendo qué elementos consideró, por qué los priorizó y cómo llegó a determinada conclusión, este requisito conecta con

el derecho a la motivación de las sentencias y con la prohibición de decisiones arbitrarias o carentes de razonamiento.

Asimismo, la explicabilidad es un requisito ético, jurídico y político, ético, porque compromete la autonomía del juez como garante de derechos; jurídico, porque asegura el cumplimiento del principio de legalidad procesal y material y político, porque permite que la ciudadanía pueda auditar el funcionamiento del poder judicial en una democracia, de esta manera, la IA no puede operar como una caja negra, sino como una herramienta transparente que se somete al escrutinio constitucional y al control democrático.

#### **5.2.16. Entrenamiento de datos con control judicial previo**

El proceso de entrenamiento de un sistema de IA judicial debe realizarse bajo estricta supervisión de un órgano judicial autónomo e interdisciplinario, que garantice que los datos utilizados no estén contaminados por sesgos estructurales ni vulneren derechos fundamentales, en este sentido, se deben excluir sentencias que contengan violaciones al debido proceso, errores judiciales graves, discriminación directa o indirecta, o que carezcan de motivación suficiente, como advierte David Alonso Roa (2021), entrenar IA con decisiones judiciales sesgadas solo perpetúa esas injusticias en cadena.

Además, debe garantizarse la anonimización de los datos judiciales usados, eliminando toda referencia personal que pueda poner en riesgo la privacidad de los individuos involucrados, esta medida está en línea con la Ley Estatutaria 1581 de 2012 sobre protección de datos personales y con los estándares internacionales del Comité de Derechos Humanos de la ONU, el uso indebido de datos sensibles en sistemas de IA podría constituir no solo una falta disciplinaria, sino incluso una violación de derechos fundamentales con consecuencias penales o administrativas.

El control judicial previo sobre los datos garantiza un uso ético y responsable de la información pública, en contextos de alta desigualdad como el colombiano, el uso indiscriminado de datos judiciales podría reforzar patrones de criminalización selectiva, por eso, el protocolo debe incluir un procedimiento de evaluación de impacto de los datos (Data Protection Impact Assessment - DPIA), que permita anticipar consecuencias negativas, ajustar metodologías de entrenamiento y excluir bases contaminadas o no representativas.

### **5.2.17. Supervisión humana continua y no delegable**

Uno de los pilares del protocolo es que la función de juzgar no puede ser delegada en una máquina, aunque la IA pueda asistir en la elaboración de proyectos de sentencia o sugerencias normativas, la decisión final siempre debe recaer en un juez humano, dotado de capacidad de deliberación ética, sensibilidad jurídica y vocación de protección de derechos, así lo ha ratificado la Corte Constitucional en sentencias como la T-220 de 2021, en las que se señala que la administración de justicia no puede convertirse en una operación automatizada sin reflexión crítica.

En esta línea, el protocolo exige que toda recomendación emitida por la IA deba ser evaluada, interpretada y, en su caso, rechazada por el juez con criterios jurídicos claros, el juez debe ejercer un control racional sobre cada sugerencia, motivando su aceptación o rechazo de forma escrita y registrada en el expediente, este procedimiento garantiza la vigencia del principio de independencia judicial y evita una “heteronomía tecnológica” en la función jurisdiccional, es decir, que las decisiones emanen de un código y no de una conciencia jurídica.

Por lo tanto, la IA debe concebirse como un auxiliar técnico, nunca como un decisor autónomo. Su rol es contribuir a la eficiencia, no sustituir la función deliberativa, este principio también debe ser observado por las fiscalías y defensores, quienes podrán usar IA como herramienta de análisis, pero nunca para definir estrategias o formular imputaciones de forma

automatizada, la centralidad del juicio humano es la garantía última de que la justicia siga siendo un espacio de humanidad, no de cálculo ciego.

### **5.2.18. Registro digital obligatorio de intervenciones de IA**

Para garantizar la transparencia y trazabilidad del uso de inteligencia artificial en la función judicial, el protocolo contempla la creación de un “Acta Digital de Asistencia Algorítmica”, documento obligatorio que debe anexarse al expediente en cada caso donde la IA haya tenido alguna intervención, ya sea en forma de sugerencia de jurisprudencia, recomendaciones de redacción o análisis predictivo, este documento deberá especificar el nombre del módulo de IA utilizado, la fecha, los datos procesados, los resultados ofrecidos, las observaciones del operador judicial y la decisión tomada respecto al uso de esa información, esta acta busca no solo la trazabilidad interna, sino el control posterior por parte de instancias superiores o disciplinarias.

Este registro también cumple con una función constitucional clave: preservar el derecho de defensa y el principio de contradicción, al conocer que una herramienta algorítmica participó en alguna fase del proceso decisorio, la defensa podrá solicitar la revisión de esa intervención, impugnar sus fundamentos técnicos o jurídicos, y exigir explicaciones al juez sobre su grado de influencia, como lo establece la jurisprudencia constitucional (Sent. C-252/19 y T-033/22), toda parte en un proceso debe tener acceso completo a los elementos que influyan en la decisión judicial, incluyendo aquellos que provengan de sistemas computacionales.

El “Acta Digital de Asistencia Algorítmica” se convierte así en una herramienta de rendición de cuentas, esencial para evitar la opacidad tecnológica, también representa un límite material al uso irrestricto de estas tecnologías, puesto que obliga al juez a declarar abiertamente su uso y a justificar su peso dentro de la sentencia, esta exigencia permite mantener la lógica del

control de convencionalidad y bloque de constitucionalidad, al someter todo acto judicial a una revisión desde los principios del debido proceso, la imparcialidad y la motivación sustancial.

### **5.2.19. Auditoría externa e independiente del sistema**

Para garantizar el respeto a los derechos fundamentales, el protocolo establece que todo sistema de inteligencia artificial utilizado en la justicia penal deberá ser objeto de auditorías externas, independientes y periódicas, realizadas por universidades públicas, organizaciones de derechos humanos, colegios de abogados y comités técnicos interinstitucionales estas auditorías deben evaluar tanto el funcionamiento técnico del sistema como su impacto jurídico y ético en los procesos penales, la Corte Constitucional ha reconocido la necesidad de estos mecanismos de fiscalización en entornos digitales (T-443/20), en donde la tecnología puede imponer dinámicas autorreferenciales, difíciles de controlar por los operadores judiciales.

Además, las auditorías deben considerar la posibilidad de sesgos algorítmicos, fallos en el aprendizaje automático o efectos discriminatorios indirectos, especialmente en contextos donde el sistema ha sido entrenado con datos históricos plagados de selectividad penal o desigualdad estructural, esto implica verificar si la IA presenta disparidades en su rendimiento según el grupo étnico, social o territorial de los procesados, estos informes deberán ser públicos, discutidos en audiencias académicas y remitidos a la Defensoría del Pueblo, con el fin de propiciar correctivos cuando se identifiquen desviaciones graves del marco jurídico constitucional.

En caso de detectarse irregularidades relevantes o impactos negativos no corregidos, el protocolo activa una cláusula de reversibilidad tecnológica, por la cual se suspende inmediatamente el uso del sistema hasta que se ajusten sus deficiencias, este principio de reversibilidad, además de obedecer a la precaución jurídica, se basa en la noción de tecnología al servicio del derecho y no al revés, como lo advierte Cristian David Salazar (2022), quien

sostiene que la innovación no puede constituirse en excusa para el debilitamiento de las garantías procesales fundamentales.

#### **5.2.20. Protocolo de objeción y exclusión por parte de la defensa**

Con el fin de preservar el principio de contradicción (Art. 29 CP), el protocolo garantiza a los procesados y sus defensores el derecho a objetar o solicitar la exclusión de cualquier insumo procesal derivado de sistemas de inteligencia artificial, esta facultad podrá ser ejercida si se considera que dicho insumo es impreciso, sesgado, opaco o contrario a los derechos fundamentales, una vez presentada la objeción, el juez deberá resolverla mediante auto motivado, con posibilidad de solicitar peritajes externos o realizar audiencias específicas de contradicción tecnológica.

Este mecanismo responde a la necesidad de proteger la asimetría tecnológica existente entre el Estado y los procesados, particularmente cuando estos últimos pertenecen a sectores sociales con escaso acceso a recursos jurídicos y tecnológicos, permitir que un procesado pueda controvertir activamente la participación de un sistema de IA en su caso es una forma de garantizar el juicio justo, la igualdad procesal y la efectividad del derecho de defensa, en este sentido, se acoge el principio de máxima protección del imputado, ampliamente desarrollado en la doctrina del constitucionalismo penal (Ferrajoli, 2001).

Además, la objeción permite que los sistemas de IA no se consoliden como autoridades infalibles dentro del proceso, lo cual sería contrario al carácter dialógico y deliberativo del juicio penal, el derecho a disentir de una propuesta algorítmica, exigir transparencia y demandar justificación es, en última instancia, un derecho a ser tratado como persona y no como objeto del proceso, esta visión es coherente con los estándares de la Corte IDH sobre participación efectiva y defensa sustancial (Caso López Lone y otros vs. Honduras, 2015).

### **5.2.21. Limitación de usos permitidos según fase procesal**

El protocolo establece una delimitación clara respecto a las etapas del proceso penal en las que se puede usar IA, con el fin de evitar abusos o delegaciones indebidas, se prohíbe expresamente el uso de IA en las fases de imputación, imposición de medidas cautelares, evaluación de testimonios, y determinación de responsabilidad penal, estas fases requieren un juicio valorativo, ético y dialógico que la IA no puede suplir sin afectar la presunción de inocencia, el principio de inmediación y la dignidad procesal del acusado.

En cambio, el uso podrá permitirse únicamente en las fases posprobatorias, específicamente en el apoyo a la redacción de sentencias o en la búsqueda de precedentes judiciales, incluso en estos casos, el uso debe ser declarado y justificado, sometido a los controles que el protocolo establece, el fundamento de esta limitación es que, en el proceso penal, no basta con aplicar lógicas estadísticas, sino que debe preservarse el principio de individualización de la justicia, lo cual exige ponderaciones humanas, sensibles a las particularidades del caso concreto.

El uso en fases sensibles podría implicar una automatización del poder punitivo, violando el principio de legalidad y convirtiendo la función judicial en una cadena de producción automática, como lo recuerda la Corte Suprema en la SP1445-2021, la valoración probatoria debe ser racional, transparente y abierta al debate, la IA, si bien es cierto puede ayudar a sistematizar información, no está en capacidad de reemplazar el juicio humano sobre la verdad procesal.

### **5.2.22. Inclusión de un módulo de aprendizaje supervisado**

El protocolo permite el uso de sistemas con aprendizaje automático (machine learning), pero únicamente si se adoptan mecanismos estrictos de supervisión humana y control judicial, cada ciclo de aprendizaje debe estar previamente autorizado por una instancia de validación

interdisciplinaria que evalúe qué datos se usarán, con qué finalidad y bajo qué criterios de evaluación, además, deberá implementarse un sistema de versionado, en el cual cada nueva iteración del sistema sea revisada antes de su puesta en operación.

Este enfoque permite evitar la evolución no controlada del algoritmo, fenómeno conocido como “drift algorítmico”, que puede provocar que el sistema aprenda patrones perjudiciales o discriminatorios sin intervención humana, en justicia penal, esto podría traducirse en decisiones más punitivistas, predictivamente selectivas o simplemente erróneas, por ello, se exige que todo aprendizaje posterior a la implementación inicial sea limitado, reversible y documentado.

Además, el protocolo contempla que el aprendizaje supervisado sea acompañado de evaluaciones de impacto jurídico, que midan cómo los cambios en el algoritmo afectan los resultados judiciales, especialmente en relación con grupos vulnerables, de esta manera, se mantiene un equilibrio entre innovación y control constitucional, asegurando que la IA permanezca subordinada al interés superior de la justicia y a la garantía de los derechos humanos.

### **5.2.23. Estándares de interoperabilidad con el sistema judicial**

Toda solución tecnológica aplicada a la administración de justicia penal debe ser plenamente interoperable con los sistemas ya existentes, como LEX, el sistema de gestión judicial de la Rama Judicial, o las plataformas del Ministerio Público y la Fiscalía General, esta interoperabilidad busca evitar duplicidades, inconsistencias de datos, pérdida de información y, sobre todo, garantizar que la IA funcione como una herramienta integrada al ecosistema judicial, no como un sistema paralelo sin control institucional.

La interoperabilidad también implica cumplir con los estándares de seguridad digital y protección de datos, en consonancia con la Ley 1581 de 2012, la Ley de Habeas Data judicial y los lineamientos de la Corte Constitucional sobre privacidad y acceso a la información pública

(T-015/2020), los sistemas de IA deben contar con protocolos de cifrado, control de acceso, trazabilidad de usuarios y copias de respaldo, de manera que se garantice la integridad de los datos procesales y la confianza en el sistema de justicia.

Además, el diseño del sistema debe considerar la usabilidad para todos los actores del proceso judicial, incluyendo jueces, abogados, fiscales, defensores y partes procesales, no se trata de imponer tecnologías complejas o restrictivas, sino de facilitar el acceso a herramientas que verdaderamente complementen el trabajo judicial sin reemplazar su dimensión ética y deliberativa, la interoperabilidad es, por lo tanto, no solo una condición técnica, sino un principio democrático de inclusión, accesibilidad y gobernanza digital.

#### **5.2.24. Revisión judicial reforzada cuando haya intervención algorítmica**

El protocolo exige que toda sentencia en cuya elaboración haya intervenido un sistema de IA deba mencionar explícitamente dicha intervención, indicando el módulo utilizado, el contenido de la sugerencia y la valoración que el juez le otorgó, esta referencia permitirá a los tribunales de segunda instancia realizar un control más riguroso de legalidad y constitucionalidad, no solo de la decisión misma, sino del uso de tecnología en la construcción del fallo.

Además, esta revisión reforzada contribuye a garantizar el control difuso de convencionalidad, principio que exige a los jueces examinar la conformidad de sus decisiones con los tratados internacionales de derechos humanos (Corte IDH, OC-18/03), si se detecta que la intervención algorítmica afectó negativamente el debido proceso, la imparcialidad, o la motivación sustancial, el fallo deberá ser anulado o devuelto para su reformulación, esta medida previene la consolidación de fallos viciados por delegación indebida en sistemas no humanos.

El principio de revisión reforzada contribuye a desarrollar una cultura institucional de rendición de cuentas tecnológica, en la que los jueces no solo respondan por sus decisiones

jurídicas, sino también por las herramientas que eligen para respaldarlas, así se afianza la idea de que la IA no es neutral ni infalible, y que debe estar siempre sujeta al imperio del derecho, tal como corresponde a todo instrumento del poder público en un Estado constitucional.

#### **5.2.25. Mecanismo ciudadano de vigilancia y observación tecnológica**

El protocolo propone la creación del Observatorio Nacional de Justicia Algorítmica, una instancia de carácter mixto (académico, judicial, social y técnico) encargada de monitorear la implementación, operación y consecuencias del uso de IA en el sistema judicial penal colombiano, este observatorio tendrá funciones de veeduría, producción de informes, formulación de alertas tempranas y recomendaciones de política pública, su existencia materializa el principio de participación ciudadana en el control del poder judicial, especialmente en contextos de transformación tecnológica.

Además, este observatorio será un espacio para la articulación entre academia, sociedad civil y órganos del Estado, permitiendo una evaluación interdisciplinaria de los efectos éticos, jurídicos y sociales de la automatización judicial, su tarea no es oponerse al uso de IA, sino garantizar que dicha implementación sea coherente con los valores democráticos, los derechos fundamentales y las exigencias del debido proceso, como ha señalado Miró (2022), la digitalización de la justicia no puede ser tecnocrática, sino dialogada, regulada y sometida a vigilancia pública.

El observatorio servirá como canal para recoger denuncias ciudadanas, elaborar indicadores de impacto y promover foros públicos de discusión, convirtiéndose en un pilar de la justicia digital abierta, transparente y controlada, este tipo de iniciativas ha sido respaldada por organismos internacionales como el Alto Comisionado de Naciones Unidas para los Derechos Humanos (A/HRC/49/52), que advierten sobre la necesidad de mecanismos participativos para gobernar el uso de tecnologías en funciones públicas esenciales.

## 6. CONCLUSIONES

La revisión sistemática del marco normativo y doctrinal evidencia que la implementación de inteligencia artificial en la elaboración de sentencias penales en Colombia requiere un análisis profundo de los alcances, límites y vacíos legales existentes, los principios fundamentales del derecho, como la motivación judicial, la transparencia, la explicabilidad y el control humano significativo, emergen como elementos esenciales para garantizar que la IA opere conforme a los parámetros del debido proceso y del Estado Social de Derecho, sin la observancia de estos principios, cualquier intento de automatización podría comprometer la legitimidad de la función jurisdiccional.

El análisis comparativo de la jurisprudencia nacional e internacional permite identificar estándares mínimos de constitucionalidad y buenas prácticas trasladables al contexto colombiano, casos y resoluciones internacionales muestran que la IA solo puede intervenir como herramienta de apoyo, bajo supervisión activa del juez, con trazabilidad completa de sus decisiones y con mecanismos de revisión y corrección de sesgos, estas experiencias proporcionan referentes concretos para estructurar un modelo de implementación que respete los principios jurídicos, garantizando la transparencia, la rendición de cuentas y la explicabilidad de las decisiones judiciales.

La valoración crítica de los riesgos y beneficios revela que, aunque la IA puede optimizar procesos judiciales, mejorar la gestión documental y facilitar la consulta de jurisprudencia, también plantea amenazas significativas, como la despersonalización de la justicia, la reproducción de sesgos estructurales y la reducción del juicio crítico del juez, por ello, la

adopción tecnológica debe estar subordinada a criterios operativos claros, asegurando que la IA nunca sustituya la deliberación humana ni comprometa la motivación jurídica de las sentencias, respetando así los principios de imparcialidad y juez natural.

El diseño del protocolo de implementación propuesto responde directamente a la necesidad de establecer controles ex ante y ex post, gobernanza de datos, evaluación de impacto algorítmico, límites a la automatización y métricas de seguimiento, incluyendo revisión de sesgos y verificación de motivación, este enfoque asegura que la IA actúe como herramienta auxiliar y no sustitutiva del juez, fortaleciendo la legitimidad democrática, la transparencia operativa y la protección de los principios jurídicos fundamentales en el ámbito penal.

Con relación a la pregunta problema, se concluye que la compatibilidad de los sistemas de IA con la función jurisdiccional penal colombiana depende del respeto estricto a los principios de transparencia, explicabilidad y control humano, solo bajo estas condiciones la tecnología puede integrarse al proceso judicial sin vulnerar el debido proceso ni los principios esenciales del Estado Social de Derecho, garantizando un equilibrio sólido entre innovación tecnológica y respeto irrestricto a los fundamentos del derecho.

### **6.1. Principales hallazgos de la investigación**

Uno de los hallazgos más relevantes de esta investigación es la ausencia de una regulación específica en Colombia sobre el uso de inteligencia artificial en el proceso penal, especialmente en la elaboración de sentencias, a pesar de algunos avances normativos generales en materia de transformación digital del Estado, no existe un marco legal que delimite el uso de algoritmos en funciones jurisdiccionales, lo que genera un vacío normativo con potencial impacto sobre principios fundamentales del derecho, particularmente en contextos donde se ejerce el poder punitivo del Estado con mayor intensidad.

Se evidenció que el ordenamiento constitucional colombiano no es ajeno a la transformación tecnológica, pero establece límites claros en la función judicial, entre ellos el principio de motivación de las decisiones, la imparcialidad del juez, el respeto al juez natural y la dignidad humana como guía de toda actuación estatal, la Corte Constitucional ha señalado en sentencias como T-443 de 2020 y T-033 de 2022 que la incorporación de tecnologías debe subordinarse a estos principios, evitando que la deliberación ética y jurídica del juez se vea sustituida por procesos automatizados.

Asimismo, se identificó que la motivación judicial constituye un principio esencial que permite evaluar la razonabilidad de las decisiones, la implementación de IA sin criterios claros y sin trazabilidad algorítmica puede derivar en fallos descontextualizados, impersonales y carentes de fundamentación, contraviniendo los principios de transparencia, explicabilidad y control humano que sustentan la legitimidad de la decisión judicial.

La investigación también reveló que la justicia digital puede amplificar desigualdades existentes, en contextos de vulnerabilidad estructural, la aplicación de sistemas algorítmicos entrenados con datos históricos podría reproducir sesgos, estigmatizar patrones sociales o reforzar estereotipos, advertencia respaldada tanto por doctrinantes como David Alonso Roa (2021) como por organismos internacionales como el Consejo de Europa, este hallazgo subraya la importancia de diseñar protocolos que mitiguen impactos sobre la igualdad procesal.

En la revisión de jurisprudencia internacional se observó que, aunque no hay decisiones que aborden directamente la IA en sentencias penales, tanto la Corte Interamericana de Derechos Humanos como el Tribunal Europeo de Derechos Humanos han insistido en la necesidad de motivación individualizada, contradicción efectiva y participación procesal, casos como Hirsi Jamaa y otros vs. Italia (TEDH, 2012) enfatizan que las decisiones automatizadas no pueden

reemplazar la deliberación personalizada exigida por la justicia penal, un principio que debe trasladarse al contexto colombiano.

Otro hallazgo relevante es que, en Colombia, existen líneas jurisprudenciales que permiten anticipar un marco de referencia preventivo, la Corte Suprema ha enfatizado la necesidad de verificabilidad y transparencia en medios probatorios tecnológicos (SP1445-2021), lo que puede aplicarse como estándar ante sistemas de IA, asegurando que la automatización respete los principios de control, trazabilidad y responsabilidad judicial.

Desde el punto de vista teórico, se constató que la dogmática penal y el constitucionalismo principialista no han sido suficientemente considerados en los debates sobre justicia digital, gran parte de la literatura se concentra en aspectos técnicos o de política pública, sin analizar críticamente los límites del acto de juzgar, este estudio propone un marco de análisis centrado en la deliberación ética, la motivación razonada y la autonomía del juez como principios irrenunciables frente a la automatización.

Se observó que el debate entre tecnocracia judicial y justicia humanizada está en un punto crítico, mientras algunos modelos internacionales promueven la “justicia predictiva” como sinónimo de eficiencia, en Colombia es imprescindible preservar al juez como sujeto central de la función jurisdiccional, asegurando que la IA solo opere como apoyo y no sustituya la deliberación basada en principios jurídicos fundamentales.

Se concluye que la implementación de IA en la justicia penal solo será compatible con el orden constitucional colombiano si se realiza bajo principios sólidos: motivación, imparcialidad, juez natural, transparencia, explicabilidad y control humano, la incorporación de tecnología debe ir acompañada de formación especializada para los operadores judiciales y de mecanismos de

supervisión que aseguren que la función jurisdiccional siga siendo legítima, ética y centrada en el respeto de los principios del derecho.

## **6.2. Aportes teóricos y prácticos**

Uno de los aportes teóricos centrales de esta investigación radica en haber abordado de manera crítica y articulada la relación entre inteligencia artificial, derecho penal y principios jurídicos fundamentales, mientras buena parte de los estudios sobre IA y justicia se concentran en aspectos técnicos o administrativos, este trabajo sitúa el problema en el corazón del acto de juzgar, entendido como un ejercicio profundamente humano que no puede ser sustituido por lógica computacional sin afectar su legitimidad, esta perspectiva se fundamenta en autores como Ferrajoli (2001), Zaffaroni (2011) y De Luis García (2021), quienes enfatizan la centralidad de los principios en la administración de justicia.

Asimismo, el trabajo delimita de manera precisa los alcances y límites de la tecnología en el ámbito judicial, concentrándose en el momento resolutorio de la sentencia penal, esta focalización permite analizar cómo los principios de motivación, imparcialidad y juez natural pueden ser tensionados por la automatización, y visibiliza los riesgos de que la introducción de IA ocurra sin criterios normativos claros, especialmente en contextos donde el poder punitivo del Estado se ejerce con mayor intensidad.

Desde el plano doctrinal, la investigación integra reflexiones de la dogmática penal, el constitucionalismo principialista y la ética jurídica, estableciendo un puente necesario entre el análisis técnico de la IA y los principios jurídicos sustantivos, esta integración permite superar la visión instrumental de la tecnología y replantear su sentido en la administración de justicia, incorporando conceptos clave como transparencia algorítmica, control humano efectivo, motivación judicial y reversibilidad procesal.

En términos de innovación académica, se propone una tipología de riesgos derivados del uso de IA en el ámbito penal, diferenciando entre riesgos epistémicos (opacidad algorítmica), éticos (despersonalización de la justicia) y jurídicos (afectación de los principios procesales), esta clasificación facilita una comprensión estructurada del problema y ofrece herramientas para evaluar y diseñar protocolos normativos que respeten los principios del derecho.

Otro aporte relevante es el análisis comparado de jurisprudencia internacional, que examina decisiones del Tribunal Europeo de Derechos Humanos, la Corte Interamericana de Derechos Humanos y tribunales de países como Estados Unidos, Brasil, Argentina y Francia, esta comparación permite identificar modelos que integran tecnología en el proceso penal sin desdibujar la función del juez, ofreciendo lecciones aplicables al contexto colombiano.

En el plano práctico, la investigación propone un protocolo de implementación de IA en la elaboración de sentencias penales basado en principios jurídicos, con fases que cubren evaluación de impacto, control ex ante y ex post, trazabilidad, reversibilidad y explicabilidad, esta estructura se alinea con estándares internacionales, como los del Consejo de Europa (2019) y el Comité de Derechos Humanos de la ONU (2021), garantizando que la IA opere como instrumento auxiliar y no sustitutivo del juez.

De manera complementaria, se visibiliza la necesidad de formación interdisciplinaria para los operadores judiciales, incorporando contenidos sobre ética algorítmica, lógica de programación y análisis de riesgos tecnológicos, esta preparación asegura que los principios de motivación, imparcialidad y juez natural se mantengan vigentes en entornos asistidos por tecnología.

Asimismo, se plantea la creación de un Observatorio Nacional de Justicia Algorítmica, concebido como un ente autónomo y consultivo que monitoree el uso de tecnologías en la justicia, genere alertas tempranas y formule recomendaciones de política pública, esta iniciativa

refuerza los principios de control, transparencia y rendición de cuentas en la función jurisdiccional.

La investigación ofrece herramientas metodológicas para estudios futuros, como la matriz de análisis de jurisprudencia, que permite sistematizar casos y detectar patrones en la interpretación de principios jurídicos en contextos tecnológicos, en conjunto, los hallazgos y propuestas de este trabajo contribuyen a una modernización judicial que respete los principios del derecho, asegurando que la IA sea un instrumento que potencie, pero no sustituya, la función humana, ética y deliberativa del juzgamiento penal..

### **6.3. Líneas futuras de investigación**

A partir de los hallazgos de esta investigación, se evidencia una necesidad urgente de desarrollar estudios empíricos que analicen cómo los sistemas judiciales están incorporando herramientas de inteligencia artificial en las distintas etapas del proceso penal en Colombia, si bien es cierto, este trabajo se centró en el momento resolutivo, es imprescindible explorar la presencia y los efectos de la IA en etapas como la investigación preliminar, la formulación de imputación o la fijación de medidas cautelares, dado que el sesgo algorítmico puede impactar anticipadamente la presunción de inocencia y la libertad personal.

Una línea futura prioritaria es el análisis crítico de los modelos predictivos de reincidencia o riesgo, como los que se han implementado en Estados Unidos y algunos países europeos, estos sistemas, basados en big data y aprendizaje automático, presentan serios cuestionamientos éticos y jurídicos, pues tienden a estandarizar perfiles criminales sin considerar el contexto social ni las particularidades del individuo, en Colombia, su uso sería especialmente problemático dada la informalidad institucional, las brechas de datos y los riesgos de discriminación estructural.

También se abre una línea interesante en torno al control judicial de algoritmos, ¿Qué estándares debe exigir el juez al momento de evaluar decisiones asistidas por IA? ¿Cómo se puede garantizar la contradicción efectiva y la explicación técnica en un entorno opaco?, estudios interdisciplinarios que aborden la relación entre epistemología jurídica, teoría de la prueba y transparencia algorítmica podrían aportar criterios sólidos para responder a estas preguntas desde un enfoque garantista.

Otra dirección necesaria es la revisión del impacto de la IA sobre los derechos colectivos y los pueblos étnicos, en el contexto colombiano, marcado por la diversidad cultural y la vulnerabilidad de comunidades afrodescendientes e indígenas, la automatización judicial sin enfoque diferencial podría invisibilizar sus cosmovisiones jurídicas, imponer lógicas de juicio ajenas y profundizar la exclusión del sistema penal, esta línea demanda investigaciones interculturales que crucen derecho, sociología y antropología.

Igualmente, se requiere estudiar el papel de la formación judicial frente a los desafíos tecnológicos. ¿Qué competencias deben adquirir los jueces en la era digital? ¿Cómo equilibrar la alfabetización tecnológica con la reflexión jurídica crítica?, investigaciones futuras pueden diseñar perfiles formativos y metodologías pedagógicas adaptadas a los nuevos contextos, incluyendo cursos obligatorios en ética de la tecnología, lógica computacional y derechos digitales en las escuelas judiciales.

Una línea poco explorada pero fundamental es el análisis del consentimiento procesal frente a decisiones automatizadas, ¿Puede una parte en un juicio aceptar voluntariamente la intervención de un sistema algorítmico? ¿Qué límites establece el derecho al libre desarrollo de la personalidad en este contexto?, el consentimiento informado, ampliamente debatido en el

derecho médico y el derecho del consumidor, puede ofrecer una clave para entender los márgenes de autonomía en el proceso penal tecnológico.

Asimismo, es urgente desarrollar estudios sobre responsabilidad jurídica por errores algorítmicos, si una IA comete un error que afecta la libertad o los derechos de una persona, ¿quién responde? ¿El Estado, el programador, el juez?, la doctrina penal y administrativa colombiana no ha resuelto aún estas tensiones., futuras investigaciones deben abordar esta problemática desde la teoría de la imputación, el principio de legalidad y la garantía de reparación integral.

Otra línea clave está relacionada con la auditoría algorítmica como función pública, el diseño de mecanismos institucionales que permitan evaluar, auditar y desmontar sistemas algorítmicos en la justicia debe ser objeto de estudio, ello implica investigaciones jurídicas y técnicas sobre modelos de auditoría externa, estándares de rendición de cuentas, certificación de sistemas y participación ciudadana en el control de la digitalización judicial.

Se recomienda profundizar en estudios de comparación jurídica con países de América Latina que ya estén avanzando en estas discusiones, como Brasil, Argentina o México, estas comparaciones pueden permitir construir un modelo latinoamericano de justicia algorítmica basado en el garantismo penal, el pluralismo jurídico y los derechos sociales, en contraste con el modelo anglosajón centrado en eficiencia predictiva y control social.

Una línea relevante para la política pública es el seguimiento legislativo y constitucional de los desarrollos tecnológicos en la rama judicial, esta línea de investigación debe indagar en la legitimidad democrática de las normas que regulan la IA, el papel del Congreso en su diseño y control, y la necesidad de una ley estatutaria que garantice los derechos fundamentales frente a decisiones automatizadas.

A nivel internacional, una línea de investigación pertinente es la armonización normativa entre derechos humanos y tecnologías judiciales, observando cómo los sistemas regionales (como la Corte IDH o el TEDH) están construyendo estándares transnacionales que orienten a los Estados en el desarrollo de políticas tecnológicas sin sacrificar las garantías procesales, esta línea contribuiría a construir una base jurídica común para enfrentar la globalización tecnológica del poder judicial.

Finalmente, se sugiere explorar el papel de las tecnologías libres y abiertas en la construcción de una justicia digital más transparente y democrática, frente a la opacidad de los algoritmos privados y las restricciones de propiedad intelectual, el uso de software abierto puede contribuir a una mayor auditoría ciudadana, autonomía tecnológica estatal y soberanía jurídica en el diseño de sistemas judiciales automatizados.

## 7. BIBLIOGRAFÍA

- Acosta, M. C. (2021). *Inteligencia artificial y derecho: Entre el desarrollo tecnológico y la garantía de derechos*. Editorial Temis.
- Alexy, R. (2002). *Teoría de los derechos fundamentales*. Madrid: Centro de Estudios Constitucionales.
- Alonso Roa, D. (2021). *Inteligencia artificial y derecho penal: retos y perspectivas*. Bogotá: Editorial Jurídica.
- Alpa, G. (2010). *La cultura jurídica europea: Historia y evolución*. Marcial Pons.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica.
- Aristizábal, C. A. (2023). *Constitución, tecnología y justicia predictiva: retos desde el neoconstitucionalismo*. *Revista Derecho y Cambio Social*, 57, 79–101.
- Asamblea General de las Naciones Unidas. (2019). *Informe sobre el impacto de las nuevas tecnologías en los derechos humanos*. A/HRC/41/35.
- Baratta, A. (2004). *Derecho y nuevas tecnologías: Entre la amenaza y la oportunidad*. Editorial Tirant lo Blanch.
- Bercovitz, A. (2018). *Transformación digital y seguridad jurídica: el papel del derecho en la sociedad de la información*. *Revista de Derecho Privado*, 32(2), 11–28.
- Bernal, C. (2005). *El derecho a una sentencia motivada: contenido y límites en el Estado constitucional*. *Revista Derecho del Estado*, 15, 27–51.
- Boaventura de Sousa Santos, B. (2009). *Refundación del Estado en América Latina: Perspectivas desde una epistemología del sur*. Siglo XXI Editores.
- Cabrera, L. M. (2020). *Debida diligencia algorítmica y principios de justicia digital*. *Revista Colombiana de Derecho*, 46(3), 210–233.
- Carbonell, M. (2011). *Neoconstitucionalismo(s)*. Editorial Trotta.
- Cassagne, J. C. (2010). *El principio de legalidad y la actividad administrativa en tiempos de informatización*. *Revista Iberoamericana de Derecho Público*, 23(1), 9–28.

- Cerda, A. (2019). Ética y gobernanza de la inteligencia artificial: Un enfoque desde los derechos humanos. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, 14(42), 55–78.
- Cifuentes, A. (2020). Tecnología y derecho procesal penal: Una aproximación crítica a la evidencia digital. *Revista Justicia y Derecho Penal*, 12(1), 123–147.
- Consejo de Europa. (2019). *Guía sobre inteligencia artificial y derechos humanos en el ámbito judicial*. Estrasburgo: Consejo de Europa.
- Contissa, G., Lagioia, F., & Sartor, G. (2018). The ethics of judicial profiling and sentencing: The Italian debate on predictive justice. In *AI and Ethics*, 1(1), 45–63
- Corte Interamericana de Derechos Humanos. (2012). *Caso Hirsi Jamaa y otros vs. Italia*. Serie C No. 277
- David Alonso Roa, M. (2021). *Justicia digital y desigualdad estructural: desafíos de la inteligencia artificial en el derecho penal*. Bogotá: Editorial Jurídica Nacional.
- De Gregorio, G. (2021). The rise of digital constitutionalism in the European Union. *International Journal of Constitutional Law*, 19(1), 41–70.
- De Luis García, E. (2023). Justicia, inteligencia artificial y derecho de defensa. *IDP: Revista de Internet, Derecho y Política*, 39, 1–20. Recuperado de <https://idp.uoc.edu>
- De Luis García, E. (2023). Detección del discurso de odio online a través de la inteligencia artificial: virtualidades del machine learning. En M. J. Catalán Chamorro (Ed.), *Data-based justice. Artificial Intelligence and Data in Justice* (pp. 1–20). Universidad del País Vasco. Recuperado de <https://www.uik.eus/en/node/7914/pdf>
- De la Maza, I. (2022). Imparcialidad algorítmica y administración de justicia: ¿ficción o posibilidad?. *Revista Derecho y Tecnología*, 3(2), 88–111.
- Domingo, R. (2011). *El espíritu del Derecho romano*. Thomson Reuters Aranzad
- De Luis García, E. (2021). *Justicia, inteligencia artificial y derecho de defensa*. Universitat de València.
- De Luis García, E. (2023). *Digitalización y algoritmización de la justicia*. En F. Miró (Coord.), *Colección Monográfica*.
- Elias, J. (2021). Procesos penales automatizados: ¿sueño de eficiencia o pesadilla de garantías?. *Revista Penal & Sociedad*, 6(4), 59–83.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

- Ferrajoli, L. (2001). Derecho y razón: Teoría del garantismo penal. Editorial Trotta. *Revista de Derecho Penal Económico*, 7(1), 109–137.
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Fundación Karisma. (2022). Justicia algorítmica y derechos digitales en América Latina: ¿quién controla al algoritmo? Bogotá: Karisma.
- Garapon, A., & Lassègue, J. (2018). *Justice digitale: Révolution graphique et rupture anthropologique*. Presses Universitaires de France.
- Gómez, D. A. (2020). La inteligencia artificial en la administración de justicia: análisis de su aplicación en el derecho penal colombiano. *Revista Justicia*, 26(40), 78–95.
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI Magazine*, 38(3), 50–57.
- González, J. D. (2023). Constitución y tecnologías emergentes: retos para el control de convencionalidad. *Revista Constitucional*, 25(3), 67–90.
- Guerrero, R. (2020). La función judicial frente al auge de los algoritmos. *Revista del Instituto Colombiano de Derecho Procesal*, 15, 221–240.
- Hildebrandt, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing.
- Hildebrandt, M. (2022). *Law for computer scientists and other folk*. Oxford University Press
- Hirsi Jamaa y otros vs. Italia, Tribunal Europeo de Derechos Humanos (TEDH), Sentencia de 23 de febrero de 2012.
- Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society*, 4(2).
- Koopman, C., & Dotan, Y. (2020). The algorithmic judge: Justice in the era of machine learning. *Law & Ethics of Human Rights*, 14(1), 1–32.
- Latonero, M. (2018). *Governing artificial intelligence: Upholding human rights & dignity*. Data & Society Research Institute.
- Lessig, L. (1999). *Code and other laws of cyberspace*. Basic Books.
- Malgieri, G. (2020). The concept of fairness in EU data protection law: A review of the jurisprudence. *Computer Law & Security Review*, 36.

- Márquez, J. A. (2022). Algoritmos y sesgo judicial: Retos para la justicia penal en Colombia. *Revista Estudios Socio-Jurídicos*, 24(2), 117–138.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Miró, F. (Coord.). (2023). *Monográfico: Digitalización y algoritmización de la justicia*. Universitat de València.
- Moreno, D. A. R. (2021). *Justicia penal, tecnología y derechos fundamentales: una mirada garantista*. Editorial Universidad del Rosario.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
- Olaya, A. (2021). Debido proceso y transformación digital del sistema judicial en Colombia. *Revista Derecho Penal Contemporáneo*, 35, 63–85.
- Pardo, R. (2020). Transparencia algorítmica y control judicial: Una exigencia del Estado de Derecho. *Revista de Derecho Público*, 101, 95–120.
- Roa, D. A. (2021). Derechos fundamentales y automatización judicial: Desafíos en el contexto colombiano. *Revista Derecho y Sociedad*, 18(3), 211–229.
- Salazar, C. D. (2022). Inteligencia artificial, derecho penal y garantías constitucionales: ¿Un nuevo modelo de justicia? *Revista Iusta*, 56, 144–165.
- Sandoval, A. L. (2022). La ética en el desarrollo de inteligencia artificial para la justicia. *Revista de Filosofía del Derecho*, 12(1), 43–66.
- Sánchez, J. M. (2019). La administración de justicia en tiempos de inteligencia artificial: oportunidades y riesgos. *Revista Derecho, Estado y Sociedad*, 42, 117–134.
- Santos, B. de S. (2003). *Crítica de la razón indolente: Contra el desperdicio de la experiencia*. Desclée de Brouwer.
- Schmitt, C. (2009). *Teología política*. Trotta.
- Supiot, A. (2005). *Homo juridicus: Ensayo sobre la función antropológica del Derecho*. Fondo de Cultura Económica.
- Taruffo, M. (2018). *La motivación de la sentencia*. Editorial Marcial Pons.
- Teubner, G. (2006). The anonymous matrix: Human rights violations by “private” transnational actors. *Modern Law Review*, 69(3), 327–346.

Tribunal Europeo de Derechos Humanos. (2012). *Hirsi Jamaa y otros vs. Italia*. Estrasburgo: TEDH.

UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. París: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

Urueña, R. (2020). Inteligencia artificial, justicia y poder judicial: perspectivas desde América Latina. En R. Uprimny (Ed.), *Tecnología, justicia y derechos humanos* (pp. 35–59). Dejusticia.

Vargas, L. F. (2022). Debido proceso y algoritmos en la función judicial: Tensiones y límites en el contexto colombiano. *Revista Nueva Época de Derecho Constitucional*, 28(2), 56–83.

Zaffaroni, E. R. (2011). *Derecho penal: Parte general*. Buenos Aires: Ediar.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

## SENTENCIAS

Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura. (2024). Límites al uso de inteligencia artificial en la Rama Judicial.

Corte Constitucional de Colombia. (2003). Sentencia C-037/03.

Corte Constitucional de Colombia. (2020). Sentencia T-443/20.

Corte Constitucional de Colombia. (2022). Sentencia T-033/22.

Corte Constitucional de Colombia. (2024). Sentencia T-323 de 2024 (Debido proceso y uso de ChatGPT).

Corte Suprema de Justicia de Colombia, Sala Penal. (2021). Sentencia SP1445-2021.

Directrices internacionales (UNESCO y Comisión Europea). (2020-2023). Recomendaciones sobre el uso de IA en sistemas judiciales.

Documento CONPES 4144 de 2025. Estrategia nacional de inteligencia artificial.

Ley 1581 de 2012. Por la protección de datos personales (Habeas Data).

Proyecto de Ley 154 de 2024. Sobre inteligencia artificial en Colombia.