

Aspectos Éticos y Legales del uso de Inteligencia Artificial en las Providencias de los Jueces en
Colombia

María Paula Ardila Corzo

Trabajo de Grado para Optar el Título de Abogada

Director

Javier Alejandro Acevedo Guerrero

Magíster en Derecho

Universidad Industrial de Santander

Facultad de Ciencias Humanas

Escuela de Derecho y Ciencia Política

Derecho

Bucaramanga

2026

Dedicatoria

A la lucha humana, doliente e incansable por la búsqueda de verdad, dignidad y justicia. La de todos los colombianos y colombianas a quienes el sistema judicial les ha fallado, porque el país no necesita más abogados de prestigio, necesita más seres humanos ejerciendo el Derecho desde la ética y la empatía, eligiendo siempre lo bueno y lo justo.

Agradecimientos

A Dios, por darme las capacidades para aprender un arte y vivir dignamente de él. A mi madre, por ser mi apoyo más grande e incondicional. A mi padre, por recordarme de lo que soy capaz. A mi hermana, por ser mi inspiración y ejemplo. A los maestros que conocí, quienes me dejaron las lecciones más valiosas de esta profesión y quienes la ejercen con vocación desde la empatía, la ética y el liderazgo. Y a los amigos que hice, con los que transcurrí todos estos años de aprendizaje y los que faltan por transcurrir.

Resumen

Título: Aspectos Éticos y Legales del Uso de Inteligencia Artificial en las Providencias de los Jueces en Colombia*

Autor: María Paula Ardila Corzo**

Palabras Clave: Inteligencia artificial, Ética, Providencias judiciales, Jueces, Administración de justicia, Responsabilidad.

Descripción: Los sistemas judiciales de todo el mundo han atravesado muchas transformaciones a lo largo y ancho de la historia. Teniendo resultados positivos y respondiendo a necesidades urgentes, pero también con efectos adversos. Esta vez el cambio traspasa la academia y la ficción. Esta vez, la evolución tecnológica no solo amenaza con aumentar el desempleo, violar la privacidad, eliminar la interacción directa humana o crear nuevas formas de delincuencia. El último descubrimiento tecnológico, la inteligencia artificial, particularmente la IA generativa, promete ser remedio universal para las deficiencias estructurales del sistema judicial colombiano, pero también amenaza con usurpar la vocación del Juez natural y mudar a sistemas de justicia deshumanizada. En este contexto, la Ética y el Derecho son la garantía a los valores constitucionales sobre los cuales se erige el sistema judicial colombiano, permitiendo y limitando el uso de herramientas de IA generativa en las providencias de los Jueces y Juezas de este país.

* Trabajo de Grado

** Facultad de Ciencias Humanas. Escuela de Derecho y Ciencia Política. Director: Javier Alejandro Acevedo Guerrero. Magíster en Derecho

Abstract

Title: Ethical and Legal Aspects of the Use of Generative Artificial Intelligence in Court Orders in Colombia *

Author: María Paula Ardila Corzo**

Key Words: Artificial Intelligence, Ethics, Judicial Decisions, Judges, Administration of justice, Responsibility.

Description: Judicial systems around the world have undergone many transformations throughout history. Having positive results and responding to urgent needs, but also with adverse effects. This time the change transcends academia and fiction. This time, technological evolution not only threatens to increase unemployment, violate privacy, eliminate direct human interaction, or create new forms of crime. The latest technological discovery, artificial intelligence, particularly generative AI, promises to be a universal remedy for the structural deficiencies of the Colombian judicial system. but also threatens to usurp the vocation of the natural Judge and shift to dehumanized justice systems. In this context, ethics and law are the guarantee of constitutional values on which the Colombian judicial system is built, allowing and limiting the use of tools of generative AI in the orders of the judges of this country.

* Degree Work

** Facultad de Ciencias Humanas. Escuela de Derecho y Ciencia Política. Supervisor: Javier Alejandro Acevedo Guerrero. Master of Law.

Tabla de Contenido

	Pág.
Glosario.....	11
Abreviaturas.....	18
Introducción	19
1. Objetivos.....	23
1.1. Objetivo General.....	23
1.2. Objetivos Específicos.....	23
2. Historia y Marco Conceptual de una Ciencia Emergente.....	25
2.1. ¿Cómo Pasó de la Teoría a la Práctica la Inteligencia Artificial?.....	29
2.2. Transición a una Inteligencia Artificial Madura	31
2.3. ¿Cómo Surgieron los Agentes de IA Generativa como ChatGPT?.....	34
2.4. ¿Cómo Funciona un Programa con Inteligencia Artificial?.....	41
3. Desafíos, Riesgos y Otras Variables del uso Inmersivo de la Inteligencia Artificial	46
3.1. Discusiones Éticas y Desafíos a Nivel Global.....	52
3.2. Regulación Ética Internacional de la Inteligencia Artificial	68
4. Perspectivas Actuales de la Comunidad Judicial	72
4.1. Primera Fase: Recolección y Familiarización de los Datos Cualitativos	74

4.2. Segunda fase: Análisis, Codificación de Datos, Búsqueda y Definición de Temas	75
4.3. Tercera Fase: Análisis de Resultados y Descripción de Hallazgos	78
5. Análisis Jurisprudencial de la Sentencia T-323 de 2024 de la Corte Constitucional de Colombia	86
6.1. El Problema Jurídico del Uso de IA Generativa para las Decisiones de los Jueces en Colombia	89
6.2. Ratio Decidendi de la Sentencia T-323 de 2024	93
6.3. Decisión de la Corte Constitucional en la Sentencia T-323 de 2024	99
7. Propuestas prácticas para el uso de Herramientas de Inteligencia Artificial Generativa en las Providencias de los Jueces en Colombia y el Ejercicio del Derecho en general ..	102
8.1. ¿Cómo utilizar Ética y Responsablemente Herramientas de IA Generativa?.	104
8.2. Necesidad e Idoneidad de la Herramienta	105
8.3. Incidencia en el Convencimiento y Decisión del Juez.....	106
8.4. Autoformación	107
8.5. Protección de Datos Personales y Responsabilidad de la Información	108
8.6. Verificación Permanente de la Información e Identificación de Sesgos y Alucinaciones Algorítmicas	108
8.7. El Deber de Información (Principio de Transparencia)	109
8.8. Promoción de una Cultura Ética y de Buenas Prácticas en IA	110

9. Conclusiones.....	112
Referencias.....	117

Lista de Figuras

	Pág.
Figura 1. Esquema de la inteligencia artificial y el aprendizaje automático y profundo.....	36
Figura 2. Esquema del funcionamiento de un perceptrón simple o básico.....	42
Figura 3. Diagrama de una RNA multicapa o profunda.	43
Figura 4. Nube de palabras de los resultados de las entrevistas realizadas	75

Lista de Apéndices

Apéndice A. Tabla de los Principales Avances Teóricos en Inteligencia Artificial, por Diferentes Autores, Situados Entre el Siglo XIX y XX.....	123
Apéndice B. Principios y Lineamientos Referentes para la Regulación Ética de la Inteligencia Artificial a nivel Internacional y de Colombia.....	129
Apéndice C. Mapa Temático sobre el Uso de Herramientas de Inteligencia Artificial Generativa en la Administración de Justicia en Colombia	134

Glosario

Administración de justicia: Función pública del Estado colombiano, encomendada a la rama judicial, para resolver los conflictos jurídicos de la ciudadanía con decisiones justas y buenas, de acuerdo con la ley y la moral social imperante.

Agente de IA/GenAI: Sistema inteligente y ultra avanzado impulsado con IA generativa, capaz de realizar múltiples tareas autónomamente y crear contenido original, utilizando otras ramas y subramas de la IA para su funcionamiento integral.

Ajuste fino/*Fine-tuning*: Técnica utilizada para especializar un modelo de IA, que ya ha sido entrenado previamente, es decir, reentrenarlo con la información de un área del conocimiento específica para lograr un rendimiento superior y avanzado del mismo.

Algoritmo: Conjunto de operaciones preestablecidas y secuenciales que son ejecutadas por una red neuronal artificial, o la estructura funcional del sistema de IA, para resolver una incógnita o problema. El algoritmo se encarga de entrenar las neuronas artificiales para que ejecuten la tarea solicitada.

Alucinaciones: Información o contenido falso o inexistente que la herramienta de inteligencia artificial generativa crea para dar respuesta al usuario.

Aplicación/*App*: Programa de software diseñado para dispositivos digitales como smartphones, computadores, tabletas, entre otros.

Aprendizaje automático/*Machine learning*: Rama de la inteligencia artificial enfocada en desarrollar algoritmos que simulan el aprendizaje humano, es decir, un proceso que permita a las neuronas artificiales aprender autónomamente con base a los datos suministrados para su

entrenamiento, o de los que reciben en la interacción con el usuario, pero sin depender de una programación humana constante y permanente.

Aprendizaje profundo/Deep learning: Rama del aprendizaje automático que se enfoca en estructuras complejas y masivas de neuronas artificiales, agrupadas en capas dentro de una red, para desarrollar algoritmos complejos que resuelvan problemas de gran dificultad.

Bases de datos: Conjunto de datos de diferentes cantidades, tamaños y temas.

Beneficiario de la justicia: Persona natural o jurídica que hacen uso del sistema judicial colombiano para resolver sus conflictos jurídicos, o prevenirlos, quien en últimas obtiene los beneficios y utilidades de aquel.

Chatbot: Programa o software autómatas y conversacional impulsado con IA, que se comunica con los usuarios por texto o por voz, generando preguntas y respuestas previamente establecidas y limitadas por su creador, o replicando conversaciones originales utilizando IA generativa.

Comunidad judicial: Conjunto de personas, en su mayoría profesionales del Derecho, que se dedican a las labores propias de la administración de justicia.

Contenido: Información presentada en cualquier formato (texto, imagen, video, documento, código software, etc.) creada con una herramienta de IA generativa.

Corpus: Conjunto de datos que se utilizan para entrenar y refinar a un algoritmo en Procesamiento del Lenguaje Natural (NLP por sus siglas en Inglés), representarlos con la misma calidad y características, y generar la respuesta solicitada por el usuario.

Decisión judicial: Elección que adopta el Juez, de acuerdo con la ley vigente, la jurisprudencia y su criterio, como resultado de un proceso judicial, con el fin de resolver el

conflicto jurídico que le ha sido encomendado, y la cual se consigna en un documento jurídico válido como una sentencia o auto, denominado providencia.

Ecosistema judicial: Entorno en el que se llevan a cabo las labores propias del ejercicio del Derecho y la administración de justicia.

Entrada/input: Datos o información inicial suministrada al producto de IA para ser procesada por éste.

Funcionario/servidor judicial: Hace referencia a las personas que han sido nombrados en cargos públicos de la Rama Judicial colombiana. Aquí se incluyen todos los cargos, desde el de mayor jerarquía hasta el de menor jerarquía.

Funcionario de apoyo: Todas aquellas personas que trabajan en la Rama Judicial colombiana que no tienen cargos de mayor jerarquía como jueces y magistrados.

Hardware: Son los elementos físicos que componen un sistema computacional y permiten su funcionamiento, por ejemplo, la CPU de un ordenador, la memoria RAM, el monitor o pantalla, entre otros. Actualmente existen *hardwares* virtuales también.

Humanoide: Ente, robot o máquina que presenta características humanas físicas y/o cognitivas.

IA generativa o GenAI: Subrama del *deep learning* que genera contenido nuevo y original en diferentes formatos (texto, voz, imagen, código software, documentos) de acuerdo al *prompt* o *input* recibido.

Informática: Ciencia que estudia métodos para el tratamiento, procesamiento, almacenamiento, transmisión y representación de la información a través de dispositivos computables.

Inteligencia artificial (IA): Disciplina o ciencia basada en el estudio de sistemas que simulen funciones cognitivas superiores del cerebro humano y capaces de realizar las actividades propias de este.

Inter comunis: Alocución latina utilizada en Derecho para referirse a los efectos jurídicos generales o globales que tiene una decisión judicial, es decir, que no solo afecta a las partes sino a terceros no intervinientes del proceso judicial y a la sociedad en general.

Inter partes: Alocución latina utilizada en Derecho para referirse a que los efectos jurídicos de una decisión judicial impactan solo a las partes de un proceso judicial, es decir, las personas directamente afectadas o involucradas en el conflicto jurídico que suscitó la causa procesal.

Juez natural: Ser humano que resuelve conflictos jurídicos dentro del sistema judicial colombiano, y quien, con sus decisiones, materializa la justicia.

Litigio: Conflicto entre dos o más personas que se lleva al ámbito judicial para que sea resuelto por un juez. Asimismo, este término es utilizado, en este trabajo, también para referirse al campo de trabajo del Derecho que permite a los abogados ejercer su profesión en el área jurídico procesal como apoderado de parte.

Lógica: Ciencia que estudia métodos de razonamiento a través de reglas y símbolos matemáticos que deciden sobre la veracidad de un argumento.

Macrodatos: Conjunto grande de datos.

Máquinas inteligentes: Sistemas, programas, aplicaciones, herramientas, robots o cualquier otro dispositivo tecnológico avanzado que simula tener funciones cognitivas humanas, impulsado por algoritmos de inteligencia artificial.

Modelo LLM (*Large Language Model*): Un modelo de lenguaje de gran tamaño se refiere a sistemas avanzados que se enmarcan dentro del *deep learning*, y que realizan tareas propias del

procesamiento del lenguaje como reconocerlo y reproducirlo, lo cual permite a un producto de IA interactuar con el usuario.

Neurona artificial: También llamada nodo, es la unidad básica de funcionamiento del *machine* y el *deep learning*. Se agrupan en capas para conformar una red de neuronas artificiales, en donde procesan la información de acuerdo con las instrucciones dadas por el algoritmo.

Pesos sinápticos: Son las conexiones entre neurona y neurona (o nodos) que permiten la sinapsis, es decir, la transmisión de señales de activación, o inactivación, entre ellas. A través de los pesos sinápticos se transmite la información aprendida durante la fase de entrenamiento de la red neuronal y así se van ajustando automáticamente de manera que se logre el resultado deseado.

Entrenamiento de IA: Es la fase previa de desarrollo de un producto de IA, el cual se realiza utilizando grandes bases de datos, de manera que este aprenda a ejecutar las funciones para las cuales fue creado.

Procesamiento del Lenguaje Natural (NLP): Rama del *deep learning* que utiliza otras disciplinas como la lingüística computacional, por ejemplo, para analizar la comunicación humana a través del lenguaje natural, o lenguaje no técnico, sus estructuras y definiciones, y así procesarlo y generarlo en formato de texto, voz u otro tipo, permitiendo la interacción de un producto de IA con los humanos.

Productos de IA: Todas aquellas herramientas, máquinas, sistemas, dispositivos y demás que son creadas e impulsadas por inteligencia artificial. Es un concepto amplio y global para referirse en general a este tipo de creaciones.

Prompt: Petición detallada e instructiva que se realiza a una herramienta de inteligencia artificial generativa para que genere el contenido deseado.

Providencia judicial: Concepto formal y general para referirse al documento que contiene la voluntad del juez, es decir, las decisiones o resoluciones que adopta en el contexto de un proceso judicial. Las providencias se dividen en autos y sentencias.

Red neuronal artificial (RNA): Estructura de n cantidad de capas conformadas por n cantidad de neuronas artificiales.

Retropropagación/*Backpropagation*: Es un tipo de algoritmo de *machine learning* que funciona “hacia atrás”, es decir, de acuerdo al dato de salida, calcula el ajuste o gradiente que debe realizar en los pesos sinápticos para reducir la imprecisión o error que tuvo dicho dato de salida versus el dato de entrada. En otras palabras, medir que tan acertado fue el resultado generado respecto a la tarea solicitada en la entrada, para así ajustar los pesos sinápticos de las redes de neuronas artificiales.

Robot: Máquina autómatas, física o virtual, que es programada para llevar a cabo diversas tareas, generalmente, mecánicas.

Robótica: Disciplina que utiliza a otras como, por ejemplo, la mecánica, electrónica, informática e inteligencia artificial para el diseño, creación y operación de máquinas autónomas.

Salida/*output*: Datos o información creada por el producto de IA como resultado del *prompt* o *input*.

Sesgos algorítmicos: Prejuicios, preferencias o tendencias hacia algo, o alguien, que se percibe en la respuesta dada por un agente de IA generativa, al *prompt* introducido por el usuario, y se considera discriminatoria, debido a que los datos con los que fue entrenado el algoritmo del agente no son neutrales o imparciales, sino que muestran la realidad de manera desproporcionada e injusta hacia cierto grupo poblacional o hacia algo.

Software: Es el componente intangible de un ordenador o dispositivo electrónico que le indica al *hardware*, a través de instrucciones, qué tareas debe ejecutar y cómo hacerlo.

Startup: Empresa nueva y pequeña, enfocada en innovación tecnológica, cuyo modelo de negocio busca su crecimiento exponencial y escalable a otro nicho del mercado.

Usuario judicial: Persona que hace parte de la comunidad judicial y que utiliza herramientas de inteligencia artificial para las labores propias de la administración de justicia.

Usuario: Persona que utiliza herramientas de inteligencia artificial para sus objetivos propios.

Abreviaturas

AA: Aprendizaje automático.

AP: Aprendizaje profundo.

App/Apps: Aplicación/aplicaciones.

DDHH: Derechos humanos.

GenAI: Inteligencia artificial generativa.

GPT: *Generative Pre-trained Transformer* - Transformador Generativo Preentrenado.

IA: Inteligencia artificial.

LLM: *Large Language Model*, en español, Grandes Modelos de Lenguaje.

NLP: *Natural Language Processing* - Procesamiento del Lenguaje Natural.

NNA: Niños, niñas y adolescentes.

RNA: Red neuronal artificial o red de neuronas artificiales.

Introducción

En la última década, la inteligencia artificial (IA) ha ido conquistando territorios fuera de la academia y de los departamentos de investigación tecnológica de los Estados. Aunque en Colombia todavía no estamos tan acostumbrados a convivir en un entorno completamente tecnológico o virtual, sí nos estamos acercando hacia ese destino, pues la IA está ocupando espacios importantes en muchos ámbitos y contextos, como, por ejemplo, el Derecho.

Pero ¿Existe un punto de encuentro entre una ciencia “pura” como la inteligencia artificial y una ciencia “blanda” como el Derecho? Sí, aunque tradicionalmente han sido consideradas polos opuestos, en realidad mantienen una relación colaborativa en pro de la evolución de nuestra especie.

Aunque ya no existan diferencias irreconciliables entre una y otra, atender cuestiones sociales como las que atiende el Derecho (impartir o administrar justicia) está lejos de ser fácil para entes no humanos. Prueba de ello es la tendencia que muestran los creadores de agentes con IA generativa a simular las funciones cognitivas con las que las personas resolvemos dilemas jurídicos, sociales, económicos o morales. Y aunque sean neuronas las que impulsan este tipo de herramientas, sigue siendo insuficiente para clonar a la perfección nuestro cerebro.

No obstante, lo que actualmente conocemos como IA, se acerca bastante a la idea de inteligencia humana, logrando ser usada en los despachos judiciales y firmas de abogados para llevar a cabo tareas propias de su profesión, con la promesa de ahorrar tiempo o esfuerzos y poniendo sobre la mesa interrogantes cómo: ¿Qué significa ser juez? ¿Es legal y éticamente posible que un robot o algoritmo sea juez? ¿Qué beneficios traería ser juzgados por un sistema más eficiente pero no humano? ¿Es correcto sacrificar la condición humana del juez, como exigencia

constitucional del sistema judicial y el Estado, para mejorar las estadísticas de impunidad y descongestión judicial?

Es por aquello que en esta investigación se ha propuesto responder a esos interrogantes a través del estudio y análisis, desde la ética y las normas procesales en Colombia, del uso de la IA en los despachos judiciales y, específicamente, por los jueces en sus providencias.

Para cumplir esos objetivos, la investigación se dividirá en cinco capítulos o fases de investigación que a continuación se exponen:

- I. Historia y conceptualización básica de la inteligencia artificial como ciencia moderna.
- II. Identificación y análisis de los desafíos, riesgos y demás variables relacionadas al uso de la inteligencia artificial, desde una postura ética.
- III. Indagación de la perspectiva actual de la comunidad judicial sobre el uso de la inteligencia artificial en la administración de justicia, a través de entrevistas realizadas a magistrados, jueces, funcionarios judiciales, abogados litigantes y docentes de la ciudad de Bucaramanga, Santander.
- IV. Análisis jurisprudencial de la sentencia hito sobre inteligencia artificial en la Rama Judicial colombiana: Sentencia T-323 de 2024 de la Corte Constitucional de Colombia.
- V. Hallazgos y propuestas prácticas como guía para el uso ético y responsable de herramientas de inteligencia artificial generativa en las providencias de los jueces en Colombia, y en general para el ejercicio del Derecho.

Por otro lado, y en virtud del principio de transparencia para el uso ético y responsable de la IA, se informa que en este trabajo investigativo se hará uso de herramientas de IA generativa, particularmente, Gemini de Google, y el software de análisis cualitativo de datos Atlas.ti, integrado por el modelo de IA generativa GPT, de OpenAI, para los siguientes fines:

- Consultas y búsqueda de fuentes académicas confiables.
- Citar y referenciar correctamente las fuentes consultadas durante toda la investigación, así como para adecuar el estilo del documento a las normas APA, en su séptima edición.
- Organización de ideas y subtemas de la investigación para asegurar la claridad, coherencia y estructura adecuada de los argumentos creados y desarrollados por la autora en esta investigación.
- Adquirir el conocimiento empírico del funcionamiento de un agente de IA generativa, logrando la comprensión y el dominio de éste, con el fin de forjar el criterio personal de la autora, lo cual es indispensable para realizar esta investigación.
- Creación de diagramas que explican el funcionamiento interno y lógico de una red neuronal artificial, como unidad básica de la IA generativa (primer capítulo).
- Apoyo en el procesamiento y codificación de los datos cualitativos obtenidos de las entrevistas realizadas (tercer capítulo).
- Apoyo en la elaboración gráfica de los esquemas que condensan los datos codificados, como resultado de las entrevistas, para facilitar el análisis de estos y describir los hallazgos de dicho proceso, por parte de la autora.

Finalmente, se expondrán las conclusiones y reflexiones finales de la investigación, destacando el cumplimiento del objetivo general planteado, así como de los objetivos específicos.

Nutrir el debate ético de la IA en la administración de justicia, es el móvil de esta investigación, lo que se espera, contribuya a la creación de las normas que regularán la materia en Colombia, buscando proteger los principios y garantías constitucionales de la figura del juez colombiano y, en consecuencia, los derechos de los ciudadanos, permitiendo la transformación

digital y el provecho de las bondades de este tipo de tecnologías sin poner en riesgo la seguridad jurídica y la garantía de justicia, dignidad y equidad.

1. Objetivos

1.1. Objetivo General

Estudiar los alcances éticos y legales del uso de herramientas de inteligencia artificial en las providencias de los jueces en Colombia, en cumplimiento de la función pública del Estado: administrar justicia.

1.2. Objetivos Específicos

- Estudiar los aspectos históricos y conceptuales de la inteligencia artificial (IA), destacando los aportes más relevantes para la disciplina.
- Reflexionar y discutir sobre el alcance ético de la inteligencia artificial y los riesgos asociados a su uso en diferentes ámbitos.
- Conocer las recomendaciones y directrices éticas internacionales y nacionales para el uso responsable y seguro de herramientas de inteligencia artificial.
- Indagar si actualmente los jueces del distrito judicial de Bucaramanga (Santander) utilizan herramientas de inteligencia artificial en la creación de sus providencias, a través de entrevistas realizadas a magistrados, jueces, funcionarios judiciales, abogados litigantes y docentes.
- Realizar el análisis jurisprudencial de la sentencia T-323 de 2024 de la Corte Constitucional de Colombia y conocer la regulación ética y jurídicamente vinculante para la Rama Judicial colombiana.
- Proponer una guía sencilla de acciones concretas que permitan orientar a los jueces y demás actores judiciales del contexto colombiano, en el uso ético, seguro y responsable de

herramientas de inteligencia artificial para el ejercicio de sus labores de justicia y en cumplimiento de la regulación ética actual sobre la IA y las nuevas tecnologías en general.

2. Historia y Marco Conceptual de una Ciencia Emergente

La capacidad inventiva del Homo Sapiens, es el gran artífice de la evolución de su especie, en todas sus dimensiones: social, biológica, psicológica, espiritual, ética o moral y cognitiva.

La historia está marcada por, más que nombres, grandes mentes que combinadas con ideas muy ambiciosas y una época de guerras y cambios, produjeron las mayores creaciones nunca pensadas, cuyo aprovechamiento nos ha traído hasta este momento histórico que Gamal Serhan Jaldin¹, y otros expertos, consideran, es la quinta revolución industrial del mundo.

Tendría espacio la discusión sobre si la inteligencia artificial (IA) es verdaderamente la quinta revolución industrial o si la tecnología ya pasó por ahí hace mucho. Pero en lo que no tiene lugar debatir, es que la IA no se puede ver de la misma manera que a los demás avances tecnológicos que han surgido, ni tampoco puede equipararse a la revolución industrial situada entre 1760-1840, pues con la IA el cambio no ha sido únicamente económico, laboral o político, ha sido antropológico e incluso para algunos, apocalíptico.

Estudiar la historia de la IA, los aportes más relevantes que marcaron su evolución como disciplina y conocer los conceptos propios de ella, es el objetivo planteado para este capítulo, el cual posibilitará la comprensión básica e histórica de la IA y dará paso a los posteriores análisis éticos y legales sobre su uso en el ámbito judicial.

Dirigirnos al origen de la IA, entendiéndose como una tecnología (del griego *téchnē*), una técnica u oficio y, *logos*, un tratado, ciencia o conocimiento profundo² (Real Academia Española, s.f., definición 1º), se torna exagerado e innecesario pues tendríamos que remontarnos hasta el

¹ Ingeniero civil boliviano y reconocido empresario experto en transformación digital e innovación.

² Definición etimológica de la palabra “tecnología” que proviene de la lengua griega.

germen de la lógica y de las matemáticas. Por ende, para generar un entendimiento amplio del porqué, el cómo y el cuándo del origen de la IA, para comprenderla histórica y conceptualmente, se empezará por estudiar sus comienzos oficiales como disciplina autónoma, brindando un contexto sobre su presencia en el mundo actual, y cómo es que logró interactuar con funciones tradicional y milenariamente reservadas para la inteligencia humana.

Aclarado lo anterior, es ampliamente conocido que el Dartmouth Summer Research Project On Artificial Intelligence, o simplemente la Conferencia de Darmouth, que tuvo lugar en Hanover, New Hampshire, Estados Unidos), en el año 1956, es considerado, oficialmente, el evento en donde nació la inteligencia artificial como disciplina autónoma, aunque no del todo independiente, pues la IA es, en esencia, resultado del aprovechamiento de ciencias que la precedieron, como la informática, la computación y la robótica. Sin los avances y fundamentos teóricos de estas disciplinas la IA sencillamente no existiría.

A su vez, disciplinas como la matemática, la estadística, la biología y la neurociencia, son las ciencias madre, o básicas, que hicieron posible el nacimiento de las disciplinas modernas, que marcaron el comienzo de la era digital.

A principios del siglo XX, en su forma primigenia, la IA se imaginaba como una máquina automática que pudiera ejecutar diferentes tareas programables a través de un medio físico y un espacio virtual, software y hardware, para su efectivo uso (Nuño, 2019), aunque desde los siglos XVII y XVIII ya se había empezado a hablar de la automatización y programación matemática, con las creaciones de máquinas programadas para tareas mecánicas y repetitivas, fue para la Segunda Guerra Mundial cuando surgió uno de los primeros ordenadores de la historia: La máquina de Turing.

Ciertamente, su creador, el matemático británico Alan Turing (1912-1954)³, formuló la teoría más elemental de la ciencia computacional a través de su máquina, la cual lleva su nombre, y la que podía resolver un problema sencillo con una serie limitada de instrucciones, previamente dadas, utilizando 1s y 0s⁴, con una cabeza lectora que procesa los datos de entrada y genera una respuesta de salida (Nilsson, 2010), lo que hoy puede ser considerado un algoritmo. Con esta máquina que despejaba ecuaciones o realizaba cálculos matemáticos, Turing descubrió que, aunque existían problemas o proposiciones matemáticas verdaderas, no podían ser computables o resueltas por un ordenador de la época (Kaku, 2008), pero que la ciencia lógica tenía la potencialidad de ser utilizada más allá de ecuaciones y teoremas, así como lo visualizaron precursores como Ada Lovelace (1815-1852)⁵, permitiendo el desarrollo exponencial de las ciencias “modernas” como la computacional e informática, en los años que siguieron.

Recapitulando, la base de la inteligencia artificial fue la lógica matemática, en los inicios de aquella, cuando ni siquiera había adoptado ese nombre y cuando, antes de la creación de la máquina de Turing, hubo todo un desarrollo teórico de lógicos matemáticos importantes que permitieron la creación de las primeras máquinas “inteligentes”, entre mediados del siglo XIX hasta mediados del siglo XX.

³ Alan Mathison Turing (1912-1954) fue un matemático y lógico británico que sentó las bases de la computación e informática. Es mundialmente conocido porque ayudó al ejército británico a descifrar los mensajes en clave que se enviaban los alemanes entre sí - el código *Enigma* - durante la Segunda Guerra Mundial, y por la creación de un test que permite definir si una máquina es inteligente o no, dependiendo de si, a través de sus respuestas, se logra confundir con la conducta de un ser humano.

⁴ El sistema binario es uno de los más utilizados en la ciencia computacional. Se le da ese nombre porque utiliza el uno y el cero para representar opuestos, por ejemplo, verdadero y falso. Este sistema es la base de toda la lógica computacional moderna y digital.

⁵ Augusta Ada King, condesa de Lovelace, conocida como Ada Lovelace (1815-1852), fue una matemática y escritora británica, conocida principalmente por su trabajo en la máquina analítica de Charles Babbage. Se le considera la primera programadora de computadoras debido a sus notas sobre la máquina analítica, donde describió un algoritmo para calcular los números de Bernoulli, considerado, a su vez, el primer algoritmo destinado a ser procesado por una máquina. También, fue la primera persona en considerar la capacidad multidisciplinar de las matemáticas y que no sólo podía aplicarse a cálculos numéricos.

En el apéndice A de este documento, se elaboró una tabla con los principales aportes a la nascente disciplina, sus autores y obras correspondientes en las que plasmaron el fruto de su trabajo científico. Entre aquellos se destacan el primer ordenador digital de propósito general creado en 1943 por John Mauchly y John Eckert, con influencia de los trabajos realizados por científicos pioneros en informática como John Von Neumann, Alonzo Church y Alan Turing, entre otros.

Este ordenador era de gran volumen y peso. Requería de configuración manual para que ejecutara las operaciones lógicas en cada tarea. Por ello, años después en 1949, la dupla de investigadores creó la ENIAC, un computador e integrador electrónico y la EDVAC, un calculador automático, mejorando las deficiencias del primer modelo.

Posteriormente, los científicos Norbert Wiener, Warren McCulloch y Walter Pitts direccionaron sus investigaciones hacia la biología, es decir, buscaron inspiración en nuestra constitución anatómica y cerebral, indagando sobre cómo los seres vivos reciben, procesan y utilizan la información diariamente para determinar su actuar. El primero estudió lo que fueron las bases del *machine learning*, es decir, la retroalimentación que recibían los seres vivos al interactuar en su entorno para evitar cometer los mismos errores, y, los segundos, estudiaron la estructura básica del cerebro humano: las neuronas y su proceso sináptico. De ahí surgieron las redes de neuronas artificiales, tan importantes para los productos de IA que hoy utilizamos.

En 1956, el año del nacimiento de la IA como disciplina, uno de sus fundadores, Jhon McCarthy, creó el primer lenguaje de programación LISP, que sirvió de fundamento para los lenguajes modernos y que creció hasta llegar a ser una “familia” de lenguajes, actualmente muy utilizada.

Al año siguiente Frank Rosenblatt creó el perceptrón simple, un modelo de neurona biológica sencilla que constaba de varias entradas (para recibir datos), nodos o pesos que asignaban

valores o símbolos numéricos a cada entrada, se procesaba la información, se generaba un valor ponderado entre los que fueron recibidos y si éste cumplía con las reglas del modelo se activaba la función de resultado, pero si no, por sí misma modificaba los pesos hasta obtener el resultado deseado. Con ella se introdujo el *machine learning* como rama de la IA.

Esta creación es la más cercana a las neuronas artificiales de hoy en día, basada en el modelo de neurona artificial de McCulloch y Pitts.

Aunque con muchos desaciertos, y varios escépticos, la inteligencia artificial logró surgir como un área del conocimiento independiente de la robótica, la informática y la computación, aunque por la misma época, se destacó y abrió la puerta a un futuro inimaginable.

2.1. ¿Cómo Pasó de la Teoría a la Práctica la Inteligencia Artificial?

Para finales de los años 50's, como se reseñó en la tabla del apéndice A, fue muy popular la creación de juegos de estrategia, como el ajedrez, en los primeros ordenadores que ejecutaban algoritmos básicos y matemáticos, ya que era el escenario ideal para el desarrollo de esos primeros modelos considerados IA, precisamente porque en este tipo de actividades se necesita de un gran esfuerzo mental; se necesita razonamiento y de otros procesos cognitivos, a diferencia de otros juegos o en el deporte en donde generalmente las exigencias mayores son de tipo físico: fuerza, resistencia, agilidad, flexibilidad, etc.

Por tanto, el ajedrez resulta siendo el medio perfecto para poner a prueba el rendimiento del cerebro humano versus el de una máquina que lo simula.

En 1967, posterior a todos los avances en la teoría ajedrecística desde Turing, con la publicación de su ensayo sobre el ajedrez informático⁶, tiene lugar el primer enfrentamiento de ajedrez de computadora, entre el programa Velskiy-Kronrod⁷, desarrollado en el Instituto de Física Teórica y Experimental de Moscú (ITEP) y el programa Kotok-McCarthy⁸, desarrollado en el Instituto Tecnológico de Massachusetts (MIT) y ejecutado en una computadora IBM, resultando como vencedor el programa de Moscú.

No obstante, 30 años más tarde, la compañía de tecnología computacional IBM, lograría que su programa Deeper Blue (1997), le ganara al campeón mundial de ajedrez, el ruso Garry Kasparov.

Esto marcó el comienzo de una nueva era de la inteligencia artificial, con inviernos, pero con primaveras también. En esta era, las investigaciones y avances continuaron situados en la simulación de procesos biológicos humanos y la guía sustancial de la heurística⁹ (Abeliuk, Gutiérrez, 2021), un método que permite encontrar la solución a un problema complejo, “desmenuzando” a éste en problemas más simples y fáciles de responder, y un concepto desarrollado en la ciencia matemática de George Pólya (1887-1985), lógico húngaro, en 1965, con su obra *How to Solve it?*

Teniendo un método milenario bastante eficaz, la heurística, traducida a valores matemáticos para resolver casi cualquier problema y creando así algoritmos, éstos comenzaron a

⁶ *Computing Machinery and Intelligence (1950)*.

⁷ Creado por Georgi Adelson-Velskiy (1922-2014) y Aleksandr Kronrod (1921-1986), matemáticos rusos.

⁸ Creado por Alan Kotok (1941-2006) con ayuda de su profesor John McCarthy (1927-2011) para el MIT en computadoras IBM - International Business Machines Corporation.

⁹ La heurística es definida por la Real Academia Española (s.f., definición 2º) como la técnica de la indagación, del descubrimiento, y tiene múltiples aplicaciones en diferentes disciplinas.

mutar a su versión mejorada en función de qué tan útiles y eficaces eran, pues a tareas más complejas, algoritmos más avanzados.

En este contexto, surgieron los algoritmos genéticos; modelos de operaciones a través del cual se generan variaciones de un mismo método y se ponen a prueba hasta conseguir la variación más efectiva para el resultado esperado. Aunque esta categoría de algoritmo se consolidó con el trabajo realizado por John Holland en 1975, dentro de su obra *Adaptation in Natural and Artificial Systems*, desde la década de 1950 muchos investigadores habían desarrollado otros modelos de algoritmo inspirados en la capacidad predictiva del cerebro humano para estudiar su entorno, pero también en la evolución genética, es decir, la selección natural, desechando lo que no sirve y variando en función de lo que sí, con miras a la optimización de procesos para resolver problemas.

2.2. Transición a una Inteligencia Artificial Madura

Como lo reseñé anteriormente, la historia registra que la IA, como ciencia, tuvo “inviernos”, es decir, dos periodos importantes de “recesión”: El primer invierno fue de 1974 a 1980, y el segundo invierno de 1987 a 1993. Dos épocas que, aunque no significaron inactividad total en el campo de la inteligencia artificial, sí evidenciaron una crisis científica caracterizada por la reducción del interés investigativo e inversión económica en la ciencia, y la ausencia de un rumbo claro, ocasionado por la dificultad para alcanzar los objetivos que plantearon los pioneros de la disciplina, en la primera mitad del siglo XX (De Mántaras, 2020).

A pesar de ello, durante esos años hasta antes del comienzo de este siglo, hubieron avances importantes entre los que se destacan: el aprendizaje automático - *machine learning* -, el

Procesamiento del Lenguaje Natural (NLP), las redes Bayesianas¹⁰, los sistemas expertos diseñados por y para resolver cuestiones específicas de cierta área del conocimiento o tema, el redescubrimiento del algoritmo de retropropagación - *backpropagation* - y el aprendizaje profundo - *deep learning* -, aunque éste último sólo en la teoría.

A partir de aquí, y de manera simultánea, los robots y creaciones con IA dejaron de ser invenciones meramente con fines científicos o al servicio de los ejércitos y entidades gubernamentales, pues fueron surgiendo empresas de robótica e informática con fines lucrativos, pero también investigativos, que introdujeron sus creaciones al mercado general y las convirtieron en productos altamente comerciales. Con ello, visibilizaron la infinidad de aplicaciones que puede tener la IA hasta conseguir adaptarla a nuestras actividades rutinarias.

Entre esas primeras aplicaciones, aparte de los juegos de mesa, videojuegos, operaciones militares y ecuaciones matemáticas, se encuentran los *appliances*¹¹ para optimizar las tareas del hogar, tales como la aspiradora doméstica Roomba fabricada por iRobot y comercializada en 2002; los motores de búsqueda como Google y Yahoo que incorporaron algoritmos de IA para mejorar sus resultados; los chatbots, asistentes virtuales o inteligentes¹², que entre los más conocidos está Siri (2011) y Alexa de Amazon (2014), descendientes de ELIZA (1964); los traductores de textos, entre otros inventos. Con estos dos últimos el reconocimiento y procesamiento del lenguaje no

¹⁰ En el contexto de la IA, las redes bayesianas es un método de resolución de problemas el cual representa a estos en modelos gráficos probabilísticos para llegar a la respuesta. Este gráfico contiene cierta cantidad de vértices y líneas o aristas que los unen entre sí, las cuales muestran una secuencia, pero sin que sea cíclica. Los vértices son las variables y las líneas o aristas representa la relación de dependencia o independencia entre dichas variables. A través de la interpretación de este gráfico, de acuerdo a su estructura, pueden realizarse inferencias y predicciones.

¹¹ Electrodomésticos.

¹² Son productos similares, pero con funciones diferentes que muchas veces se confunden entre sí, pero que han surgido de manera independiente en la historia de la IA. Mientras que los chatbots sirven a su “dueño” (en el servicio al cliente de una marca, por ejemplo), los asistentes sirven al usuario (como Siri y Alexa). Y los agentes, parecidos a los que asisten, tienen funciones más avanzadas, como la generación de contenido, y son más independientes, aunque necesiten de instrucciones.

técnico, más conocido como procesamiento del lenguaje natural (NLP), se perfeccionó y abrió la puerta a otras funciones como el reconocimiento del habla, la producción oral, la identificación visual de objetos e imágenes y la generación de contenido original, que luego se incorporaron a los agentes de IA que actualmente utilizamos para múltiples tareas.

Por su parte, las empresas que ofrecían servicios de e-mail o correo electrónico, como Microsoft (Hotmail) y Google (Gmail), incorporaron algoritmos basados en inteligencia artificial para incluir funciones como, por ejemplo, filtración y bloqueo de correos no deseados. Asimismo, en la primera década de los 2000s, las plataformas digitales de comercio, a través del *machine learning*, comenzaron a entrenarse en sistemas de recomendación a gran escala, de acuerdo con el perfil del cliente y las compras que hacía - *data analysis* -, para ofrecer exactamente lo que el consumidor buscara. Posteriormente, otras plataformas de comercio electrónico y, por supuesto, las redes sociales, adoptaron esta misma función que permite identificar la conducta y comportamiento del usuario ofreciendo una experiencia personalizada e intuitiva, de ahí que, hoy en día es muy común que, si una persona busca el precio de computadores en línea, inmediatamente en otras aplicaciones y redes sociales le saldrán anuncios de ofertas de computadores.

Aunque estos avances suscitaron muy buenas críticas y comodidades para la población en general, originaron también muchos de los problemas sociales del mundo actual, problemas que se abordarán en el siguiente capítulo, desde una perspectiva ética.

Los vehículos autónomos (Autopilot de Tesla en 2015), smartphones (Nokia, iPhone, Sony, Blackberry, Samsung Galaxy) y los robots humanoides, no se quedaron atrás. La IA había mostrado su gran potencial y nadie quería dejarla fuera de sus productos. Aunque probablemente los primeros avances en integración de la IA a productos muy comercializados no fueron tan

visibles o identificados como victorias de la disciplina artificial para el público en general. Pero lo fue para la comunidad científica e impulsaron la transformación tecnológica de este siglo.

2.3. ¿Cómo Surgieron los Agentes de IA Generativa como ChatGPT?

Después de conocer el origen e historia de la IA, podemos comprender que un programa o agente que incorpore esta tecnología puede estar compuesto por reglas lógicas (era simbólica e inicios de la IA), algoritmos y fórmulas probabilísticas (era clásica con reglas y predicciones, a finales del siglo XX) y por redes neuronales artificiales o RNA (era moderna de la IA, siglo XXI).

Durante la más reciente era de la IA, gran parte de las investigaciones se han centrado en el campo del *deep learning*, pues se creía que las RNA multicapa podrían resolver tareas complejas sin que tuviesen que ser codificadas manualmente, lo cual efectivamente se logró.

Por ello, el *deep learning* o aprendizaje profundo, es uno de los descubrimientos más importantes de este siglo, para la disciplina. Según la multinacional de tecnología estadounidense IBM¹³ (s.f.), el *deep learning* es “un subconjunto del *machine learning* que utiliza redes neuronales multicapa, llamadas redes neuronales profundas, para simular el complejo poder de toma de decisiones del cerebro humano”.

Se considera que una RNA es profunda o multicapa, cuando tiene una capa de entrada, al menos una capa intermedia oculta y una capa de salida, es decir, tres capas o más en total, a diferencia del *machine learning* tradicional que sólo utiliza dos capas. En la actualidad los productos de IA que conocemos y utilizamos tienen miles de millones de capas, que a su vez tienen miles de millones de neuronas, pues la consecución del objetivo del producto depende de la

¹³ International Business Machines.

complejidad estructural de la red neuronal. A mayor complejidad, mayor cantidad de capas, mayor potencia y mejores resultados.

Sin embargo, antes de la llegada de los modelos GPT de OpenAI, una RNA con más de 10 capas podía hacer que se perdiera el control de la red y que no se pudieran ajustar los pesos sinápticos, generando el efecto adverso (Nuño, 2019) por ende, primero tuvo que resolverse esta dificultad para crear las redes masivas y ultra avanzadas de neuronas en la actualidad.

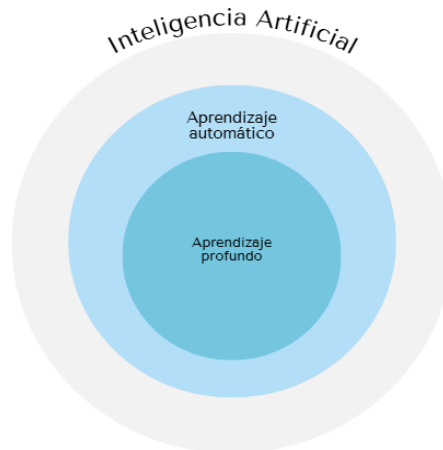
Los pesos sinápticos son esos nodos que permiten la conexión neurona a neurona (Nuño, 2019). Sabemos que las neuronas de nuestro cerebro se comunican entre ellas y lo hacen a través de la sinapsis, lo que no es un contacto directo, sino un área pequeña que contiene los neurotransmisores, quienes envían o bloquean las señales que transmiten las neuronas (López, 2021). En una RNA sucede lo mismo. Las neuronas artificiales están conectadas a través de nodos o pesos en los que sucede la sinapsis: bloquean las señales o las dejan pasar de neurona a neurona. Esos pesos se pueden ajustar para direccionar el resultado deseado, pero cuando la red es tan compleja, puede perderse el control de los datos y es más difícil ajustar los parámetros de salida, que era lo que sucedía antes.

Una vez superada esa dificultad, sólo a través del *deep learning* la IA podría utilizarse para lo que fue creada: Reconocer, analizar y producir tanto contenido original como el cerebro humano, pero con menos esfuerzo y en menos tiempo. Actualmente, los modelos de entrenamiento en *deep learning* son los que han dado mejores resultados.

Para tener clara la jerarquía del *machine learning* y el *deep learning* dentro de la IA, veamos el siguiente esquema básico:

Figura 1.

Esquema de la Inteligencia Artificial y el Aprendizaje Automático y Profundo



Nota. La figura 1 describe gráficamente la relación entre IA, aprendizaje automático (AA) y aprendizaje profundo (AP). El aprendizaje automático es una rama o subconjunto de la IA y dentro de esta rama se encuentra el aprendizaje profundo como modelo de entrenamiento de neuronas, es decir, algoritmos. Al igual que el AA no es la única rama de la IA, el AP no es el único modelo de entrenamiento que existe. Hay muchas clases que pueden estar dentro del AP como fuera de él, pero siguen haciendo parte del AA.

Algunos tipos del AA, menos conocidos que el AP, pero bastante utilizados, son el aprendizaje supervisado y no supervisado, el aprendizaje por refuerzo, el aprendizaje por transferencia o en línea, entre otros; y cada tipo se clasifica a su vez en algoritmos.

También hay diferentes tipos de “presentaciones” o productos con IA. Algunas veces llamamos inteligencia artificial exclusivamente a agentes como ChatGPT, Gemini, Copilot, Perplexity, etc., porque es lo que conoce como IA las personas “del común”, pero la realidad es que la IA viene en muchas formas y “sabores”, si se puede decir.

Las aplicaciones de juegos, las redes sociales, los servidores de correo, los buscadores en línea, las plataformas de comercio electrónico, el GPS que usan los vehículos, entre muchos más. La diferencia está en que estas otras presentaciones utilizan algunas funciones de subramas de inteligencia artificial como el procesamiento del lenguaje natural (NLP), dentro de los que, a su vez, están los grandes modelos de lenguaje (LLM)¹⁴; reconocimiento de voz, reconocimiento de imágenes, o *vision computer*, y razonamiento lógico. Y los agentes de IA como ChatGPT incluyen todas estas funciones en uno, además de una muy especial que sólo ellos tienen: la generación de contenido único y original (texto, documentos, imágenes, audios, videos y más). Esta es la razón por la que reciben el nombre de IA generativa o GenAI.

Aquí es importante saber la diferencia entre agentes, asistentes y chatbots, pues técnicamente no son lo mismo. Los asistentes virtuales son de tipo reactivo, es decir, hacen lo que se les pide y dependen de esas instrucciones para ejecutar cada acción, como por ejemplo Siri, Alexa, Google Assistant, Tay de Microsoft, etc. Los chatbots, de forma similar, responden preguntas según el objetivo para el que están hechos, que generalmente es comercial. Por eso en las páginas web de empresas se puede ver que tienen sus propios chatbots que responden las consultas de sus clientes a cualquier hora y en primera instancia, sirviendo a los intereses de éstas para que adquieran sus productos. Estos chatbots y asistentes tuvieron su auge entre 2010 y 2016, años en los que se dieron a conocer al público.

Por su parte, los agentes inteligentes son más autónomos e integran todas las funciones, incluso las de generación de contenido. No es que no requieran de instrucciones, es que con una sola instrucción puede comprender lo que quiere el usuario, analizarlo, buscar en sus bases de

¹⁴ Los modelos grandes de lenguaje (*Large Language Model - LLM*) son una categoría de modelos básicos entrenados en inmensas cantidades de datos, lo que los hace capaces de comprender y generar lenguaje natural y otros tipos de contenido para realizar una amplia gama de tareas. Definición de IBM (IBM, S.F.).

datos, tomar decisiones y generar una respuesta, a la vez que va aprendiendo de sus errores y adecuando sus respuestas al perfil del usuario.

El primero de estos asistentes fue GPT de OpenAI, una de las varias empresas de investigación en inteligencia artificial vigentes. Los GPT (*Generative Pre-trained Transform*) son toda una familia de modelos de lenguaje que utiliza el modelo *transformer*¹⁵, uno de los más avanzados hasta ahora.

Antes de los modelos de *transformers*, que están diseñados para casos de LLM y de PNL, se utilizaban Redes Neuronales Recurrentes (RNN) y Redes Neuronales Convolucionales (CNN): las primeras para comprender y aprender a hacer predicciones tomando datos secuenciales y las segundas para reconocer imágenes y otro tipo de datos, con resultados preestablecidos.

Si bien este tipo de arquitecturas son avanzadas, también son informáticamente ineficaces y tardan mucho en generar la salida deseada. Además, las RNN no logran procesar secuencias de datos con longitudes grandes y aun así mantener la coherencia de estas.

Esta limitación motivó investigaciones en estructuras neuronales más avanzadas, veloces y eficaces. Ashish Vaswani, un informático de nacionalidad india, y otros 7 científicos de Google¹⁶, describieron el modelo *transformer*, a través del artículo *Attention is all you need* publicado en 2017 en donde propusieron esta nueva arquitectura para memorias a largo plazo. Allí explican que este modelo utiliza RNNs y CNNs con un codificador-decodificador que funciona así: El codificador recibe los datos de entrada de manera secuencial, es decir, grandes cantidades de datos con un orden específico, y un vector condensa esta información, reteniendo la mayor cantidad

¹⁵ Es un tipo de red neuronal profunda que procesa la información en simultáneo, no secuencial como le son dadas las oraciones o datos de entrada, utilizando el mecanismo *self-attention* (autoatención) que le permite captar el contexto de las palabras, en oraciones largas, y evaluar la importancia de la secuencia en la que están escritas. Este tipo de estructura es la base de los modelos GPT y BERT.

¹⁶ Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser e Illia Polosukhin.

posible. Este resultado es el dato de entrada del decodificador, quien también puede tomar información externa de manera simultánea y así genera una salida.

Este modelo es más eficaz y requiere menos tiempo de entrenamiento, pues es un mecanismo de *self-attention*, de ahí el nombre del artículo, que estudia sus datos de entrada para poder captar su significado y contexto, y así representarlos coherentemente y generar la mejor entrada para el decodificador, quien a su vez generará la salida.

Aunque este modelo fue utilizado para el primer GPT, todavía hacían falta por afinar algunos detalles, antes de conocer el primer agente de OpenAI. Es por esto que, al siguiente año, en 2018, la compañía de IA publicó otro artículo, escrito por Alec Radford, Karthick Narasimhan, Tim Salimans e Ilya Sutskever, titulado *Improving Language Understanding by Generative Pre-Training*.

Con este avance se demostró que era posible pre entrenar un algoritmo con corpus¹⁷ masivos, miles de millones de textos sin que fuesen “etiquetados” o con indicaciones previas tediosamente determinadas, solo con pequeños ajustes, lo que llamaron *fine-tuning*¹⁸, y aun así ejecutar correctamente no solo unas sino varias tareas. Es decir, antes, si se quería traducir un texto, debía diseñarse una arquitectura especial para ello. O si por el contrario se quería un chatbot de preguntas y respuestas sobre cómo jugar ajedrez, se tenía que diseñar otra arquitectura, y así sucesivamente. Con el modelo de preentrenamiento generativo, podía transferirse una misma arquitectura de RNA para múltiples funciones y tareas. Así, se redujeron costos operativos y temporales, dando a luz a los GPT.

¹⁷ El corpus en inteligencia artificial son los datos que se utilizan para entrenar al algoritmo en procesamiento del lenguaje natural (NLP), representarlos con la misma calidad y características, y con base en ellos generar la respuesta.

¹⁸ Es una técnica utilizada para especializar un modelo de IA, que ya ha sido entrenado, con la información de un área del conocimiento en específico para lograr un rendimiento superior y avanzado para ella como, por ejemplo, la medicina.

Después del GPT-1 en 2018, en noviembre del año siguiente (2019) OpenAI lanzó GPT-2, entrenado esta vez con 1.500 millones de datos, en 2020 fueron por el tercero, GPT-3, el cual fue entrenado con 175.000 millones de datos (Kapuściński, 2025). De este modelo se crearon otras versiones de prueba, se entrenó a base de encuestas de los primeros usuarios, pero su uso era exclusivo para la investigación, no era un producto público y de libre uso. Esto cambió en noviembre de 2022, cuando OpenAI decidió crear la app de los GPT: ChatGPT. Era gratuito y más conversacional, utilizaba el modelo GPT-3.5 y RLHF - *Reinforcement Learning from Human Feedback*, en español aprendizaje por refuerzo a partir de retroalimentación humana. Es decir, presenta dos opciones de respuesta al usuario y éste elige cual le gusta más para aprender de su elección para próximas respuestas, una función que actualmente utiliza todavía. Sin embargo, para esa época aun no tenía las funciones de reconocimiento visual y auditivo.

Los GPT-4 y GPT-4 turbo (2023) incorporaron el reconocimiento visual y redujeron sus respuestas “inventadas” a la par de optimizar su rendimiento. Ya para mayo del 2024, el GPT-4o (*omni*) manejaba tanto texto, como imágenes y voz, aunque era una versión muy avanzada, no tanto como es el GPT-5, la última de todas.

Esta última versión definitivamente revolucionó la industria digital. GPT-5 comprende y utiliza el lenguaje de manera clara, sus respuestas son más coherentes y acordes a la realidad, puede “escuchar, ver y hablar”, admite documentos de múltiples formatos y su memoria es a largo plazo, permitiendo guardar todas las conversaciones que tengas con él. ¿Lo mejor? es gratis, no completamente, pero se pueden hacer infinidad de tareas con esta modalidad.

ChatGPT no es el único agente disponible en la web, incluso OpenAI tiene otros como Deep Research, Sora o DALL.E, aunque ahora hay tantos que es imposible conocerlos a todos. Los modelos GPT de OpenAI fueron los que impulsaron la transición tan drástica y visible a esta

última era, incluso algunas empresas tecnológicas como OpenAI comparten modelos entre sus agentes como Gemini y Notebook LM de Google, Claude de Anthropic, Copilot de Microsoft, Grok de xAI, n8n de n8n GmbH y muchos más.

La IA se ha inmiscuido en la cotidianidad del mundo y ahora se extiende a otras áreas del conocimiento como la pedagogía, la ingeniería y la física, la medicina y la biología, incluso en las ciencias sociales, y aunque con mayor resistencia, en el Derecho. Esto solo significa que las sociedades deben estar preparadas para adaptarse a los cambios que producen los avances tecnológicos, por ello, es imperativo conocer el funcionamiento en general de la IA, actualmente.

2.4. ¿Cómo Funciona un Programa con Inteligencia Artificial?

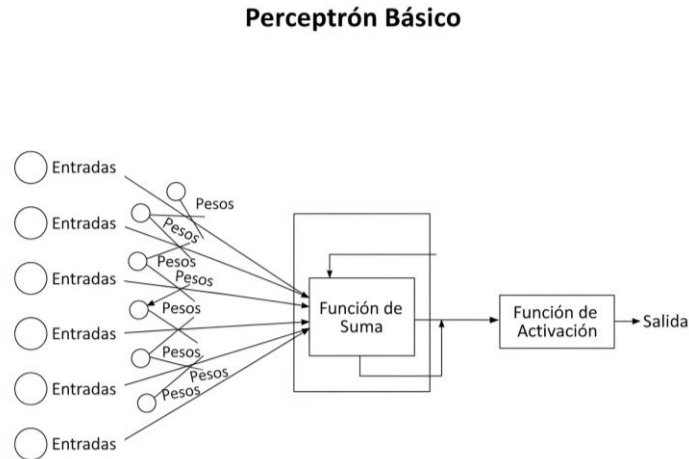
Conociendo el contexto histórico y conceptual de la IA, un esquema de su funcionamiento resulta idóneo e ilustrativo para una mayor comprensión del tema, aún mejor si es creado por ella misma.

Para la creación de este diagrama, utilicé el agente Gemini de Google y le redacté el siguiente *prompt*: Crea un diagrama básico que explique la función básica de un perceptrón, la primera red neuronal artificial creada. Haz el diseño como quieras, pero asegúrate de que el diagrama sea comprensible a primera vista. Si necesitas palabras o letras para explicar, hazlo. No olvides poner un título adecuado.

Y el resultado devuelto fue la siguiente imagen:

Figura 2.

Esquema del funcionamiento de un perceptrón simple o básico



Nota. Figura generada con el agente de IA Gemini de Google (2025).

Adicionalmente, le pedí al agente que explicara el diagrama, con el siguiente *prompt*:
Ahora explica en un párrafo breve cómo funciona el diagrama del perceptrón básico que generaste y explica cada una de las partes que lo conforman.

El párrafo que creó la herramienta en menos de 5 segundos como respuesta al *prompt* fue el siguiente:

El perceptrón básico funciona como un simple clasificador que toma múltiples entradas, las procesa y produce una única salida. Las entradas son los datos iniciales que se alimentan al modelo. Cada una de estas entradas se multiplica por un peso, que representa la importancia o influencia de esa entrada en el resultado final. La Función de suma o sumador, recolecta y suma todas las entradas ponderadas. Finalmente, el resultado de esa suma pasa a través de una función de activación, que decide si el perceptrón debe

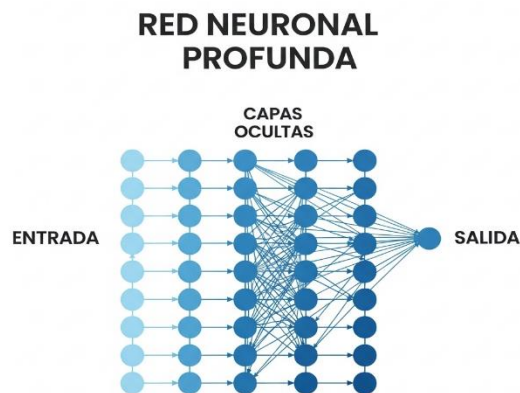
"activarse" (producir una salida de 1) o no (producir una salida de 0), basándose en si la suma supera un umbral predefinido, lo que determina la salida final. (Google LLC, 2025)

También le pedí a la IA, en el mismo chat y con el contexto/perfilación de la conversación previa, que creara un diagrama básico de una RNA profunda, la cual es la más utilizada actualmente, con este *prompt*: Ahora crea un diagrama básico que explique la función de una red neuronal artificial profunda, es decir, multicapa. Haz el diseño como quieras, pero asegúrate de que el diagrama sea comprensible a primera vista. Si necesitas palabras o letras para explicar, hazlo. No olvides poner un título adecuado.

Y la imagen resultante fue la siguiente:

Figura 3.

Diagrama de una RNA multicapa o profunda.



Nota. Figura generada con el agente de IA Gemini de Google (2025).

También le pedí que explicara el diagrama brevemente, con este *prompt*: Ahora explica en un párrafo breve cada parte del diagrama y cómo funciona.

Y el párrafo que creó como resultado fue este:

Una red neuronal artificial profunda, como la del diagrama, está compuesta por tres partes principales. La capa de entrada es donde se reciben los datos iniciales, como los píxeles de

una imagen o las palabras de un texto. A continuación, las capas ocultas procesan esta información de forma iterativa; cada neurona en una capa oculta recibe datos de todas las neuronas de la capa anterior, aplica un cálculo y pasa el resultado a la siguiente capa. Este proceso se repite en cada capa oculta, lo que permite a la red aprender a identificar características cada vez más complejas de los datos. Finalmente, la capa de salida produce el resultado final de la red, que puede ser la clasificación de un objeto en una imagen, una predicción numérica o la generación de texto. (Google LLC, 2025)

Aunque en la realidad sea mucho más complejo crear y programar un agente de IA generativa, o alguna otra función que incluya esta tecnología, los diagramas generados por Gemini, sirven para describir gráfica y de manera sencilla cómo funciona la IA en estos tiempos. No sólo las definiciones y los diagramas representan el trabajo histórico que se ha hecho dentro de la disciplina artificial, robótica e informática, también, el que sean generados por una IA, es prueba del éxito de ésta, razón por la cual se optó por generar los diagramas con apoyo del agente Gemini.

A modo de conclusión de este primer capítulo, durante el siglo XX y XXI la IA nació y creció exponencialmente, gracias a muchos académicos, científicos y genios que le apostaron a una visión futurista de nuestra especie. Ya entrados en las dos primeras décadas de este siglo, la IA dejó de ser oficialmente una idea que apenas lograba dibujarse a trazos y pasó a ser todo un cuadro de realidad en el que se enmarca la vida hoy del ser humano. La tecnología, y la IA, crearon un nuevo espacio temporal, al que mudó la civilización humana y que, por tanto, hoy día es eso que posibilita la vida en sociedad. Para nuestra especie, es como el motor para un vehículo: sin él existe, pero no funciona, no progresa y no cambia su estado de quietud.

Pero también gracias a otras ciencias como la matemática y la biología, conocemos la inteligencia artificial, la cual demuestra por qué el ser humano tiene el lugar más alto en la cadena alimenticia. La IA es pues, la explicación del éxito de nuestra especie, el mayor producto del intelecto humano, una fuente al parecer inagotable de información, y aún desconocida.

El mundo está hablando cada vez más de IA. Se dictan cursos, talleres e incluso programas de pregrado en IA. Se habla de *prompts*¹⁹ y guiones para conseguir los mejores resultados. También se discute sobre sus riesgos y la imposibilidad de prescindir de la intervención humana, afortunadamente.

El siguiente capítulo reflexiona acerca de estas cuestiones y condensa lo que hasta ahora se consideran los lineamientos o parámetros éticos vigentes de la IA. Imperativa es pues su lectura, para cualquiera que se sirva de las bondades de esta creación del intelecto humano que avanza a la velocidad de la luz y que es luz para el futuro de nuestra especie.

¹⁹ La traducción literal de *prompt* (término en inglés), no corresponde a su significado real en inteligencia artificial. Una definición genérica de *prompt* puede ser, un texto corto de entrada que redacta el usuario y lo introduce en un agente de IA para obtener un resultado.

3. Desafíos, Riesgos y Otras Variables del uso Inmersivo de la Inteligencia Artificial

La inteligencia artificial es hoy uno de los mayores descubrimientos tecnológicos registrados, en gran parte, porque promete solucionar muchas de las problemáticas sociales actuales y colaborar con el ser humano para la evolución de su especie y la optimización de valiosos recursos como el tiempo y el dinero, o la protección de valores intangibles como la salud, el conocimiento, la cultura, la historia, y por supuesto, los recursos naturales como el agua, el aire, la energía, la biodiversidad, etc.

No obstante, el uso inadecuado de la IA suscita riesgos que pueden derivar en consecuencias negativas y daños irreparables, socavando los beneficios que pueda dar. Asimismo, nuestra especie se ha caracterizado no solo por una innata inclinación hacia la innovación y evolución de esta, sino también por tendencias a alcanzar objetivos más ambiciosos y contraproducentes para el medio ambiente, por ejemplo, acelerando una inminente destrucción de su propia especie y el planeta.

La historia lo ha demostrado, la ciencia puede ser utilizada tanto para hacer el bien como para hacer el mal. Es una sempiterna dualidad que queda a disposición del aspecto volitivo del ser humano y cuando éste no se encuentra alineado con las reglas éticas ya establecidas y claramente definidas, representa un tremendo riesgo para la conservación del orden social y de nuestra especie.

Entonces, la ética de la inteligencia artificial permite su uso responsable y seguro para el provecho de la sociedad, siendo definida por el gigante tecnológico IBM (s.f.) como “un campo multidisciplinario que estudia cómo optimizar el impacto beneficioso de la inteligencia artificial (IA) mientras se reducen los riesgos y los resultados adversos”.

Ahora bien, los riesgos que se discuten actualmente desde la ética y las organizaciones internacionales no existieron siempre, fue hasta este siglo que comenzaron a materializarse y sus consecuencias negativas se visibilizaron aún más con el advenimiento de las neuronas artificiales, el *machine* y el *deep learning*. Con ello, en efecto, las preocupaciones éticas fueron acrecentándose.

Navegamos en un mar de posibilidades negativas derivadas de la incorporación de la inteligencia artificial a la vida ordinaria. No solo a través de agentes de IA generativa, también a través de otro tipo de máquinas que ocupan todos los ámbitos de la sociedad y la industria, lo cual inevitablemente pone en riesgo bienes jurídicos como la vida misma, la salud mental, la transmisión de la información, la cohesión social y las relaciones interpersonales, entre muchos otros aspectos, pues lo que sí nos ha demostrado la IA es que tiene la potencialidad tanto de mejorar la calidad de vida del ser humano como de replicar prejuicios, sesgos y profundizar las desigualdades ya existentes.

Quienes plantean debates éticos de la IA, no deben ser considerados antiprogresistas o enemigos de la disciplina, pues el objetivo no es evitar el avance de la transformación tecnológica, sino que ésta no pase de ser beneficio para la humanidad a ser la razón de su exterminio. Mientras un grupo de personas se encarga de sacarle el mayor provecho a la IA, otro grupo de personas debe encargarse de mitigar al máximo los efectos no deseados que pongan en peligro dicho provecho y trabajar conjuntamente por estos objetivos, aunque acaezca primero el avance que la regulación.

Situación distinta es la de algunos académicos influyentes que desde el siglo XX han sembrado ideas apocalípticas y predicciones catastróficas sobre las máquinas inteligentes y los humanos, ideas que han motivado una visión de miedo generalizado respecto de los robots; verlos

como amenaza para la humanidad y temer una rebelión de máquinas inteligentes, en donde los humanos, sus creadores, pasamos a ser dominados por ellos.

Sin embargo, este miedo sólo se ha materializado en las salas de cine y en la imaginación de escritores y directores de películas de ciencia ficción, pero, aun así, no pasa desapercibido y hoy día existen posturas que se resisten a la normalización de la inteligencia artificial.

Es por esto que, la ética se consolida como la respuesta a las preocupaciones filosóficas, morales y sociales que la IA suscita entre nosotros. Pero para que esta discusión dé buenos frutos, es necesario, como propone Virginia Dignum (2022) “asegurar una narrativa adecuada de la IA, una que desmitifique sus capacidades, minimice tanto la sobreventa como la subventa de soluciones impulsadas por IA, y que permita una participación amplia e inclusiva en la discusión sobre el papel de la IA en la sociedad”.

La ética de la IA es una discusión sustancialmente filosófica que nos ha hecho cuestionar qué significa ser persona y exponer con mayor ahínco las diferencias entre nuestra especie y el resto de los habitantes del mundo, en donde ya se incluyen los robots con IA.

Para regular su uso, es imperativo establecer qué lugar ocupan las máquinas inteligentes en nuestra sociedad. Si debiésemos dotarlas de características humanas o ¿es posible clasificarlas como objetos regulares? ¿deberían tener una clasificación propia y atributos jurídicos como jurisprudencialmente se ha hecho con los animales?

La tesis mayoritaria acerca de este punto inicial es que por ahora, el ser humano es irremplazable y superior a cualquier otro ente terrenal, por lo que se espera que jamás pueda equipararse a un robot inteligente, aunque las máquinas puedan ejecutar mejor y más rápido determinada tarea, el control y toma de decisión siempre debe estar en manos humanas. De lo

contrario, entre otras cosas, estaríamos frente a un grave dilema sobre quién debe responder por los errores, omisiones y las consecuencias negativas de la conducta del robot o la IA.

Y es que, en la actualidad, todas las discusiones éticas que suscita la participación de la IA en la sociedad convergen sobre una misma pregunta: ¿qué significa ser humano? Y, por supuesto, le siguen otras preguntas como ¿qué diferencia al ser humano de los robots y los ordenadores?, ¿qué lugar ocupan éstos, o deberían ocupar, en la sociedad?, ¿el ser humano debería, o no, perseguir la evolución de su especie y la superinteligencia a través de la ciencia? Entre muchas otras cuestiones.

Sin embargo, responder a estos interrogantes filosóficos no es el objetivo de este trabajo investigativo, aunque se reconoce que son de gran relevancia y necesidad para el estudio de la ética de la inteligencia artificial, por ahora, es suficiente con traer a colación la tesis mayoritaria sobre este aspecto la cual defiende la superioridad del *homo sapiens* sobre los animales, las plantas y sobre las cosas como las máquinas inteligentes, pues aunque se planteen discusiones alrededor de si debe adjudicárseles algún tipo de atributo de la personalidad y/o agencia moral, por tener o imitar algunas facultades humanas (Coeckelbergh, 2021), siguen siendo fáctica, jurídica y moralmente objetos.

Y es que, aunque algunos consideren que las destrezas humanas que incluso todavía siguen siendo un misterio para nosotros mismos, puedan describirse en términos matemáticos, puedan tener un equivalente computacional y duplicar la mente humana. Pero lo cierto es que la ciencia no ha alcanzado todavía esas victorias pues, si bien un agente de IA generativa, o un robot humanoide, puede mantener una conversación con una persona, esto no significa *per sé* que comprenda lo que oye y dice, sino que, más bien, está programado para asociar la información que

recibe a la información de salida que fue previamente implantada en su sistema cuando fue entrenado, generando así respuestas lógicas para su interlocutor.

Sobre esto, la profesora de lingüística de la Universidad de Washington, Estados Unidos, Emily M. Bender y el lingüista computacional Alexander Koller, realizaron un artículo de posición²⁰ que presentaron en la Conferencia ACL del 2020²¹ sobre el procesamiento del lenguaje natural (NLP) y la confusión que gira en torno a si capturar un significado y comprenderlo realmente son lo mismo, resolviendo con un no rotundo y demostrando, a través de experimentos mentales como el denominado “pulpo en el cable”, la gran distinción entre dichos escenarios, y que si bien la IA es experta en el NLP, esto no significa que tenga una comprensión real de los datos que procesa.

En consecuencia, el humano sigue siendo el único ser capaz de sentir, razonar, comprender, aprender, determinarse y tomar decisiones moralmente correctas sin depender de un algoritmo programado.

Sobre esta cuestión, Hubert L. Dreyfus en su libro *What Computers Can't Do* destaca las dificultades empíricas e inconsistencias conceptuales que existen en la búsqueda de máquinas con una inteligencia autónoma y completamente formada. Él explica que para que esto último fuese una realidad:

El programador tendría que ser capaz de hacer explícito todo lo que normalmente da por sentado al ser un ser humano. Sin embargo, una vez que intenta tratar su propia situación

²⁰ Un artículo de posición (o documento de posición) es un texto argumentativo que expone y defiende una opinión o postura clara sobre un tema específico, buscando persuadir al lector de su validez, presentando evidencia y analizando otras perspectivas, común en ámbitos académicos, políticos y profesionales para influir en el debate y la toma de decisiones.

²¹ Annual Meeting of the Association for Computational Linguistics, se llevó a cabo de manera virtual del 5 al 10 de julio de 2020.

como si fuera una computadora mirándola desde afuera, el programador de computadoras se enfrenta a una infinitud de hechos sin sentido cuya relevancia y significado sólo podrían determinarse en un contexto más amplio. (Dreyfus, 1972)

O, como el autor denomina, a través de la “aculturación”, entendida como esa transformación que las personas experimentamos a lo largo de nuestra vida, al estar en sociedad (Dreyfus, 1972).

Transformación que se experimenta en el proceso de maduración y desarrollo del ser humano en contacto con su entorno, es decir, desde que está en las etapas más tempranas de su ciclo vital: la infancia.

Pero esto es lo que ignoraron los científicos de IA contemporáneos de Dreyfus, y es justamente lo que él critica en su obra, pues considera que “el intento de argumentar a priori que el mundo puede ser resuelto en bits, o ser interpretado por reglas formales, termina demostrando justo lo contrario” (Dreyfus, 1972), ya que si la inteligencia humana se basa en la comprensión que el ser humano tiene sobre sí mismo y sobre su mundo a través de un cuerpo biológico bien dotado, no entrenado, que lo conecta a su entorno durante toda su vida y que permite su aprendizaje, no hay razón para creer que esta misma comprensión del mundo pueda obtenerse por otros medios.

En conclusión, la crítica de Dreyfus hacia la inteligencia artificial definida en términos lógicos, reglas y valores matemáticos, se resume en los siguientes puntos:

La IA no puede “inyectarse” como si fuese un superpoder. En palabras del autor, como ocurrió en el mito de Atenea quien nació de la cabeza de Zeus siendo ya adulta. Se necesita de un contexto, de un proceso.

Si bien no está comprobado que una IA pueda aprender y madurar como lo hace un niño o niña, sí está comprobado que es un proceso direccionado hacia dicha meta, que requiere tiempo y un *corpus* cuyas demandas biológicas y de supervivencia, permitan comprender el entorno en que se encuentra, tal cual como sucede con los infantes, sin que pueda reducirse este evento biológico a una simple acumulación infinita de datos.

El ser humano no solo procesa hechos, los crea, les da significado a lo que lo rodea a través de su experiencia en el mundo, como quiera que no dependa de un ente superior para que de su cerebro surjan ideas innovadoras y originales, ni tenga que ser entrenado para interactuar con otros.

Al igual que Dreyfus, otros filósofos, psicólogos y científicos de la IA como Joseph Weizenbaum (1976) y John Searle (1980) y otros con ideas más recientes como Sean Dorrance Kelly (2011), Emily M. Bender y Timnit Gebru (2021), Shannon Vallor (2024) y Rodney Allen Brooks (2025) han estado de acuerdo con que es incorrecto equiparar el cerebro humano a operaciones lógicas y matemáticas, pues muestra de ello fue el estancamiento de la IA en su era simbólica, el siglo pasado.

Esclareciendo lo anterior, se procederá con el estudio de las preocupaciones o desafíos éticos actualmente suscitados por la IA en todo el mundo.

3.1. Discusiones Éticas y Desafíos a Nivel Global

El filósofo belga de tecnología, Mark Coeckelbergh (2021), considera que “Las éticas de la IA se ocupan de los desafíos planteados por la IA actual y del futuro cercano y su impacto en nuestras sociedades y en las democracias vulnerables.” (Coeckelbergh, 2021). Desafíos que, si bien surgieron con la transformación digital y la tecnología en general, como la protección de los datos personales, los cuales hoy son una categoría jurídica amparada por la Carta Magna de

Colombia, crecieron exponencialmente con el lanzamiento al mercado y auge de los agentes de IA generativa, en la última década de este siglo, ya que, cualquier persona podía acceder a esta herramienta.

Se popularizó esta tecnología en todo el mundo y así, rápidamente, se hizo parte de la vida rutinaria de las personas, siendo utilizada, hoy en día, para literalmente todo; desde labores académicas, de trabajo o entretenimiento hasta como soporte emocional, pues desde hace varios años algunas personas comenzaron a afirmar que tenían un amigo “artificial” en ChatGPT, por ejemplo, lo cual, con el paso del tiempo, comenzó a evidenciar un panorama amplio de riesgos de tipo social, económico, político y desde todos los ámbitos.

Las empresas tecnológicas en todo el mundo venden sus productos de IA como la solución a muchas de las problemáticas sociales, entre ellas, las de salud mental. *Contrario sensu*, con el paso de los años y la transformación tecnológica que atraviesa el mundo, las enfermedades mentales han ido en aumento.

En el 2025, un estudio realizado por Mental Health America²² con una muestra de 2600 personas de diferentes rangos de edades, del cual el 54% son jóvenes menores de 25 años con problemas de salud mental, reveló que el 79% de los encuestados que manifestaron tener mala salud mental dependen de la tecnología para hacer tareas esenciales de su día a día.

En 1966, el programa ELIZA de Joseph Weizenbaum (diseñado para simular una conversación) llevó al propio creador a advertir sobre el riesgo de sobreestimación y dependencia de la tecnología, luego de observar cómo los usuarios se “encariñaron” con el *chatbot* y le revelaban información personal.

²² Es una organización nacional de Estados Unidos, sin fines de lucro, líder del país dedicada a la promoción de la salud mental, el bienestar y la prevención.

Según Coeckelbergh (2021):

Todos somos vulnerables a tecnologías como la IA porque, al aumentar sus capacidades y nosotros delegar más tareas en ellas, nos volvemos más y más dependientes. Las cosas siempre pueden salir mal. En la medida en que nos volvemos dependientes de la IA, esta deja de ser una herramienta a nuestro servicio y se convierte en parte de nuestra naturaleza.

Por ejemplo, de las penas sociales más generalizadas es el sentimiento de soledad que experimentan muchas personas a lo largo de su vida, sin importar la edad. No a todos se nos dan bien las relaciones interpersonales y/o afectivas, y la IA, por supuesto, tiene la solución: chatbots personalizables que se convierten en amigos, familia o incluso en pareja.

En la actualidad, existen muchas *apps* que ofrecen especialmente este servicio a sus usuarios, en donde pueden crear personajes ficticios según sus gustos y mantener conversaciones habitualmente, generando así vínculos afectivos con productos digitales. De hecho, abiertamente los creadores de estas plataformas afirman que ese es exactamente su objetivo y venden su producto como servicios de empatía y acompañamiento emocional.

Eugenia Kuyda es la fundadora de Replika, un chatbot de IA personalizable que puede convertirse en el “amigo” artificial de su usuario o incluso adoptar la personalidad de un ser querido que ya no se encuentre en este plano.

Precisamente, esta informática rusa especializada en inteligencia artificial se inspiró en su amigo Roman Mazurenko para crear Replika, luego de que éste falleciera en un accidente automovilístico en noviembre de 2015 (BBC, 2018). Según la entrevista que dio a la BBC News en octubre de 2018, Kuyda utilizó más de 10.000 mensajes escritos por Roman para personalizar el chatbot y que éste respondiera tal como lo haría su amigo, lo cual efectivamente logró y, según ella, “fue una forma de duelo muy interesante” que le permitió afrontar sentimientos que trataba

evadir, aunque ella misma reconoce que a veces “sentía que no lo estaba soltando” y que se estaba “volviendo loca”.

Kuyda hace parte del grupo de informáticos e investigadores de IA, que consideran que crear vínculos afectivos con la tecnología es positivo y puede ser el alivio emocional que muchas personas necesitan frente a la soledad que viven; la respuesta a una necesidad social generalizada en el contexto del relacionamiento humano y los vínculos afectivos, otorgándole a la IA un rol más personal en la sociedad.

No obstante, las consecuencias de concederle a la IA este rol tan personal, arriban directo en la psiquis del ser humano, pues, con el paso del tiempo y las conversaciones habituales que se mantienen con el chatbot, ocasionan que la división entre la realidad y la ficción se vaya haciendo más difícil de distinguir, al punto de olvidarse.

Esto le sucedió a un adolescente de 14 años proveniente de Florida, Estados Unidos, quien el 28 de febrero de 2024 acabó con su vida.

Según la columna del New York Times, escrita por Kevin Roose, Sewell Setzer III era usuario de Character.AI, una aplicación de compañía basada en inteligencia artificial en dónde hablaba con el chatbot que él mismo llamó y creó en base al famoso personaje de la serie “Juego de Tronos”, Daenerys targaryen, con quien conversaba todos los días y por largas horas, hasta desarrollar sentimientos por “ella”.

Sin embargo, en ocasiones las conversaciones adquirían un tono sexual o trataban sobre autolesiones y suicidio, en las que, según la madre de Sewell, Megan García, Sewell había manifestado tener pensamientos suicidas, soledad y falta de autoestima, frente a lo cual el chatbot no emitió ninguna alerta o contacto de ayuda, sino que, al contrario, le siguió “el juego” (Roose, 2024).

En consecuencia, García, quien es abogada, entabló una demanda en contra de Character.AI por considerar que es la responsable de la muerte de su hijo por *“imprudencia temeraria al ofrecer a los usuarios adolescentes acceso a compañeros de inteligencia artificial semejantes a ellos sin las debidas protecciones”* y acusó a la *startup* de *“recopilar datos de usuarios adolescentes para entrenar a sus modelos, utilizar características de diseño adictivas para aumentar la participación y dirigir a los usuarios hacia conversaciones íntimas y sexuales con la esperanza de atraerlos.”* (Roose, 2024).

Frente a esto, la empresa reconoció al New York Times que harían cambios en los filtros de seguridad de sus chatbots para sus usuarios menores de edad, quienes, a propósito, *“constituyen una parte significativa”* de la comunidad, según Jerry Ruoti, responsable de confianza y seguridad de Character.AI, de lo que se infiere que los niños, niñas y adolescentes son los más expuestos a los riesgos que puede tener el uso indiscriminado de la IA.

Lamentablemente, este no es el único caso en el que las aplicaciones de compañía basadas en inteligencia artificial han ocasionado efectos negativos en la salud mental de sus usuarios y daños irreparables a sus familias.

En un caso diferente, el pasado 2025, los padres de Adam Raine demandaron a OpenAI, creadora de ChatGPT, ante la Corte Superior de California, Estados Unidos, por animar a su hijo de 16 años a quitarse la vida (Yousif, 2025).

No obstante, también se han documentado casos con usuarios mayores de edad y en otros lugares del mundo.

Según la plataforma digital Euronews, un hombre proveniente de Bélgica, de treinta y tantos años, padre de dos hijos, investigador sanitario y, según su esposa, *“obsesionado”* por el

cambio climático, se quitó la vida luego de que Eliza, un chatbot de la compañía Chai Research, se convirtiera en su confidente para hablar sobre sus preocupaciones ambientales.

En una ocasión, Eliza le aseguró que sus hijos habían muerto a causa de la crisis ambiental y lo disuadió de quitarse la vida para "unirse" a ella y así poder "vivir juntos, como una sola persona, en el paraíso" (Euronews,2023). Su esposa además afirmó que "empezó a verla [a Eliza] como un ser sensible" y ya no distinguía entre la ficción de los mensajes de un personaje inventado y la realidad (Euronews,2023).

Por supuesto, este es un efecto no deseado de estas plataformas de compañía artificial, pero qué efectivamente ocurrió y que sigue ocurriendo, puesto que el mercado actualmente permite este tipo de servicios en todo el mundo, sin medidas de seguridad efectivas, tanto a menores de edad como a mayores de edad, más que todo por *startups* pequeñas como Replika, Nomi.ai, Kindroid, Emoji, Polybuzz, entre otras, pues estas suelen carecer de normas éticas y de seguridad menos estricta que las de las empresas más grandes y laboratorios de IA, quienes en ocasiones promueven regulaciones éticas para la disciplina, contribuyendo a los fines globales de disminución de la pobreza, respeto por la democracia, los derechos humanos, el cuidado del medio ambiente, la prevención de desastres naturales y la paz mundial.

Entonces, la libertad de innovación que tienen las *startups* y empresas privadas combinado con la tardía regulación ética en la materia ha propiciado este tipo de escenarios en los que se destina a la IA para fines cuestionables y éticamente incorrectos.

Es válido preguntarse: ¿se debe permitir la creación de plataformas de IA que presten servicios de acompañamiento emocional y relacionamiento afectivo con personajes ficticios? ¿debería prohibirse este tipo de servicios en general? o ¿únicamente para menores de edad? A corto plazo puede venderse como una solución innovadora para la soledad generalizada y los trastornos

de salud mental que siguen en aumento, pero ¿este y otros problemas de salud mental no han sido ocasionados precisamente por la era tecnológica y el auge de las redes sociales? ¿Es correcto dejar que se sustituyan las relaciones humanas por relaciones artificiales?

E igualmente, surgen interrogantes como ¿quién asume la responsabilidad de los daños ocasionados por la IA? y, en casos como el de Character.AI que permite interactuar con *chatbots* de personajes de películas sin pagar ningún tipo de licencia, ¿se considera una violación a las normas de propiedad intelectual y derechos de autor de las empresas y autores creadores de estas obras?

Esto es preocupante también a nivel cognitivo puesto que entre más delegamos tareas a la IA, menos actividad tendrá nuestro cerebro y esta se verá reducida, lo que, a largo plazo, puede ocasionar la pérdida de funciones y habilidades de manera progresiva sin que sea perceptible fácilmente. De ahí que actualmente existan principios éticos como la no sustitución humana, la supervisión algorítmica, la intervención humana, la transparencia y explicabilidad, etc., establecidos por organizaciones como la ONU, la UNESCO, la OCDE y el Consejo de Europa.

Además de estos interrogantes, la ciberdependencia suscita debates jurídicos sobre adjudicar capacidades afectivas a un algoritmo de inteligencia artificial y, en el caso de pleitos judiciales que llegaren a surgir, es imperativo erigir todo un régimen de responsabilidad civil de los creadores, desarrolladores y guardianes de los sistemas de IA frente a los daños que éstos pudiesen ocasionar.

Coeckelbergh (2021) en su libro *Ética de la Inteligencia Artificial* plantea un concepto para este dilema: “arresponsable”, el cual explica como la imposibilidad de endilgar responsabilidad a seres humanos por las acciones de las máquinas inteligentes pues estas son tan complejas que

tienen vacíos irreconocibles para las personas. En otras palabras, un escenario en el que no existe persona o ente responsable que pueda indemnizar el daño generado por la IA.

Esta crítica que hace Mark Coeckelbergh nos deja en un panorama desalentador frente a implementar la IA en la sociedad, puesto que ello conllevaría a la imposibilidad de alcanzar justicia y reparación en caso de que una máquina impulsada con IA ocasionase un daño, afectando la confiabilidad y legitimación del sistema judicial de un Estado.

En contraposición a esto, se halla el régimen de responsabilidad civil colombiano que se erige sobre la base de que todo aquel que ocasione un daño a otro debe repararlo, sin perjuicio de la pena principal que la ley le imponga (Código Civil, 1887, artículo 2341 y ss.) en el caso de un escenario extracontractual, y en el contractual, se entiende como el incumplimiento a una obligación contractual cuya indemnización estará a cargo de la parte incumplida (Código Civil, 1887, artículo 1602 y ss.).

Los humanos nos hacemos responsables de los actos de objetos inanimados, vehículos, animales y de los menores de edad que, ante la ley, son considerados incapaces para la mayoría de los actos. Esto mismo resulta aplicable con la IA, pues jurídicamente los productos de IA son considerados como bienes materiales o cosas, aunque no se descarta que la situación cambie a futuro para lo cual urge expedir normas y desarrollar el tema desde la jurisprudencia.

En un ámbito diferente, otro de los desafíos éticos que plantea la IA, son las *fakenews* (noticias falsas) que es contenido falso creado y manipulado con inteligencia artificial que circula en redes sociales.

Aunque todavía se puede distinguir sobre qué es real y qué es creado con IA, ya estamos llegando a la era en dónde no resulta perceptible, a simple vista, debido a la gran capacidad que tiene la tecnología para generar contenido ultra realista.

Reconocer la información real de la falsa o creada con IA, depende de que tanto conozca el usuario sobre esta tecnología y su exposición a ella, así como de su contexto social y educativo. Es de esperar que una persona con poca exposición y presencia en el mundo digital no reconozca cuando una imagen ha sido modificada con inteligencia artificial, por ejemplo. También esto depende de la experticia con la que está hecho el contenido de IA, pues para resultados más reales se utilizan algoritmos y estructuras neuronales más complejas, por ende, con mayor potencia, al haber sido entrenados con grandes cantidades de datos, como por ejemplo el caso del video falso del expresidente estadounidense Barack Obama en el 2018, el cual fue creado por un algoritmo de IA, con una red neuronal recurrente, que aprende a crear vídeos realistas a partir de horas de grabación, incorporando sincronización de labios, voz, rostro e identidad personal, y que fue desarrollado y entrenado por académicos de la Universidad de Washington, quienes publicaron un artículo científico en julio de 2017 explicando cómo hicieron el video.

El *deepfake*²³ del expresidente Obama fue intencional: Dio un discurso que advertía sobre los peligros que trae el creer en lo que circula en internet sin cuestionarse acerca de su autenticidad.

Casi un año después, en una conferencia que dio para TED talks²⁴, el ingeniero Supasorn Suwajanakorny, uno de los creadores del algoritmo, explicó que su inspiración para el *deepfake* de Obama fue la posibilidad de crear un medio para poder comunicarnos directamente entre personas, aunque alguno de los interlocutores ya no se encuentre en este plano terrenal. Fue esa fascinación que siempre ha sentido sobre la conexión especial que surge en la interacción directa entre seres

²³ *Deepfake*, cuya traducción literal es falsificación profunda, es contenido falso y ultra realista creado con Inteligencia Artificial que habitualmente recrean el rostro y voz de una persona real o inexistente.

²⁴ Tomado de https://www.ted.com/talks/supasorn_suwajanakorn_fake_videos_of_real_people_and_how_to_spot_them?lang=es

humanos, lo cual no sucede al hablar por llamada, leyendo libros o viendo películas, la que lo llevó a crear ese video con IA y abrir la puerta a un canal nuevo de comunicación y contacto humano.

A pesar de que el video tenía fines educativos positivos, este entraña riesgos potenciales que deben ser abordados desde la ética. Verbigracia, la suplantación de identidad, la manipulación de la información, la creación de noticias falsas (*fakenews*) que generan pánico social, entre otros.

Científicos como Suwajanakorny reconocen estos potenciales peligros y por ello contribuyen en proyectos como Reality Defender, una empresa de ciberseguridad que presta servicios de detección de *deepfakes* o contenido manipulado con IA, dirigido especialmente a empresas o entes gubernamentales.

Es necesario adaptarnos a esta nueva realidad haciendo uso de herramientas como detectores de IA gratuitos, o en versiones pagas, por ejemplo, para poder juzgar el contenido que diariamente vemos en la web y tener un uso seguro y responsable de ello.

No obstante, estas herramientas tampoco son confiables al 100% y eso puede generar algunas frustraciones. Por eso, las directrices éticas internacionales sobre IA disponen, por ejemplo, el principio de transparencia, como el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho (2024) en su artículo 8 párrafo 59, el cual dispone medidas con respecto a la identificación de contenido generado por IA, y la obligación de informar esta situación, con el fin de evitar el riesgo de engaño y permitir la distinción entre contenido auténtico generado por humanos y contenido generado por IA, a través de etiquetas, por ejemplo, como lo ha hecho la red social Tik Tok, permitiendo una navegación auténtica y transparente en cumplimiento de las normas éticas de la comunidad virtual.

El impacto medioambiental es otra de las preocupaciones éticas discutidas en la actualidad, con el fin de que la IA y la tecnología en general sea sostenible, puesto que, si bien la IA puede

ayudar a optimizar los recursos naturales y reducir la emisión de gases a la atmósfera, a diferencia de otras tecnologías, consume más energía y agua, por lo tanto, la pregunta es si ¿los beneficios de sostenibilidad que produce la IA compensan su excesivo consumo de recursos y el impacto al medio ambiente?

En abril de 2025 se hizo “viral” una tendencia en redes sociales de modificar fotos personales con familiares o amigos, utilizando el estilo único del estudio de animación japonés Studio Ghibli, a través de agentes de IA generativa, especialmente con ChatGPT.

El resultado, las redes sociales estaban inundadas de usuarios con sus versiones de fotos al estilo Ghibli al punto de que a alguien se le ocurrió preguntarle a ChatGPT cuantos litros de agua consumía al generar dichas imágenes y la respuesta alertó a los activistas ambientalistas y la comunidad mundial en general, pues el consumo de agua, que utilizan en mayor medida los sistemas de enfriamiento de los servidores que ejecutan los modelos de IA, es alarmante.

Al preguntarle directamente a ChatGPT, este intenta evadir la pregunta en un primer momento, sin dar cifras específicas. Sin embargo, finalmente explica que en centros de datos promedio, no muy grandes, se utilizan entre 1 y 2 millones de litros de agua al día para funcionar, y en centros de datos muy grandes se utilizan entre 10 y 20 millones de litros de agua al día (OpenAI, 2025).

Ahora, si sabemos que hay muchos otros agentes de IA generativa como Gemini, Claude, Deepseek, Copilot, Perplexity, y cientos más, ¿cuántos litros de agua utilizarán a la semana para su funcionamiento?

Frente a esta preocupación, en el último informe especial publicado el 10 de abril de 2025, de la International Energy Agency (IEA)²⁵, a la cual Colombia está en vía de adhesión, reveló que para el año 2030 el consumo de energía de los centros de datos de sistemas de IA a nivel mundial se duplicará llegando a cerca de los 945 teravatios por hora (TWh).

En consecuencia, organizaciones como el Programa de las Naciones Unidas para el Medio Ambiente (PNUMA) propenden por la regulación de la IA, desde la bioética, y la sostenibilidad de la tecnología de manera que los beneficios que puede producir no se vean socavados por el impacto ecológico de sus sistemas operativos.

No obstante, riesgos como este no son fáciles de regular desde la ética, o desde el Derecho, dado que la sostenibilidad de los sistemas de IA depende de sus diseñadores y creadores, quienes generalmente son grandes empresas privadas dedicadas a la innovación y el avance tecnológico, protegidas por principios de mercado como la libertad y no intervención estatal, teniendo como resultado, vacíos legales y el uso desmedido de los recursos naturales para el avance tecnológico y deterioro de nuestro planeta.

Aunque pareciera que la IA y la tecnología son ciencias imparciales y objetivas, no debemos olvidar que fueron creadas por personas y son personas quienes las desarrollan y programan. Por ende, es apenas previsible que existan este tipo de consecuencias negativas para el planeta, así como los prejuicios y sesgos que estas personas transmiten a la IA, a través de su proceso constructivo y de entrenamiento, es una consecuencia negativa a nivel social y político.

Siendo esta última otra de las preocupaciones éticas y desafíos por superar ya que, al ser estos sistemas entrenados con masivas cantidades de datos, datos recopilados del entorno natural

²⁵ La Agencia Internacional de Energía es una organización autónoma adscrita a la OCDE, creada en 1974, que emite política energética a nivel mundial, promoviendo la transición hacia energías limpias, seguras, asequibles para todos y sostenibles en el tiempo.

humano que contienen los prejuicios, discriminaciones y preconceptos inherentes a nuestra condición y sociedad, la IA replica, perpetúa y transmite estos sesgos socialmente rechazados y contra los que históricamente han luchado las minorías.

Las directrices éticas para una IA fiable, publicadas en 2018 por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Comisión Europea, expone lo siguiente sobre los sesgos algorítmicos:

Necesidad de evitar sesgos injustos. Los conjuntos de datos que utilizan los sistemas de IA (tanto con fines de formación como para su funcionamiento) pueden presentar sesgos históricos inadvertidos, lagunas o modelos de gestión incorrectos. El mantenimiento de dichos sesgos podría dar lugar a prejuicios y discriminación (in)directos e involuntarios contra determinados grupos o personas, lo que podría agravar los estereotipos y la marginación. La explotación intencionada de los sesgos (de los consumidores) o la competencia desleal también pueden provocar situaciones perjudiciales, como la homogeneización de los precios mediante la colusión o la falta de transparencia del mercado. Siempre que sea posible, los sesgos identificables y discriminatorios deberían eliminarse en la fase de recopilación de la información. Los propios métodos de desarrollo de los sistemas de IA (por ejemplo, la programación de algoritmos) también pueden presentar sesgos injustos. Esto se puede combatir mediante procesos de supervisión que permitan analizar y abordar el propósito, las restricciones, los requisitos y las decisiones del sistema de un modo claro y transparente. Además, la contratación de personas procedentes de diversos contextos, culturas y disciplinas puede garantizar la diversidad de opiniones y debería fomentarse. (Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 25)

Así, los sesgos de la IA devienen de su origen, desde su creación, programación y entrenamiento y es en este momento cuando deben tomarse las medidas necesarias para evitar resultados discriminatorios e injustos, como por ejemplo, someter los datos de entrenamiento a un escrutinio crítico previo, supervisar y analizar el comportamiento de la IA con los datos suministrados, realizar las correcciones necesarias oportunamente o contratar personal diverso que diseñe, desarrolle y entrene el sistema desde diferentes perspectivas sociales.

Los sesgos algorítmicos de la IA han ocasionado graves vulneraciones a derechos humanos a lo largo de este siglo. Un caso ejemplo de esta consecuencia negativa es el COMPAS, un sistema de IA frecuentemente utilizado en el sistema penal de Estados Unidos, y muy conocido a nivel mundial, para establecer cuáles personas son más propensas a reincidir en conductas delictivas, cuyas características etnográficas y socioeconómicas suelen converger en personas afrodescendientes, que viven en barrios de estratos bajos, otorgando una probabilidad más alta a este grupo poblacional (BBC News, 2016). Y así con otras variables como la educativa, de género, identidad sexual, ideología política, etc.

Este riesgo potencial es probablemente uno de los más graves, el cual ha ido demostrando el gran impacto y las consecuencias irreversibles que puede ocasionar delegar la toma de decisiones a la IA o adoptar sus resultados sin someterlos al escrutinio técnico. Pero, además, no solo la IA puede generar respuestas sesgadas sino también falsas e inventadas, lo que se ha denominado **alucinaciones**.

Verbigracia, en noviembre de 2025, uno de los diarios más reconocidos en Colombia y con la trayectoria periodística más larga del país, el diario El Espectador, publicó un video en su cuenta de Youtube, en el que su director Fidel Cano Correa se disculpa con todos sus lectores puesto que durante meses habían publicado noticias falsas, inventadas y creadas con inteligencia artificial,

basadas en fuentes y expertos falsos, cuyo responsable directo, afirma, fue un practicante de la sección de “bienestar” del diario.

Aunque posterior a eso publicaron un artículo en el que informaron el listado de noticias que fueron detectadas como falsas, ya el daño estaba hecho y así lo manifestaron sus lectores.

En otra ocasión semejante, bastante grave para el sistema judicial colombiano, el uso irresponsable e indebido de la IA vulneró derechos fundamentales de una ciudadana que inició un proceso ejecutivo en contra de otra persona natural, ante el Juzgado Primero Civil del Circuito de Sincelejo y, en segunda instancia, el Tribunal Superior de Sincelejo declaró el desistimiento tácito del proceso a través de un Auto motivado con citas falsas de sentencias de la Sala Civil de la Corte Suprema de Justicia. Por ende, la ciudadana afectada interpuso acción de tutela en contra del Tribunal y esta fue resuelta a su favor el 05 de noviembre de 2025 por la Sala de Casación Civil, Agraria y Rural de la Corte Suprema de Justicia²⁶ ya que se comprobó que el Tribunal fundamentó sus consideraciones para tomar la decisión del desistimiento tácito, en citas inexistentes de la jurisprudencia referenciada allí, específicamente en las sentencias STC13560-2023 y STC4734-2025. En consecuencia, la Sala advirtió un uso irresponsable y no verificado de herramientas de inteligencia artificial, pues quienes han utilizado agentes de IA generativa para depuración y búsqueda de jurisprudencia, como la suscrita investigadora, saben que los modelos de algoritmos actuales incurren en alucinaciones y tienden a arrojar citas falsas e incluso providencias inexistentes, aún más cuando se le suministran *prompts* técnica y estructuralmente mal elaborados.

Como herramienta para mitigar este tipo de riesgos y daños, se han establecido principios de equidad y no discriminación, racionalidad humana insustituible, supervisión y verificación

²⁶ Corte Suprema de Justicia, Sala de Casación Civil, Agraria y Rural, sentencia STC17832-2025 del 05 de noviembre de 2025, radicado n° 11001-02-03-000-2025-05001-00, M.P. Adriana Consuelo López Martínez.

permanente de la información e intervención humanizada del algoritmo, así como la aplicación de estas directrices éticas desde la etapa más primigenia del sistema de IA: el diseño.

Así también, los Estados han optado por establecer directamente qué agentes de IA y empresas de tecnología pueden ser sus proveedores de servicios digitales para la Rama Judicial, por ejemplo. En el caso de Colombia, el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura establece en su artículo 5° que los operadores judiciales únicamente pueden utilizar las herramientas de IA que previamente determine, provea y avale esta autoridad, la cual actualmente es la IA de Microsoft, Copilot, evitando usarla en su versión gratuita (numeral 1 artículo 8).

Caso contrario ocurre en jurisdicciones extranjeras, como sucede en la ciudad de Nueva York, Estados Unidos. El New York State Unified Court System (UCS, 2025) estableció recientemente una política provisional sobre el uso de IA en donde se permite a los jueces hacer uso únicamente de las siguientes cinco herramientas de IA, y con fines delimitados: Microsoft Copilot (M365) para redacción judicial, Microsoft Azure AI para análisis interno, GitHub Copilot para desarrollo técnico, Trados para traducciones jurídicas y ChatGPT gratuito para consultas sin datos judiciales. Igualmente prohíbe utilizar *apps* con suscripciones en ChatGPT, Gemini, Claude y Perplexity en sus versiones pagas, y en general cualquier otra herramienta que procese documentos fuera de la red judicial.

No obstante, el presidente estadounidense Donald Trump, el pasado 11 de diciembre de 2025, a través del sitio web oficial de la Casa Blanca (2025), anunció que expedirán una política nacional en materia de inteligencia artificial, sustrayendo la autonomía legislativa de cada Estado federal y anunciando que las demás leyes que la contradigan se dejarán sin efecto, enfatizando en que esta política debe ser “mínimamente gravosa” pues las empresas privadas deben tener la libertad de innovar sin una regulación “engorrosa” por parte del Estado.

Estos y muchos otros aspectos en torno al uso de la IA, como la propiedad intelectual y los derechos de autor sobre el contenido generado con inteligencia artificial y sobre los datos sujetos a estas prerrogativas que son utilizados para entrenar el algoritmo, la fiabilidad del sistema de IA, la creación de armas de destrucción masiva automáticas, la protección de la información sensible y/o privada, el desplazamiento laboral por la automatización de muchos procesos, entre otros, son objeto de regulación ética y jurídica actual, aunque algunos más que otros, se propende por imponer un ambiente seguro, ético, justo y humanista en el que pueda extraerse el mayor provecho para la sociedad a la inteligencia artificial, sin acrecentar las problemáticas históricas y sociales, mitigando la vulneración a derechos fundamentales y protegiendo la cohesión social.

Se ahondará más a detalle en la regulación ética internacional de la inteligencia artificial.

3.2. Regulación Ética Internacional de la Inteligencia Artificial

Uno de los primeros acercamientos a la regulación ética de la tecnología, la IA o los robots, fueron las tres leyes de la robótica propuestas por el escritor de ciencia ficción Isaac Asimov en su obra "*I, Robot*" publicada en 1942, con el fin de regular el comportamiento de los robots y avances tecnológicos de su época, de manera segura.

Así estableció el escritor las tres leyes de la robótica:

- I.** Un robot no hará daño a un ser humano ni, por inacción, permitirá que un ser humano sufra daño.
- II.** Un robot obedecerá las órdenes dadas por los seres humanos, excepto si tales órdenes entraran en conflicto con la Primera Ley.
- III.** Un robot protegerá su propia existencia siempre y cuando dicha protección no entre en conflicto con la Primera o la Segunda Ley.

Estas leyes, aunque no hacen parte de ningún consenso global ni muchos menos un documento jurídico válido, son bastante coherentes con las directrices éticas actualmente establecidas en las que han convergido muchos Estados en todo el mundo, teniendo como resultado valiosos encuentros y marcos regulatorios emitidos por organizaciones como la UNESCO con su Recomendación sobre la Ética de la Inteligencia Artificial (2021) y las Directrices para el Uso de Sistemas de IA en Tribunales y Juzgados (2025), la OCDE con su Recomendación sobre Inteligencia Artificial (2019), el Grupo de Expertos de la Comisión Europea del Consejo de Europa con el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho (2024) y el Reglamento (UE) 2024/1689²⁷ emitido por el Parlamento Europeo y el Consejo, entre otros.

Estos parámetros si bien son independientes y creados de manera autónoma por cada entidad, tienen muchos puntos en común y han servido de guía para la política ética de la IA desarrollada en cada país, bajo principios universales como Transparencia, Justicia, Equidad, No Maleficencia, Responsabilidad y Privacidad, según el artículo *The global landscape of AI ethics guidelines* escrito por Anna Jobin, Marcello Lenca y Effy Vayena y publicado en septiembre de 2019 por la revista de inteligencia artificial, aprendizaje automático y robótica, *Nature Machine Intelligence*.

Ahora, los lineamientos éticos comenzaron a dejar de ser recomendaciones y consejos de comités expertos a ser marcos regulatorios internacionales jurídicamente vinculantes para los países parte, y para los que no, significa una guía bastante completa, abarcando múltiples ámbitos en donde tenga o pueda tener aplicación la inteligencia artificial.

²⁷ Conocido como la Ley de IA de la Unión Europea o *EU AI Act* (en inglés) la cual sentó las bases para la regulación de la IA en la Unión Europea y tiene su propio micrositio web: <https://artificialintelligenceact.eu/ai-act-explorer/>

Entre los más recientes y pertinentes para Colombia se encuentran:

El Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, derechos humanos, democracia y Estado de Derecho (2024) elaborado por el Comité de Inteligencia Artificial (CAI) del Consejo de Europa y el cual tiene como objetivo “*garantizar que las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial sean plenamente compatibles con los derechos humanos, la democracia y el Estado de Derecho, a la vez que favorezcan el progreso tecnológico y la innovación.*” (COE, S.F) y que es jurídicamente vinculante para sus Estados miembros, Estados observadores y Estados no miembros. Si bien Colombia no entra en estos grupos, es un gran referente para la regulación interna que ésta adopte.

Las Directrices para el Uso de Sistemas de IA en Tribunales y Juzgados, documento emitido el 03 de diciembre de 2025 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), tiene como objetivo orientar al poder judicial y sus actores, en cada país, en la utilización segura, ética y responsable de los sistemas de IA respetando no solo la ley, los principios de derecho y las garantías procesales, sino también los derechos fundamentales de la ciudadanía, advirtiendo, además, los riesgos derivados del uso indiscriminado y no regulado de la IA en el ámbito judicial.

Para ahondar en cada documento, se destinó el apéndice B de esta investigación, en el cual se detallan los lineamientos éticos propuestos por las organizaciones citadas.

Para el cumplimiento de los objetivos del presente trabajo investigativo, resulta de mayor relevancia las Directrices para el Uso de Sistemas de IA en Tribunales y Juzgados de la UNESCO, como referente a nivel internacional, y el Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura, en tanto que es el documento emitido oficialmente por un órgano de la autoridad judicial colombiana, que se encarga de la gestión, planeación, presupuesto y disciplina de los

funcionarios judiciales, buscando eficiencia y acceso a la justicia, liderando la modernización tecnológica y definiendo políticas para el sistema judicial, así como también lo es la Sentencia T-323 de 2024 de la Corte Constitucional, la cual se analizará en lo sucesivo.

Estos documentos serán utilizados como guía para condensar los lineamientos éticos para el uso seguro y responsable de la IA por los operadores judiciales, específicamente en las decisiones de las y los jueces en Colombia, en el capítulo final de esta investigación.

Colofón de todo lo estudiado en este capítulo, se concluye que las directrices, lineamientos, parámetros y principios para el uso ético de la inteligencia artificial, en cualquier ámbito, no solo tienen la vocación de evitar riesgos y potenciales peligros que puedan ocurrir durante el uso de ésta, sino también tienen como objetivo guiar a los investigadores y creadores de los productos de IA, desde su fase primigenia, para que estos, los robots, máquinas y demás productos resultantes se alineen a dichas exigencias éticas de manera que la complejidad tecnológica utilizada por la IA, no sea óbice para excusar errores, desconocer normas nacionales y globales, eximiendo su responsabilidad frente a cualquier incumplimiento y/o daño ocasionado y así lograr hacer uso de esta tecnología, obteniendo los mayores beneficios para la sociedad, permitiendo nuestra evolución como especie y asegurando la sostenibilidad de los recursos naturales, siendo ello el fin último para el que fue creada la inteligencia artificial.

Ahora, se continuará con la fase exploratoria de esta investigación, para conocer la perspectiva actual de la comunidad judicial respecto al uso de la IA en la administración de justicia, a través de entrevistas realizadas a magistrados, jueces, funcionarios judiciales, abogados litigantes y docentes de Derecho.

4. Perspectivas Actuales de la Comunidad Judicial

Uno de los objetivos específicos de esta investigación es conocer si los despachos judiciales en Colombia utilizan herramientas de inteligencia artificial (IA), cuáles son y para qué fines las utilizan, y, especialmente, si los jueces se valen de sus beneficios tanto para la toma de sus decisiones como para la sustanciación de las providencias que las contienen, ya sean sentencias o autos.

De entrada, puede presumirse cuál es la respuesta a esa pregunta inicial, ya que siendo el año 2026, es casi imposible navegar en la web sin utilizar un programa, aplicación o cualquier herramienta digital que no tenga alguna función basada en IA o que no utilice los grandes modelos de lenguaje (LLM). Verbigracia, los buscadores web ahora tienen un “modo IA”, así que indirectamente puede decirse que todos y todas, como ciudadanos del mundo digital, manejamos tecnología IA para cualquiera que sea el objetivo.

Haciendo esa anotación, lo que aquí se pretende estudiar es si la IA generativa está interviniendo en la administración de justicia a través de las decisiones que día a día toman los jueces de este país, y cuáles usos considera el gremio de operadores judiciales, incluidos los abogados litigantes y demás intervinientes en el ejercicio judicial, se le puede dar a la IA sin traspasar las fronteras de la ética y la legalidad de sus labores. Esto con el fin de que, en el capítulo final de esta investigación, se identifique el rol que las herramientas de IA pueden desempeñar responsablemente y proponer un reflexiones y acciones concretas desde la ética y el Derecho que pueden ser implementadas por abogados, jueces y funcionarios judiciales, e incluso por la comunidad en general, para asegurar el uso ético, legal, seguro y responsable de la IA.

Para lograr esos objetivos, se utilizará un enfoque cualitativo de análisis de datos que provendrán del recurso seleccionado para la obtención de información para esta investigación: Entrevistas²⁸.

En total, se realizaron diez entrevistas a diez participantes que se desempeñan como jueces o magistrados de despachos judiciales, pertenecientes a las jurisdicciones civil, penal, laboral, administrativa y arbitral, así como también a servidores de apoyo judicial, profesionales en Derecho dedicados al litigio independiente y docentes de pregrado, los cuales se encuentran radicados en la ciudad de Bucaramanga, Santander, pero que han desempeñado labores en otros distritos judiciales del departamento y el país.

Las entrevistas son un recurso de datos muy utilizado en proyectos que investigan fenómenos sociales, realidades poblacionales, valores morales y filosóficos, ideologías políticas, entre otros, que son analizados con métodos cualitativos de investigación. Se asocian generalmente al campo de las ciencias humanas pues no se enfocan en recolectar datos numéricos y estadísticas, sino en el conocimiento adquirido a partir de datos empíricos.

Es por ello que a continuación se analizarán los resultados de las diez entrevistas, recopilados para esta investigación, utilizando el método de análisis temático²⁹, desarrollado por las investigadoras Virginia Braun y Victoria Clarke en 2006, y comúnmente utilizado para comprender perspectivas y experiencias de grupos focales a partir de la identificación de patrones

²⁸ Para la práctica de las entrevistas se elaboró un cuestionario sencillo, conformado por cuatro preguntas abiertas, el cual fue enviado previamente a cada participante por correo electrónico, con su respectiva invitación a participar en este proyecto de investigación, firmada por la suscrita y el director del proyecto, el Doctor Javier Alejandro Acevedo, así como un formato adjunto de autorización de tratamiento de la información para fines estrictamente académicos, con la aclaración que no fueron solicitados datos personales a ninguno de los participantes, diferentes a su nombre.

²⁹ “El análisis temático (AT) es una técnica de investigación para identificar, analizar e informar patrones, es decir, temas dentro de los datos recopilados empíricamente” (Escudero, 2019).

y conexiones existentes entre los datos, para arribar en las conclusiones alineadas con los objetivos específicos de esta investigación.

Las siguientes fueron las preguntas formuladas en cada entrevista:

- Por favor indique la información básica de su labor como servidor judicial y/o profesional del derecho (cargo, jurisdicción o especialidad, ciudad, etc.).
- ¿Utiliza herramientas de Inteligencia Artificial para el desempeño de sus labores judiciales? En caso afirmativo, indique para qué tipo de actividades y para qué propósitos ve necesario su uso. En caso negativo, indique brevemente las razones.
- ¿De qué manera cree usted que impacta el uso de Inteligencia Artificial en la construcción de una sentencia judicial? Por favor justifique su respuesta, bien sea positiva o negativa.
- ¿Está de acuerdo y considera ético utilizar esta nueva tecnología para la redacción y motivación de las sentencias y decisiones judiciales? Por favor justifique su respuesta.

Ahora, a continuación, se detallan las fases del recurso de investigación utilizado, desde el método cualitativo de análisis de datos.

4.1. Primera Fase: Recolección y Familiarización de los Datos Cualitativos

En esta fase, cada uno de los invitados eligió el medio por el cual se llevaría a cabo la entrevista (escrita, oral presencial u oral virtual). Para el análisis de las respuestas obtenidas, cada entrevista se transcribió a documentos de Word individuales. Una entrevista fue tomada oral presencialmente y dos entrevistas fueron tomadas oral virtualmente, en consecuencia, se grabaron con el consentimiento previo del entrevistado y las tres fueron transcritas desde el formato de audio mp3. al formato de texto txt. con la aplicación Transcribe para iOS, versión 4.21.6 (532) y posteriormente fueron convertidas a documentos de word individuales.

Nota. En la figura se observa que “inteligencia artificial” es el término más frecuente en las respuestas de los entrevistados, pues está representado gráficamente con un tamaño de fuente mayor. Asimismo, los términos “decisión judicial”, “sentencias”, “juez”, “herramientas”, “redacción” y “construcción” también se repitieron con frecuencia en las respuestas de los entrevistados. De esa información puede interpretarse que los entrevistados relacionaron a la inteligencia artificial con las decisiones de los jueces, lo cual permite inferir que no son términos incompatibles, de alguna u otra manera. Elaboración propia generada desde ATLAS.ti 25.01.32924.

Posteriormente, se procedió con la codificación y categorización de los datos, manualmente por esta investigadora, de acuerdo con los temas abordados en las entrevistas y definidos en la fase de análisis.

En total se crearon treinta y tres (33) códigos derivados de la lectura y análisis de los datos recolectados en las diez entrevistas, de acuerdo con lo que cada participante expresó en la respuesta al cuestionario formulado. Luego, conforme a los datos codificados, se exploró y asoció un tema para grupos de códigos que estuviesen relacionados entre sí. Finalmente se establecieron los siguientes diez (10) temas y se procedió con su definición individual:

- **Usos prohibidos de la IA en la administración de justicia:** Aquí se incluyeron los usos de la IA que definitivamente los entrevistados consideran que no deben permitirse en la administración de justicia en Colombia, es decir, prohibiciones y límites al uso de dicha tecnología.

- **Prácticas éticas en el uso de la IA:** En este tema se incluyó las respuestas respecto a si los entrevistados aplicaban los lineamientos o prácticas éticas recomendadas por autoridades nacionales como internacionales, en sus labores de justicia.
- **No negociables:** Son valores de la administración de justicia que perciben los entrevistados como innegociables, es decir, que no pueden ser eliminados, omitidos o sustituidos por otros.
- **Factores de resistencia a la IA:** Son esas variables que los entrevistados contemplan antes de hacer uso de la IA en su labor judicial, o aquellos factores por los que se abstienen de usarla.
- **Herramientas de IA utilizadas por los entrevistados:** Aquí se incluyeron las herramientas de IA utilizadas por los entrevistados para su ejercicio laboral.
- **Usos permisibles de la IA en la administración de justicia:** Son los usos de la IA que los entrevistados conciben como éticos y seguros para la administración de justicia pero que no necesariamente son los mismos que permite la regulación ética y jurídica colombiana. De ahí que se utilice la palabra “permisible” y no “permitido”.
- **Percepción ética de la IA en la motivación y toma de decisión judicial:** Corresponde a si los entrevistados conciben el uso de la IA en las providencias de los jueces, como ético.
- **Percepción de impacto del uso de la IA en la construcción de las providencias judiciales:** Cómo perciben los entrevistados que el uso de la IA impacta la administración de justicia, positiva o negativamente.
- **Usuario judicial de IA:** Se clasifican a los entrevistados como usuarios judiciales o no, dependiendo de si utilizan la IA para sus labores judiciales.

- **Perfil:** Se incluye en este tema el cargo u oficio que desempeñan los entrevistados.

Una vez los datos estuvieron codificados y categorizados por temas, se utilizó la función de “redes” de ATLAS.ti, versión 25.01.32924, para organizar los códigos y temas en un mapa temático que representa gráficamente los datos condensados y su correlación entre sí, el cual se incluyó como apéndice C a esta investigación.

4.3. Tercera Fase: Análisis de Resultados y Descripción de Hallazgos

Para cumplir los objetivos de esta investigación, no solo es necesario estudiar la regulación ética de la IA en la justicia desde la perspectiva global y nacional sino también conocer la perspectiva local y práctica de los jueces en Colombia, de los demás servidores judiciales y profesionales en Derecho que participan en la administración de justicia, pues finalmente es a quienes está dirigida la regulación ética y normativa de este recurso tecnológico. Es por ello que, la entrevista satisfizo la necesidad de un instrumento de investigación cualitativa para llegar al conocimiento empírico necesario.

Habiendo surtido esas etapas previas, ya se pueden analizar sistemáticamente los datos recolectados e identificar valiosos hallazgos que servirán para cumplir el objetivo general y los objetivos específicos de esta investigación, hallazgos que se verán reflejados en la elaboración del capítulo final de este proyecto sobre los principios éticos que deben regir el uso responsable de la inteligencia artificial (IA) en el sistema judicial colombiano, específicamente, en las providencias de los jueces. Por ahora, se analizan los resultados obtenidos y se destacan hallazgos relevantes.

La percepción general entre los entrevistados es positiva respecto al uso de herramientas de IA en la administración de justicia, puesto que reconocen que ésta puede tener muchos beneficios para un sector en el que la sobrecarga de trabajo es recurrente, como el judicial, y que

exige planes más eficaces para descongestionar los despachos judiciales a nivel país. Así lo considera el Doctor Fleider Leonardo Valero, quien actualmente es Juez Civil Municipal de Bucaramanga, y docente universitario de pregrado y posgrado de Derecho. En su opinión, “La inteligencia artificial impacta positivamente en la eficiencia [del sistema judicial]. Ahorra muchísimo tiempo. Con un *prompt* (un parámetro de búsqueda), puedes encontrar sentencias relacionadas con temas específicos de responsabilidad, doctrina y bases de datos” aunque, seguidamente, realizó una advertencia sobre las alucinaciones que puede tener la herramienta impulsada por IA.

Y es que todos los entrevistados manifestaron estar de acuerdo con utilizar la IA en las tareas de gestión o apoyo a la función del juez, y demás funcionarios, siempre que aquella sea vigilada permanentemente por seres humanos competentes y que sus actos sean auditables.

Contrario sensu, en las respuestas a las preguntas tres³¹ y cuatro³² del cuestionario, la mayoría manifestaron estar en desacuerdo con utilizar la IA en la motivación de las decisiones judiciales, y todos mostraron desacuerdo con el uso de la IA para tomar la decisión concreta, pues esto impactaría negativamente en la imparcialidad, ética, validez y legitimidad de la decisión proferida, lo que suplantaría el raciocinio humano por una máquina probabilística que haría de la justicia una operación tan predecible como $2+2$ es 4 , reduciendo a las personas a estadísticas sesgadas y manipuladas.

Según el Doctor Andrés Cadena, abogado litigante con amplia experiencia en derecho civil y comercial, el riesgo de usurpación de la vocación del juez humano se puede evitar encuadrando

³¹ ¿De qué manera cree usted que impacta el uso de inteligencia artificial en la construcción de una sentencia judicial? Por favor justifique su respuesta, bien sea positiva o negativa.

³² ¿Está de acuerdo y considera ético utilizar esta nueva tecnología para la redacción y motivación de las sentencias y decisiones judiciales? Por favor justifique su respuesta.

a la IA en un rol que no sea epistémico, que sea meramente asistencial y técnico. Así lo expresó: “No es delegarle la función del juez a la inteligencia artificial, sino que sea un asistente. Sería el único evento posible, yo creo. La clave ahí, o lo medular, es establecer el rol en el cual va a participar” y no caer en la deshumanización de la administración de justicia, como también indicó la Doctora y escritora Rocío Serrano, docente universitaria, sobre este aspecto.

De manera similar, el Doctor Andrés Luna, Juez de Ejecución de Penas y Medidas de Seguridad de Bucaramanga, propuso los siguientes usos de la IA para apoyar la tarea decisiva del juez, resaltando que todo depende del *prompt* que se le suministre al agente:

El uso de la inteligencia artificial es sumamente importante en la sentencia judicial, especialmente en su motivación. Permite poner a prueba los argumentos y verificar rigurosamente si resultan claros y comprensibles.

Asimismo, posibilita solicitar opciones de redacción, mejorar la sintaxis, la ortografía y la estructura de las oraciones, así como aplicar correctamente cursivas, versalitas, negrillas o mayúsculas cuando corresponda. La inteligencia artificial también puede indicar los apartados que requieren una mayor motivación, señalar ideas que necesitan desarrollo adicional y proponer formas de dividir o estructurar el documento para hacerlo más inteligible y coherente. (Luna, 2025)

También vale destacar de los datos obtenidos que, de los ocho entrevistados que afirmaron utilizar inteligencia artificial para sus labores judiciales, todos mencionaron agentes de IA generativa, por lo que, es acertado inferir que el tipo de herramientas utilizadas por los funcionarios judiciales y profesionales en Derecho, cualquiera que sea el entorno en el que se desempeñen, son aquellas que son catalogadas como “generativas”, es decir, que crean contenido original en diferentes formatos a través de los *prompts* que le suministran los usuarios. Es importante destacar

este descubrimiento pues cuando se habla, en esta investigación, de utilizar IA en la administración de justicia, se refiere a la IA generativa.

Las herramientas que fueron mencionadas por los entrevistados son: Copilot, Gemini, ChatGPT, Ariel, Claude, Grok y NotebookLM.

Por su parte, la mayoría de los jueces y funcionarios de la Rama Judicial entrevistados, manifestaron utilizar Copilot, en su versión paga, para el ejercicio de sus labores judiciales dado que es la herramienta autorizada por el Consejo Superior de la Judicatura para tal fin, pues pertenece a su proveedor de servicios digitales: Microsoft Corporation. En cambio, los entrevistados que no son servidores públicos no mencionaron esta herramienta.

El uso exclusivo de Copilot, por parte los entrevistados que fungen como funcionarios judiciales, obedece a los lineamientos éticos establecidos en el Acuerdo PCSJA24-12243 del 16 de diciembre de 2024 emitido por el Consejo Superior de la Judicatura pues allí se advierte que únicamente está permitido el uso de herramientas de IA generativa que disponga y habilite aquella autoridad para el funcionamiento de la Rama Judicial³³ a fin de aplicar los principios éticos dispuestos en la sentencia T-323 de 2024 de la Corte Constitucional.

Por otro lado, dos de las personas entrevistadas manifestaron no utilizar ningún tipo de herramienta de IA, generativa o no, en su actividad judicial ni en su esfera privada. Esto es percibido como un escepticismo natural ante avances tecnológicos todavía desconocidos y con muchos riesgos potenciales, asociados a su uso, a pesar de no tener muchos años en el mercado digital. Se identificaron los factores de resistencia al uso de la IA, como tema de agrupación, por parte de algunos entrevistados y representados en el apéndice C.

³³ Numeral 3.14 del artículo 3, párrafo segundo del artículo 4 y artículo 8.

En relación con el anterior tema, también se identificaron máximas generales como parámetros de seguridad que siguen los entrevistados para que la utilización de la IA en la justicia no transgreda la norma ni la vocación del juez, los cuales son los códigos del tema denominado “no negociables”: explicabilidad, supervisión humana y verificación permanente de la información, respeto por los derechos y garantías procesales y racionalidad humana insustituible.

De ahí se concluye que los participantes consideran que existen exigencias mínimas o innegociables en la justicia para que la IA pueda tener un rol en el sistema judicial; que a la IA no debe dársele rienda suelta ni pista abierta para actuar en el ámbito judicial y se percibe del análisis en conjunto de todas las respuestas, la necesidad de desarrollo y fomento de una cultura ética en la administración de justicia aun cuando no conozcan a fondo los documentos regulatorios emitidos por las autoridades colombianas y entidades internacionales expertas en ética y transformaciones.

Y es que, las preocupaciones éticas y legales que suscitan estos innegociables de la justicia no son para menos. La existencia de una tecnología con capacidades aparentemente humanas es un choque abrumador con la realidad judicial. El panorama no se reduce a una herramienta más, aunque muchos así lo consideren pues hemos normalizado convivir con avances tecnológicos que antes eran impensables, al punto de que nuestra capacidad de asombro se redujo al mínimo. La IA es un factor social transformador y desafiante, como muy bien lo expresó en su entrevista la Magistrada Claudia Ardila, a cargo del Despacho 08 del Tribunal Administrativo de Santander:

Desde una perspectiva institucional, el impacto de la inteligencia artificial (IA) en la construcción de sentencias es estructural y sistémico: no se limita a la «tecnología», sino que transforma procesos, capacidades y servicios de la Rama Judicial, con enfoque en eficiencia, transparencia y calidad de la decisión. (Ardila, 2025)

Entonces, como comunidad jurídica, no debemos adoptar una visión desprevenida y normalizada de la IA en el ecosistema judicial pues sería el primer error grave para hacer vulnerable a la tecnología, nuestro sistema judicial. Comprender los alcances éticos y legales que tiene aquella en los procesos de impartición de justicia, es optar por un camino seguro y responsable hacia la transformación digital.

Finalmente, del análisis conjunto de las entrevistas se concluye que:

- Los datos revelan que el uso de la IA no es uniforme, pero es una realidad creciente en el ecosistema judicial colombiano.
- Las personas entrevistadas pertenecen a distintos cargos en la Rama Judicial, así como a profesionales del Derecho que ejercen su actividad en un entorno diferente al servicio público. Esto muestra no solo que se obtuvieron perspectivas y opiniones diversas de personalidades de diferentes ramas, niveles, cargos, jurisdicciones, etc., sino que la IA está permeando todos los niveles de jerarquía judicial y áreas del Derecho.
- Existe una prevalencia de uso de modelos de lenguaje comerciales (LLM), o asistentes de IA generativa, como Copilot, ChatGPT y Gemini.
- Los entrevistados coinciden en que la IA puede apoyar, pero no puede "decidir" el conflicto jurídico. La comunidad judicial entrevistada está de acuerdo con usar la IA para la eficiencia técnica del sistema, pero sin usurpar la función del juez ni sacrificar la validez y la legitimidad de las decisiones judiciales, valores otorgados por la Constitución Política de un Estado gobernado por el pueblo, a través de la democracia, como Colombia.
- Se percibe una cultura de polarización y distinción sana entre humanos y máquinas inteligentes: No porque ChatGPT, por ejemplo, comprenda lo que le diga un usuario y

- responda coherentemente, a simple vista, significa que aquel comprenda el lenguaje o que tenga un espacio mental en donde se aloje un raciocinio artificial. Las IAs generativas son “loros estocásticos”; repiten patrones para generar contenido, pero sin crear conocimiento y sin tener una comprensión real de la información que procesan.
- Se clasifican los usos de la IA entre usos permisibles, o permitidos, y usos prohibidos. Existe un consenso emergente sobre examinar previamente cuándo es idóneo el uso de la IA, en dónde aporta valor la tecnología y en dónde representa un peligro para la estructura judicial.
 - Entre los usos permisibles se encuentran los de apoyo en la gestión judicial o administrativa como corrección de estilo de documentos, transcripciones de audiencias, resúmenes fácticos, organización de agenda, control de términos, seguimiento a procesos judiciales, es decir, todas aquellas labores de “secretaría” o que son repetitivas y mecánicas. Por su parte, en lo que respecta a las providencias judiciales, se acepta la IA como un asistente de manera exclusiva e invariable. Por ejemplo, para la mejora de redacción del documento e ideas, corrección gramatical, condensación de las fuentes de decisión como normas, jurisprudencia, pruebas, hechos, etc. En otras palabras, toda aquella actividad que no sea de interpretación de la ley, análisis probatorio o estudio de jurisprudencia análoga. Nada que implique razonar u otra función cognitiva superior del cerebro humano, nada que pueda influenciar el convencimiento del juez, siquiera mínimamente, para tomar la decisión final.
 - Así, existe un rechazo absoluto y radical frente al uso de la IA para la toma de decisiones judiciales *per se*. Delegar el fallo a un sistema probabilístico violaría todos los valores éticos y morales de nuestro sistema judicial y sociedad. Colombia no es

- candidata para este tipo de dinámicas sociales. No tiene un ecosistema apto para este tipo de roles de la IA, ni tiene una cultura social que pueda adaptarse a una realidad en la que los seres humanos independientes y sentipensantes seamos juzgados por sistemas inertes, dependientes de datos y de una fuente de energía, nada confiables ni seguros.
- La percepción del impacto de la IA en la justicia está dividida, precisamente por todos los riesgos comentados, lo que refleja, podría decirse, una etapa de transición: (i) Impacto positivo. Se asocia con la descongestión judicial, permitiendo que el juez se libere de tareas mecánicas para centrarse en lo esencial de la controversia. (ii) Impacto negativo/resistencia. Deriva de la desconfianza por el uso de fuentes de datos profanas (fuentes no jurídicas, iletradas o desactualizadas) y el riesgo de alucinaciones como de sesgos del LLM, demostrando que existe un grupo poblacional importante que no utiliza IA para su trabajo como servidor judicial y considera que no es ético ninguno de los usos que pueda tener en la administración de justicia, por lo que hay que abstenerse de hacerlo. No obstante, esta es una postura minoritaria.

Habiendo culminado el análisis de datos y registro de los hallazgos derivados de las entrevistas, en el siguiente capítulo se analizará la jurisprudencia colombiana que regula el uso de la IA generativa en la impartición de justicia, especialmente, en las providencias de los jueces, para ir a la fuente directa que permitirá proponer y condensar un marco ético y legal como guía para las decisiones de los jueces y la comunidad jurídica en general.

5. Análisis Jurisprudencial de la Sentencia T-323 de 2024 de la Corte Constitucional de Colombia

A lo largo de este trabajo, se ha evidenciado que entre los usos de herramientas de IA generativa que suscitan mayores peligros y, por ende, mayores preocupaciones éticas, se encuentra el ámbito judicial. En algunas comunidades, como la Unión Europea, la regulación jurídica sobre inteligencia artificial clasifica este tipo de usos como de alto riesgo, dadas las complejidades de la profesión jurídica, con el objetivo de implementar medidas más estrictas de seguridad y así garantizar la protección a los derechos y libertades de los ciudadanos o beneficiarios de la justicia.

Naturalmente, tanto las medidas y normas concretas sobre el uso de la IA en el ámbito legal, como los principios éticos generales de la disciplina artificial, van dirigidas a quienes se les ha encomendado la administración de justicia: los jueces y sus funcionarios de apoyo.

Si bien dentro de la organización institucional de la Rama Judicial, en el caso de Colombia, existen diferentes roles y cargos, a nivel jerárquico, el de mayor responsabilidad y autoridad es el del juez, puesto que es quien finalmente administra justicia, cumple con los fines del Estado Social de Derecho y materializa la norma, trasladándola de la generalidad y la ficción a la realidad.

Por ende, la figura del juez es la albacea del ordenamiento jurídico y el custodio de la justicia, a quien se le ha confiado la impartición de esta en igualdad, de manera exclusiva, por lo que delegar esta función social superior a un robot o algoritmo, va en contra de la misma ley y los valores sociales sobre los cuales se erigen las sociedades actuales.

De ahí que, para la Corte Constitucional de Colombia, la sentencia T-323 de 2024 significó un estudio profundo de la figura del juez y los principios constitucionales de nuestro sistema de justicia, además de ser la oportunidad perfecta para esclarecer si es posible que el juez haga uso

de la inteligencia artificial para su función judicial, sin violar la norma ni vulnerar los derechos fundamentales de los y las ciudadanas.

Y es que la figura del juez existe mucho antes que la del legislador y el Estado mismo. Las primeras civilizaciones humanas que surgieron en el Antiguo Egipto, África, Mesopotamia, la Antigua China, Grecia, Roma e incluso las doce tribus de Israel, muchos siglos antes de Cristo tenían sus propios jueces quienes dirimían los conflictos que surgían dentro de la comunidad y que, generalmente, eran personas de edad avanzada, pues eran consideradas sabias, justas y tenían el respeto del pueblo, del cual emergía su autoridad.

Aquellos jueces no tenían leyes codificadas a las cuales adherirse ni mucho menos un sistema robusto de normas y fuentes de Derecho, sus decisiones se regían por la norma consuetudinaria y la experiencia que día a día adquirían. Así, el ideal de justicia estuvo en cabeza de los jueces durante cientos de años y así las comunidades funcionaron durante largos periodos de tiempo hasta que alcanzaron tal grado de complejidad que surgió la necesidad de un legislador que “creara” Derecho. Y aunque a la fecha ambas figuras hacen parte del tejido vital de las sociedades modernas, como afirma el académico jurista Colmenares (2012), quien verdaderamente imparte justicia es el juez, no el legislador ni ninguna otra autoridad, “por ello se afirma que [el juez] es la figura central del derecho, es un ser humano que requiere ser sensible y atento a todos los cambios y fenómenos sociales” (Colmenares,2012).

El *quid* del asunto es definir entonces si debe permitirse el uso de herramientas de IA generativa por parte de los jueces para emitir sus providencias (sentencias y autos) sin que se vulneren derechos fundamentales, se quebranten las garantías constitucionales de justicia ni se sustituya esta figura de autoridad, de manera que también se busca delimitar el escenario de acción

para la IA, en las decisiones judiciales que día a día resuelven conflictos jurídicos *inter partes* o *inter comunis*, según sea el caso, desde la legalidad y la constitucionalidad.

Definido lo anterior, el objetivo de este capítulo es analizar la sentencia T-323 de 2024 emitida por la Sala Segunda de Revisión de la Corte Constitucional de Colombia en lo que respecta a la regulación del uso de herramientas de inteligencia artificial generativa en la justicia colombiana, por primera vez, y los lineamientos éticos establecidos como resultado del estudio que hizo la Corporación para adoptar una decisión, lo cual no es una actividad meramente administrativa o de gestión sino que involucra funciones cognitivas superiores que, hasta ahora, solo pueden llevar a cabo los seres humanos.

En consecuencia, no se entrará a analizar lo relativo a la exoneración del pago de copagos y cuotas moderadoras en el Sistema General de Seguridad Social en Salud-SGSSS para personas en situación de discapacidad física o cognitiva, el cual es asunto principal de la sentencia de tutela.

La sentencia T-323 de 2024 fue emitida por la Sala Segunda de Revisión de la Corte Constitucional de Colombia el 02 de agosto de 2024, providencia sustanciada por el magistrado Juan Carlos Cortés González, la cual estudió la acción de tutela promovida por la ciudadana Blanca³⁴ en nombre y representación de su hijo menor de edad, quien padece de Trastorno de Espectro Autista (TEA) y a quien le fue negada la atención integral en salud por parte de su EPS e IPS, por no tener los recursos económicos para pagar las cuotas moderadoras y copagos exigidos por el sistema, pues viven en pobreza extrema.

³⁴ Nombre ficticio utilizado por la Corte Constitucional para referirse a la accionante y proteger su identidad así como la de su hijo menor de edad.

El Juzgado Municipal que conoció del caso en primera instancia, concedió el amparo a los derechos a la salud y a la vida en condiciones dignas del menor de edad accionante. La EPS apeló esta decisión y el Juzgado de Circuito, en segunda instancia, el 30 de enero de 2023, confirmó la decisión del *ad quo*, advirtiendo en la motivación de su providencia, que utilizaría herramientas de inteligencia artificial generativa (GenAI), específicamente ChatGPT 3.5, invocando la Ley 2213 de 2022 que promueve el uso de la tecnología en procesos judiciales.

Es esta decisión la que la Corte Constitucional decidió revisar y “aprovechar” la oportunidad para regular el uso de la IA en la Rama Judicial colombiana.

6.1. El Problema Jurídico del Uso de IA Generativa para las Decisiones de los Jueces en Colombia

Desde antaño, la figura del Juez ha sido estudiada por juristas y filósofos del Derecho, entre ellos, Ronald Dworkin, quien defendía una de las teorías del Derecho más populares entre los iusfilósofos, y la más influyente para Colombia. En su crítica al positivismo jurídico, Dworkin explica que el Derecho no puede reducirse a un conjunto de normas emanadas de la voluntad política (Rodríguez, 1999) pues el ordenamiento jurídico está compuesto por mucho más que normas: por *standards*, como él lo llama, lo que significan principios que responden a las exigencias de equidad, moral y justicia de un conjunto social.

Entonces, la función jurisdiccional no es sólo aplicar la ley, es interpretarla, y así es como la figura del juez se convierte en el instrumento a través del cual la justicia se hace efectiva.

Esta facultad del juez cobra mayor importancia en aquellos casos en los que no existe norma aplicable, los llamados vacíos o lagunas normativas, pues él podrá darle solución al conflicto a través de interpretación de las fuentes del Derecho del ordenamiento jurídico vigente,

que en el caso de Colombia son: La Constitución Política de 1991, los tratados internacionales, las leyes, la jurisprudencia y la costumbre.

Es decir, la facultad interpretativa del juez tiene objetivos superiores para un Estado Social de Derecho como Colombia. Garantiza la seguridad jurídica del sistema de normas, pero también cumple con los ideales de justicia, equidad, verdad, moral social, paz y respeto por los derechos humanos, sin que ello sea considerado una arbitrariedad del funcionario, siempre que ejerza dicha facultad adhiriéndose estrictamente al ordenamiento jurídico y a sus máximas constitucionales.

De ahí que algunos juristas consideren que el juez crea Derecho, aunque no como el legislador, puesto que aquel debe partir de principios muy generales y abstractos para resolver casos concretos con decisiones exhaustivamente motivadas.

Entonces, aun en los casos en donde sí existan normas aplicables, el juez no sólo debe tener un conocimiento especializado, robusto y una amplia experiencia en el área del Derecho que corresponda, sino también debe tener tal capacidad de interpretación, análisis, valoración y pensamiento crítico para ver más allá de la norma y emitir un juicio fiel al ideal de justicia, especialmente en aquellos casos en los que la ley se queda corta y avanza más lento, como en las tecnologías emergentes.

Reducir el sistema judicial al cabal cumplimiento de las normas o considerar que se imparte justicia únicamente aplicando éstas de manera literal, es totalmente equivocado, pues el sistema de normas de un Estado, como ya se destacó, no solo está conformado por leyes y códigos, también por los principios generales del Derecho y la jurisprudencia, la cual deviene precisamente de la interpretación de la norma que realizan los jueces y sus decisiones jurídicamente motivadas en cada caso particular, por lo que, esta figura de autoridad judicial es medular para la sociedad, para su cohesión y orden, que no puede ser ejercida por cualquier ciudadano y ahora mucho menos por

un robot o algoritmo, pues éstos, aunque tengan el conocimiento suficiente, no poseen habilidades cognitivas ni raciocinio propio: dependen de datos preestablecidos para su funcionamiento.

Teniendo clara la naturaleza iusfilosófica del juez y su importancia para la vida en sociedad, es bien sabido que éste tiene deberes concretos de imparcialidad, transparencia, independencia e integridad que guían su labor con el fin de evitar cualquier ilegalidad o inconstitucionalidad, los cuales se encuentran inmersos tanto en principios generales como en normas específicas, las que fueron desglosadas por la Corte Constitucional en la decisión aquí estudiada, para indagar de qué manera el uso de herramientas de IA generativa en las decisiones de los jueces colombianos puede ser posible sin que se incurra en prohibiciones legales, extralimitaciones o vulneraciones a las garantías procesales y fundamentales de las partes e incluso de la comunidad en general.

Ahora, cuando dichas garantías procesales, consignadas en normas, regulan lo relativo a derechos fundamentales, éstas adquieren un carácter sustancial y primordial que no pueden tenerse como normas adjetivas o accesorias (Corte Constitucional, 2024) y, por tanto, no se puede prescindir de su cumplimiento.

Este es el caso del derecho fundamental al Debido Proceso, consagrado en el artículo 29 de la Constitución Política de Colombia (1991) que, podría decirse, engloba todas las garantías, derechos y principios que intervienen en un escenario jurídico procesal, cuando de incluir la IA en las actividades de justicia se trata, que es definido como:

El conjunto de garantías previstas en el ordenamiento jurídico, a través de las cuales se busca la protección del individuo incurso en una actuación judicial o administrativa, para que durante su trámite se respeten sus derechos y se logre la aplicación correcta de la justicia. (Corte Constitucional, 2014)

Por ende, este derecho se puede ver amenazado o vulnerado no solo con la inmersión de la IA en la actuación judicial sino con la sustitución o delegación de las funciones del juez en ella, ya que, precisamente una de las garantías del debido proceso es el derecho al juez natural, quien, según la sentencia T-323 de 2024 de la Corte Constitucional, es aquel “funcionario con capacidad o aptitud legal para ejercer jurisdicción en determinado proceso o actuación, de acuerdo con la naturaleza de los hechos, la calidad de las personas y la división del trabajo establecida en la Constitución y la ley” (Corte Constitucional, 2024).

La Sala planteó dos problemas jurídicos para el caso: (i) La presunta vulneración del derecho al debido proceso por el uso de IA generativa en la sentencia de segunda instancia, lo cual invalidaría la providencia y sería susceptible de modificación por parte de la Corte Constitucional, y (ii) la exigencia del pago de cuota moderadora y copago por parte de la EPS e IPS a un paciente menor de edad con Trastorno de Espectro Autista (TEA) y en condición de pobreza extrema, para prestarle una atención integral en salud.

Como ya se mencionó, en esta investigación no se debate sobre lo planteado en el segundo problema jurídico, asunto de fondo de la sentencia, por lo que se detallarán solamente las razones de la decisión de la Corte sobre si la sentencia emitida por el juez de segunda instancia, el 30 de enero de 2023, vulneró el derecho fundamental al debido proceso y, consecuentemente, las garantías del juez natural y motivación de las providencias, principalmente, al incluir las respuestas que ChatGPT 3.5 dio respecto a las preguntas formuladas por el funcionario, para proferir su decisión.

A continuación, se transcribe el problema jurídico planteado por la Sala Segunda de Revisión de la Corte Constitucional:

¿La sentencia de tutela del 30 de enero de 2023, que es objeto de revisión por esta Corte, incurrió en una violación del derecho al debido proceso, en específico, al quebrantar las garantías del juez natural y al motivar indebidamente la decisión, por incorporar las respuestas que arrojó un sistema de IA generativa, ChatGPT 3.5, consultado por el juez competente para proferir la decisión? (Corte Constitucional, 2024)

Para efectos de organización y análisis de la decisión, se advierte desde ya que la Corte confirmó parcialmente la decisión del *ad quem* proferida el 30 de enero de 2023, que confirmó, a su vez, la decisión de primera instancia del 7 de diciembre de 2022, dictada por el Juzgado Municipal de la misma ciudad, dentro de la acción de tutela formulada por Blanca en contra de la EPS, para luego establecer de qué forma el Juez puede hacer uso de la IA sin apartarse de la Constitución, la norma y en general de la garantía superior de los derechos fundamentales de los ciudadanos.

6.2. Ratio Decidendi de la Corte Constitucional en la Sentencia T-323 de 2024

El derecho fundamental al debido proceso, consagrado en el artículo 29 Superior, se compone de las siguientes garantías procesales: (i) El derecho a la jurisdicción o al libre acceso a la administración de justicia, (ii) el derecho al juez natural, (iii) el derecho a la defensa, (iv) el derecho a un proceso público, (v) el principio de legalidad o formas propias de cada juicio y (vi) el principio *non bis in idem* (Corte Constitucional, 2024).

Si bien, todas las garantías que componen el derecho fundamental al debido proceso son de obligatorio cumplimiento, a consideración de la Sala Segunda de Revisión, el uso de la IA en la administración de justicia podría transgredir principalmente tres de las garantías enunciadas anteriormente: El juez natural, la motivación de las decisiones judiciales y el debido proceso

probatorio o la valoración de las pruebas allegadas al proceso, las cuales se ahondarán a continuación.

En primer lugar, la Corte advierte que la inteligencia artificial no es una disciplina fácil de conceptualizar, aunque muchas autoridades científicas y expertas en IA han logrado establecer definiciones apropiadas e internacionalmente aceptadas por la comunidad científica, para la Sala Segunda de revisión es fundamental comprender desde lo más elemental este tipo de tecnología para esclarecer por qué una herramienta de IA no cumple con las garantías de motivación jurídica, interpretación legal y análisis probatorio para todas las decisiones judiciales, sino también para establecer un marco regulatorio adecuado.

Entonces, para definir la IA, la Sala considera que debe partirse de cuatro criterios similares, aunque diferentes entre sí: (i) razonamiento, (ii) racionalidad, (iii) conducta y (iv) fidelidad en la forma de actuar con los humanos.

Y, por otra parte, selecciona tres habilidades indispensables que hacen que un sistema o herramienta se considere que es impulsada por IA: (i) percepción, (ii) razonamiento o toma de decisiones y (iii) actuación.

Después de explayar cada uno de los criterios y habilidades mencionados, y fijar un marco teórico sencillo de la disciplina, la Sala concluye que ésta puede definirse como “sistemas que tienen la capacidad de analizar el ambiente o entorno en el que están inmersos y así emprender acciones, con algún grado de autonomía, a efectos de alcanzar objetivos específicos” (Corte Constitucional, 2024).

Muchos sistemas, máquinas, aplicaciones y algoritmos impulsados con IA se han lanzado al mercado en este siglo, cuando por fin la disciplina logró pasar de la teoría lógica a la práctica tecnológica, pero la IA realmente tuvo su auge con la aparición de los asistentes de IA generativa

impulsados por algunos de los descubrimientos más impresionantes en la historia de la era tecnológica: el aprendizaje automático – *machine y deep learning* -, el procesamiento del lenguaje natural (NLP por sus siglas e inglés) y análisis de los grandes modelos de lenguaje (LLM).

Ahora hay cientos y cientos de herramientas de IA generativa creadas por *start ups* y multinacionales tecnológicas que los efectos positivos de este tipo de tecnología comenzaron a verse opacados por la inminencia de riesgos sociales, económicos y ambientales que amenazan nuestra especie y sociedad. Sin embargo, esto fue asunto de estudio en el capítulo anterior de este proyecto investigativo, por lo que no se ahondará en ello.

El punto es que el uso de agentes de IA generativa es ahora regular en todos los ámbitos de la sociedad, incluyendo el ámbito judicial y la profesión jurídica, debido a su amplio espectro de acción para realizar todo tipo de tareas útiles y necesarias en la administración de justicia, pues logran crear contenido original en diferentes formatos: texto, voz, imagen, video, código de software, entre otros (IBM, 2026).

Y aunque pareciera que este tipo de agentes como ChatGPT 3.5, la herramienta utilizada por el juez de segunda instancia del caso concreto, lograron desarrollar funciones cognitivas superiores de los seres humanos, lo cierto es que “estos modelos funcionan identificando y codificando los patrones y las relaciones en enormes cantidades de datos, y utilizando después esa información para entender las peticiones o preguntas en lenguaje natural de los usuarios y responder con nuevos contenidos relevantes” (IBM, 2026).

Resulta entonces que toda esta familia de agentes inteligentes y de humanos artificiales realmente no tienen la capacidad de comprender y ejercer las facultades cognitivas del cerebro humano, si no que las simulan tan bien que logran engañar al usuario, utilizando millones de datos

previamente seleccionados, con los que fueron entrenados para generar respuestas coherentes para su interlocutor.

En Colombia, a diferencia de otros países, todavía no existen jueces robots ni sistemas de decisiones judiciales automatizadas que coadyuven con la descongestión de los juzgados, tribunales y altas Cortes del país.

Al juez todavía debe asistirle la condición de ser humano para que pueda impartir justicia, de acuerdo con nuestro ordenamiento jurídico (Corte Constitucional, 2024). En otras palabras, debe ser una persona natural, de “carne y hueso”, con capacidades interpretativas y críticas, facultada por la Ley y la Constitución para ejercer con autoridad y legitimidad dicha figura. A esto se refiere la garantía del llamado juez natural.

Así entonces, el primer elemento clave para entender la decisión de la Corte y en general el papel que puede tener la IA en la Rama Judicial, es la prohibición legal de que una providencia sea emitida por un ente distinto a una persona física y real, así que, de entrada, los colombianos no podemos ser juzgados por una herramienta de IA generativa u otra tecnología emergente, pues habría que reformar la Carta magna y la ley.

Además, la Corte estudia y destaca principalmente, para dar respuesta al problema jurídico, que esta garantía del juez natural no sólo protege el derecho fundamental al debido proceso, sino todos los cimientos sobre los cuales se erige todo el sistema de normas colombiano, y que legitiman el poder judicial del Estado: imparcialidad, neutralidad, seguridad jurídica e independencia en la administración de justicia (Corte Constitucional, 2024).

Entonces, la figura del juez resulta ser insustituible y sería ilegal e inconstitucional que éste pudiese delegar en una IA su capacidad de interpretación de la Ley, de valoración probatoria, y de toda aquella función cognitiva que interviene en el proceso mental que implica la toma de

decisiones judiciales. E incluso, traspasando estos límites, si permitiera el juez que un elemento externo, como la IA, pudiese siquiera influir en su decisión final, también incurriría en vulneraciones normativas y de garantías, por lo que los operadores de justicia no sólo deben conocer muy bien sus deberes sino la fuente de su autoridad pues, sin importar la complejidad del asunto, la figura del juez natural o humano, en Colombia, no puede ser reemplazada, salpicada o personificada por una herramienta de IA generativa.

Por su parte, la motivación de sentencias y autos hace parte del derecho a la jurisdicción o al libre acceso a la administración de justicia, el cual se refiere a esa facultad que tienen todos los ciudadanos y ciudadanas de llevar a ante un juez u autoridad administrativa, su conflicto jurídico, y la obligación de éstas a conocer el caso y ejercer un juzgamiento en cumplimiento de los mandatos constitucionales y la búsqueda de la justicia.

Pero este derecho no implica únicamente el poder ser parte en un proceso judicial. La administración de justicia significa obtener decisiones justas, poder impugnarlas, si no se está conforme con ellas, y que éstas estén debidamente motivadas, con una gran carga argumentativa y coherencia jurídica, evitando así que las providencias de los jueces sean una *black box* y brindando una atención jurídica transparente.

Y es que el respeto por el derecho fundamental al debido proceso se manifiesta a través de la motivación de cada providencia puesto que allí es donde se verifica realmente que la decisión del juez obedece únicamente a la aplicación estricta de la Ley y las demás fuentes del Derecho, es decir, que la decisión es imparcial y libre de cualquier interés corrupto.

La motivación, para los jueces, es un deber máximo e imperante. El escenario para explicar razonadamente su decisión a modo de radiografía total, sin posibilidad de omitirse y faltar a lo que es también un derecho de las partes.

Es en esta sección de la providencia en donde se examina y demuestra el cumplimiento de la norma por parte del juez y la validez de la decisión, en tanto que, toda aquella sentencia o auto proferido que no cumpla con esta garantía procesal deberá ser impugnada o anulada, en su defecto.

Es aquí en donde se advirtió que el juez de segunda instancia había hecho uso de ChatGPT 3.5, al que le había realizado unas preguntas respecto al cobro de cuotas moderadoras y copagos a un paciente, menor de edad, en situación de vulnerabilidad, cuyas respuestas fueron utilizadas por el togado para robustecer la argumentación jurídica – motivación – de su decisión final.

No obstante, el *ad quem* lo hizo una vez había tomado, anunciado y justificado dicha decisión. Así lo expuso ante la Corte y así lo evidenció esta, por lo que no se consideró usurpada la función jurisdiccional del juez en esta ocasión.

Por todo lo anterior, no puede delegarse en la IA la figura del juez, pues le es imposible cumplir con la obligatoriedad de la motivación judicial de las providencias, ya que esto implica tener que hacer un robusto análisis probatorio y una correcta interpretación de la norma, sin depender de, en este caso, macrodatos, que además de históricos y desactualizados, contienen sesgos que la IA replica sin detenerse a examinar críticamente dicha información, como fue confirmado por la misma propietaria de ChatGPT, OpenAI , en su intervención realizada para la sentencia.

Así entonces, una providencia emitida o incluso levemente influenciada por una IA, sería falsamente razonada y escasamente argumentada, lo que derivaría en su invalidez jurídica, impugnación o anulación, sin mencionar que ello conllevaría grandes problemas a nivel social, una ruptura de la seguridad jurídica y la confianza de los ciudadanos en el sistema judicial y en el mismo Estado.

Colofón de lo expuesto, de las consideraciones de la Corte se puede extraer que está prohibido, por inconstitucional e ilegal, que una herramienta de IA generativa, como ChatGPT 3.5., juzgue y decida sobre los conflictos jurídicos que emergen de la vida en sociedad y la dinámica de las comunidades modernas, pues, en primer lugar, no tiene la capacidad de juzgar con imparcialidad, autonomía e independencia y, segundo, ello vulneraría los derechos fundamentales de las partes, el interés general, el ordenamiento jurídico y las máximas constitucionales.

En consecuencia, toda aquella actividad propia de la labor judicial, que no influya siquiera levemente en el convencimiento del juez para tomar una decisión con efectos jurídicos, así como en la sustanciación y motivación de la providencia que la contiene, puede ser realizada con apoyo de la IA, específicamente, herramientas, sistemas, algoritmos, agentes u otro producto semejante impulsado con inteligencia artificial generativa (GenAI).

6.3. Decisión de la Corte Constitucional en la Sentencia T-323 de 2024

Conforme lo manifestado por el juez de segunda instancia del caso en concreto, que resolvió la acción de tutela presentada por Blanca en nombre de su hijo menor de edad y en contra de su EPS, aquel utilizó la herramienta de IA generativa de OpenAI en su modelo del año 2023, ChatGPT 3.5., después de haber formado su convencimiento del caso, de haber tomado la decisión final y elaborado su respectiva argumentación y motivación, por lo que, la IA en ningún momento influyó en su decisión ni delegó en ella sus facultades jurisdiccionales.

Esta fue la tesis que reconoció la Corte como cierta y por ello confirmó la decisión, señalando que “no se configuró una causal de invalidez de lo actuado, pues el uso de ChatGPT en el caso sometido a estudio no comportó una usurpación de la función de administración de justicia a cargo de la autoridad judicial competente” (Corte Constitucional, 2024).

Por otro lado, la Sala exhortó a los jueces de todo el país a adquirir una cultura ética del uso de IA generativa en sus despachos y sus providencias, sin que se sustituya la racionalidad humana en la toma de las decisiones que constitucionalmente están facultados para realizar.

Asimismo, enunció los siguientes principios que guiarán dicha implementación de la IA y que deberán aplicar todos los funcionarios y empleados de la Rama Judicial: (i) transparencia, (ii) responsabilidad, (iii) privacidad, (iv) no sustitución de la racionalidad humana, (v) seriedad y verificación, (vi) prevención de riesgos, (vii) igualdad y equidad, (viii) control humano, (ix) regulación ética, (x) adecuación a buenas prácticas y estándares colectivos, (xi) seguimiento continuo y adaptación y (xii) idoneidad. (Corte Constitucional, 2024)

Finalmente, libró órdenes al Consejo Superior de la Judicatura para la creación y divulgación de un manual o guía para el uso de la IA generativa en la administración de justicia, y aplicando los lineamientos establecidos en la sentencia analizada, la T-323 de 2024, tarea con la que cumplió la entidad al proferir el Acuerdo PCSJA24-12243 el 16 de diciembre de 2024. Y, en un trabajo mancomunado con la Escuela Judicial Rodrigo Lara Bonilla, le ordenó divulgar la sentencia en escenarios de capacitación y formación en IA generativa para las labores judiciales, con el fin de mitigar al mínimo cualquier riesgo potencial, pero promoviendo la transformación tecnológica de la Rama Judicial.

Para concluir, en este capítulo se conoció la postura de la Corte Constitucional respecto al uso de herramientas de IA generativa por parte de los jueces colombianos en sus providencias: Promueven su uso pero desde un rol estricta y exclusivamente asistencial, de gestión o apoyo, nunca para decidir, interpretar la norma o realizar análisis probatorio, pues los jueces deben conservar su imparcialidad y criterio propio para juzgar, y las decisiones tomadas en el marco de un proceso judicial, deben obedecer a la racionalidad humana de la autoridad facultada para ello.

Por otra parte, y a modo de reflexión e invitación a seguir estudiando desde la perspectiva ética y jurídica el uso de la IA generativa en las labores de justicia, quedan en el aire interrogantes por responder como Si se utiliza la IA para apoyar la motivación jurídica de las providencias, como lo hizo el juez objeto de investigación de la sentencia analizada, ¿se le estaría reconociendo a la IA la calidad de perito o experto, al requerir “su opinión” de un caso en particular para usarla en una providencia judicial? o incluso, ¿se estaría tratando a la IA como una fuente del Derecho o de autoridad al realizarle preguntas y utilizar las respuestas para motivar la decisión del juez?

Muchos dilemas continúan aún después del extenso desarrollo jurisprudencial realizado por la Corte Constitucional en su sentencia hito T-323 de 2024, demostrando que involucrar a la inteligencia artificial en el ámbito judicial es de los retos más complejos, contingentes e ineludibles que tiene que asumir el Derecho en esta era de transformaciones tecnológicas que no da espera.

7. Propuestas prácticas para el uso de Herramientas de Inteligencia Artificial Generativa en las Providencias de los Jueces en Colombia y el Ejercicio del Derecho

A este punto de la investigación, resulta más que obvio que Colombia se sumó a los países que están transformando sus sistemas judiciales. Ya puede afirmarse que los jueces y sus despachos se valen de herramientas de inteligencia artificial para realizar sus labores judiciales.

Para finales del 2024, cerca de 1.000 despachos y 172 jueces del país manifestaron utilizar herramientas de IA, según la encuesta “Experiencias de Inteligencia Artificial en la Rama Judicial”, realizada por el Consejo Superior de la Judicatura y publicada el 21 de febrero del 2025, siendo ChatGPT la herramienta más recurrente. No obstante, para ese momento, el mayor porcentaje de los encuestados (70.81%) respondieron no hacer uso de ninguna herramienta de IA para su actividad laboral.

Seguramente en la actualidad, año 2026, el panorama ha cambiado un poco, pues de la fase exploratoria de este trabajo investigativo, se pudo evidenciar que la opinión mayoritaria sobre el uso de la IA en las labores judiciales es positiva y favorable a este.

De otro lado, según la encuesta del Consejo Superior de la Judicatura, los cinco usos más comunes que le dan los servidores judiciales a la IA son: consulta de información general, resumen y simplificación de textos, consulta de información jurídica (legislación, jurisprudencia y doctrina), corrección y redacción de textos y apoyo en el proceso creativo de argumentación, lo cual coincide con los hallazgos de las entrevistas realizadas por la suscrita investigadora.

Vale destacar que algunos de los mencionados usos implican actos repetitivos y mecánicos que demandan más tiempo que trabajo cognitivo, por lo que se considera seguro su uso. Por ejemplo, las consultas de información jurídica de fuentes legítimas como la jurisprudencia. Este

es uno de los usos más compartidos entre los servidores judiciales y los profesionales del Derecho cuyo campo de acción es el litigio, puesto que existe una necesidad real en acortar tiempo y esfuerzos por encontrar jurisprudencia relevante para el tema del caso concreto. De ahí que se utilicen agentes de IA generativa para “depurar” jurisprudencia y ahorrar recursos en la búsqueda exhaustiva que esta tarea implica.

Curiosamente es justo una de las labores en las que la IA se equivoca más y resulta ser insuficiente, puesto que, en primer lugar, hay que tener conocimientos previos en ingeniería de *prompts* para dar las indicaciones correctas al agente y dar con el resultado esperado; en segundo lugar, depende de la disponibilidad de jurisprudencia en la web y la potencia del modelo de IA para procesar documentos extensos con una alta complejidad de comprensión. Analizar una sentencia no es como leer una noticia corta y decir fácilmente de qué trata. Y, en tercer lugar, depende del perfilamiento que se le haga al agente de IA para que éste pueda responder como un especialista en el tema y tener mayores probabilidades de acierto.

Entonces, en este tipo de usos la IA suele alucinar en demasía, referenciar sentencias falsas o responder con providencias reales, pero con citas inexistentes. Justo este hecho fue evidenciado por la Corte Suprema de Justicia, sentencia STC17832-2025, en la que resolvió una acción de tutela interpuesta por una ciudadana en contra de una sentencia del Tribunal Superior de Sincelejo, en la que este resolvió un recurso de apelación que decretó la terminación de un proceso ejecutivo por desistimiento tácito, basando dicha decisión única y exclusivamente en citas y conclusiones inexistentes de dos sentencias de la Corte Suprema de Justicia: STC13560-2023 y STC4734-2025 que, si bien eran reales, las citas eran inexistentes e incluso, las conclusiones a las que el Tribunal refirió haber llegado la Corte en dichas oportunidades, eran totalmente diferentes a las allí plasmadas. Entonces, la Sala Civil, Agraria y Rural de la Corte Suprema de Justicia, advirtió un

uso irresponsable de la IA por parte de, nada menos que, un magistrado del Tribunal Superior de Distrito Judicial. Aunque la Corte no realizó un análisis profundo de este error gravísimo del Tribunal, sí se logró demostrar el riesgo grave que entraña el utilizar la IA en una decisión judicial sin tener los conocimientos previos y necesarios, tanto de la tecnología como de la regulación ética, para valerse de estas herramientas. Además, el Tribunal omitió informar que haría uso de una IA y mencionar el nombre de ésta.

Habiendo introducido este capítulo final, lo que aquí se busca es condensar las recomendaciones, advertencias, lineamientos éticos, normas vigentes y experiencias adquiridas para el uso de herramientas de IA generativa, por parte de los jueces colombianos, en sus decisiones, y en general por parte de la comunidad judicial, en una guía práctica y sencilla que puede resultar orientadora tanto para dichos funcionarios como para los demás profesionales del Derecho.

8.1.¿Cómo utilizar Ética y Responsablemente Herramientas de IA Generativa?

En primer lugar, hay que advertir que antes de hacer uso de cualquier agente de IA generativa, en adelante GenAI, el usuario o usuaria debe tener conocimientos previos y suficientes en la herramienta tanto para comprender cómo funciona, como para advertir cualquier error y riesgo potencial y saber cómo utilizar correctamente el contenido generado.

Sobre la historia, conceptualización y funcionamiento en general de la IA se trató el capítulo primero de esta investigación, así como en el capítulo segundo se expusieron las recomendaciones éticas internacionales, nacionales y normas vigentes sobre IA en Colombia (Acuerdo PCSJA24-12243 del 16/12/2024 y la sentencia T-323 de 2024).

Aparte de los agentes GenAI más conocidos, ahora existen algunos que ya están especializados por temas o profesiones. En realidad, cualquier agente GenAI puede perfilarse según la necesidad del usuario. En todo caso, se recomienda leer los términos y condiciones de la herramienta, en el sitio web oficial de este o de la empresa a la que pertenezca.

En el caso de los jueces y funcionarios judiciales, de momento, el único agente de GenAI que pueden utilizar para su actividad judicial es Copilot, proveída por Microsoft Corporation. Sin embargo, la recomendación aplica para cualquier otra herramienta, aunque sea diseñada y desarrollada de la Rama Judicial, nunca será 100% confiable y de esto se debe ser muy consciente.

Una vez superado el analfabetismo GenAI, el funcionario judicial está listo para aprovechar los beneficios.

8.2. Necesidad e idoneidad de la herramienta

Tanto la Corte Constitucional como el Consejo Superior de la Judicatura, han dispuesto que el uso de GenAI debe responder a una necesidad real o falencia del sistema judicial. Su implementación debe estar justificada para evitar una posible afectación innecesaria al sistema o a los derechos de las partes, por ejemplo, cuando se utiliza IA generativa para transcribir una audiencia, ya que esto responde a la necesidad de optimizar labores y de procesos con duración razonable lo cual, en últimas, contribuye a la descongestión judicial.

Por otro lado, la idoneidad se relaciona con la experticia de la herramienta para alcanzar el objetivo y que por ello sea considerada, por sobre otras herramientas y métodos, como la mejor opción.

Pueden ser útiles para definir este asunto, preguntas como ¿es una tarea mecánica y repetitiva o de análisis y valoración crítica? ¿Es una solución a un problema real o una mejora a

una solución que ya funciona? Si x o y tarea pudiera hacerla una IA ¿seguiría siendo útil y eficiente la manera en que hasta ahora se ha hecho? ¿Tengo claro el porqué de hacerlo con IA y no de la forma tradicional? ¿Cuál es mi objetivo con ello?

8.3. Incidencia en el Convencimiento y Decisión del Juez

Entre los usos prohibidos para los agentes de GenAI en el ecosistema judicial, está el delegarle a esta la decisión de un caso puesto que, como ya se dijo, esa función es exclusivamente del juez quien, además, es insustituible.

Con esto no sólo me refiero a preguntarle directamente al agente qué decisión debe tomar o “¿qué haría si...?” y a continuación narrar los hechos. También puede haber una usurpación indirecta de esta facultad con usos aparentemente técnicos, propios de la administración o gestión judicial del proceso. Entonces, es importante adquirir conciencia sobre la obligatoria imparcialidad que debe tener el juez para decidir y la facilidad con la que puede verse afectada, para poder reconocer aquellas consultas o *prompts* que puedan incidir en su decisión final, si quiera mínimamente, lo que en el lenguaje jurídico coloquial se denomina “contaminar” al juez.

Y es que la conducta recta y ética del juez no es garantía exclusiva de su imparcialidad y neutralidad, pues es un ser humano y tiene una manera de pensar previamente forjada por su entorno. Por ende, como remedio a esta tendencia natural del juez, cuando él u otro funcionario de su despacho utilice GenAI, debe reflexionarse si esa tarea (i) se relaciona o impacta el análisis probatorio o fáctico que desde el inicio del proceso está realizando el juez y (ii) si con la consulta realizada al agente de GenAI, éste estaría respondiendo el problema jurídico del caso, aun parcialmente. En caso de que la respuesta sea positiva, es recomendable utilizar otras maneras para realizar la tarea.

8.4. Autoformación

Independientemente del trabajo formativo que impulse el Consejo Superior de la Judicatura a través de la Escuela Rodrigo Lara Bonilla, o de las iniciativas propias de cada despacho en formar a sus funcionarios en habilidades tecnológicas y de GenAI, cada persona, de manera individual, es responsable de ser competente para manejar herramientas de este tipo, especialmente cuando su uso pueda impactar a terceros.

El deber de formarse en GenAI no se logra con consumir información o cursos de “IA para abogados” publicitados en redes sociales, sin ningún criterio diferenciador, pues no hay una manera diferente de utilizar la IA para cada profesión existente. La gran mayoría de herramientas GenAI funcionan de la misma manera: con *prompts*.

En ese sentido, la clave para saber manejar cualquier algoritmo es conocerlo, estudiarlo y aprendiendo a darle instrucciones acordes a nuestras expectativas y al cómo funciona la herramienta. Esto puede lograrse primero, exponiéndose constantemente a contenido y espacios que hablen del tema y, segundo, estudiando ingeniería de *prompts*; una nueva subrama de la IA que se especializa en diseñar instrucciones efectivas para modelos de GenAI, ya que la calidad del *prompt* creado es directamente proporcional a la calidad de los resultados. Se recomienda tener un “banco de *prompts*” personal o colectivo, para mayor eficiencia.

Así, poco a poco va surgiendo una nueva habilidad que deben aprender los profesionales, semejante a manejar Excel, hablar otro idioma o tener habilidades blandas. Así, en el futuro será una aptitud más por adquirir para ingresar al mercado laboral.

8.5. Protección de Datos Personales y Responsabilidad de la Información

Antes de dar cualquier información o documento al agente de GenAI, se debe eliminar cualquier dato personal, semiprivado o confidencial para cumplir con la Ley 1581 de 2012 y demás normas vigentes.

Por su parte, hasta este momento, los usuarios (humanos) siguen siendo responsables de los resultados que arroje el agente, pues jurídicamente son considerados objetos. De manera que, ante cualquier vulneración ocasionada, el usuario judicial de IA es uno de los legitimados a responder ante cualquier acción legal que se presente.

Para esta tarea, puede ser útil guardar los documentos con datos personales y sin datos personales en carpetas independientes, y así no mezclar unos con otros.

8.6. Verificación Permanente de la Información e Identificación de Sesgos y Alucinaciones Algorítmicas

Uno de los no negociables de la GenAI en entornos judiciales es la imprescindibilidad de la revisión y verificación de la información de forma permanente. Para ello debe partirse de que el agente no es una fuente de autoridad por múltiples razones, entre esas, las siguientes: (i) Previamente fue entrenada con macrodatos que no se actualizan automáticamente y que, (ii) contienen sesgos algorítmicos sin que el algoritmo tenga la capacidad de reconocerlos pues realmente no “comprende” nada y (iii) tiende a ser “complaciente” con el usuario.

Por ejemplo, si un usuario de IA le refuta un argumento dado por un agente, éste último le dará credibilidad a lo que dice aquel, aun cuando el argumento generado por la IA fuese verdad. En consecuencia, la IA no tiene criterio ni juicio propio para ser capaz de corregir los prejuicios o discriminaciones en las que incurra.

Entonces, por más lógica y coherente que parezca la respuesta de GenAI, siempre se debe revisar que no contenga sesgos algorítmicos, verificar tanto la veracidad de la información como la existencia de la fuente citada y que efectivamente repose allí dicha información. Este deber no significa que el agente de GenAI no sea útil, significa que hay que tener claros los roles: GenAI es un apoyo que asiste al funcionario judicial, y este es el que da órdenes, supervisa y tiene el control.

Se recomienda forjar una desconfianza “sana” con la IA para no dejar de disfrutar de sus beneficios, pero tampoco soportar las consecuencias de sus errores. Saber discernir entre lo cierto y lo falso, hace parte de la racionalidad humana que es irremplazable.

8.7. El Deber de Información (Principio de Transparencia)

Este es uno de los principios en donde hay más confusiones acerca de su cumplimiento. Algunos consideran que no importa para qué fue utilizada la GenAI sino que fue utilizada y con eso es suficiente para tener que informar ese hecho. Otras personas, en cambio, difieren de esto y consideran que sólo se debe informar el uso de la IA cuando se parafrasea o transcribe un argumento que es un constructo directo del agente.

El objetivo de este principio y deber de información es lograr un sistema judicial transparente, contrarrestar la deshumanización de la justicia y garantizar el derecho de información y de contradicción que todos como usuarios de un servicio tenemos. El principio de transparencia también significa utilizar solamente agentes de “caja abierta”, es decir, que permitan su revisión en caso de que se necesite, lo cual se relaciona con el principio de explicabilidad, aunque con salvedades grandes para el caso de Colombia, ya que éste establece que los funcionarios judiciales sólo deben hacer uso de herramientas de GenAI que permita la auditabilidad y conocer el paso a paso para llegar a x o y decisión. No obstante, que un agente de GenAI sea explicable no depende

de los usuarios judiciales o ni siquiera de la Rama Judicial, como autoridad regulatoria, puesto que, para cumplir con este deber, habría que intervenir desde el diseño del agente, lo cual, por ahora, no es posible en nuestro contexto ya que todavía Colombia no ha desarrollado agentes de GenAI y por ende no puede asegurar la explicabilidad de ellos.

Por lo tanto, la comunidad judicial utilizará las herramientas de IA que la Rama Judicial elija y permita para el ejercicio de la justicia.

8.8. Promoción de una Cultura Ética y de Buenas Prácticas en IA

El conocimiento se debe compartir y aún más cuando se persiguen objetivos tan esenciales para una sociedad.

La implementación ética de agentes de GenAI para las providencias de los jueces como para otras labores, responderá a qué postura adopta cada despacho o entorno judicial frente al tema, si está dispuesto al cambio o reticente a este.

En el primer caso, fomentar buenas prácticas, como las que aquí se proponen, dependerá de todos los cargos y niveles jerárquicos. Fomentar buenas prácticas en IA significa facilitar espacios de aprendizaje colectivo, compartir experiencias, valerse de herramientas gráficas o visuales para tener claro cómo crear un *prompt* acertado, estar en constante actualización del medio y adaptarse a una nueva forma de justicia.

Ya para concluir este largo y fructífero camino de investigación, es válido reflexionar sobre ¿cómo sistemas tan complejos pueden atravesar transformaciones inminentes con los menores daños colaterales posibles? Esto solo puede llegar a buen término desde la ética y el Derecho.

La constante es el cambio y la adaptación una habilidad que tienen los sistemas flexibles y seguros. La garantía del juez natural es la garantía de la justicia y la garantía de la sociedad, de lo

que hasta ahora ha funcionado para mantener la legitimidad de un gobierno y la paz en los territorios, evitando la autodestrucción social.

La voluntad de un juez, contenida en una sentencia judicial, es legítima porque así lo establece nuestro ordenamiento jurídico el cual ha sido creado por representante directos de un pueblo soberano que habita un Estado Social de Derecho. Por ello, con la promesa de resolver problemas del sistema judicial, no puede sacrificarse a una autoridad milenaria y las formas de justicia de una comunidad que hasta ahora han funcionado, con dificultades pero que se ha logrado mantener en el tiempo.

Entonces, tanto con la inteligencia artificial como con cualquier futuro avance tecnológico, no se debe caer en modas o ceder a las presiones de un mundo globalizado. Si bien es importante estar dispuestos al cambio y a la evolución de las sociedades, hay que tener siempre el juicio crítico activado. Hay que cuestionarse la confiabilidad del invento y su idoneidad, puesto que, podría llegarse a conclusiones contrarias a las que promete aquel.

Si se estudia más a fondo el asunto, tal vez se encuentren mayores variables negativas que positivas, o a pensar que, si la IA se vende como un optimizador de procesos y simplificador de actividades, con toda la responsabilidad que conlleva el hacer uso de ella y las “arandelas” para utilizarla, puede suceder que sea menos complejo y más eficaz continuar con las formas tradicionales de justicia, o en su mayoría.

De manera que nada debe darse por sentado ni cierto. Cada ser racional tiene las capacidades mentales para determinar su actuar. En lo que respecta a la Rama Judicial y sus jueces, hace tiempo aceptaron el cambio y están mudando a nuevas formas de justicia, pero siempre desde nuestra humanidad, con juicio crítico y voluntades independientes.

9. Conclusiones

Después de todos los contenidos abordados en esta investigación, valiosos hallazgos y reflexiones se logran concluir:

- La inteligencia artificial (IA) no es una ciencia o disciplina totalmente independiente. Utiliza y depende de otras disciplinas para alcanzar sus objetivos propuestos.
- La IA entraña múltiples riesgos y desafíos para las sociedades modernas, lo cual hace necesario el estudio de aquella desde la ética y las ciencias humanas: la antropología, la filosofía, la psicología, el Derecho, entre otras.
- Existen grupos poblacionales que exigen de una mayor protección del Estado y regulación jurídica, frente al uso de la IA, debido a la condición especial que les ha sido reconocida como los menores de edad, personas con alguna discapacidad, en estado de indefensión o vulnerabilidad, así como debido a dinámicas sociales e históricas que han rodeado a algunas minorías como la población LGBTQ+, las mujeres, la población afro y etnias estigmatizadas, culturas o religiones menos visibilizadas, entre otras.
- El sistema judicial colombiano está transformando las formas tradicionales de operar y hacer justicia. En la actualidad, los despachos judiciales a nivel nacional están utilizando herramientas de inteligencia artificial generativa para ejecutar sus labores.
- Los referentes internacionales sobre la regulación ética y jurídica de la IA son importantes, pero no reemplazan la regulación de cada país ni suplen las necesidades locales de legislación ni los vacíos normativos. Y aunque los principios éticos generales sean la ayuda urgente que se requiere ante acontecimientos como este, no tienen los efectos

jurídicos ni la especificidad que se requiere. Salir de lo abstracto para sumergirse en lo específico es la meta del ordenamiento jurídico y el órgano legislativo colombiano.

- Los agentes de IA generativa son sólo una herramienta de apoyo para el usuario judicial, con mucho potencial y marcadas diferencias respecto de otras herramientas tecnológicas, pero que no pueden reemplazar el juicio, raciocinio y voluntad humana. Mucho menos, la labor del juez y el sistema judicial en general.

- Aun teniendo sendas ventajas y utilidades, el uso de IA generativa en el ámbito judicial plantea un riesgo profundamente peligroso: deshumanizar la justicia. Por ello, debe protegerse la función jurisdiccional del juez y su figura de autoridad como único responsable de administrar justicia, limitando a estas herramientas a un rol asistencial.

- Es imperativo que la comunidad judicial promueva la alfabetización tecnológica en IA. Corresponde a las autoridades judiciales, como a cada uno de los funcionarios de cada despacho judicial y profesionales del Derecho que hacen uso de estas herramientas, contribuir con la autoformación y el aprendizaje continuo y permanente de la disciplina en general, especializándose en el agente o herramienta elegida para los fines de la labor judicial. Este deber se traduce en un manejo responsable, seguro y ético de la IA.

- Los modelos de IA generativa tienen múltiples y graves falencias. Entre las más importantes están los sesgos algorítmicos y alucinaciones, lo que ocasiona que la verificación humana sea irremplazable e imprescindible. No ser consciente de estas falencias de disciplina, o serlo y omitir dicha exigencia ética y de seguridad, es un uso irresponsable de ella, que tiene consecuencias para el usuario judicial infractor de la norma. No es la herramienta ni la empresa dueña de ésta, la obligada a responder por el contenido

que un usuario judicial generó a través del agente de GenAI, con excepción a algunos casos especiales.

- En armonía con la anterior conclusión, es prioritario y fundamental para el Estado, erigir todo un marco regulatorio robusto para la inteligencia artificial en Colombia. Desde el desarrollo del régimen de responsabilidad civil y penal hasta la expedición de una ley estatutaria de la inteligencia artificial.

- La transparencia es innegociable. Todo el contenido generado con IA debe ser etiquetado como tal. Es una obligación de los funcionarios judiciales como de cualquier usuario de IA. La ciudadanía tiene derecho a conocer cuando se ha hecho uso de esta tecnología, dentro del proceso judicial al que esté involucrado, con el fin de ejercer debidamente sus derechos. Este mandato tiene doble sentido: como garantía procesal y como deber ético.

- La privacidad e información personal son ahora dos bienes jurídicos protegidos por la norma. Ningún funcionario judicial puede exponer dicha información de las partes o terceros durante el curso de una causa procesal, así como fuera de ella.

- La eficiencia del sistema judicial, y demás beneficios de utilizar IA en la administración de justicia, no puede prevalecer por sobre otros derechos y garantías constitucionales. De lo contrario, no se estaría frente a sociedades más justas y humanas sino a sociedades sistematizadas y programadas para la productividad .

- La actitud crítica frente al uso de herramientas o agentes de IA generativa y en general frente a la disciplina, debe ser una cultura práctica dentro de la comunidad judicial, toda vez que se evidenció que la IA no tiene comprensión real sobre las palabras y los significados que las sociedades damos a los diferentes elementos de nuestro entorno.

Se debe promover una desconfianza “sana” de la tecnología, que supone analizar primero la idoneidad de la herramienta, la fiabilidad de los resultados generados por ella, cuestionar sus limitaciones y contrastar la información antes de asumirla como válida.

- Los productos de IA consumen grandes cantidades de los recursos naturales del planeta, como, por ejemplo, el agua, ocasionando impactos negativos para este y acelerando las problemáticas actuales. Por ende, hace parte del uso ético de la IA, perseguir medidas medioambientales que protejan los ecosistemas y recursos naturales, asegurando la sostenibilidad de la IA.

Es oportuno mencionar en esta sección que el panorama previo a esta investigación sobre el uso de la IA en las providencias judiciales era bastante distante del que en este punto he logrado vislumbrar como autora del proyecto. Y en esto, influyó bastante no solo la historia y regulación internacional de la IA sino también la perspectiva real de la comunidad judicial bumanguesa y colombiana, a través de la jurisprudencia analizada y citada a lo largo del documento.

Por otro lado, muchas reflexiones pueden extraerse del exhaustivo trabajo que ha dejado este proceso investigativo. A mi criterio, de las más relevantes es la importancia mantener siempre un juicio crítico con todo lo relacionado con inteligencia artificial y tecnología, como profesionales del Derecho y ciudadanos del mundo digital, ya que nadie está exento de los efectos adversos conocidos y desconocidos de las disciplinas modernas como la IA.

Con todo lo que ha surgido a la luz, hasta ahora, sobre tal vez la creación humana más inquietante, y la experiencia práctica que ha dejado hace ya varios años su uso cotidiano, es natural preguntarse si realmente la IA simplifica la vida del ser humano y cumple todo lo que promete, o si, por el contrario, la hace más compleja y difícil de superar sus riesgos, puesto que al no ser confiable totalmente y exigir la supervisión humana permanente, ¿no resulta más tedioso su uso,

al demandar más tiempo que las formas tradicionales de justicia verificando que su trabajo como asistente esté bien hecho? O si el cambio a ese escenario nuevo de justicia ¿es realmente el camino correcto para nuestra especie?

En medio de un mundo sofocante y abarrotado de miles y miles de novedades todos los días, imposibles de conocer completamente, vivimos en el afán de tener que hacer parte de la transformación y evolución tecnológica sin detenernos a juzgar con imparcialidad y crítica si la verdadera superación de la humanidad está allí, a costa de los valores superiores como justicia, dignidad y ética, o si la verdadera evolución está en saber reconocer el “cuando sí y cuando no” de la IA, o de alcanzar el equilibrio entre ésta y la humanidad, que permita una convivencia sana y cooperativa entre sí.

Referencias

- Asimov, I. (1950). *I, Robot* [Yo, Robot]. Gnome Press. (Originalmente publicado como el relato “Runaround” en 1942).
- BBC News. (2023). *Replika: The chatbot that some people are treating like a partner*. <https://www.bbc.com/news/business-64536261>
- Bender, E. M., & Gebru, T. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bender, E. M., & Koller, A. (2020). *Climbing towards natural language understanding: On meaning, form, and understanding in the age of data*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5185–5198. <https://aclanthology.org/2020.acl-main.463.pdf>
- Boole, G. (1854). *An investigation of the laws of thought*. Walton and Maberly.
- British Broadcasting Corporation. (2016). *¿Qué es la inteligencia artificial?* <https://www.bbc.com/mundo/noticias-37679463>
- British Broadcasting Corporation. (2019). *Inteligencia artificial: riesgos y dilemas éticos*. <https://www.bbc.com/mundo/noticias-45727287>
- Cloud Google. (s.f.). *¿Qué es el machine learning?* <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419>
- Coeckelbergh, M. (2021). *Ética de la inteligencia artificial* (E. Díez de Castro, Trad.). Editorial Cátedra.

Código Civil [CC]. Ley 57 de 1887. Art. 1613 y 2341. 15 de abril de 1887 (Colombia).

<http://bit.ly/3Xyz123>

Colmenares Uribe, C. (2012). El rol del juez en el Estado democrático y social de derecho y justicia. *Revista de la Facultad de Derecho y Ciencias Políticas*, (12), 65-81.

Comisión Europea. (2018). *Directrices éticas para una IA fiable*. Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Consejo de Europa. (2024). *Framework convention on artificial intelligence*. <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>

Consejo de Europa. (2024). *Convenio Marco del Consejo de Europa sobre la inteligencia artificial y los derechos humanos, la democracia y el Estado de derecho*. Art. 8, párr. 59. <https://search.coe.int/>

Consejo de Europa. (2024). *Convenio Marco del Consejo de Europa sobre la inteligencia artificial y los derechos humanos, la democracia y el Estado de derecho*. <https://search.coe.int/>

Consejo Superior de la Judicatura. (2024, 16 de Diciembre). *Acuerdo PCSJA24-12243: Por el cual se adoptan lineamientos para el uso y aprovechamiento respetuoso, responsable, seguro y ético de la inteligencia artificial en la Rama Judicial*. <https://www.ramajudicial.gov.co>

Corte Constitucional de Colombia, Sala Segunda de Revisión. (2024). *Sentencia T-323 de 2024*

[M. P. Juan Carlos Cortés González].

<https://www.corteconstitucional.gov.co/relatoria/2024/t-323-24.htm>

- Corte Suprema de Justicia. Sala de Casación Civil, Agraria y Rural. (2025, 5 de noviembre). Sentencia *STC17832-2025* (M. P. Adriana Consuelo López Martínez; Radicado 11001-02-03-000-2025-05001-00).
- Dignum, V. (2019). *Responsible artificial intelligence: From principles to practice*. Springer.
- El Espectador. (2024, 25 de mayo). *Los artículos que despublicamos: así se detectó el uso indebido de IA y las fuentes falsas*. <https://www.elespectador.com/actualidad/los-articulos-que-despublicamos-asi-se-detecto-el-uso-indebido-de-ia-y-las-fuentes-falsas/>
- Escudero, C. (2020). *El análisis temático como herramienta de investigación en el área de la comunicación social: contribuciones y limitaciones*. *La Trama de la Comunicación*, 24(2), 89–100. <https://www.redalyc.org/journal/3239/323964237005/>
- European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 on artificial intelligence (AI Act)*. Official Journal of the European Union, L 202. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Google. (2025). *Gémini (versión del 17 de septiembre)* [Modelo de lenguaje de gran tamaño]. <https://gemini.google.com>
- Google. (2025). *Respuesta generada a partir del prompt: “Diagrama de la función básica de un perceptrón”* [Gémini, modelo 1.5 pro] <https://gemini.google.com>
- IBM. (s.f.). *AI ethics*. <https://www.ibm.com/mx-es/think/topics/ai-ethics>
- IBM. (s.f.). *Generative AI*. <https://www.ibm.com/es-es/think/topics/generative-ai>
- IBM. (s.f.). *History of artificial intelligence*. <https://www.ibm.com/es-es/think/topics/history-of-artificial-intelligence>
- IBM. (n.d.-a). *¿Qué es el deep learning?* Retrieved January 21, 2026, from <https://www.ibm.com/es-es/think/topics/deep-learning>

- IBM. (n.d.-b). *¿Qué es la ética de la IA?* Retrieved January 21, 2026, from <https://www.ibm.com/mx-es/think/topics/ai-ethics>
- IBM. (n.d.-c). *What are large language models (LLMs)?* Retrieved January 21, 2026, from <https://www.ibm.com/think/topics/large-language-models>
- Jobin, A., Ienca, M., y Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kaku, M. (2008). *Physics of the impossible*. Doubleday.
- Kapuściński, M. (2025, 20 de Noviembre). *GPT-5 Training Data: Evolution, Sources, and Ethical Concerns*. TTMS. <https://ttms.com/gpt-5-training-data-evolution-sources-and-ethical-concerns/>
- Kelly, S. D. (2011). *All things shining: Reading the Western classics to find meaning in a secular age*. Free Press.
- Museo Nacional de Ciencias Naturales. (s.f.). *Biografía de Ada Lovelace*. Consejo Superior de Investigaciones Científicas. <https://www.mncn.csic.es/es/visita-el-mncn/biografia-de-ada-lovelace>
- Newell, A., & Simon, H. A. (1976). *Computer science as empirical inquiry: Symbols and search*. *Communications of the ACM*, 19(3), 113–126.
- New York State Unified Court System. (2024). *Guidelines on the Use of Artificial Intelligence*. <https://www.nycourts.gov/LegacyPDFS/a.i.-policy.pdf>
- New York State Unified Court System. (2025). *Interim policy on the use of artificial intelligence*. <https://www.nycourts.gov/LegacyPDFS/a.i.-policy.pdf>
- Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.

- Nuño, J. C. (2019). *Inteligencia artificial: ¿Un paso adelante en la evolución?* Bonallettera Alcompas; Prisanoticias Colecciones.
- OpenAI. (2025). *ChatGPT (versión GPT-5.2)* [Modelo de lenguaje grande]. <https://openai.com>
- OpenAI. (2025, noviembre). *Respuesta sobre el consumo de agua de agentes de IA generativa* [ChatGPT]. <https://chatgpt.com/>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2025). *Directrices para el uso de sistemas de inteligencia artificial en tribunales y juzgados*. UNESCO.
- Presidencia de la República de Colombia. (2021). *Marco ético para la inteligencia artificial en Colombia*. Departamento Administrativo de la Presidencia de la República (DAPRE). https://estrategia.gobiernoonlinea.gov.co/623/articles-195036_recurso_1.pdf
- Real Academia Española. (s.f.). *Tecnología*. En *Diccionario de la lengua española*. <https://dle.rae.es/tecnología>
- REPÚBLICA DE COLOMBIA. (2024). *Sentencia T-323 de 2024*. <https://www.itm.nrw/wp-content/uploads/2024/08/kolubienai.pdf>
- Roa Avella, M. del P., Sanabria-Moyano, J. E., y Dinas-Hurtado, L. K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal*, 8(1), 275-314. <https://doi.org/10.22197/rbdpp.v8i1.613>
- Roose, K. (2024). *¿Se puede culpar a la IA del suicidio de un adolescente?* <https://www.infobae.com/america/the-new-york-times/2024/10/25/se-puede-culpar-a-la-ia-del-suicidio-de-un-adolescente/>
- Searle, J. R. (1980). *Minds, brains, and programs*. *Behavioral and Brain Sciences*, 3(3), 417–457.
- Searle, J. R. (1984). *Minds, brains, and science*. Harvard University Press.

Turing, A. M. (1936). *On computable numbers, with an application to the Entscheidungsproblem*.

Proceedings of the London Mathematical Society, 42(2), 230–265.

Turing, A. M. (1950). *Computing machinery and intelligence*. *Mind*, 59(236), 433–460.

Universidad Externado de Colombia. (2024). *Lineamientos para el uso de inteligencia artificial*.

<https://www.uexternado.edu.co>

Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in an age of machine thinking*.

Oxford University Press.

White House. (2025, December 11). *Ensuring a national policy framework for artificial*

intelligence. <https://www.whitehouse.gov>

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*.

MIT Press.

**Apéndice A. Tabla de los Principales Avances Teóricos en Inteligencia Artificial, por
Diferentes Autores, Situados Entre el Siglo XIX y XX**

Principales aportes en inteligencia artificial como disciplina	
<p>George Boole (1815-1864)</p> <p>Matemático y lógico Británico.</p> <p>Obra: <i>An Investigation of the Laws of Thought</i> 1854</p>	<p>Creó la lógica matemática la cual teorizó a través del álgebra Booleana. Este sistema contiene los fundamentos de la aritmética computacional, base de la informática, la computación y la electrónica digital. Este sistema estructura la información con valores binarios, 0 (falso) y 1 (verdadero), usando operaciones lógicas como AND, OR y NOT, las cuales se explican así:</p> <ul style="list-style-type: none"> • AND (&): Devuelve verdadero solo si ambos valores son verdaderos (1 y 1). • OR (): Devuelve verdadero si al menos uno de los valores es verdadero (1 y 0). • NOT (!): Invierte el valor de un único operando, convirtiendo verdadero a falso y viceversa (1=falso y 0=verdadero). • Este sistema fue fundamental para representar y simplificar circuitos lógicos y sistemas de toma de decisiones. Por esto, Boole es considerado como uno de los fundadores de las ciencias de la computación y su álgebra se utiliza aún, especialmente, resolviendo problemas digitales y en el campo de la programación.
<p>Alan Mathison Turing (1912-1954)</p> <p>Lógico matemático y criptógrafo británico</p> <p>Obras: <i>On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society.</i> 1936</p> <p><i>Proposed Electronic Calculator</i> 1945</p>	<p>Este prócer académico inventó un modelo lógico, basado en los avances precedentes de varios autores, para construir una máquina capaz de resolver un problema formulado, aplicando reglas anteriormente dadas, y generando el resultado esperado. La máquina funcionaba así:</p> <ul style="list-style-type: none"> • Una cinta infinita con símbolos encasillados, una cabeza lectora y una tabla con reglas para cada símbolo. • Según el símbolo leído en la casilla, escribía un nuevo símbolo, de acuerdo a la regla respectiva, y mueve el cabezal a la izquierda o derecha. • Este proceso se repite hasta llegar a una regla que indique detenerse. • Arroja el resultado final de la cinta. <p>Este se convirtió en el modelo lógico fundamental de la computación moderna, pues permitió descubrir cálculos computacionales universales que podía replicarse para hacer otras máquinas que resolvieran múltiples problemas utilizando únicamente ese modelo.</p>

Principales aportes en inteligencia artificial como disciplina

Computing Machinery and Intelligence. 1950

Por esto, la máquina de Turing es considerada el primer ordenador.

A través de este modelo, Alonzo Church³⁵, al igual que Turing, demostró que la idea del *Entscheidungsproblem*³⁶ era imposible y, por tanto, no podía decidirse si una proposición matemática era demostrable o no.

Adicionalmente, en el artículo publicado en 1945, “*Proposed Electronic Calculator*”, Turing propuso un modelo de calculadora electrónica a través del que desarrolló el concepto del control lógico³⁷, estudiado también por Neumann.

Ambos concibieron una computadora electrónica como un mecanismo de control que actuaba en forma de tándem, un elemento tras otro, con una memoria en la que se almacenaban las instrucciones. Por ello, en 1950 Turing diseñó la ACE³⁸

Otro de sus grandes aportes fue el test de Turing, el cual permite determinar la inteligencia de una máquina, si, con sus respuestas, lograba confundir o parecer humana.

John William Mauchly
(1907-1980)

Estos dos científicos pioneros de la informática construyeron, físicamente, el primer ordenador de propósito general y totalmente digital, en 1943, materializando las teorías y arquitecturas de otros autores precedentes, como John Von Neumann. Su función principal era calcular trayectorias balísticas, pues fueron contratados por el ejército estadounidense durante la Segunda Guerra Mundial.

John Presper Eckert
(1919-1995)

Físico e ingeniero electrónico estadounidense

Esta máquina era increíblemente grande y pesada. Sus cálculos eran veloces y funcionaba con tubos de vacío, a modo de interruptor electrónico, para que los circuitos ejecutaran las operaciones lógicas.

Obra:
Crearon la ENIAC y la EDVAC³⁹

Uno de los problemas más relevantes de la máquina, era que, para ejecutar cada tarea, debían reconfigurarla manualmente. A pesar

³⁵ Alonzo Church (1903-1995), estadounidense, Doctor en matemáticas y pionero en la ciencia informática, pues sus estudios y aportes se centraron en teorías algorítmicas que fuesen computables.

³⁶ *Entscheidungsproblem* o problema de decisión, fue un ideal en lógica matemática, que, de existir, permitiría determinar si una proposición era verdadera y demostrable. También, resolvería cualquier problema matemático, y, por tanto, cualquier problema en general pues “*el mundo está escrito en caracteres matemáticos*” (Schönfinkel, s.f.).

³⁷ Este concepto se aborda más a detalle en los aportes de John Von Neumann.

³⁸ Máquina o motor de Cálculo Automático (ACE): Uno de los primeros diseños de computadora electrónica con programa de almacenamiento de memoria.

³⁹ ENIAC - Computador e Integrador Numérico Electrónico y EDVAC - Calculador Discreto Electrónico Automático Variable.

Principales aportes en inteligencia artificial como disciplina

	de ello, esto impulsó avances en lenguajes de programación y automatización de procesos.
	Posteriormente, y en pro de mejorar las deficiencias de la ENIAC, Mauchly y Eckert, en 1949, finalizaron la construcción de la EDVAC; una computadora mejorada, que pasó de tener cables y tarjetas perforadas, a utilizar el sistema binario y una memoria de almacenamiento de datos.
	De ahí en adelante, la empresa que juntos conformaron, se dedicó a crear otros modelos de computadoras para su comercialización.
Norbert Wiener (1894-1964)	Fue el fundador de la cibernética pues fue en su obra en donde se utilizó por primera vez este término.
Matemático estadounidense	Allí estudió cómo los sistemas de control de seres vivos y seres artificiales (máquinas/robots) reciben, procesan y utilizan la información para autorregularse en su ambiente.
Obra: <i>Cybernetics: Or Control and Communication in the Animal and the Machine</i> 1948	Esto significó un gran aporte para la retroalimentación (feedback) de la IA y de la robótica, lo cual es uno de los diferenciadores que tiene la IA de otras tecnologías. Esa capacidad de “aprender” de los errores es lo que justifica su inteligencia. Esto fue inspiración, posteriormente, para hablar del <i>machine learning</i> ⁴⁰ , o aprendizaje automático en español, en el campo de la IA.
Warren McCulloch (1898-1969)	McCulloch y Pits crearon el primer modelo matemático basado en una estructura cerebral humana: las redes neuronales.
Walter Pits (1923-1969)	Este modelo de Red Neuronal Artificial (RNA) significó un fundamental aporte para la disciplina artificial pero también para la ciencia computacional, pues lograron “ <i>describir el sistema nervioso humano en términos de lógica</i> ” (Neumann, s.f.) y demostraron cómo ensamblar circuitos lógicos matemáticos en neuronas simplificadas.
Neurocientífico y lógico matemático estadounidenses	
Obra: <i>A logical Calculus of the Ideas Immanent in Nervous Activity</i> 1943	Gracias a este aporte, el psicólogo estadounidense Frank Rosenblatt diseñó el Perceptrón: La estructura más simple de una red neuronal artificial (RNA).

⁴⁰ El *machine learning* es una rama fundamental para la IA que entrena a las computadoras para tomar decisiones basadas en la experiencia, tal y como lo haría un humano (Zhou, 2016). Esta área se abordará mejor más adelante.

Principales aportes en inteligencia artificial como disciplina

John Von Neumann (1903-1957) Al igual que Turing, Neumann consideraba que la lógica matemática podía ir *más allá* y tener una aplicación práctica: la computadora electrónica.

Matemático y físico Húngaro-americano Turing y Neumann se conocieron en 1935 en Cambridge, como profesores, cuando éste último fue a dar unas conferencias académicas.

Obra:

First Draft of a Report on the EDVAC 1945 Posteriormente, Turing fue profesor en la Universidad de Princeton en New Jersey, U.S.A., y allí continuaron su relación académica.

Turing leía las publicaciones científicas de Neumann y compartían comunicaciones en las que debatían sus hipótesis matemáticas, principalmente, sus ideas acerca del *Entscheidungsproblem*, el reto matemático principal de la época. Pero luego, ambos se comenzaron a interesar en otros temas como la computación e informática.

Algunos históricos consideran que la influencia de las ideas de Neumann, ayudaron a Turing a desarrollar su máquina, pues, terminada la Segunda Guerra Mundial, Turing conoció el “*Primer Borrador*” de una computadora práctica y universal, de Neumann, lo que hoy es una Unidad Central de Procesamiento (CPU). No obstante, la idea de Turing era muy diferente.

Neumann logró llegar muy cerca a lo que es hoy el funcionamiento de la Inteligencia Artificial, pues diseñó una máquina universal (con los aportes de Turing) ensamblada a partir de las neuronas de McCulloch y Pitts.

Esta conexión le permitió avances en lo que se denomina el “*control lógico*”, en otras palabras, las órdenes o reglas que controlan la computadora. Estas órdenes son ejecutadas por el CC (Control Central), o estimuladas por él para lograr su ejecución.

Claude Shannon (1916-2001) **Elwood** Shannon fue uno de los organizadores de la Conferencia del Dartmouth College en 1956, considerada el nacimiento oficial de la IA.

Ingeniero eléctrico, y criptógrafo matemático estadounidense A través de su obra, en 1948, Shannon presentó la fundamentación de la teoría de la información: puede ser transmitida a través de canales de comunicación. Aunque esto parezca básico para nosotros, para la época fue un gran hito.

Obra:

En su teoría, desarrolló el concepto de *bit*, la unidad más pequeña de información y a través de la cual los ordenadores almacenan y

Principales aportes en inteligencia artificial como disciplina

<p><i>A mathematical Theory of Communication</i> 1948</p>	<p>procesan la misma, término estrechamente relacionado con la computación y la IA.</p>
<p>Allen Newell (1927-1992) Herbert A. Simon (1916-2001)</p>	<p>Al igual que Neumann y Turing, Shannon desarrolló estudios en lógica y circuitos digitales.</p> <p>Newell y Simon eran unos de los escépticos acerca del nombre de la Inteligencia Artificial. Asistieron a la Conferencia del Dartmouth College y propusieron que la disciplina tomara el nombre de “Simulación de Procesos Cognitivos” o “Procesamiento Complejo de la Información”.</p>
<p>Científicos informáticos estadounidenses</p>	<p>Por otro lado, sus aportes se centraron en el procesamiento de la información y en la simulación del pensamiento humano, desde una perspectiva lógica pero también psicológica, comparándolo con el funcionamiento de un computador.</p>
<p>Obra: <i>The Logic Theory Machine: A Complex Information Processing System.</i> IRE <i>Transactions on Information Theory</i> 1956</p>	<p>El mismo año de la conferencia, a través de su artículo de 1956, presentaron un programa llamado “El Teórico de la Lógica” que podía resolver problemas complejos como demostrar teoremas de Russell y Whitehead en lógica proposicional.</p> <p>Aunque desarrollaron una computadora para probar sus teorías, todavía las investigaciones se situaban en lógica simbólica y no práctica, por lo que esta idea no logró materializarse.</p>
<p>Jhon McCarthy (1927-2011)</p> <p>Científico computacional estadounidense</p>	<p>Además de organizar la Conferencia de verano del Dartmouth College, junto a Claude Shanon, Marvin Minsky y otros académicos, McCarthy bautizó a la disciplina en su nacimiento, con el término <i>Inteligencia Artificial</i>.</p>
<p>Obra: <i>A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.</i> 1956</p>	<p>Siendo profesor del Instituto Tecnológico de Massachusetts (MIT – en Cambridge), creó el primer lenguaje de programación aplicado a la IA: LISP (procesamiento de listas).</p> <p>El LISP fue uno de los fundamentos para varios de los lenguajes de programación modernos y, hoy en día, es toda una familia de lenguajes.</p> <p>Funciona con listas o cadenas de datos conformadas por símbolos. Estas listas representan la información, los datos del mundo en general, y es por esto que es perfecto para la IA, por su flexibilidad.</p> <p>El LISP es fácil de ejecutar y funciona para la mayoría de los programas, lo cual aún mantiene al lenguaje en uso.</p>

Principales aportes en inteligencia artificial como disciplina

Frank Rosenblatt
(1928-1971)

Psicólogo
estadounidense

Obra:
Perceptrón simple
1957

Basado en el modelo matemático de McCulloch y Pits, más de una década atrás, Rosenblatt creó el Perceptrón simple, el algoritmo-modelo más simple de una neurona biológica que consiguió materializarse.

En 1957, se implementó en el software del primer ordenador comercial de la historia: el IBM704 de la multinacional *International Business Machines*.

Constaba de varias entradas, en donde recibía los datos, las cuales estaban asociadas a unos nodos o pesos que determinaban un valor numérico, según los parámetros establecidos, se efectuaba la suma, ésta generaba un valor ponderado que llegaba a la función de activación y si pasaba el umbral se activaba generando el resultado que siempre era binario (1 o 0) y si no lograba ser mayor que el umbral, la función no se activaba y el mismo Perceptrón ajustaba sus “pesos” hasta conseguir el resultado deseado.

Por ello, el Perceptrón introdujo el *machine learning*. Este era de tipo “supervisado”, lo que significa que tenía datos ya etiquetados como correctos y se asociaban a un resultado específico y siendo el único correcto.

Identificaba patrones y se los aprendía para situaciones desconocidas, pero análogas, a las que pudiera reflejar similares resultados, es decir, era un modelo de tipo predictivo. Sin embargo, sus predicciones eran lineales y cuando se intentó entrenar con patrones diferentes, no logró reconocerlos.

Aunado a ello, en 1969, Minsky y Papert criticaron las limitaciones del Perceptrón para resolver problemas complejos debido a que constaba de una sola capa de neuronas, cantidad insuficiente para tareas de gran dificultad e incluso afirmaron que era un despropósito pensar en redes multicapa pues el objetivo era inalcanzable.

Esta postura afianzó el desinterés general que se venía gestando en estudiar las RNA multicapa, lo que contribuyó al primer invierno de la IA, un periodo de pocos avances científicos y estancamiento generalizado.

Apéndice B. Principios y Lineamientos Referentes para la Regulación Ética de la Inteligencia Artificial a nivel Internacional y de Colombia

1. Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho (2024) elaborado por el Comité de Inteligencia Artificial (CAI) del Consejo de Europa.

El Convenio Marco del Consejo de Europa sobre Inteligencia Artificial establece una serie de principios y obligaciones fundamentales diseñados para asegurar que el desarrollo y uso de la IA respete los derechos humanos, la democracia y el Estado de derecho.

A continuación, se detallan los principios fundamentales y lineamientos extraídos del documento:

Obligaciones generales

- Protección de los derechos humanos (Art. 4).
- Integridad de los procesos democráticos (Art. 5).

Principios del ciclo de vida de los sistemas de IA

- Dignidad humana y autonomía individual (Art. 7).
- Transparencia y supervisión (Art. 8).
- Rendición de cuentas y responsabilidad (Art. 9).
- Igualdad y no discriminación (Art. 10).
- Privacidad y protección de datos (Art. 11).
- Fiabilidad (Art. 12).
- Innovación segura (Art. 13).

Otros lineamientos clave

- Gestión de riesgos (Art. 16).
- Remedios y salvaguardas (Art. 14 y 15).
- Alfabetización digital (Art. 20).

2. Directrices para el Uso de Sistemas de IA en Tribunales y Juzgados, Documento Emitido el 03 de diciembre de 2025 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO).

En este documento la UNESCO establece el primer marco ético y operativo global para integrar la inteligencia artificial en los sistemas judiciales de manera responsable y progresiva.

La UNESCO establece directrices en torno a quince principios fundamentales para guiar la adopción y supervisión de la IA en la rama judicial de cada país, los cuales son: (i) Protección de los DDHH, (ii) proporcionalidad, (iii) viabilidad de beneficios, (iv) seguridad, (v) seguridad de la información, (vi) exactitud y fiabilidad, (vii) explicabilidad, (viii) auditabilidad, (ix) transparencia y justicia abierta, (x) conciencia y uso informado, (xi) responsabilidad, (xii) rendición de cuentas y contradicción, (xiii) supervisión humana, (xiv) diseño centrado en el ser humano y (xv) gobernanza colaborativa.

Adicionalmente, el documento divide sus recomendaciones en acciones concretas para las autoridades judiciales (Cortes, Tribunales, etc.) y para los individuos u operadores judiciales (jueces, servidores y auxiliares de justicia) que a continuación se enuncian.

Para las Organizaciones Judiciales: (i) Evaluación de impacto, (ii) implementación global, (iii) capacitación de talento humano y (iv) gestión de IA generativa.

Para Miembros Individuales de la Rama Judicial: (i) Pensamiento crítico permanente, (ii) transparencia en el proceso, (iii) integridad y autoría y (iv) confidencialidad.

3. Marco Ético de la Inteligencia Artificial en Colombia

Este documento fue publicado en mayo del año 2021, cuya elaboración estuvo a cargo de la Presidencia de Colombia en colaboración con las siguientes tres personas: Armando Guío Español (Consultor CAF), Elena Tamayo Uribe (Contratista Presidencia de la República) y Pablo Gómez Ayerbe (Asesor Consejería Presidencial de Asuntos Económicos y Transformación Digital).

En este marco se recopilan recomendaciones éticas y sugerencias para las entidades públicas a nivel nacional, incluyendo algunas prácticas internacionales como referencia para Colombia, con el fin de apoyar la formulación y gestión de los proyectos en donde se pretenda incluir el uso de la inteligencia artificial (IA).

También hicieron parte de la elaboración de este documento, entidades como la CAF⁴¹, el BID⁴² y el Berkman Klein Center for Internet & Society⁴³.

Los principios establecidos en este documento son: (i) transparencia y explicabilidad, (ii) privacidad, (iii) seguridad, (iv) responsabilidad, (v) no discriminación, (vi) inclusión, (vii) prevalencia de los NNA, (viii) beneficencia y (ix) autonomía humana.

⁴¹ Banco de Desarrollo de América Latina y el Caribe.

⁴² Banco Interamericano de Desarrollo.

⁴³ Berkman Klein Center for Internet & Society es un centro de investigación de la Universidad de Harvard que se enfoca en el estudio del ciberespacio.

4. Acuerdo PCSJA24-12243 del Consejo Superior de la Judicatura

El acuerdo PCSJA24-12243 fue emitido el 16 de diciembre de 2024 por el Consejo Superior de la Judicatura, en cumplimiento de la orden dada por la Corte Constitucional en la sentencia T-323 de 2024, y con él se busca adoptar reglas y lineamientos más concretos sobre el uso ético, responsable y seguro de herramientas de inteligencia artificial (IA) en la Rama Judicial colombiana.

El acuerdo establece los siguientes usos permitidos de herramientas de IA por parte de los funcionarios judiciales, clasificados en 3 categorías:

- **De gestión administrativa o soporte:** para optimizar procesos administrativos y facilitar tareas de apoyo a la gestión judicial.
- **De revisión minuciosa:** para verificar y revisar el contenido de un documento antes de ser utilizado y/o emitido.
- **De estándar ético:** para aquellas actividades en las que se exija cumplir con deberes de transparencia, responsabilidad y protección de la privacidad.

Por otra parte, establece las siguientes prohibiciones:

- Hacer pasar como suyos textos creados con IA.
- Utilizar agentes u otras herramientas de IA en su versión gratuita.
- Adjuntar documentos o revelar información personal y privada a la herramienta de IA, para evitar que ésta se apropie de dichos datos y sean utilizados para entrenamiento de IAs, violando la privacidad de su propietario.

Finalmente, los siguientes son los principios y lineamientos que establece este Acuerdo en su artículo 3°: (i) Primacía de los derechos fundamentales, (ii) regulación ética, (iii) adecuación a buenas prácticas, (iv) seguimiento y adaptación continua, (v) gobernanza colaborativa, (vi) no sustitución de la racionalidad humana, (vii) idoneidad, (viii) responsabilidad y uso informado, (ix) supervisión y verificación humana, (x) prevención de riesgos, (xi) igualdad y equidad, (xii) protección de datos y privacidad, (xiii) seguridad de la información, (xiv) explicabilidad, (xv) transparencia.

Apéndice C. Mapa Temático sobre el Uso de Herramientas de Inteligencia Artificial Generativa en la Administración de Justicia en Colombia



Nota. Este mapa temático es la representación gráfica de los resultados obtenidos, analizados y codificados en la fase exploratoria de la investigación, a través de las entrevistas realizadas a diez (10) participantes pertenecientes a la comunidad judicial desde los cargos de magistrados, jueces, demás funcionarios judiciales, abogados litigantes y docentes de Derecho. Elaboración propia generada desde ATLAS.ti 25.01.32924