

DESAFÍOS ÉTICOS EN LA IMPLEMENTACIÓN DE INTELIGENCIA ARTIFICIAL GENERATIVA EN LA JUSTICIA: SEGSOS DE DATOS E INJUSTICIA ALGORÍTMICA

ETHICAL CHALLENGES IN THE IMPLEMENTATION OF GENERATIVE ARTIFICIAL INTELLIGENCE IN JUSTICE: DATA BIASES AND ALGORITHMIC INJUSTICE

JOSEPH COLMENARES RAMÍREZ *

RESUMEN: Este artículo se propone evaluar los desafíos éticos que plantea la implementación de inteligencia artificial generativa (IAG) en la justicia, frente a los riesgos de injusticia algorítmica y sesgos de datos en su uso. Para responder a ello, el análisis se centró en dos aspectos principales: i) las implicaciones de un sistema judicial en el que se utilice IAG y, a partir de allí; ii) determinar los lineamientos éticos que rijan su funcionamiento en la labor judicial. Con metodología documental-descriptiva, se exploró el alcance de esta nueva tecnología como parte de la modernización del Derecho en su aspecto más adjetivo. Se concluyó una serie de criterios orientadores que funcionarán como referencia para futuros proyectos de *lege ferenda* en los sistemas jurídico-procesales de Latinoamérica.

PALABRAS CLAVE: Inteligencia artificial generativa, Derecho Procesal, justicia, sistema judicial, algoritmos, sesgos.

ABSTRACT: This research article aims to evaluate the ethical challenges posed by the implementation of generative artificial intelligence (GAI) in justice, in view of the risks of algorithmic injustice and data bias in its use. In order to respond to this, the analysis focused on two main aspects: i) the implications of a judicial system in which GAI is used and, from there; ii) determine the ethical guidelines that govern its operation in judicial work. Thus, through a documentary-descriptive methodology, the scope of this new technology was explored as part of the modernization of law in its most adjective aspect. Finally, we concluded with a series of guiding criteria that will serve as a reference for future *lege ferenda* projects in the legal-procedural systems of Latin America.

KEYWORDS: Generative artificial intelligence, Procedural Law, justice, judicial system, algorithms, biases.

RESUMEN / ABSTRACT

* Estudiante de Quinto año de Derecho, Universidad Católica del Táchira, Venezuela. Miembro Investigador del Centro de Estudios de Derecho Procesal (CEDEPRO) y Miembro Estudiante del Instituto Colombiano de Derecho Procesal (ICDP). Correo electrónico: josephcol2002@gmail.com.

I.- INTRODUCCIÓN

En la actualidad se está experimentando lo que diversos autores han descrito como la “Cuarta Revolución Industrial”¹, una nueva era en la que tecnologías digitales, físicas y biológicas convergen en un mundo cada vez más globalizado. Mientras las máquinas se encargan de realizar las tareas más desagradables, monótonas o peligrosas del día a día, los humanos se encargan de supervisarlas y centrarse en sus labores más esenciales. Ello ha significado un cambio de paradigma para distintas áreas de la vida, entre ellas, el entretenimiento, el arte o el trabajo; pero debe plantearse su influencia en el Derecho, específicamente en el proceso judicial.

En virtud de ello, los países de la región latinoamericana han centrado sus esfuerzos en desarrollar reformas en las que destaca el uso de Tecnologías de la Información y la Comunicación (TIC), en favor de la modernización de la justicia. Esta visión se ha centrado en la búsqueda de eficacia y la eficiencia en el uso de los recursos destinados al funcionamiento del sistema judicial, destacándose el papel que puede cumplir la Inteligencia Artificial (IA) en la justicia. De tal manera, es necesario valorar las implicaciones regulatorias que tendría su implementación en las formas del proceso, así como los desafíos éticos en la automatización de ciertas labores en el Poder Judicial.

La IA es un tipo de tecnología capaz de emular procesos de la inteligencia humana, tales como la escritura, reconocimiento de imágenes y el análisis de datos, a través de un conjunto de algoritmos pre-entrenados para realizar dichas tareas. En particular, se destaca un subcampo de ésta, denominada Inteligencia Artificial Generativa (IAG), la cual se centra en la creación de ideas y contenidos nuevos a partir de ciertas órdenes denominadas *prompts*. De tal modo que permite componer audios, crear imágenes o videos e, incluso, escribir textos casi indistinguibles a una creación humana con tan solo pedirlo en una oración.

Tal capacidad de creación se fija como una atractiva respuesta al retardo judicial tan característico de los sistemas judiciales latinoamericanos, el cual se enfatiza en el retraso injustificado en la producción de decisiones judiciales. Por otro lado, ha de tomarse en cuenta cómo esta poderosa herramienta ha sido puesta en tela de juicio, por tanto, compone un medio bastante útil para la creación de contenidos, pero que han sido capaces de reflejar prejuicios sociales o respuestas irrespetuosas mediante sesgos de datos. De modo que hace cuestionar la viabilidad de su implementación en la labor judicial, además de las posibles violaciones que podría componer que un juez se inhiba de sus labores al automatizarlas con IA, afectándose la garantía al debido proceso y el principio del juez natural.

Por todo lo anteriormente dicho, es oportuno, entonces, cuestionarse como pregunta problema: ¿Cuáles son los desafíos éticos que enfrenta la implementación de IAG en la justicia, en consideración a los riesgos asociados a sesgos de datos e injusticia

¹ SCHWAB, Klaus (2016), *La cuarta revolución industrial*. Editorial Debate.

algorítmica? Para responder a ello, esta investigación se encargó de componer una serie de objetivos específicos capaces de aproximar una respuesta a dicho cuestionamiento, de tal modo que se propuso: i) definir qué es la IAG y cómo opera; ii) evaluar la implementación de la IAG en el ámbito judicial; y iii) determinar los lineamientos éticos necesarios para su implementación.

Para desarrollar estas ideas, la investigación contó con un apartado metodológico compuesto de un enfoque documental-cualitativo que, a través de la revisión exhaustiva de fuentes bibliográficas relevantes, empleó una propuesta jurídico-orientativa a la formulación de parámetros de aplicabilidad de estas nuevas tecnologías en lo jurisdiccional.

"Cualquier tecnología suficientemente avanzada es indistinguible de la magia"
— Arthur C. Clarke.

II.- LA INTELIGENCIA ARTIFICIAL

Esta tecnología ha transformado silenciosamente el mundo durante los últimos años, haciéndose presente en motores de búsqueda, redes sociales y tiendas *online* usados en el día a día. A pesar de sus beneficios, existe un común descontento debido a la posible automatización de la vida, junto a la amenaza del desempleo masivo, similar a lo ocurrido en la Revolución Industrial, y una consecuente degradación en la calidad educativa. Metafóricamente, este miedo recuerda a momentos históricos en los que la aparición de la escritura hizo temblar a la oralidad de epopeyas grandiosas en la Antigua Grecia; donde la imprenta hizo surgir una horda de detractores que preferían copiar a mano sus libros; y donde la llegada del internet hizo escandalizar a los amantes de los medios impresos.

Su irrupción ha de compararse con otros grandes descubrimientos a lo largo de la historia, tal como la fotografía, que permitió capturar momentos de manera más accesible, relegándose la necesidad de contratar a un pintor que interpretara la realidad en un lienzo. De manera similar, la IAG se ha consolidado como una herramienta informática que ofrece la posibilidad de crear contenidos de forma automática, a través de breves órdenes denominadas *prompts*.² No obstante, persisten preocupaciones significativas sobre los métodos empleados para generar estas creaciones y el impacto que podrían tener en el futuro.

² Puede ser definido como una instrucción o texto inicial que se le proporciona a una herramienta de IA generativa para guiar su generación de respuestas o resultados.

La Organización para la Cooperación y el Desarrollo Económico (OCDE), recientemente redactó una serie de Recomendaciones sobre Inteligencia Artificial, donde definió los sistemas de IA como:

“...una máquina que puede, de acuerdo con un conjunto de objetivos definidos por humanos, realizar predicciones, recomendaciones o tomar decisiones que tengan una influencia sobre ambientes reales o virtuales. Los sistemas de IA son diseñados para operar con distintos niveles de autonomía”.³

Conforme a lo anterior, en palabras más simples, se tiene a la IA como: “...la tecnología que permite a ordenadores y máquinas simular el aprendizaje, la comprensión, la resolución de problemas, la toma de decisiones, la creatividad y la autonomía humanas”⁴. Tales definiciones dejan avisar algunos rasgos acerca de la utilidad de esta herramienta. Para comprender mejor el poder detrás de esta novedosa tecnología es necesario ver más allá y distinguirla según sus capacidades, las cuales, por lo general, tienden a clasificarse en tres oleadas:

- a. La IA simbólica: consiste en programar las reglas lógicas para que un sistema pueda actuar de forma inteligente, como jugar al ajedrez o diagnosticar una enfermedad.
- b. La IA basada en datos: utiliza técnicas de aprendizaje automático para extraer patrones y conocimiento a partir de grandes cantidades de datos, y es la responsable del auge actual de la IA.
- c. La IA basada en contextos: es la base de la superinteligencia artificial, un hipotético sistema capaz de comprender el mundo real y de ofrecer soluciones innovadoras a problemas nuevos o desconocidos.⁵

Asimismo, como señala Franganillo (2023), se diferencian por el grado de complejidad y autonomía de la IA, que se sub-clasifica en IA específica o débil e IA general o fuerte⁶. En concreto, la IAG se identifica como un modelo de IA general o fuerte, la cual permite la producción automatizada de contenidos basados en texto, imagen, sonido y video de alta calidad a través de sus redes neuronales. Estos modelos, además, son capaces de asemejarse a una hipotética IA basada en contextos, ya que abarcan vastos

³ Recommendation of the Council on Artificial Intelligence. Organización para la Cooperación y el Desarrollo Económico, 2019. OECD Legal Instruments. Disponible en: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

⁴ STRYKER, Cole; KAVLAKOGLU, Eda. “What is artificial intelligence (AI)?”. 16 de agosto de 2024. International Business Machine Corporation (IBM). Disponible en: <https://www.ibm.com/topics/artificial-intelligence>

⁵ BOUCHER citado por FRANGANILLO, Jorge. “La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos”, *Methaodos revista de ciencias sociales*, 2023, 11(2). p. 2. Disponible en: <http://dx.doi.org/10.17502/mrcs.v11i2.710>

⁶ *Ibidem*. La IA específica o débil se trata de un sistema diseñado para realizar tareas específicas y predecibles. Por otro lado, la IA general o fuerte tiene capacidad de realizar tareas más allá de las que se han programado.

conjuntos de datos que les permite identificar patrones y crear nuevas obras basadas en ellos. De tal forma que, a medida que se vuelven más sofisticados estos sistemas, son capaces de generar contenido casi indistinguible del creado por humanos.

De acuerdo al modelo de IA que se trate, pueden emularse ciertas tareas que podría realizar un humano, la diferencia se encuentra en la eficiencia en realizarlas. Mientras que a una persona le tomaría una hora realizar un inventario, con tal solo proporcionarle los datos, la IA es capaz de reconocerlos, ordenarlos, clasificarlos y suministrarle un valor a cada uno de esos datos según las necesidades del mismo. No obstante, esta es solo una de las muchas funciones que tiene la IA, por lo que dependerá del modelo que lo realice.

2.1.- Modelo

Los modelos de IA son algoritmos diseñados para emular procesos del pensamiento, tales como la escritura, la creación de imágenes o audio, aprendizaje de datos, reconocimiento de patrones y toma de decisiones de manera automatizada. En pocas palabras, son distintas estructuras de código que permiten realizar ciertas labores a través de IA, como las redes neuronales profundas, máquinas de soporte vectorial, árboles de decisión, entre muchos otros. De tal modo, por ejemplo, las redes neuronales profundas (Deep Neural Networks, DNN) se caracterizan por ser una clase de algoritmos que imita el aprendizaje de un cerebro y, a través del cual los grandes bancos de datos con los que son entrenadas las IAs son procesados.⁷

Por otro lado, las máquinas de soporte vectorial (Support Vector Machines, SVM), tienen la principal función de tomar grandes cantidades de datos y separarlas, para así clasificarlas según distintos valores⁸. Asimismo, los *decision trees*, que se encargan de formular un esquema jerárquico que se alimenta de una cantidad de preguntas (nodos) que le permite realizar una predicción; una forma bastante sencilla de comprenderlo es a través del análisis de un caso sobre responsabilidad civil en un accidente de tránsito:

⁷ AOUICHAOU, A. et. al. "Comparison of Group-Contribution and Machine Learning-based Property Prediction Models with Uncertainty Quantification". *Computer Aided Chemical Engineering*. Elsevier, 2021, Vol.50. p. 755. ISBN 9780323885065. Disponible en: <https://doi.org/10.1016/B978-0-323-88506-5.50118-2>.

⁸ International Business Machine Corporation. "Funcionamiento de SVM". 17 de octubre de 2021 Disponible en: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>.

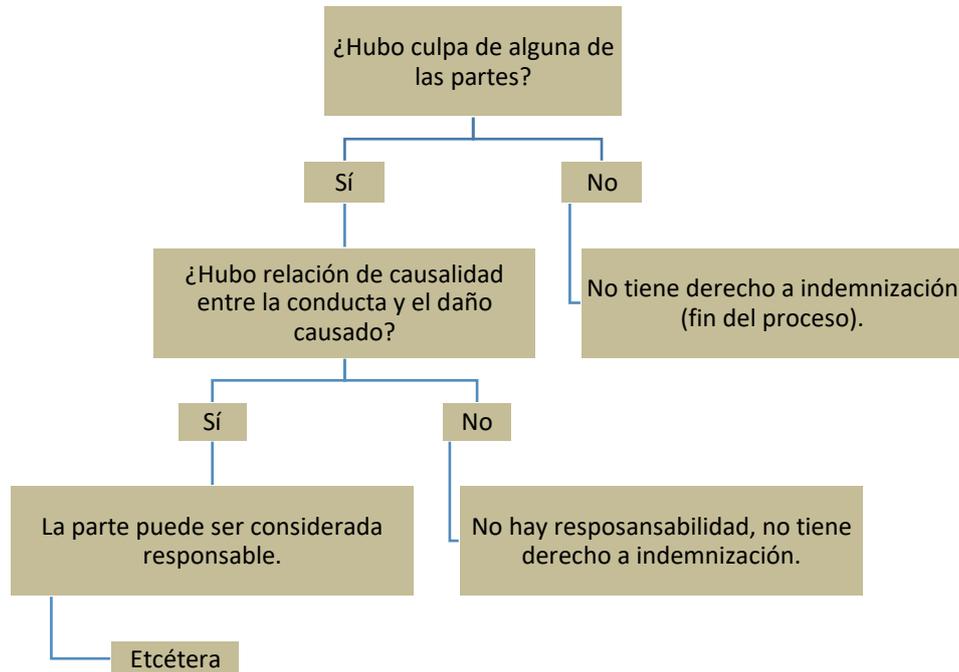


Tabla 1: Elaboración propia (2024).

Este último ejemplo es una versión bastante simplificada de la aplicación que podría dársele a los *decision trees* a través de algoritmos, ya que estos se componen de una cantidad mucho mayor de preguntas, a las cuales denomina “preguntas binarias”. El caso anterior podría ampliarse mucho más, haciendo preguntas como: “¿La parte culpable actuó con negligencia, imprudencia o incumplió una norma de tránsito?” o “¿La víctima contribuyó al accidente con su propio comportamiento?”. De esta forma funciona esta herramienta que, reforzada a través de algoritmos y aprendizaje automático, es capaz de tomar decisiones automatizadas bastante eficientes, al responder cada una de las preguntas con base a un conjunto de datos con los que ha sido preentrenada; en este caso, sería toda la legislación y jurisprudencia en materia de responsabilidad civil y tránsito.

La demostración anterior es uno de los casos que *prima facie* pueden pensarse como la aplicabilidad de estas nuevas herramientas en la justicia, pero es necesario entender cómo son concebidas éstas. Para ello, es importante identificar cuáles son los modelos de IAG que hacen posible la realización de estas tareas, entre ellos se pueden clasificar en tres grandes grupos según su utilidad: a. Modelos de Redes Adversarias Generativas (GAN); b. Modelos de Codificadores variacionales automáticos (VAE); y c. Modelos de Aprendizaje por refuerzo general (GRL).

2.1.1. Modelos GAN.

Los modelos GAN o adversariales se caracterizan por la interacción entre dos redes neuronales: una red generadora y una red discriminadora, las cuales compiten entre sí. En términos más técnicos Goodfellow et al. (2014) definen este modelo de la siguiente forma:

“En el marco de trabajo de redes adversarias propuesto, el modelo generativo se enfrenta a un adversario: un modelo discriminativo que aprende a determinar si una muestra pertenece a la distribución del modelo o a la distribución de los datos. El modelo generativo puede considerarse análogo a un equipo de falsificadores que intenta producir moneda falsa y utilizarla sin ser detectado, mientras que el modelo discriminativo es análogo a la policía, que intenta detectar la moneda falsificada. La competencia en este juego impulsa a ambos equipos a mejorar sus métodos hasta que las falsificaciones sean indistinguibles de los artículos genuinos”.⁹

Siguiendo el ejemplo propuesto por los autores, el modelo adversarial se encarga, a través de una red que genera y otra que discrimina, de buscar la forma más cercana a lo que se quiera generar con esta tecnología. El ejemplo más ilustrativo de este tipo de tecnología es su aplicación en la página web “<https://thispersondoesnotexist.com/>”, la cual es capaz de generar rostros bastante realistas a través de algoritmos y que utiliza una versión más sofisticada a la propuesta por Goodfellow et al. (2014), llamada StyleGAN2¹⁰, la cual fu desarrollada por ingenieros de Nvidia y Aalto University.

⁹ GOODFELLOW, Ian; POUGET-ABADIE, Jean. et al. “Generative Adversarial Nets”. *Neural Information Processing Systems*, 2014, p. 2. En su idioma original: “In the proposed adversarial nets framework, the generative model is pitted against an adversary: a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine article” (Traducción propia). Disponible en: <https://arxiv.org/abs/1406.2661>

¹⁰ KARRAS, Tero. et al. *Analyzing and Improving the Image Quality of StyleGAN*. 2019. Disponible en: <https://arxiv.org/abs/1912.04958>



Fuente: Karras et al. y Nvidia. StyleGAN2 (2019)¹¹.

Estos modelos no son sólo capaces de generar este tipo de rostros, sino además se han desarrollado más tipos, con los que pueden crearse ubicaciones, autos, mapas e incluso, caballos a través de StyleGAN2¹². Lo anterior ha generado preocupación, ya que en muchos casos estas creaciones son indistinguibles a la realidad para algunas personas y podría llegar a ser un conflicto frente a la generación de pruebas falsas en juicios por la facilidad con las que pueden generarse estas creaciones. Ello será uno de los puntos a abarcar cuando se trate el uso responsable de estas herramientas en lo judicial. Su utilidad en la justicia se enfoca en las ventajas que aportaría este modelo a la detección de este tipo de imágenes.

2.1.2. Modelos VAE.

Este modelo suele ser bastante asociado a los adversariales, por cuanto su mayor utilidad es la generación de imágenes a través de un proceso similar. Mientras en los GAN se encuentra una red generadora y una discriminadora, en los modelos VAE se halla un codificador y un decodificador, los cuales son utilizados para crear una red neuronal de aprendizaje no supervisado¹³. Este proceso, en palabras más simples, toma un *input* el cual es presentado al codificador para ser comprimido a una representación reducida, muestra que luego el decodificador tomará para realizar lo contrario, al tomar esa

¹¹ Imágenes generadas automáticamente a través de StyleGAN2. Disponible en: <https://thispersondoesnotexist.com>

¹² Los últimos años han aparecido gran cantidad de servicios que mediante StyleGAN2 y sistemas similares han logrado generar creaciones automáticas de distintos objetos y animales, estos pueden ser consultados en la página web de This X does not exist que recopila algunos de ellos. Disponible en: <https://thisxdoesnotexist.com>

¹³ CONTRERAS, Óscar. Modelos Generativos en el Aprendizaje Automático y su aplicación a la generación de Imágenes Digitales. *Journal Boliviano De Ciencias*, 2021, 17(51). p. 97. Disponible en: <https://revistas.univalle.edu/index.php/ciencias/article/view/110>

información reducida y generar nuevos datos a partir de ella¹⁴. Es por ello que se denomina “variacional”, ya que a través del *input* es capaz de reconstruir variaciones a partir del mismo, tal como en el ejemplo siguiente:



Fuente: Departamento de Ciencias de la Computación e Inteligencia Artificial.
Universidad de Sevilla (s.f.).¹⁵

Frente a los modelos GAN, la mayor limitación de estas creaciones está en que las imperfecciones son más comunes en los VAE. De todas formas, su aplicación es mucho más útil en el procesamiento de datos para decisiones probabilísticas, tal como predecir la generación de terremotos mediante la creación de simulaciones realísticas en su proceso de compresión y descompresión¹⁶. De ahí que uno de sus usos más comunes es la reconstrucción de fenómenos climáticos¹⁷ y en la justicia podría aportar una herramienta útil en la generación de variables de casos similares.

2.1.3. Modelos GRL.

Por último, los modelos de aprendizaje por refuerzo se encargan de mejorar la toma de decisiones mediante la interacción con un entorno, observar consecuencias y ajustar las acciones para ser más eficiente a través de la experiencia acumulada. Una explicación mucho más técnica y que se considera adecuada para aprender de qué se tratan los modelos de GRL es la propuesta por Sutton *et. al.* (2018):

¹⁴ *Ibidem*.

¹⁵ CAPARRINI, Fernando. “Variational AutoEncoder”. *Blog Matemática*. Universidad de Sevilla. Departamento de Ciencias de la Computación e Inteligencia Artificial. s.f. Disponible en: <https://www.cs.us.es/~fsancho/Blog/posts/VAE.md>

¹⁶ KINGMA, Diederik. *et al.* An Introduction to Variational Autoencoders. *Foundations and Trends R in Machine Learning*. Universiteit van Amsterdam. 2019. p. 3. Disponible en: <https://arxiv.org/pdf/1906.02691>

¹⁷ SZWARCMAN, Daniela *et. al.* “Quantizing reconstruction losses for improving weather data synthesis”. *Science Reports*. 2024, vol.14, 3396. Disponible en: <https://doi.org/10.1038/s41598-024-52773-2>

“El aprendizaje por refuerzo es diferente del aprendizaje supervisado, el tipo de aprendizaje estudiado en la mayoría de las investigaciones actuales en el campo del aprendizaje automático. El aprendizaje supervisado consiste en aprender a partir de un conjunto de ejemplos etiquetados proporcionados por un supervisor externo experto. Cada ejemplo es una descripción de una situación junto con una especificación -la etiqueta- de la acción correcta que el sistema debe realizar ante esa situación, que a menudo consiste en identificar una categoría a la que pertenece la situación. El objetivo de este tipo de aprendizaje es que el sistema extrapole, o generalice, sus respuestas para que actúe correctamente en situaciones no presentes en el conjunto de entrenamiento. Se trata de un tipo de aprendizaje importante, pero por sí solo no es adecuado para aprender de la interacción. En los problemas interactivos suele ser poco práctico obtener ejemplos del comportamiento deseado que sean a la vez correctos y representativos de todas las situaciones en las que tiene que actuar el agente”.¹⁸

El más claro ejemplo de este modelo es el utilizado en juegos como el ajedrez o el Go, donde esta tecnología se encarga de analizar cada una de las múltiples jugadas posibles para elegir la más conveniente. En el ámbito judicial, su utilidad yacería en la asistencia en la toma de decisiones, al ser entrenados con una serie de decisiones y jurisprudencia que les permita encontrar patrones y recomendar decisiones con base a ellos.

2.2.- ¿Cómo funcionan estas herramientas?

La IAG no tiene capacidad de “pensamiento”, pues, aunque su arquitectura pretenda simular una red neuronal de aprendizaje continuo, similar a la de un cerebro humano, la forma en que genera respuestas se ve determinada por un proceso de continuas predicciones a través de *tokens*¹⁹. Dichos *tokens* son calculados en función de

¹⁸ SUTTON, Richard; BARTO, Andrew. “Reinforcement Learning: An Introduction”. *The MIT Press*. Londres. 2014. p. 2. En su idioma original: “Reinforcement learning is different from supervised learning, the kind of learning studied in most current research in field of machine learning. Supervised learning is learning from a training set of labeled examples provided by a knowledgeable external supervisor. Each example is a description of a situation together with a specification—the label—of the correct action the system should take to that situation, which is often to identify a category to which the situation belongs. The object of this kind of learning is for the system to extrapolate, or generalize, its responses so that it acts correctly in situations not present in the training set. This is an important kind of learning, but alone it is not adequate for learning from interaction. In interactive problems it is often impractical to obtain examples of desired behavior that are both correct and representative of all the situations in which the agent has to act”. (Traducción propia). Disponible en: <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>.

¹⁹ En programación, es la forma de referirse a una pequeña unidad de código que compone los aspectos más básicos de un programa. Para quienes no manejan lenguajes de programación como Python o C, es más simple definirlo como una pequeña pieza que forma un rompecabezas, en este caso un programa. En el contexto particular de la IAG, un *token* representa una unidad básica de texto, como una palabra, un carácter o incluso solo signos de puntuación. Estos permiten descomponer el lenguaje humano en

datos e instrucciones con los que han sido entrenadas estas tecnologías²⁰. De esta forma funcionan los sistemas basados en GPT-4.0²¹, los cuales son incapaces de sustituir el razonamiento humano, pues la respuesta que ofrecen al *prompt* no es más que el cálculo de un algoritmo que toma una infinidad de datos para predecir la combinación correcta.

Ahora bien, para determinar estas predicciones, estos sistemas establecen una serie de valores para cada *token*, el cual puede ser simples caracteres, palabras o fragmentos de palabras²². Zhang *et. al.* (2022) establece que para preprocesar el texto, usualmente “... (i) dividimos el texto en tokens; (ii) creamos un vocabulario para mapear cadenas de tokens a índices numéricos; y (iii) convertimos los datos de texto en índices de tokens para que los modelos los manipulen”.²³ En otras palabras, se le asigna valores al texto para formar lo que se convertirá en un modelo de lenguaje, ello a través de la ley de Zipf, que ayuda a determinar la frecuencia de ciertas palabras en un idioma²⁴.

Posterior a la transformación del texto sin procesar en datos de secuencia o *tokens*, es posible crear un modelo de lenguaje que estime la probabilidad conjunta de una serie de datos; es decir, que arme el rompecabezas y comunique una respuesta al *prompt*. La formulación de este modelo de lenguaje es la que permite, justamente, que la creación de texto a través de predicciones secuenciales se diferencie de aquella paradoja del mono con una máquina de escribir, ya que el texto formulado por este algoritmo podría ser confundido por lenguaje natural y no es una simple combinación aleatoria de caracteres.²⁵

En virtud a las explicaciones anteriores, comprender cómo se formulan los sistemas de generación de texto permite avisar, entonces, de qué formas pueden utilizarse éstas para la automatización de ciertas diligencias y autos en el proceso. Asimismo, comprender que la creación de texto a través de algoritmos de predicción secuencial deja en evidencia los riesgos que significa sustituir el razonamiento humano por la transcripción de textos sintéticos, ya que estos están aún muy lejos de la reflexión necesaria en escritos como demandas, contestaciones y más aún, en las sentencias.

elementos más manejables para que las máquinas puedan analizarlos, entenderlos y procesarlos. MICROSOFT LEARN. “Tokens and character sets”. 8 de febrero de 2021. Disponible en: <https://learn.microsoft.com/en-us/cpp/cpp/character-sets?view=msvc-170>.

²⁰ ZHANG, Anston. *et. al. Dive into Deep Learning*. Amazon Science. 2022. p. 341. Disponible en: <https://d2l.ai/d2l-en.pdf>.

²¹ Modelo diseñado por OpenAI y que corresponde a las siglas de “Generative Pre-trained Transformer” (transformador generativo pre-entrenado)

²² ZHANG, *Cit.* (n. 9) p. 341.

²³ *Ibíd.* p. 341-342. En su idioma original: “... (i) split text into tokens; (ii) build a vocabulary to map token strings to numerical indices; and (iii) convert text data into token indices for models to manipulate”. (Traducción propia).

²⁴ *Ibíd.*

²⁵ *Ibíd.*

III. ÉTICA EN LA IA

Frente a la gran cantidad de virtudes y posibilidades presentadas con anterioridad en la implementación de IA en la justicia, lo cierto es que esta herramienta dista aún de ser perfecta. Por lo tanto, es necesario hacer la salvedad de valorar su implementación, pues, de nada sirve traer esta herramienta a la justicia sin estimar los riesgos que puede causar al funcionamiento de un sistema judicial dentro del Estado de Derecho. Es por tanto que, la implementación de sistemas basados en IA, tal como los modelos repasados con anterioridad, presentan una serie de desafíos éticos. De modo que, es necesario abarcar cuáles son aquellos retos a los que se enfrenta en su adecuación como herramienta en los sistemas de justicia.

3.1.- Sesgos de datos y de dónde provienen

Una de las mayores preocupaciones en la implementación de estos sistemas son los riesgos en los sesgos de datos, pues, al ser entrenados con ciertos conjuntos de datos, existe un riesgo a que las respuestas a los *prompts* puedan ser ofensivas o parcializadas²⁶. El auge de estos sistemas de datos ha llevado a la digitalización masiva de los datos, los cuales sirven para tomar decisiones y entrenar a la IA. No obstante, muchos de estos datos suelen ser incompletos, dejándose de lado ciertos grupos de personas que son menos representados en los conjuntos de datos, lo cual crea sesgos algorítmicos²⁷. Por lo tanto, las respuestas que surjan de este conjunto de datos estarán siempre sesgadas.

Aunque no lo parezca, al no tener una cantidad suficiente de datos o al contener un conjunto de datos no representativos, puede resultar en que los algoritmos realicen predicciones que señalen discriminatoriamente a ciertos grupos étnicos²⁸. Al respecto, puede señalarse el caso de un estudio centrado en las bases de datos de imágenes de las fuerzas del orden en los Estados Unidos, donde se demostró que las personas afrodescendientes tenían más probabilidades de ser señaladas erróneamente en las

²⁶ AGUDELO MEJÍA, Dimaro; AGUDELO MEJÍA, Norman, Toma de decisiones asistida por inteligencia artificial: el Chat GPT en el derecho procesal. *La justicia en la era de la revolución tecnológica*. Institución Universitaria de Envigado. 2023, p. 265. Disponible en: https://www.iue.edu.co/fondo_editorial/la-justicia-en-la-era-de-la-revolucion-tecnologica/

²⁷ Asamblea General de las Naciones Unidas. *Resolución A/HRC/56/68*. Formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia. Consejo de Derechos Humanos 56° período de sesiones 18 de junio a 14 de julio de 2024 Tema 9 de la agenda Racismo, discriminación racial, xenofobia y formas conexas de intolerancia: seguimiento y aplicación de la Declaración y el Programa de Acción de Durban. p. 4. <https://documents.un.org/doc/undoc/gen/g24/o84/23/pdf/g2408423.pdf>

²⁸ *Ibidem*.

redes de reconocimiento facial²⁹. Lo anterior se debía a errores en la identificación facial de ese grupo étnico, además de la sobrerrepresentación en las bases de datos, lo cual concluía en respuestas erróneas debido a un sesgo de datos, el cual provenía de un sesgo histórico humano.

Por ello, es interesante tomar en cuenta al respecto lo siguiente: “Un elemento central del aprendizaje automático es hacer predicciones sobre el futuro a partir de datos del pasado. Sin embargo, si los datos del pasado están sesgados en contra de ciertos grupos, en particular por criterios raciales y étnicos, los modelos informáticos pueden reproducir y amplificar esos sesgos. El uso de datos sesgados o erróneos para tomar decisiones en la vida real puede perjudicar aún más a los grupos raciales y étnicos marginados, porque el uso de esos datos en el contexto de la inteligencia artificial crea más datos, que luego se utilizan para sustentar decisiones futuras. Estos sistemas que se refuerzan a sí mismos pueden reproducir y profundizar las disparidades existentes”.³⁰

De esta forma, es justo señalar que, aunque en la vida real puede advertirse como las decisiones humanas son sesgadas por prejuicios sociales, da una mayor situación de seguridad que éstas puedan ser, al menos, el producto de un proceso reflexivo de otro ser humano, mismo razonamiento que puede ser aun revisado por otra persona, distinta a la primera, en virtud de las decisiones judiciales, y no dependiendo un enunciado probabilístico carente de razonamiento contextual; pues, aunque la decisión tomada por la IA sea correcta, deben tomarse en cuenta los riesgos que rodea la idea de un “juez autómatas”, el cual se basa en una serie de datos preentrenados distintos a los ocurridos en la realidad.

3.2.- Injusticia algorítmica y sistemas predictivos

Aunque cada avance pueda sugerir una mayor capacidad de la IA para resolver problemas lógicos y mejorar su precisión, es importante tener en cuenta sus limitaciones pues, tal como señala Castrillón (2023):

“... tratándose de decisiones mediadas por algoritmos automatizados, existe el riesgo de que la decisión mediada por IA implique consecuencias injustas para quien se vea afectado por dicha decisión, *aunque el resultado de la aplicación del algoritmo sea correcto de acuerdo con su programación*”.³¹ (Resaltado añadido)

²⁹ *Ibidem*.

³⁰ *Ibid.* p. 4-5.

³¹ CASTRILLÓN, Eddison. Lineamientos éticos para la implementación de la inteligencia artificial en la decisión judicial: la política de modernización de la justicia civil y los retos éticos frente al riesgo de injusticia algorítmica. *La justicia en la era de la revolución tecnológica*. Institución Universitaria de Envigado. 2023. p.

A este fenómeno se le ha llamado “injusticia algorítmica”, el cual es el mayor riesgo que presenta la implementación de IA en las decisiones judiciales; pues, al ser automatizadas pueden ser objeto de situaciones que encrudezcan las desigualdades sociales por la ausencia de razonamiento abstracto y contextual, propio del pensamiento humano. Por ello, es necesario señalar que “...un algoritmo no es pre reflexivo, ni tampoco realiza juicios de razonabilidad, pero puede orientarse a establecer patrones comunes entre los datos que le sirven de insumo para tomar decisiones con la mayor probabilidad de acierto”³². Por lo cual, es correcto señalar que una decisión algorítmica es producto de un proceso matemático de probabilidad basado en datos pre entrenados, y no un proceso reflexivo de los hechos presentados en juicio.

IV.- IMPLEMENTACIÓN DE LA IA EN LA JUSTICIA

La implementación de IA es objeto de discusiones, donde para muchos los avances son motivo de entusiasmo, para otros, es motivo de preocupación frente a la posibilidad de que éstas tecnologías sean utilizadas sin consideraciones éticas de por medio. Recientemente, la Corte Constitucional de Colombia se encargó de estudiar este último caso, con el fin de establecer una serie de criterios orientadores para su uso responsable en la Rama Judicial. De tal modo, revisó una sentencia en la que un juez de Cartagena hizo uso de ChatGPT para ampliar los argumentos de su decisión³³.

La Corte realizó profundas consideraciones sobre la protección de derechos fundamentales en un sistema judicial que utiliza IA, así como los impactos del uso de estas herramientas en la sociedad y su estado en Colombia. Para responder a ello, en su decisión subrayó la imperiosa necesidad de una regulación, centrada en la protección del derecho a la privacidad, el principio del juez natural y el proceso debido como sus pilares fundamentales. Por lo tanto, consideró necesario establecer una serie de criterios orientadores para el futuro uso de estas tecnologías en el ámbito judicial³⁴.

A partir del apogeo de la Inteligencia Artificial (IA) generativa durante los últimos años, las propuestas para su implementación en la Rama Judicial son cada vez más

207. Disponible en: https://www.iue.edu.co/fondo_editorial/la-justicia-en-la-era-de-la-revolucion-tecnologica/

³² *Ibíd.* 218

³³ Corte Constitucional de Colombia, 2 de agosto de 2024. Sentencia T-323/2024. Disponible en: <https://www.corteconstitucional.gov.co/relatoria/2024/T-323-24.htm>

³⁴ En particular, su fundamento jurídico 423 establece los siguientes criterios: a. Transparencia; b. Responsabilidad; c. Privacidad; d. No sustitución de la racionalidad humana; e. Seriedad y verificación; f. Prevención de riesgos; g. Igualdad y equidad; h. Control humano; i. Regulación ética; j. Adecuación a buenas prácticas y estándares colectivos; k. Seguimiento continuo y adaptación; l. Idoneidad. *Ibíd.*

comunes, destacándose en cada una de ellas las ventajas que traería la eficacia de los procesos algorítmicos como respuesta al constante retardo judicial que sufre el acceso a la justicia, y que particularmente se enfatiza en el retraso injustificado en la producción de sentencias, pero es necesario preguntarse: ¿Es viable la implementación de IA generativa en la decisión judicial?

4.1.- En labores de gestión administrativa y corrección

Una de las propuestas más interesantes es la de colocar las tecnologías a disposición de los actos de mero trámite. En el caso de las TIC, éstas ya cuentan con gran utilidad al agilizar audiencias a través de medios telemáticos, introducir escritos a través de medios electrónicos, además de contar con un ecosistema digital que permite consultar las decisiones más importantes de las altas cortes de Colombia. A pesar de ello, la utilización de sistemas basados en IA reforzarían los medios ya existentes.

Una de las aplicaciones más interesantes para la gestión administrativa sería a través de los ya repasados árboles de decisión, los cuales podrían ser utilizados para la formación de autos admisorios. En concreto, la IA puede intervenir en labores de mero trámite que faciliten la consecución de los procesos tanto para abogados como para servidores judiciales. Al respecto, Quiñones ha formulado un proyecto con algunas de las aplicaciones de esta tecnología en el proceso para la automatización en la presentación de demandas:

“1. Plantillas dinámicas: utilizar IA para crear plantillas de demandas que se adapten según el tipo de caso y jurisdicción. El software puede guiar al abogado paso a paso, asegurándose de que todos los elementos requeridos estén presentes.

2. Verificación de requisitos: antes de que la demanda sea finalizada, la IA puede revisarla y asegurarse de que cumple con todos los requisitos necesarios, señalando cualquier omisión o error para su corrección.

3. Análisis comparativo: la IA puede comparar la demanda con otras demandas similares previamente presentadas, identificando posibles inconsistencias o elementos que podrían fortalecer el uso.

4. Revisión de legislación y jurisprudencia: integrando una base de datos legal, el sistema podría verificar automáticamente que todas las referencias legales y jurisprudenciales estén actualizadas y sean aplicables al caso.

5. Optimización del lenguaje: mediante el análisis lingüístico, la IA puede sugerir modificaciones en la redacción para hacerla más clara, concisa y persuasiva.

6. Gestión de plazos: integrando un calendario, la IA podría recordar automáticamente los plazos relevantes para la presentación y seguimiento de la demanda, reduciendo el riesgo de incumplimientos.

7. Simulación de escenarios: basándose en casos similares y su desenlace, la IA podría simular posibles respuestas a la demanda y recomendar estrategias o modificaciones.

8. Interfaz de usuario amigable: una plataforma intuitiva que guíe al abogado en el proceso de elaboración de la demanda, evitando errores comunes y asegurando la inclusión de todos los elementos necesarios.

9. Integración con sistemas judiciales: si el sistema judicial lo permite, la IA podría presentar la demanda electrónicamente de forma automática una vez que haya sido revisada y aprobada por el abogado.

10. Actualizaciones constantes: la plataforma debería actualizarse automáticamente con cambios en la legislación, formularios, requisitos y jurisprudencia para mantener su relevancia y eficacia”.³⁵

De este modo, se logra una mayor economía procesal al ser revisados de manera automática los requisitos de las demandas a través de revisiones automáticas con IA y la conversión de los escritos en conjuntos de datos que pueda tratar como un LLM³⁶. Estas propuestas van mucho más allá de la eliminación del papel en la labor judicial, pues esta justicia que ahora está “digitalizada” conforma un conjunto de datos con un potencial tratable para diversos fines. A través del *Machine Learning*³⁷ puede desarrollarse un sistema que complemente la búsqueda de fuentes legales y jurisprudenciales para jueces, auxiliares judiciales o abogados. Así como la creación de *chatbots*³⁸ capaces de compartir información legal en lenguaje sencillo para las personas que quieran consultar un caso legal³⁹.

Asimismo, la IA en su apartado generativo puede lograr agilizar la generación de providencias de manera automática, especialmente aquellas en las que no hay

³⁵ QUIÑONES, María. Justicia digital e inteligencia artificial: reducción de tiempos procesales. *XLV Congreso Colombiano de Derecho Procesal: El futuro de la administración de justicia y los métodos de resolución de conflictos* (45: 30 septiembre - 04 octubre: Bucaramanga, Colombia). Memorias. Bucaramanga, 2024. pp. 163-164. Disponible en: https://drive.google.com/file/d/1_IgM07HaxuqKpUSmFcVvdQjDoau-3rj4/view

³⁶ Corresponde a las siglas de “Large Language Model” (Modelo lingüístico grande), el cual caracteriza a los tipos de IA capacitados con enormes conjuntos de datos. Brown, Tom et al. Language models are few-shot learners. *arXiv*. 2020. p. 11. Disponible en: <https://arxiv.org/abs/2005.14165>.

³⁷ “El Machine Learning (aprendizaje automático) es una rama de la inteligencia artificial que permite que las máquinas aprendan ciertas tareas sin ser programadas de manera específica para ellas”. BBVA. “Machine learning: ¿qué es y cómo funciona el maestro en reconocer patrones?” Comunicaciones. 15 de julio de 2024). Disponible en: <https://www.bbva.com/es/innovacion/machine-learning-que-es-y-como-funciona/>

³⁸ Estos son *apps* o interfaces basadas en IA que pueden mantener conversaciones similares a las humanas mediante el procesamiento de lenguaje natural y el *Machine Learning*. Las empresas desarrollan *chatbots* para solucionar problemas o dudas de sus clientes sin la necesidad de operadores humanos.

³⁹ European Commission for the Efficiency of Justice (CEPEJ). “CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment”. 2018. p. 63 Disponible en: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>

razonamiento alguno del juez, como notificaciones y citaciones. Por otro lado, en su aparato predictivo puede servir como un apoyo en las técnicas de justicia predictiva, como las usadas por aseguradoras que calculan y evalúan las posibilidades de éxito en un litigio o una conciliación⁴⁰. Estos últimos pueden conformarse a través de sistemas tan sencillos como los *decision trees* y ser reforzados con IA, de tal modo que, al igual que las listas de chequeo propuestas por Quiñones⁴¹, sea capaz de determinar si una demanda cumple con todos aquellos requisitos de forma establecidos en la ley, dejándole al juez la tarea fundamental de examinar el fondo del asunto.

4.2.- En la valoración de pruebas.

En Colombia se ha desarrollado un sistema experto capaz de asistir al juez en la valoración probatoria. Esta propuesta, encabezada por Vargas, es una herramienta pensada para los estudiantes de Derecho en su formación como futuros jueces de la República, incorporándose a su desarrollo profesional el uso de IA⁴². Este sistema se encarga de apreciar y entrelazar los hechos que le son presentados, lo que le permite valorar, de forma individual y conjunta, todas las pruebas propuestas por las partes en el proceso, tal como señala el autor⁴³.

Si bien, es aclarado por el mismo autor que este sistema no pretende reemplazar al juez, los sistemas de IA anticipan un impacto en la forma que se valoran las pruebas. Lejos de establecer un sistema de automatismos, la IA podría contribuir a ofrecer resultados basados en datos, tal como señala Nieva⁴⁴. Por tanto, ofrecen la posibilidad de una decisión apoyada de sistemas objetivables de los cuales disponga el juez para tratar grandes cantidades de datos. Esto solo es posible, tal como señala Castellano, si se comprende la forma en que se valoran las pruebas, de modo que una fórmula algorítmica sea capaz de emular ese proceso⁴⁵, pero no de sustituirlo, del mismo modo que se concibe el sistema experto diseñado por Vargas.

Aun así, la posibilidad de una valoración asistida por IA de parte del juez permanece como una propuesta interesante, y tal como señala Valentín:

⁴⁰ CASTELLANO, Pere. Inteligencia Artificial y valoración de la prueba: las garantías Jurídico-Constitucionales del órgano de control. *THEMIS Revista De Derecho*, (79). 2021. p. 286. Disponible en: <https://revistas.pucp.edu.pe/index.php/themis/article/view/24878/24922>

⁴¹ QUIÑONES, cit. (n. 14), p. 165.

⁴² Ámbito Jurídico. "El sistema experto 'juez inteligente' asiste en la valoración de la prueba judicial". Entrevista a Orión Vargas Vélez. *Legis*. 18 de agosto de 2022. Disponible en: <https://www.ambitojuridico.com/noticias/en-ejercicio/el-sistema-experto-juez-inteligente-asiste-en-la-valoracion-de-la-prueba>

⁴³ *Ibidem*.

⁴⁴ NIEVA, Jordi Citado por CASTELLANO, cit. (n. 15), p. 288.

⁴⁵ *Ibidem*.

“La respuesta resumida es que los jueces juzgan a partir de su experiencia, sus recuerdos, sus intuiciones, sus emociones y preferencias (incluso ideológicas), sus problemas, su diálogo con el subsistema social que integran o aquellos con los que se relacionan (el foro, la academia, las redes, etc.), etc.

Pues bien: la IA nos puede proporcionar herramientas para expulsar o mitigar los efectos que producen en las salas de justicia algunos atávicos heurísticos vinculados con aquellos factores, como emociones, gustos, prejuicios, preferencias, aficiones, etc. De esa forma, en una mirada preliminar, parece que la IA puede utilizarse para potenciar las virtudes del sistema de valoración racional y a la vez expulsar o al menos mitigar sus principales defectos”.⁴⁶

Si bien es cierto que la IA es capaz de mitigar los sesgos personales que tenga el operador de justicia, también es justo señalar que esta misma tecnología es totalmente capaz de reforzarlos y recrudecerlos. De tal modo, que aun cuando deja de lado las emociones, gustos o prejuicios, los sistemas con los que ha sido entrenada podrían estar repleto de ellos, por lo que pasaría como en el caso repasado en título anterior, donde un sistema de reconocimiento facial podría llegar a ser discriminatorio por sesgos en sus bases de datos. De ahí que, de acuerdo a Valentín, su aplicación deberá orientarse como una herramienta que asista al juez en ciertas labores, más no sea quien tome la decisión *per se*, la cual a él le corresponde emitir.⁴⁷

4.3.- En la decisión judicial

La propuesta de un “juez autómatas” a través de la implementación de IA en las decisiones judiciales se asemeja a la premisa del Juez mecánico que refiere Colmenares: “la ley dispone y el juez obedece”.⁴⁸ En este orden de ideas, es preciso recordar lo señalado por Barona: “Más allá de conocer las normas y la jurisprudencia, la función de ser juez implica capacidad de interpretarlo con emociones, percepciones, intuiciones, o

⁴⁶ VALENTÍN, Gabriel. “JUSTITIA EX MACHINA: ¿PUEDE LA INTELIGENCIA ARTIFICIAL AYUDARNOS EN LA VALORACIÓN RACIONAL DE LAS PRUEBAS?”. *XLV Congreso Colombiano de Derecho Procesal: El futuro de la administración de justicia y los métodos de resolución de conflictos* (45: 30 septiembre - 04 octubre: Bucaramanga, Colombia). Memorias. Bucaramanga, 2024. p. 116-117. ISBN digital: 978-958-52944-5-5. Disponible en: https://drive.google.com/file/d/1_IgM07HaxuqKpUSmFcVvdQjDoau-3rj4/view

⁴⁷ *Ibíd.* p. 127-128

⁴⁸ COLMENARES, Carlos. El rol del juez en el estado democrático y social de derecho y justicia. *Revista Academia & Derecho*, Número 5. 2012. p. 65. Disponible en: <https://dialnet.unirioja.es/ejemplar/504669>

lo que es lo mismo, las sensibilidades subjetivas. El juez humano contextualiza. El juez robot, no”.⁴⁹

Por otro lado, aunque en la Sentencia T-323 la Corte admitió el uso de herramientas como ChatGPT, éstas no pueden sustituir la actividad judicial, pues significaría una violación al principio del juez natural⁵⁰. Asimismo, la Rama Judicial no controla los datos en los cuales se basan las respuestas obtenidas de esta herramienta, por lo que genera el riesgo a sesgos de datos. Pues, sus decisiones no se basan en una ponderación de la realidad de quienes acuden al proceso, sino en una serie de probabilidades que se generan a través de datos.

Por lo tanto, es crucial, además de un marco regulatorio robusto y de garantías, un sistema de IA generativa propio. De modo que, al estar especializado en la jurisprudencia y normativa patria, se encargue de mejorar el acceso a la información de criterios jurídicos. Además, esta herramienta le proporcionaría al juez un sistema creado para la labor que ejerce y así no recurra a herramientas externas como ChatGPT que en sus políticas prohíbe de manera expresa su uso para asesoramiento jurídico o toma decisiones, en las cuales incluye la justicia⁵¹.

Es evidente que en la actualidad no es posible permitir, dentro del marco del Estado Social de Derecho, implementación de IA generativa en la decisión judicial debido al riesgo de la injusticia algorítmica. En particular, esta investigación ha abordado lo que no es objeto de automatización en la actividad judicial, pero no puede ignorarse la posibilidad que presenta para mejorar el acceso a la justicia en otras áreas como el manejo y gestión de expedientes. En el caso particular de la decisión judicial, se encuentran perfectamente explicadas sus razones por las que su automatización no es idónea, pero en el caso de los autos de mero trámite, deberán generarse las políticas necesarias que busquen optimizar su eficacia.

Ante este panorama, el proyecto de modernización de la justicia tanto en Colombia como en los demás países de Iberoamérica deberá tomar en cuenta estas consideraciones en la inevitable implementación de estas nuevas tecnologías en sus sistemas judiciales, haciendo a su vez, las necesarias mejoras a nivel de infraestructura que permitan su correcto funcionamiento. Se estima que la implementación de IA podrá ser beneficiosa siempre que ésta no intente sustituir la función judicial, ya que, de acuerdo a Carnelutti: “Un ordenamiento jurídico se puede pensar sin leyes, pero no sin jueces”.

⁴⁹ BARONA, Silvia. El algoritmo en la prueba y en la decisión judicial: ¿instrumental o funcional? *La justicia en la era de la revolución tecnológica*. Institución Universitaria de Envigado. 2023. p. 31. Disponible en: https://www.iue.edu.co/fondo_editorial/la-justicia-en-la-era-de-la-revolucion-tecnologica/

⁵⁰ Corte Constitucional. 2 de agosto de 2024. Sentencia T-323/2024.

⁵¹ OpenAI. *Políticas de Uso*. Construir con ChatGPT, punto 2.b. 10 de enero de 2024. Disponible en <https://openai.com/policies/usage-policies>. Esta regulación en específico fue añadida en su actualización del 15 de febrero de 2023.

V.- LINEAMIENTOS ÉTICOS PARA LA IMPLEMENTACIÓN DE IA EN LA JUSTICIA

En el caso analizado, la Corte determinó que era correcto el uso de IA de manera responsable, como un medio de apoyo a la labor judicial, pero dio por alto algunos de los errores cometidos por la IA. Es el caso que, en su sentencia, el juez realizó una serie de *prompts* a ChatGPT para clarificar algunas dudas en la motivación de su decisión. De tal manera que en el caso relatado se realizó la siguiente pregunta:

Prompt ingresado por el juez.	Respuesta del aplicativo de IA.
¿Menor autista está exonerado de pagar cuotas moderadoras en sus terapias?	Sí, es correcto. De acuerdo con la normativa en Colombia, los menores con diagnóstico de autismo están exonerados de pagar cuotas moderadoras en sus terapias. Este beneficiario está regulado por la Ley 1753 de 2015 y se aplica a todas las instituciones prestadoras de servicios de salud públicas o privadas que reciban recursos del sistema general de seguridad social en salud ⁵² .

Tabla 2: Elaboración Propia

A simple vista, parece no haber error alguno e incluso la respuesta dada por ChatGPT luce bastante correcta, pero ha de señalarse que la ley citada por el aplicativo es un caso más de alucinación por la IA. Al consultar el Gestor Normativo de manera manual, puede evidenciarse que la Ley 1753 de 2015 no contempla expresamente la exoneración que menciona, sino que se encuentran regulada en el Decreto N° 1652 de agosto de 2022 expedido por el Ministerio de Salud y Protección Social⁵³.

Este tipo de alucinaciones son bastante comunes, tanto así que en Estados Unidos, Sudáfrica y Brasil se han denunciado casos de jueces o abogados que emitieron resoluciones judiciales o presentaron documentos jurídicos que incluían referencias a

⁵² Este cuadro contiene una transcripción de la pregunta hecha en: Corte Constitucional. 2 de agosto de 2024. Sentencia T-323/2024.

⁵³ Decreto N° 1652 (agosto de 2022). Ministerio de Salud y Protección Social. Decreto "Por el cual se adiciona el Título 4 a la parte 10 del Libro 2 del Decreto 780 de 2016 relativo a la determinación del régimen aplicable para el cobro de pagos compartidos o copagos y cuotas moderadoras a los afiliados al Sistema General de Seguridad Social en Salud". Colombia. Disponible en: https://www.minsalud.gov.co/Normatividad_Nuevo/Decreto%20No.%201652%20de%202022.pdf

sentencias inexistentes debido al uso de chatbots de IA.⁵⁴ Al respecto, el caso más representativo es *Mata v. Avianca Inc.* del 22 de junio de 2023, conocido por el Tribunal de Distrito Sur de Nueva York, donde un abogado fue objeto de una medida disciplinaria por presentar una demanda hecha por ChatGPT, la cual contenía precedentes judiciales inexistentes.⁵⁵

Por ello, es necesario generar una serie de lineamientos éticos que permitan la utilización de estas tecnologías, pero que controlen una posible sustitución de la intervención humana; es decir, que asegure que los escritos que requieran un razonamiento provengan de otra persona. En particular, la Sentencia T-323 de la Corte Constitucional se adelantó en este aspecto, al definir la pauta para la regulación de este tipo de tecnologías. Así pues, estableció unos criterios diseñados para guiar y controlar la actividad de los operadores de justicia en los que incluye: a. Transparencia; b. Responsabilidad; c. Privacidad; d. No sustitución de la racionalidad humana; e. Seriedad y verificación; f. Prevención de riesgos; g. Igualdad y equidad; h. Control humano; i. Regulación ética; j. Adecuación a buenas prácticas y estándares colectivos; k. Seguimiento continuo y adaptación; l. Idoneidad.⁵⁶

Por otro lado, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) también ha elaborado una serie de directrices para el uso de sistemas de inteligencia artificial en juzgados y tribunales. En dichas directrices se hace un repaso de herramientas como VICTOR en Brasil para revisión del cumplimiento de requisitos en apelaciones; SUVAS en la India, un programa capaz de traducir miles de documentos en inglés a otras lenguas nativas o las herramientas de procesamiento de lenguaje natural para predecir decisiones judiciales del Tribunal de Justicia de la Unión

⁵⁴ GUTIÉRREZ, Juan. *Documento de consulta pública: Directrices de la UNESCO para el uso de sistemas de inteligencia artificial en juzgados y tribunales*. UNESCO. 2024. p. 5. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000390781_spa

⁵⁵ CORTÉS, Nicolás. Abogado que lleva demanda contra Avianca usó ChatGPT y todo le salió mal. *El Tiempo*. 28 de mayo de 2023. Disponible en: <https://www.eltiempo.com/mundo/eeuu-y-canada/chatgpt-abogado-con-ayuda-del->

⁵⁶ Entre ellos, ha de destacarse el criterio “g. Igualdad y equidad”; por tanto, se encuentra alineado con la Recomendación sobre la Inteligencia Artificial de la Organización para la Cooperación y el Desarrollo Económico (OCDE), que promueve el crecimiento inclusivo de estas tecnologías “Las partes interesadas deberían poner en marcha activamente la administración responsable en aras de una IA fiable en busca de resultados beneficiosos para las personas y el planeta, como aumentar las capacidades del ser humano y estimular la creatividad, avanzar en la inclusión de poblaciones infrarrepresentadas, reducir las desigualdades económicas, sociales, de género y de otra índole, y proteger los entornos naturales, reforzando de este modo el crecimiento inclusivo, el bienestar, el desarrollo sostenible y la sostenibilidad medioambiental”. Organización para la Cooperación y el Desarrollo Económico. *Recommendation of the Council on Artificial Intelligence*. OECD Legal Instruments. 2019. Disponible en: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

Europea⁵⁷. Asimismo, sobre casos en que se ha implementado IAG por jueces para resumir decisiones en lenguaje sencillo⁵⁸.

En razón de lo expuesto anteriormente, la UNESCO se encargó de formular una serie de lineamientos que se conforman en trece principios, los cuales tienen por objeto el respeto universal por la justicia y la orientación en el uso de sistemas de IA, incluida la IAG. De tal manera que delimita estos criterios para el uso de la IA de la manera siguiente:

Directrices para el uso de sistemas de IA por parte de la judicatura		
Principios	Definición	Sub-principios
Protección de los derechos humanos	Adoptar sistemas de IA basados en el respeto de los derechos humanos que permitan al poder judicial respetar, proteger y promover los derechos humanos mientras administra justicia.	<p>a. Equidad: Adoptar sistemas de IA que garanticen un acceso inclusivo a la tecnología.</p> <p>b. No discriminación: Evitar sesgos y resultados que refuercen la discriminación.</p> <p>c. Equidad procesal</p> <p>d. Protección de datos personales: Adoptar sistemas de IA que protejan los datos personales tratados para la administración de justicia.</p>
Proporcionalidad	Adoptar sistemas de IA que persigan fines legítimos y proporcionales en el contexto de su uso.	
Protección	Adoptar sistemas de IA que eviten, aborden, prevengan y eliminen daños no deseados	
Seguridad de la información	Adoptar sistemas de IA que protejan la información confidencial de acuerdo con las normas internacionales de acceso a la información.	
Conocimiento y uso informado	Comprender las funcionalidades, los tipos de usos, los impactos potenciales, las limitaciones y los riesgos de los sistemas de IA	

⁵⁷ GUTIÉRREZ, cit. (n. 18), p. 4.

⁵⁸ *Ibíd.* p. 5.

	disponibles para tomar decisiones informadas sobre su implementación.	
Uso transparente	Informar de manera adecuada y oportuna cuándo y cómo se utilizan los sistemas de IA y cómo funcionan estas herramientas.	
Rendición de cuentas y auditabilidad	Garantizar la rendición de cuentas informando y explicando por qué el poder judicial adoptó determinadas herramientas de IA. Adoptar medidas administrativas, jurídicas y humanas para garantizar que los sistemas de IA puedan ser auditados durante y después de su despliegue.	
Explicabilidad	Adoptar sistemas de IA que sean transparentes en cuanto a su funcionamiento, su desarrollo, sus datos de entrenamiento, sus limitaciones (incluido su margen de error), sus capacidades y la finalidad de los sistemas.	
Precisión y fiabilidad	Adoptar y utilizar sistemas de IA precisos, es decir, sistemas de IA que puedan proporcionar información útil y pertinente y producir resultados y predicciones correctos.	
Supervisión humana	Mantener un nivel adecuado de control e implicación humanos en relación con todos los sistemas de IA.	
Diseño centrado en el ser humano	El desarrollo, despliegue y uso de sistemas de IA deben seguir principios de diseño centrados en el ser humano.	
Responsabilidad	Las organizaciones que despliegan sistemas de IA y las personas que los utilizan deben asumir la responsabilidad de las	

	decisiones y acciones tomadas con el apoyo de herramientas de IA.	
Gobernanza y colaboración de múltiples partes interesadas	Las organizaciones que forman parte del poder judicial deben consultar a diversas partes interesadas a lo largo del ciclo de vida del sistema de IA, especialmente a aquellas que puedan verse directa o indirectamente afectadas por su despliegue	59

Esta serie de principios conforman una guía orientadora para los países que quieran avanzar en la implementación de este tipo de herramientas, por lo que su análisis es fundamental para la presente investigación. Al tratarse de los desafíos éticos que supone su utilización en la justicia, ha de mencionarse que cada uno de los trece principios propuestos en las Directrices de la UNESCO están destinados a la implementación de estas tecnologías de manera ética.

Es esencial entender que la IA no puede reemplazar a los operadores de justicia o los abogados en muchos aspectos, por ello debe haber un uso Transparente, con Rendición de cuentas, que sea auditable y que necesariamente sea supervisado por humanos⁶⁰. Sobre la base de todos los lineamientos repasados es posible trazar una hoja de ruta para la construcción de un sistema de garantías tecnológico. Sin embargo, la realidad latinoamericana debe ser una importante consideración, en virtud de la gran desigualdad social y que el goce de los adelantos tecnológicos es un derecho humano; tal como lo recordó la Corte Suprema de Justicia de Colombia en su Sentencia SC370-2023 de 2023⁶¹.

VI.- CONCLUSIÓN

En definitiva, la modernización de la justicia es un proyecto que debe alinearse con estas tecnologías emergentes, y cada vez es más claro que los juzgados y sus operadores de justicia deberán adaptarse a ellas para ofrecer un mejor servicio. No obstante, es

⁵⁹ El cuadro mostrado corresponde a los principios formulados por *Ibid.* p. 11.

⁶⁰ *Ibidem.*

⁶¹ Corte Suprema de Justicia de Colombia. Sala de Casación Civil, Agraria y Rural. Sentencia SC370-2023, 10 de octubre de 2023. Disponible en: <https://cortesuprema.gov.co/corte/wp-content/uploads/2023/10/SC370-2023.pdf>.

necesario adoptar un enfoque integral que permita una implementación bajo una serie de lineamientos éticos que aseguren las condiciones óptimas en el acceso a la justicia.

En virtud de lo anterior, luego de haber repasado cada uno los criterios orientadores propuestos por la UNESCO, la OCDE e incluso la Corte Constitucional de Colombia es posible concluir que para enfrentar los desafíos éticos en la implementación de IAG en la justicia, su implementación deberá contener una hoja de ruta basada en: a. Protección de los derechos humanos; b. Proporcionalidad; c. Protección; d. Seguridad de la información; e. Conocimiento y uso informado; f. Uso transparente; g. Rendición de cuentas y auditabilidad; h. Explicabilidad; Precisión y fiabilidad; i. Supervisión humana; j. Diseño centrado en el ser humano; k. Responsabilidad; y l. Gobernanza y colaboración de múltiples partes interesadas.

Esto es, considerar primordialmente la igualdad en el acceso al desarrollo digital y la protección de los derechos fundamentales en su implementación.

De este modo, se reducen los riesgos a la aparición de sesgos de datos e injusticia algorítmica, mientras estas tecnologías le ofrecen al Derecho Procesal nuevas perspectivas sobre la tramitación de procesos con aspectos automatizados. Por lo tanto, las consideraciones realizadas en el presente trabajo constituyen, a su vez, una guía práctica acerca del funcionamiento de estas tecnologías y sus implicaciones jurídicas, las cuales serán objeto de futuros proyectos de *lege ferenda* en los sistemas jurídico-procesales de Latinoamérica. Así pues, la IAG en la justicia es una realidad cada vez más latente, a la cual los sistemas de justicia tarde o temprano deberán adaptarse, y lo mejor es que sea pronto, ya que de lo contrario se quedarán obsoletos y siguiendo firmemente la línea de los arcaicos juicios de papel y del retardo judicial.

BIBLIOGRAFÍA CITADA

- AGUDELO MEJÍA, Dimaro; AGUDELO MEJÍA, Norman, Toma de decisiones asistida por inteligencia artificial: el Chat GPT en el derecho procesal. La justicia en la era de la revolución tecnológica. Institución Universitaria de Envigado. 2023. Disponible en: https://www.iue.edu.co/fondo_editorial/la-justicia-en-la-era-de-la-revolucion-tecnologica/
- AOUICHAOU, A. et al. "Comparison of Group-Contribution and Machine Learning-based Property Prediction Models with Uncertainty Quantification". *Computer Aided Chemical Engineering*. Elsevier, 2021, Vol.50. ISBN: 9780323885065. Disponible en: <https://doi.org/10.1016/B978-0-323-88506-5.50118-2>.

- ASAMBLEA GENERAL DE LAS NACIONES UNIDAS. “Resolución A/HRC/56/68”. Formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia. Consejo de Derechos Humanos 56° período de sesiones 18 de junio a 14 de julio de 2024, Tema 9 de la agenda Racismo, discriminación racial, xenofobia y formas conexas de intolerancia: seguimiento y aplicación de la Declaración y el Programa de Acción de Durban. Disponible en: <https://documents.un.org/doc/undoc/gen/g24/084/23/pdf/g2408423.pdf>.
- BARONA, Silvia. El algoritmo en la prueba y en la decisión judicial: ¿instrumental o funcional? *La justicia en la era de la revolución tecnológica*. Institución Universitaria de Envigado. 2023. Disponible en: https://www.iue.edu.co/fondo_editorial/la-justicia-en-la-era-de-la-revolucion-tecnologica/.
- BROWN, Tom et al.,. “Language models are few-shot learners”, arXiv, 2020, Disponible en: <https://arxiv.org/abs/2005.14165>.
- CAPARRINI, Fernando. Variational AutoEncoder. *Blog Matemática*. Universidad de Sevilla. s.f. Disponible en: <https://www.cs.us.es/~fsancho/Blog/posts/VAE.md>.
- CASTRILLÓN, Eddison. (2023). Lineamientos éticos para la implementación de la inteligencia artificial en la decisión judicial: la política de modernización de la justicia civil y los retos éticos frente al riesgo de injusticia algorítmica. *La justicia en la era de la revolución tecnológica*. Institución Universitaria de Envigado. Disponible en: https://www.iue.edu.co/fondo_editorial/la-justicia-en-la-era-de-la-revolucion-tecnologica/.
- COLMENARES, Carlos. El rol del juez en el estado democrático y social de derecho y justicia. *Revista Academia & Derecho*, Número 5. 2012. Disponible en: <https://dialnet.unirioja.es/ejemplar/504669>
- CORTE CONSTITUCIONAL. 2 de agosto de 2024. Sentencia T-323/2024. Disponible en: <https://www.corteconstitucional.gov.co/relatoria/2024/T-323-24.htm>.
- CONTRERAS, Óscar. Modelos Generativos en el Aprendizaje Automático y su aplicación a la generación de Imágenes Digitales. *Journal Boliviano De Ciencias*, 17(51). 2021. Disponible en: <https://revistas.univalle.edu/index.php/ciencias/article/view/110>.
- EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ). “CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment”. 2018. Disponible en: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.
- FRANGANILLO, Jorge. “La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos”, *Methados revista de ciencias sociales*, 2023, 11(2). Disponible en: <http://dx.doi.org/10.17502/mrcs.v11i2.710>

- GOODFELLOW, Ian; POUGET-ABADIE, Jean. et. al. "Generative Adversarial Nets". *Neural Information Processing Systems*, 2014, Disponible en: <https://arxiv.org/abs/1406.2661>.
- IBM, "What is artificial intelligence (AI)?", International Business Machine Corporation. 16 de agosto de 2024. Disponible en: <https://www.ibm.com/topics/artificial-intelligence>.
- IBM, *Funcionamiento de SVM*. 17 de octubre de 2021. Disponible en: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>.
- KARRAS, Tero. et. al. *Analyzing and Improving the Image Quality of StyleGAN*. 2019. Disponible en: <https://arxiv.org/abs/1912.04958>
- KINGMA, Diederik. et al. *An Introduction to Variational Autoencoders. Foundations and Trends R in Machine Learning*. Universiteit van Amsterdam. 2019. Disponible en: <https://arxiv.org/pdf/1906.02691>.
- MICROSOFT LEARN. "Tokens and character sets". 8 de febrero de 2021. Disponible en: <https://learn.microsoft.com/en-us/cpp/cpp/character-sets?view=msvc-170>.
- OPENAI. *Language models are few-shot learners*. arXiv. 2020. Disponible en: <https://arxiv.org/abs/2005.14165>.
- ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICO. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD Legal Instruments. Disponible en: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
- QUIÑONES, María. "Justicia digital e inteligencia artificial: reducción de tiempos procesales". *XLV Congreso Colombiano de Derecho Procesal: El futuro de la administración de justicia y los métodos de resolución de conflictos* (45: 30 septiembre - 04 octubre: Bucaramanga, Colombia). Memorias. Bucaramanga, 2024. Disponible en: https://drive.google.com/file/d/1_IgMO7HaxuqKpUSmFcVvdQjDoau-3rj4/view
- SCHWAB, Klaus, *La cuarta revolución industrial*. Editorial Debate, 2016.
- STRYKER, Cole; KAVLAKOGLU, Eda. "What is artificial intelligence (AI)?". 16 de agosto de 2024. International Business Machine Corporation (IBM). Disponible en: <https://www.ibm.com/topics/artificial-intelligence>
- SUTTON, Richard; BARTO, Andrew. "Reinforcement Learning: An Introduction". *The MIT Press*. Londres. 2014. Disponible en: <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>.
- SZWARCMAN, Daniela et. al. "Quantizing reconstruction losses for improving weather data synthesis". *Science Reports*. 2024, vol.14, 3396. Disponible en: <https://doi.org/10.1038/s41598-024-52773-2>
- VALENTÍN, Gabriel. "Justitia ex machina: ¿puede la inteligencia artificial ayudarnos en la valoración racional de las pruebas?". *XLV Congreso Colombiano de Derecho Procesal: El futuro de la administración de justicia y los métodos de resolución de conflictos* (45: 30 septiembre - 04 octubre: Bucaramanga, Colombia). Memorias. Bucaramanga,

2024. ISBN digital: 978-958-52944-5-5. Disponible en:
https://drive.google.com/file/d/1_IgM07HaxuqKpUSmFcVvdQjDoau-3rj4/view.

ZHANG, Anston. et al. Dive into Deep Learning. Amazon Science. 2022. Disponible en:
<https://d2l.ai/d2l-en.pdf>