

**Bogotá, 8 de septiembre de 2023**

Honorable Magistrado  
Juan Carlos Cortés González  
Corte Constitucional

Ref: Expediente T-9.301.656  
Acción de tutela presentada por Beatriz Elena Chávez  
Jerónimo, en nombre y representación de su hijo menor de  
edad, contra Salud Total EPS

Cordial saludo,

Juan Diego Castañeda y Juan de Brigard, actuando en calidad director encargado y coordinador de la línea de autonomía y dignidad de la Fundación Karisma, y Ana Bejarano Ricaurte, Susana Echavarría Medina y Laura Urrego Aguilera, actuando en calidad de directora, coordinadora de proyectos y abogada de El Veinte, remitimos comentarios en el caso de la referencia, en atención a la invitación dirigida a la Fundación Karisma en auto del pasado 25 de agosto de 2023.

Para sustentar nuestra posición la presente intervención se dividirá en dos partes. En la primera de ella realizaremos un contexto sobre la digitalización de la justicia en Colombia y a las tecnologías de inteligencia artificial generativas, que servirá de marco para las respuestas a las preguntas enviadas por la honorable Corte Constitucional, que resolverán en la segunda parte. En la tercera, para concluir, propondremos una serie de solicitudes.

1. El contexto de los esfuerzos de digitalización de la justicia y el uso de la Inteligencia Artificial por parte de operadores judiciales.

Digitalización de la justicia en Colombia

Antes de abordar el caso que nos convoca, al respecto del uso de la herramienta ChatGPT para el fallo de una tutela, debemos partir del hecho de que este evento se enmarca dentro de una tendencia general del

1



Estado colombiano hacia la digitalización de los servicios judiciales. Esta transformación avanzó de manera importante a raíz del aislamiento físico al que nos vimos obligados por causa de la pandemia del Covid-19, aunque su origen es anterior a ello. En dicho periodo surgieron diversas políticas entre las que destacan el CONPES 4024 de 2021 y el Decreto 806 de 2020<sup>1</sup>, hecho permanente con la Ley 2213 de 2022, medidas que respondieron a la imposibilidad de continuar con la atención presencial en las diligencias judiciales, por una parte, y a la necesidad general de encontrar mecanismos más eficientes y expeditos para completar trámites judiciales, por otra. Esta incorporación de nuevas tecnologías en la rama judicial busca, más allá de mantener activos los servicios en las condiciones anómalas de la pandemia, descongestionar y reducir los costos asociados al acceso a la justicia<sup>2</sup>.

Y aunque llevados de manera adecuada estos esfuerzos pueden dar resultados positivos en términos de acceso a la justicia, en Colombia su avance se ve diezmado por la inmensa brecha digital (en 2021, según el DANE, sólo el 61,6% del total de los hogares cuenta con acceso a Internet, disminuyendo esta cifra a 32,5% cuando se miden centros poblados y rurales dispersos)<sup>3</sup> y por otras causas como la falta de adecuación de los procesos y la estructura judicial a la digitalización, la falta de infraestructura tecnológica y la poca claridad en los nuevos protocolos que han de adoptarse.<sup>4</sup>

Hasta el presente, la digitalización ha consistido, fundamentalmente, en permitir que ciertos procedimientos que exigían la presencia física de las

<sup>1</sup> «Decreto 806 de 2020 - Gestor Normativo - Función Pública», accedido 5 de septiembre de 2023, <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=127580>; Consejo Nacional De Política Económica y Social (CONPES) y Departamento Nacional de Planeación, «Concepto favorable a la nación para contratar una operación de crédito público externo con la banca multilateral hasta por usd 100 millones, o su equivalente en otras monedas, destinados a financiar el programa para la transformación digital de la justicia en Colombia- Fase I», 4024 § (s. f.), <https://www.minjusticia.gov.co/Sede-Electronica/Documents/TDJC/Documentos-Interes/Conpes%204024%20Transformacion%20Digital%20de%20la%20Justicia%20en%20Colombia.pdf>.

<sup>2</sup> Consejo Superior de la Judicatura, Informe de transformación digital de la Rama Judicial, 2020, <https://www.ramajudicial.gov.co/documents/10228/35666503/INFORME+TRANSFORMACION+C3%92N+DIGITAL+RAMA+JUDICIAL..PDF/53701101-e30c-466b-841a-98fa9fce8e9>

<sup>3</sup> DANE, «Encuesta de Tecnologías de la Información y las Comunicaciones en Hogares», 2022, <https://www.dane.gov.co/index.php/estadisticas-por-tema/tecnologia-e-innovacion/tecnologias-de-la-informacion-y-las-comunicaciones-tic/encuesta-de-tecnologias-de-la-informacion-y-las-comunicaciones-en-hogares-entic-hogares>.

<sup>4</sup> Ana María Ramos Serrano, «La digitalización de la Justicia en Colombia: en busca del tiempo perdido», *Revista Sistemas Judiciales*, n.º 24 (s. f.), <https://sistemasjudiciales.org/wp-content/uploads/2021/10/8.-SJ24.-Ramos-Serrano.pdf>.



personas o la radicación de documentos impresos puedan completarse por medios virtuales. Es decir, implica fundamentalmente la adopción de herramientas capaces de emular la comunicación presencial (como las videollamadas, el correo electrónico o los formularios virtuales), pero en ningún caso –salvo quizá el del sistema PRISMA de la Fiscalía General de la Nación–<sup>5</sup> se pretende reemplazar una decisión autónoma humana mediante un sistema informático.

En este sentido, tanto las condiciones propias de este caso particular, como el marco jurídico para la digitalización de los servicios judiciales del Estado, como las experiencias previas a nivel nacional, son sumamente insuficientes para evaluar el uso de herramientas de Inteligencia Artificial Generativa en general en la rama judicial. En cualquier caso, y a manera de criterio rector, es necesario resaltar que el proceso de digitalización de la justicia en Colombia, sean cuales sean las nuevas herramientas que se adopten, debe hacerse dentro de un marco de derechos humanos que brinde garantías para el acceso efectivo y la no vulneración de los derechos, sopesando los riesgos potenciales de cada nueva tecnología en este ámbito. La falta de evaluación de impacto en derechos humanos resulta preocupante en la medida en que los avances en materia de digitalización hasta ahora se han implementado sin participación pública efectiva ni espacios de discusión democrática amplia, y las decisiones de política pública en este contexto han carecido del sustento necesario. Hasta el presente, carecemos también de cifras sobre la penetración y la efectividad de la digitalización en el sistema judicial y, por lo tanto, no tenemos evidencia sobre la idoneidad de las herramientas utilizadas, ni sobre las nuevas barreras para los usuarios más vulnerables que su introducción haya podido producir.

Dicho esto, pasemos a la evaluación concreta de la herramienta utilizada en este caso y sus posibles consecuencias.

La inteligencia artificial generativa y su uso en la rama judicial

---

<sup>5</sup> Fiscalía General de la Nación y Dirección de Políticas Públicas y Estrategia, «Herramienta PRiSMA: Perfil de Riesgo de Reincidencia para la Solicitud de Medidas de Aseguramiento», 2019, <https://www.fiscalia.gov.co/colombia/wp-content/uploads/Perfil-de-riesgo-de-reincidencia-para-solicitudes-de-medida-de-aseguramiento.pdf>.



# Fundación **Karisma**

Para comprender las implicaciones del uso de inteligencia artificial generativa en decisiones judiciales es necesario entender a qué nos referimos cuando hablamos de Inteligencia Artificial (IA). Este concepto no es unívoco; por el contrario, IA es un término que se usa para describir una vasta serie de sistemas informáticos distintos en su desarrollo, sus fundamentos técnicos, los datos con que se entrenan, sus capacidades y sus usos. Para comprender este caso, nos compete hablar específicamente de un tipo específico de IA, llamado Modelo Grande de Lenguaje (LLM, por sus siglas en inglés). Este tipo de sistema, popularizado mediante la aplicación ChatGPT desarrollada por la empresa OpenAI, es lo que se conoce como una IA capaz de procesar y generar contenidos en lenguaje natural. ChatGPT es particularmente conocida por su faceta generativa, es decir, por su capacidad de producir contenidos nuevos a partir de bases de datos previamente utilizadas para su entrenamiento. ChatGPT produce, a partir de insumos de texto (*prompts*), introducidos por el usuario normalmente en forma de órdenes o preguntas, respuesta en forma de textos cortos en lenguaje natural que parecen responder a la solicitud del usuario.

Y decimos *parecen* porque lo que el modelo de lenguaje hace al procesar un insumo es calcular probabilísticamente la secuencia de palabras más “adecuada” frente a dicho insumo. Para lograr esto, el modelo ha sido alimentado con una inmensa cantidad de textos, que incluyen bases de datos desestructuradas<sup>6</sup> (textos obtenidos de Internet que –aunque es imposible saberlo con certeza– muy probablemente incluyen las sentencias de la Corte Constitucional colombiana, así como noticias al respecto, por ejemplo) y otras específicamente curadas (como por ejemplo colecciones específicas de artículos o libros). Sin embargo, este es sólo el primer paso. Posteriormente, el sistema es entrenado por seres humanos que se encargan de “afinar” las respuestas del modelo<sup>7</sup> para evitar, por ejemplo, que imprima oraciones gramaticalmente mal escritas, lenguaje vulgar o que entregue instrucciones detalladas para fabricar explosivos o hacer planes de suicidio, entre otras<sup>8</sup>.

<sup>6</sup> «How ChatGPT Works? Training Model of ChatGPT», *Edureka* (blog), 18 de febrero de 2023, <https://www.edureka.co/blog/how-chatgpt-works-training-model-of-chatgpt/>.

<sup>7</sup> «Introducing ChatGPT», accedido 8 de septiembre de 2023, <https://openai.com/blog/chatgpt>.

<sup>8</sup> «New and Improved Content Moderation Tooling», accedido 8 de septiembre de 2023, <https://openai.com/blog/new-and-improved-content-moderation-tooling>.



Esta última etapa de entrenamiento es manual y responde a los criterios (morales, de riesgo legal y lingüísticos, entre otros) que han sido definidos de manera privada por la empresa OpenAI para conducir a crear un producto comercialmente viable.<sup>9</sup> Sin embargo, este entrenamiento no está –y probablemente ni siquiera podría estar– dirigido a producir contenidos verídicos, que sean fieles a la realidad o los hechos. El modelo se construye para producir contenidos verosímiles que simulan probabilísticamente –y en lo posible sean indistinguibles de– textos escritos por seres humanos.

La manera de operar de los modelos grandes de lenguaje (LLM) implica que cuando el modelo carece de información adecuada para responder a la pregunta o esta está formulada de manera extraña, la respuesta mantiene la apariencia de ser correcta, en tanto que es gramaticalmente adecuada y generalmente coherente con el insumo, pero puede no estar anclada en ningún dato real y contener aseveraciones falsas. Por ejemplo, puede ofrecer referencias de fuentes inexistentes como artículos académicos que nunca han sido escritos ni publicados o precedentes jurídicos irreales. A este fenómeno por el cual el modelo produce un texto que aparenta ser veraz, pero que no tiene un sustento en la realidad, se lo conoce como una “alucinación”.<sup>10</sup> De manera análoga, las fuentes que se han utilizado para entrenar el modelo incluyen contenido que no pretende ser veraz ni ha sido verificado de manera alguna –como escritos de ficción, secciones públicas de comentarios en la web o material de foros, entre otros– y que de todas formas pueden reflejarse en las respuestas que ofrece la herramienta a los usuarios. Estos casos son problemáticos en la medida en que sólo pueden ser verificados (y en ese sentido corregida la información) manualmente por el propio usuario, si este es riguroso y realiza una investigación subsiguiente que permita constatar la veracidad de la información ofrecida por el modelo. No hay, en el modelo mismo, nada que permita garantizar que sus respuestas sean verídicas ni estén sustentadas en la realidad.

En este sentido, el uso de una herramienta como ChatGPT difiere fundamentalmente del uso de un motor de búsqueda –como el de Google, por ejemplo– en la medida en que el criterio último de qué información es

---

<sup>9</sup> Ver referencia anterior.

<sup>10</sup> «What Is LLMs Hallucinations», Deepchecks, accedido 5 de septiembre de 2023, <https://deepchecks.com/glossary/llms-hallucinations/>.



confiable y resulta seleccionada para ofrecer una respuesta en el primer caso está a cargo del modelo de lenguaje que entrega un texto en apariencia completo, mientras que en el segundo, está en manos del usuario del motor de búsqueda que ha de curar la información que el motor arroja. A manera de analogía, podemos pensar en el motor de búsqueda como una consulta en diversos libros de una biblioteca, en que la persona elige en último término a qué fuentes dar credibilidad sobre la base de la información disponible (quién es el o la autora, qué revista o editorial publicó el material, en qué fecha fue publicado, etc.) y de su criterio, mientras que el modelo de lenguaje funciona como la consulta a un tercero que entrega una información consolidada, con la apariencia de ser real y adecuada, pero sin garantías de ser verídica o confiable. La única manera de corroborar la veracidad y pertinencia de dicha información es una indagación adicional que lo soporte.

En segundo lugar, las respuestas ofrecidas por el modelo de lenguaje pueden incorporar sesgos que el usuario final desconoce. Por ejemplo, en la medida en que está entrenado con información de todo el mundo y el volumen de información disponible sobre el norte global es más amplio que la información del sur global<sup>11</sup>, el sistema podría estar, por ejemplo, favoreciendo líneas jurisprudenciales norteamericanas o europeas por el simple hecho de que su volumen en las bases de datos es mayor. Estos sesgos implícitos son difíciles o imposibles de notar para el usuario final, pues no existe una manera de rastrear las fuentes específicas que el modelo usa para ofrecer cada una de sus respuestas y, en la medida en que sus respuestas están escritas de manera correcta y verosímil, percibir su carácter tendencioso, si lo hay, requiere un análisis crítico.

Esto nos lleva a la principal conclusión con respecto al uso de herramientas basadas en modelos grandes de lenguaje como ChatGPT y es que la responsabilidad por los fallos judiciales –su sentido y argumentos– recae siempre e indefectiblemente en el juez. Esto sin perjuicio de la discusión sobre la responsabilidad de las empresas que ofrecen este tipo de herramientas sobre los contenidos que los modelos grandes de lenguaje producen o la forma en la que promocionan las capacidades de la herramienta.

---

<sup>11</sup> «Usage Statistics and Market Share of Content Languages for Websites, September 2023», accedido 8 de septiembre de 2023, [https://w3techs.com/technologies/overview/content\\_language](https://w3techs.com/technologies/overview/content_language).



En lo sucesivo, al responder a cada una de las preguntas de la honorable Corte Constitucional, ahondaremos en este y otros aspectos del uso de herramientas de inteligencia artificial generativa en las decisiones judiciales.

## 2. Respuesta a las preguntas enviadas por la Honorable Corte Constitucional

¿Cuáles estima son las implicaciones de utilizar herramientas de inteligencia artificial generativas en la sustanciación de decisiones judiciales? ¿Considera que el uso de este tipo de herramientas de inteligencia artificial trae beneficios en la materialización del derecho de acceso a la administración de justicia y la eficiencia en el trabajo de los jueces?

Si bien la congestión del sistema judicial en Colombia puede constituir una barrera para la garantía del derecho de acceso a la justicia<sup>12</sup>, es fundamental que los remedios utilizados para aliviarla no produzcan un impacto negativo sobre la calidad de las decisiones judiciales. En el presente caso, seleccionado por la Honorable Corte Constitucional por haberse usado el modelo de lenguaje grande ChatGPT para fundamentar una decisión judicial, el sentido de dicha decisión (la garantía del acceso a servicios de salud por parte de un menor con diagnóstico de autismo) resulta poco controversial. La tutela, en este caso, cumplió su objetivo y la argumentación que sustenta la decisión es adecuada y suficiente. Esto, porque la subregla jurisprudencial sí consistía en la exoneración de copagos para personas con autismo<sup>13</sup> y porque las sentencias de tutela son de breve motivación cuando no revoquen o modifiquen el fallo, no unifiquen la jurisprudencia constitucional ni aclaren el alcance general de las normas constitucionales<sup>14</sup>. Por esta razón, a pesar de haberse servido el juez del modelo ChatGPT, consideramos que el presente caso no ilustra de manera concreta los riesgos posibles del uso de esta tecnología en la sustanciación de decisiones.

---

<sup>12</sup> La Corte Constitucional ha mencionado que hacen parte de los derechos al debido proceso y al acceso a la justicia los hechos de obtener una respuesta oportuna frente a las pretensiones que se hayan formulado; y que no se incurran en omisiones o dilaciones injustificadas en las actuaciones judiciales. Sentencias T-286 de 2020 y SU-179 de 2021, entre otras.

<sup>13</sup> Corte Constitucional, sentencias T-674 de 2016, T-409 de 2019, T-563 de 2019, entre otras.

<sup>14</sup> Decreto 2591 de 1991, artículo 35.



De evaluarse todas las implicaciones del uso de inteligencia artificial generativa dentro de la rama judicial únicamente con base en este caso, los riesgos quedarían probablemente subestimados, en la medida en que en esta ocasión, en función de la simplicidad del fallo y de las circunstancias particulares del caso, dichos riesgos afortunadamente no se materializaron.

Un análisis superficial del presente caso podría considerarse como una instancia en que esta herramienta de inteligencia artificial generativa sirvió al juez para llegar a una decisión adecuada de manera más expedita y eficaz, pero ha de tenerse en cuenta que un uso responsable de la herramienta implica siempre que el usuario señale con claridad que usó la herramienta y cuáles fueron los resultados, y haga un control posterior consciente de los resultados arrojados por la herramienta. Este paso crucial puede resultar tanto o más exigente que elaborar la argumentación misma del fallo y su omisión puede implicar que la decisión quede sustentada, por ejemplo, en precedentes inexistentes, irrelevantes para el caso o insuficientes para su sustentación. En este sentido, los beneficios en términos de eficacia, si la herramienta se utiliza de manera responsable, probablemente no constituyen un ahorro significativo en términos de tiempo o esfuerzo, en la medida en que de todas formas tendrán que ser revisados posteriormente.

Más aún, la facilidad de uso y la apariencia de verosimilitud de las respuestas que ofrece el modelo pueden llevar a que los usuarios desestimen la relevancia de contrastar y corroborar los resultados obtenidos dándolos por ciertos, lo que hace más probable que los riesgos se materialicen.

¿Considera que, en cambio, acarrea riesgos respecto de los derechos fundamentales, en especial, el derecho al debido proceso?

Aunque no es posible responder a esta pregunta de manera absoluta, pues como se explicó más arriba los riesgos están asociados al uso que se hace del modelo antes que al modelo mismo, es importante reconocer que un mal uso sí puede acarrear consecuencias graves en términos de derechos.

Consideramos que, en la medida en que ningún sistema de IA generativa (incluido, por supuesto, ChatGPT) puede garantizar que sus productos son



veraces, libres de sesgo y adecuados al contexto y, más aún, en tanto que *no* son una fuente de derecho establecida, es claro que la responsabilidad sobre el contenido de un fallo –como sucedería con cualquier otro producto escrito con la asistencia de uno de estos modelos– permanece ineludiblemente sobre el usuario, en este caso el juez, que firma el fallo como propio. En este sentido, si el contenido de la sentencia proferida tiene errores, inconsistencias o, por ejemplo, cita fuentes inexistentes o irrelevantes, estos errores deben reconocerse como faltas del juez y, bajo ninguna circunstancia, puede este excusarse en el uso de la herramienta como un eximente de su responsabilidad.

Es indispensable comprender que cualquier situación en que el usuario desplace la responsabilidad, por negligencia, desconocimiento, descuido o cualquier otra causa al modelo de lenguaje en que se apoyó para elaborar su decisión, incrementa los riesgos de que dicha decisión sea de pobre calidad. En lo relativo al debido proceso, específicamente, consideramos que puede verse comprometido en la medida en que el juez, si desplaza la carga al modelo de lenguaje, está desatendiendo su responsabilidad de decidir y motivar la sentencia, establecidas en los numerales 6 y 7 del artículo 42 del Código General de Proceso (Ley 1564 de 2012) sobre los deberes del juez.

Finalmente es importante resaltar que el uso de herramientas de Inteligencia Artificial generativa es, en un sentido práctico, imposible de detectar con certeza. Aún cuando existen herramientas tecnológicas para determinar si un texto fue escrito utilizando una de estas herramientas o no, eludirlas y engañarlas resulta tan simple como parafrasear o retocar la respuesta obtenida. Por este motivo, una prohibición absoluta sobre el uso de estos sistemas no sólo sería inconducente, sino también imposible de verificar en la realidad e implicaría una carga aún mayor sobre la revisión posterior de los fallos proferidos. En el mismo sentido, existe un Código General Disciplinario así como una Ley Estatutaria de la Administración de Justicia que contemplan los mecanismos disciplinarios para los casos en que jueces de la república desatiendan o excedan sus responsabilidades y que por sí mismos permiten atender la mayoría de las situaciones que pueden presentarse a raíz de un mal uso de la herramienta.

Como mencionamos más arriba, este caso no es idóneo para establecer nuevos lineamientos generales con respecto al uso de estas tecnologías de



inteligencia artificial en la rama judicial en general, en la medida en que el caso concreto no constituye un ejemplo paradigmático de uso erróneo o irresponsable de la herramienta y su utilización no produjo consecuencias negativas para el demandante. Para construir soluciones más estructurales en este tema es necesario abordarlo en una discusión regulatoria más amplia, en el marco de un proceso participativo en el que diversos actores pueden hacerse parte, y en el que se analice de forma detallada el impacto en derechos humanos que pueda tener adoptar una tecnología en el ejercicio de los operadores judiciales.

¿Considera que el uso de estas herramientas afecta la independencia del juez?

Nuevamente, para responder a esta pregunta es necesario hacer un análisis preciso del uso que el juez le da a la herramienta. Tal como sucede en otros los casos en que el juez se sirve de recursos externos para tomar decisiones (como la consulta a terceros o la revisión de fuentes bibliográficas, entre otros) éste debe garantizar que la información socavada sea pertinente para tomar la determinación y no comprometa su independencia.

Si el juez está haciendo caso omiso de su propio deber como decisor, la independencia se verá afectada, el que el uso de la herramienta esté acompañado por una revisión juiciosa y responsable atenúa –aunque no anula, pues aún pueden persistir sesgos ocultos para el juez mismo– los factores que atentarían contra la independencia del juez, como la injerencia de presiones externas.

Dicho esto, se debe partir por el entendido de que la debida administración de justicia es un presupuesto del Estado Social de Derecho y de la democracia constitucional. Su materialización, además de garantizar un derecho fundamental en sí mismo, protege y hace efectivo otros derechos, libertades y garantías de quienes recurren al aparato de justicia para la solución de disputas y define las cargas y deberes en cabeza de la administración y la ciudadanía<sup>15</sup>.

---

<sup>15</sup> Corte Constitucional, Sentencia C-037 de 1996.



Esta visión, desplazada a una interpretación vigente del ordenamiento jurídico colombiano, pone al juez en el centro del proceso y le otorga las facultades de dirigir a las partes y orientar y mediar las disputas en pro de la justicia y ante la prevalencia del ordenamiento jurídico. Así, el juez abandona “su papel estático, como simple observador y mediador dentro del tráfico jurídico”, y se convierte “en un partícipe más de las relaciones diarias de forma tal que sus fallos no sólo sean debidamente sustentados desde una perspectiva jurídica, sino que, además, respondan a un conocimiento real de las situaciones que le corresponde resolver”.<sup>16</sup>

Esta perspectiva que se sostiene desde de la concepción de la administración de justicia en la jurisprudencia constitucional, evoluciona para adaptarse al contexto que se requiera para fallar, reflexivamente, en derecho. El juez se convierte entonces, para la Corte Constitucional, en un “funcionario -sin vendas- que se proyecta más allá de las formas jurídicas, para así atender la agitada realidad subyacente y asumir su responsabilidad como un servidor vigilante, activo y garante de los derechos materiales”<sup>17</sup>. De allí que, en el mandato por alcanzar la justicia material en el orden constitucional, los jueces deban procurar por: i) obtener el derecho sustancial, y ii) buscar la verdad en sus fallos.<sup>18</sup> Esta aproximación a la verdad, aunque ontológicamente inviable, jurídicamente corresponde a “un principio y un derecho constitucional que se impone a las autoridades y a los particulares”.<sup>19</sup>

De esta manera, la política judicial en Colombia debe estar orientada por las nociones sobre la materialización de la administración de justicia en los términos establecidos por la Corte Constitucional. No es gratuito que al establecer los lineamientos para la digitalización de la justicia en el Decreto 806 de 2020 y posteriormente en la Ley 2213 de 2022 se persiga el fin de “implementar el uso de las tecnologías de la información y las comunicaciones en las actuaciones judiciales” para agilizar el trámite de los procesos judiciales y flexibilizar la atención para los usuarios del sistema de administración de justicia. Este mandato, que la Corte Constitucional desarrolla en las sentencias C-420 de 2020 y C-121 de 2023, responde a la necesidad de establecer y adaptar el uso de tecnologías de la información,

---

<sup>16</sup> *Ibid.*

<sup>17</sup> Corte Constitucional, Sentencia SU-768 de 2014

<sup>18</sup> *Ibid.*

<sup>19</sup> Corte Constitucional, Sentencia C-396 de 2007.



precisamente, para la satisfacción de garantías procesales y, en consecuencia, la materialización del acceso a la administración de justicia.

No puede perderse de vista que la evolución y adopción de herramientas tecnológicas en la actividad judicial deben seguir de cerca estándares constitucionales y de derechos humanos para la garantía del derecho a la administración de justicia y el debido proceso, así como de los demás derechos que se tramitan en el curso de los procesos.

La decisión del Juzgado 1º Laboral del Circuito de Cartagena recurre al uso del aplicativo ChatGPT, chatbot de inteligencia artificial (IA en adelante) generativa, para “extender los argumentos de la decisión adoptada” bajo la justificación de la “incorporación de las TIC en los procesos judiciales” establecida en la Ley 2213 de 2022. La motivación reducida de la incorporación de una herramienta tecnológica como ChatGPT para el reconocimiento del derecho a la salud de un menor de edad con discapacidad sin dar cuenta de las implicaciones del uso de IA generativa dentro de la actividad judicial, puede llegar a desconocer los fundamentos constitucionales sobre la materialización de la administración de justicia, especialmente si su uso se extiende indiscriminadamente a fallos futuros.

El uso de IA generativa en la administración de justicia sin salvaguardas puede afectar las garantías del debido proceso al producir desconfianza sobre la fiabilidad de las decisiones, promover que fácticamente se de un reemplazo del juez por la herramienta y la cristalización de los fallos dentro de los sistemas judiciales. Los algoritmos que han sido utilizados en otros países por la administración de justicia y de los cuáles se han estudiado sus efectos, presentan una diferencia con ChatGPT, son herramientas de análisis de riesgo diseñadas para ello que se usan como apoyo en toma decisiones, aún así, también funcionan a través de aprendizaje automático y procesamiento de lenguaje natural. El uso de estas herramientas ha generado afectaciones a los derechos humanos, y el aprendizaje sobre los riesgos que representan puede ser tenido en cuenta para evaluar las condiciones en que una inteligencia artificial generativa que se soporte en LLMs no diseñada para la asistencia de procesos judiciales se use en la administración de justicia.

Es necesario partir del hecho de que los resultados que dan las tecnologías de aprendizaje automático no siempre van a ser correctos, justos o



acertados con la realidad. Las IAs generativas logran predecir las palabras que con mayor posibilidad siguen a otro conjunto de palabras por el volumen de datos y texto que se les ha suministrado durante su entrenamiento,<sup>20</sup> sus limitaciones que se manifiestan en alucinaciones (hallucinations), un fenómeno en el cual el sistema proporciona respuestas que son incorrectas, irrelevantes o sin sentido.<sup>21</sup> Estas deficiencias han sido reconocidas por el propio sistema, el cual en su plataforma de ingreso plantea que puede generar información incorrecta, producir instrucciones dañinas o con contenido sesgado y tiene un conocimiento limitado de eventos posteriores al 2021.<sup>22</sup>

Una de las primeras formas en las que se ha identificado que el uso de inteligencia artificial en el sistema judicial puede afectar derechos humanos es por los sesgos que las herramientas contienen, replican y perpetúan. Por ejemplo, en las revisiones del Harm Assessment Risk Tool (HART)<sup>23</sup>, se reveló que el hecho de que los resultados se den teniendo en cuenta datos como el lugar en donde vive el sujeto, su género o la comunidad a la que pertenece, incide en que los pronósticos individuales en realidad reflejan privaciones de la comunidad.<sup>24</sup> A su vez, al ser datos históricos estos incluyen sesgos que los sistemas policiales ya tenían como la hipervigilancia a ciertas comunidades o la focalización de la investigación a determinados delitos.<sup>25</sup> Lo mismo fue identificado en el caso *Loomis v. Wisconsin* por la Corte Suprema de Wisconsin en Estados Unidos reconociendo que, si bien la predicción de COMPAS no era determinante para la sentencia, al basar sus puntajes en datos grupales, de los cuales no se conocía la forma en la que eran analizados por la IA, había terminado prediciendo de forma desproporcionada que individuos pertenecientes a minorías raciales tenían un más alto riesgo de incidencia.<sup>26</sup>

<sup>20</sup> New York Times, “How Does ChatGPT Really Work?”, <https://www.nytimes.com/2023/03/28/technology/ai-chatbots-chatgpt-bing-bard-llm.html>

<sup>21</sup> New York Times, “ Artificial Intelligence Glossary: Neural Networks and Other Terms Explained”, <https://www.nytimes.com/article/ai-artificial-intelligence-glossary.html>

<sup>22</sup> OpenAI, ChatGPT, <https://chat.openai.com/>

<sup>23</sup> Este es un sistema utilizado por la policía de Durham en Inglaterra con el fin de predecir la posibilidad de reincidencia de una persona, basa sus decisiones en información de retenciones pasadas, historial criminal de la persona y otros datos como su género o código postal. Ver. Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality, *Information & Communications Technology Law*, 27:2, 223-250. DOI: [10.1080/13600834.2018.1458455](https://doi.org/10.1080/13600834.2018.1458455). Pág. 228.

<sup>24</sup> Marion Oswald, et al. Pág. 235.

<sup>25</sup> Marion Oswald, et al. Pág. 235.

<sup>26</sup> *State of Wisconsin v Eric L. Loomis* (n 42) Párr. 66.



Estos sesgos que generan discriminación, también se encuentran presentes en inteligencias generativas como ChatGPT. La propia plataforma ha establecido que sus resultados pueden resultar sesgados, toda vez que esta herramienta produce datos de salida a través de la generalización que le permite aplicar el conocimiento utilizado en su entrenamiento a entradas nuevas, lo que puede incidir en que se perpetúen los sesgos que están presentes en sus datos de entrenamiento produciendo respuestas basadas en estereotipos o favorecimiento de ciertos grupos o ideas.<sup>27</sup> Sobre esto, es importante añadir que ChatGPT fue desarrollado con un fin multipropósito y por un privado, como consecuencia, no es claro cuál es la información a partir de la cual fue entrenada y qué sesgos puede contener. A ello se debe sumar que la herramienta no referencia las fuentes que utiliza para brindar sus respuestas, lo que dificulta aún más la posibilidad del usuario, en el caso de estudio el juez, de identificar estos sesgos.

Los sesgos no se predicen sólo de los resultados o predicciones que puede dar una IA sino de todo su ciclo de vida, pues la creación, diseño e implementación de algoritmos implícitamente comprende juicios de valores esenciales. Por ejemplo, dentro de la creación de HART se tuvo en cuenta la posibilidad de que el sistema podría incluir errores en sus resultados, por lo cual, se les asignó un valor dependiendo del impacto que tenían al generar falsos positivos o falsos negativos.<sup>28</sup> A los errores que producían falsos positivos se les dio un menor valor con el fin de aumentar deliberadamente la probabilidad de que ocurrieran estos y disminuir la existencia de falsos negativos.<sup>29</sup> Si la herramienta fuera determinante para la toma de decisiones, un falso negativo podría reflejarse en la libertad de una persona que tiene una alta probabilidad de cometer un delito violento y grave prontamente.<sup>30</sup> Si bien tiene sentido privilegiar la seguridad, no debe perderse de vista que esa elección también posibilitaba que un número considerable de personas de bajo o mediano riesgo fueran clasificadas de forma equivocada.<sup>31</sup>

<sup>27</sup> Emilio Ferrara. (s/f). *Should ChatGPT be biased? Challenges and risks of bias in large language models*. Arxiv.org. <http://arxiv.org/abs/2304.03738>

<sup>28</sup> Marion Oswald, et al (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality. Pág. 236.

<sup>29</sup> Marion Oswald, et al (2018). Pág 236.

<sup>30</sup> Marion Oswald, et al (2018). Pág 236.

<sup>31</sup> Marion Oswald, et al (2018). Pág 236.



La elección de privilegiar un tipo de error es solamente una de las muchas decisiones que componen el ciclo de vida de una IA, y la producción de sesgos puede estar implícita en cualquiera. En este punto, es necesario considerar que ChatGPT no es una herramienta licenciada para la administración de justicia colombiana y que sus respuestas, por sí mismas, no corresponden a un ejercicio razonado de ponderación en derecho. Esto implica que no se tiene un control sobre su diseño, conocimiento sobre la información por la que fue entrenada y la forma en la que lo procesa, y acceso a conocer o modificar su algoritmo. El uso de un sistema sobre el que no se tiene ningún tipo de control y además un conocimiento limitado, representa mayores retos para la identificación de sesgos u otros riesgos para los derechos y libertades.

Otro aspecto que pone en juego la fiabilidad es que las IAs se equivocan, por ejemplo, en Reino Unido se utilizó una IA para determinar la capacidad económica de los ex-cónyuges en procesos de alimentos, por un error en su configuración entre abril del 2011 y enero del 2012 ésta interpretaba deudas como si fueran activos, y en 2.200 casos se emitieron sentencias basadas en esa información.<sup>32</sup> ChatGPT no es la excepción a esta regla, la propia plataforma reconoce que puede arrojar respuestas erróneas y que la información que utiliza para estas se encuentra limitada hasta el 2021, esto implica que no es una fuente confiable de información.<sup>33</sup> La crítica respecto de los errores que puede cometer un sistema de aprendizaje automático no se trata solo de la necesidad de solucionarlos con la búsqueda de mayor precisión, sino que son resultados que causen afectaciones graves a vidas humanas si la IA, se toma como un proveedor de información certera y fiable para el apoyo en toma de decisiones judiciales en las que están en juego otros derechos.<sup>34</sup>

Todas estas condiciones que afectan la fiabilidad de las decisiones judiciales se acentúan por la falta de transparencia y la opacidad del procesamiento que las IAs hacen con la información de entrada para llegar a las predicciones o resultados. Las IAs son generalmente desarrolladas por instituciones privadas, por lo cual, un interés sobre su propiedad impide

<sup>32</sup> A. D. Reiling, "Courts and Artificial Intelligence," *International Journal for Court Administration* 11, no. 2 (2020): 1-10. Pág. 7.

<sup>33</sup> Juan David Gutierrez. ChatGPT in Colombian Courts. En *Verfassungsblog*. Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung. <https://doi.org/10.17176/20230223-185205-0>

<sup>34</sup> Marion Oswald, et al (2018). Pág. 233.



muchas veces conocer cómo funcionan y el manejo que le dan a la información recolectada de los usuarios.<sup>35</sup> La falta de transparencia se hace particularmente importante en el ámbito de administración de justicia, pues la capacidad de razonamiento es intrínseca a las decisiones jurídicas.<sup>36</sup> Esta opacidad plantea desafíos para la capacidad del individuo de conocer los factores que su juzgador tuvo en cuenta en la toma de decisiones y, como consecuencia, retos para cuestionar el proceso.<sup>37</sup> Primero, afecta la posibilidad de las personas de conocer a su juzgador, y esto es lo que permite evaluar la competencia e identificar si debe hacerse algún tipo de recusación<sup>38</sup>, sin embargo, por la ausencia de transparencia en su uso no se tiene la capacidad de conocer los sesgos que podrían tener las consideraciones del fallo. Segundo, respecto de la impugnación, si no se conoce la información y la forma en la que fue procesada, se elevan las dificultades para cuestionar las decisiones tomadas a partir de un resultado de la IA.<sup>39</sup> La opacidad en muchos casos no solo se predica frente al individuo que es sujeto de una decisión, sino también de los operadores judiciales que por la falta de capacidad que tienen de comprender el funcionamiento del algoritmo, pueden asignar a la información que les provee un peso equivocado en su decisión.<sup>40</sup>

Se ha afirmado que para contrastar las deficiencias antes mencionadas, se exige que el algoritmo solo informe o recomiende y sea el operador judicial quien siempre tome la decisión<sup>41</sup>. Sin embargo, se ha visto con preocupación que de forma fáctica la autonomía del juez se pueda comprometer por el uso de la herramienta. Como ya se dijo, puede que el juez o tribunal no tenga conocimiento de la forma en la que un algoritmo funciona y qué factores tiene en cuenta para responder sus cuestionamientos. Esto es especialmente importante si las preguntas que el usuario hace no buscan obtener información en la cual basarse para contextualizar sus consideraciones, sino que pretenden definir cuál es el sentido del fallo. Por ejemplo, el juez de segunda instancia preguntó a

---

<sup>35</sup> A. D. Reiling, Pág 5.

<sup>36</sup> Marion Oswald, et al. Pág. 241.

<sup>37</sup> Marion Oswald, et al. Pág. 240.

<sup>38</sup> Corte IDH. *Caso Castillo Petruzzi y otros Vs. Perú*. Sentencia de 30 de mayo de 1999. Fondo, reparaciones y costas. Serie C No. 52. Párr. 133.

<sup>39</sup> Marion Oswald, et al. Pág. 240.

<sup>40</sup> Lorna McGregor; Daragh Murray; Vivian Ng, "International Human Rights Law as a Framework for Algorithmic Accountability," *International and Comparative Law Quarterly* 68, no. 2 (Abril 2019): 309-344. Pág. 317.

<sup>41</sup> Lorna McGregor; et al. Pág. 317.



ChatGPT ¿si las acciones de tutela en estos casos se deben conceder?, es una pregunta que usa la herramienta fuera de sus limitaciones y capacidades, lo que puede comprometer su criterio y los derechos de las partes. En ese sentido, ¿cómo puede el juez tomar una decisión independiente, autónoma y en derecho cuando no es consciente de los factores que intervienen en la respuesta que arroja la IA? Los operadores judiciales, por el sesgo automatización que tenemos los humanos,<sup>42</sup> pueden ver limitada su capacidad de discernimiento porque están sujetos a creer que los resultados arrojados por máquinas son automáticamente correctos.<sup>43</sup>

Ahora bien, respecto de la cristalización de las decisiones, ponemos la atención sobre que estas herramientas se están utilizando en un contexto cambiante, aún cuando, generan sus resultados a partir del procesamiento de datos sobre juicios humanos que ya se han hecho en el pasado.<sup>44</sup> La realidad humana se transforma constantemente, igual que lo hacen los valores y principios que en determinados momentos las sociedades privilegian. La información con la que las IAs son entrenadas y realizan sus predicciones no se renuevan de forma automática, eso implica que estos sistemas requieren de un proceso constante de actualización y revisión so pena de que sus predicciones inevitablemente se conviertan en anacrónicas.<sup>45</sup> Aún de existir una herramienta que se actualice con esa agilidad, los resultados del sistema de aprendizaje automático no dejan de basarse en datos históricos, lo que en últimas implica que no hay lugar a la innovación dentro del sistema.<sup>46</sup> Puede que sea cierto que ChatGPT al ser una herramienta generativa brinde respuesta ante situaciones novedosas, sin embargo lo hace a través de un proceso de generalización que le permite aplicar el conocimiento utilizado en su entrenamiento a entradas nuevas, más no por medio de innovación.<sup>47</sup>

---

<sup>42</sup> Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality, *Information & Communications Technology Law*, 27:2, 223-250. DOI: [10.1080/13600834.2018.1458455](https://doi.org/10.1080/13600834.2018.1458455). Pág. 238.

<sup>43</sup> Marion Oswald, et al. Pág. 241.

<sup>44</sup> Marion Oswald, et al. Pág. 235.

<sup>45</sup> Marion Oswald, et al. Pág. 235.

<sup>46</sup> Marion Oswald, et al. Pág. 235.

<sup>47</sup> Emilio Ferrara. (s/f). *Should ChatGPT be biased? Challenges and risks of bias in large language models*. Arxiv.org. <http://arxiv.org/abs/2304.03738>



¿Conoce ejemplos nacionales o internacionales en los que se ha promovido o ejecutado el uso de herramientas de inteligencia artificial para la sustanciación de decisiones judiciales o se hayan proferido lineamientos o protocolos al respecto? ¿En qué ha consistido esta promoción o cómo se ha regulado el uso de estas herramientas?

En la jurisprudencia comparada encontramos dos casos dignos de mención: un caso en el Perú en que un juez se sirvió de la herramienta para hacer un cálculo matemático que le permitiera determinar la responsabilidad económica de un padre con respecto a la manutención de su hija y un caso en México en el cual un magistrado del Tribunal Electoral estaba determinando si el uso de un eslogan político particular (“ya sabes quién”) podía constituir un desequilibrio en la campaña; en este caso, el magistrado se sirvió de la herramienta para consultar sobre las connotaciones de dicho eslogan en el contexto mexicano.<sup>48</sup>

Ambos casos, aunque interesantes, son apenas incipientes y difíciles de comparar con el caso presente en que el fallo del juez está directamente informado por una respuesta provista por la herramienta. Sin embargo, un par de elementos son destacables: el primero es que el uso de la herramienta sólo ha salido a luz cuando es declarado por los jueces mismos. Esto podría implicar que ya ha sido utilizada en más casos, sin que su uso haya sido explicitado. Lo anterior lleva a pensar que prohibir o desmotivar su uso por medio de represalias puede tener como consecuencia el ocultamiento de su uso y un funcionamiento menos transparente de la justicia, antes que una disminución real de las instancias en que la herramienta es efectivamente utilizada. Lo que, a su vez, tiene como consecuencia que los riesgos propios del uso de la herramienta (como los sesgos implícitos o la cristalización de las decisiones abordada más atrás) sean más difíciles de detectar y minimizar. En este sentido, la transparencia con respecto al uso es un pilar fundamental.

En segundo lugar, en ambos ejemplos se evidencia un desconocimiento por parte de los jueces de la manera en que la herramienta opera. En el caso peruano, el juez adjuntó un enlace que no conducía a la búsqueda

---

<sup>48</sup> «A jueces y magistrados de Perú y México también les cayó la “fiebre” de ChatGPT - Agenda Estado de Derecho», 18 de abril de 2023, <https://agendaestadodederecho.com/a-jueces-y-magistrados-de-peru-y-mexico-tambien-les-cayo-la-fiebre-de-chatgpt/>.



que él hizo en la plataforma –por ser esta privada– lo que implicó que no se pudiera revisar el uso que hizo de la herramienta. Por su parte, en el caso mexicano, el magistrado parece desconocer que la herramienta puede responder de manera distinta según el contexto y la formulación específica de la consulta, pues no especifica qué consulta hizo y parece desconocer el que pequeñas variaciones en los insumos pueden ir produciendo en cada caso resultados distintos. Esto da lugar a pensar que hubo una aproximación peligrosamente poco crítica a los resultados obtenidos mediante la herramienta, tomándolos como ciertos en la medida en que, en palabras del magistrado, “nos facilita una serie de información procesando obviamente bases de datos todo el conocimiento que está a disposición de los tribunales” [sic].<sup>49</sup>

Por lo pronto no parece existir jurisprudencia suficiente que de líneas claras al respecto de cómo moderar o regular el uso de estas herramientas en la Rama Judicial y, como hemos resaltado enfáticamente hasta aquí, a nuestro juicio, el presente caso no contiene los elementos necesarios para avanzar una decisión de fondo sobre el uso de esta tecnología en la rama judicial colombiana.

¿Considera pertinente el uso de herramientas de inteligencia artificial generativas en la sustanciación de sentencias judiciales? O, en cambio, ¿cree que debería restringirse o prohibirse su uso en este contexto? Explique la razón de su respuesta.

El uso que puede hacerse de una herramienta de inteligencia artificial generativa en el marco de la sustentación de una sentencia judicial es amplio y diverso: la herramienta puede utilizarse para indagar información, para revisar precedentes, para construir formatos o para redactar textos y tomar decisiones. Cada uno de estos usos implica riesgos e impactos diferentes en los derechos de las personas vinculadas con el proceso. Para cada uno de estos usos, a su vez, pueden concebirse parámetros de uso razonable y responsable que conduzcan a minimizar dichos riesgos.

Consideramos que 1) en ningún caso un sistema automatizado de decisión debe reemplazar la función del juez y que la responsabilidad individual del mismo no se ve, de manera alguna, afectada o diezmada cuando este se

---

<sup>49</sup> «A jueces y magistrados de Perú y México también les cayó la “fiebre” de ChatGPT - Agenda Estado de Derecho».



ha servido de una herramienta como ChatGPT para tomar una decisión, 2) que el uso irresponsable o acrítico de este tipo de herramientas trae consigo riesgos importantes a la garantía de derechos y por lo mismos debe desincentivarse a través de guías de buenas prácticas y mecanismos de fomento de la transparencia y 3) que la prohibición del uso de sistemas de inteligencia artificial generativa es imposible de hacer cumplir en la práctica y tendría como consecuencia el ocultamiento de su uso, lo que redundaría en una dificultad mayor para hacer control de la calidad de las decisiones judiciales. En concordancia con lo anterior, sugerimos que la restricción de su uso se de a manera de política pública de salvaguardias que permita a los funcionarios de la rama judicial conocer los riesgos que implican estas herramientas y conducirse de acuerdo a unos mínimos necesarios para reducirlos tanto como sea posible.

Es concebible que en el futuro se presenten otros casos que ameriten una regulación más específica o robusta con respecto a estas tecnologías, pero adelantarse a ello con base en unos hechos jurídicos que no dan pie a tal desarrollo puede resultar contraproducente o demostrarse insuficiente en un breve periodo.

¿Qué recomendaciones haría para desarrollar un marco destinado al buen uso de herramientas de inteligencia artificial para apoyar la sustanciación de decisiones judiciales?

En vista de las deficiencias expuestas antes, pero atendiendo al hecho de que la IA puede resultar útil en algunos escenarios para la administración de justicia, abogamos porque su uso se dé siempre bajo la aplicación de estrictas recomendaciones y salvaguardas que permitan la protección de los derechos humanos. El reto de la regulación de las IAs en el mundo ya ha tenido algunos avances importantes, organizaciones multilaterales como la UNESCO y la Comisión Europea por la Eficiencia de la Justicia del Consejo Europeo (CEPEJ) han desarrollado una serie de principios éticos para el uso de IAs. A su vez, el Parlamento Europeo se encuentra tramitando una Ley de IA con el fin de regular el diseño, creación y uso de esta tecnología en la Unión Europea.

Los tres instrumentos concuerdan en la necesidad de que salvaguardas se prediquen de todo el ciclo de vida de las IAs, en especial cuando conlleven un alto riesgo para los derechos humanos. La revisión de los impactos que



estas tecnologías tienen sobre los derechos humanos permiten ver que la afectación no solo se da en torno a los resultados de los algoritmos, sino que también puede estar presente en su diseño y proceso de entrenamiento.<sup>50</sup> Incluso, tener en cuenta las salvaguardas desde las etapas más tempranas de vida de una IA, permitirá que su creación, diseño e implementación se dé desde un enfoque preventivo de vulneración de derechos y minimicen los riesgos de afectación posteriores.<sup>51</sup>

Las IA generativas como ChatGPT representan retos particulares para la aplicación de salvaguardas, pues al ser inteligencias artificiales multipropósito se propicia su uso en distintos contextos y por una gran cantidad de usuarios. Su gran escala de uso lleva a que aún se este proceso de dimensionar el impacto que puede tener en la sociedad y en los derechos humanos. Al mismo tiempo, por la cantidad de tareas y áreas en las que puede ser utilizada su impacto termina siendo determinado por quien sea su usuario final o consumidor.<sup>52</sup> Es decir, podría ser utilizado en ámbitos en donde tiene poco o nulo impacto, por ejemplo, una persona que busca mejorar su manejo del tiempo y le pide a ChatGPT que le ayude a organizar una rutina, u otros de gran impacto, como su uso en el apoyo para toma de decisiones judiciales.

En torno al escenario antes expuesto, la Honorable Corte Constitucional debe tener en cuenta que el uso de estas herramientas en la administración de justicia debería contemplar, como mínimo y de forma no excluyente, una serie de salvaguardas consistentes en:

1. La idoneidad de las herramientas utilizadas y la necesidad de salvaguardas con enfoque en la formulación de una política pública del uso de herramientas de inteligencia artificial en la administración de justicia

Los instrumentos regulatorios antes mencionados coinciden en que el uso de herramientas de inteligencia artificial en áreas consideradas de alto riesgo, como la administración de justicia, debe

<sup>50</sup> Lorna McGregor; Daragh Murray; Vivian Ng, "International Human Rights Law as a Framework for Algorithmic Accountability," *International and Comparative Law Quarterly* 68, no. 2 (Abril 2019): 309-344.

<sup>51</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021.

<sup>52</sup> Helberger, N. & Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*, 12(1). <https://doi.org/10.14763/2023.1.1682>



estar supeditado a que estas sean idóneas para el fin propuesto.<sup>53</sup> En ese sentido, la UNESCO considera que las tecnologías de la IA no garantizan por sí sola la prosperidad de los seres humanos, por lo que la decisión de utilizar estas tecnologías debería justificarse eligiendo un método adecuado, proporcional y basado en fundamentos científicos rigurosos.<sup>54</sup> En razón a esto, se promueve que el uso de salvaguardas se de durante todo el ciclo de vida de la IA pues, como fue explicado antes, las afectaciones que pueden causar sobre derechos humanos no sólo se predicen de sus resultados finales.<sup>55</sup>

El llamado a la búsqueda de herramientas tecnológicas idóneas no es únicamente del plano internacional, sino que la propia norma que invoca el juez de segunda instancia en su fallo plantea que “se podrán utilizar las tecnologías de la información y de las comunicaciones, cuando se disponga de los mismos de manera idónea”.<sup>56</sup> Esta idoneidad debe estar orientada también a garantizar la fiabilidad e independencia de las decisiones tomadas por los jueces, y la garantía de responsabilidad en su uso. Como consecuencia, una de las primeras salvaguardias implica que el uso de estas herramientas en la administración de justicia se dé en el marco de la implementación de una política pública que instruya a los operadores judiciales sobre la utilización de los sistemas de IA velando porque se respeten los derechos humanos, la independencia judicial y se satisfagan los intereses de la justicia. Esta debería desarrollar, a lo menos, las salvaguardas que se plantean a continuación y plasmar estándares que permitan identificar y definir cuáles herramientas de IA pueden ser utilizadas por su idoneidad en la administración de justicia.<sup>57</sup>

## 2. Los deberes de transparencia

---

<sup>53</sup> Comisión Europea. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. 2022/14954. 25 de noviembre del 2022.

<https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf>

<sup>54</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 26.

<sup>55</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 63.

<sup>56</sup> Congreso de la República de Colombia. Ley 2213 de 2022. Artículo 2.

<sup>57</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 57.



La transparencia y la explicabilidad de las IAs que se usen dentro de los sistemas judiciales son una condición previa para el respeto de otros derechos fundamentales, puesto que su ausencia merma la posibilidad que tienen los operadores judiciales de fundamentar sus decisiones y la de los ciudadanos de entenderlas e impugnarlas. En razón a ello, la transparencia y explicabilidad de estas tecnologías deberá predicarse de forma irrestricta respecto de la manera en la que funciona, los casos en los que se usa y el lugar que ocupa dentro de la decisión judicial.<sup>58</sup> Adicionalmente, debe tenerse en cuenta que la información aportada por sistemas como ChatGPT deberá ser siempre corroborada de forma explícita. Con el fin de promover la transparencia debería evitarse el uso de algoritmos de *blackbox*, en donde se conocen los datos de entrada y el resultado, pero no la forma en la que la información es procesada.<sup>59</sup> Sobre ello debe tenerse en cuenta que las IAs son mayormente desarrolladas por empresas privadas, por lo que se hace necesario encontrar un balance entre la necesidad de transparencia y los derechos que se tienen sobre propiedad intelectual, priorizando cuando sea necesario los intereses de la justicia.<sup>60</sup>

### 3. La sensibilización a la ciudadanía y a la administración de justicia

La transparencia sobre la implementación de IAs no deberá entenderse de una forma pasiva, sino también desde una labor activa de la administración que se concrete en educación abierta y accesible a fin de garantizar de un lado, el uso responsable de la IA en la administración de justicia y, de otro, una participación pública efectiva.<sup>61</sup> La educación y sensibilización debería ir dirigida a los sujetos de las decisiones judiciales (ciudadanía en general) y los operadores judiciales, e incluir como mínimo información que permita conocer la forma en la que funciona, limitaciones y riesgos que presenta para los derechos humanos.<sup>62</sup> Particularmente, si se

<sup>58</sup> European Commission for the Efficiency of Justice (CEPEJ). European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. Diciembre de 2018. Principio 4.

<sup>59</sup> Marion Oswald, et al (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality. ALGO-CARE.

<sup>60</sup> Marion Oswald, et al (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality. ALGO-CARE.

<sup>61</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 44.

<sup>62</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 45.



usa una IA no licenciada para la administración de justicia debería hacerse especial énfasis en: (1) los sesgos que la herramienta tiene para que el operador judicial los pueda identificar y se abstenga de que se reproduzcan en sus decisiones,<sup>63</sup> y (2) la posibilidad de error y la necesidad de verificar de una forma visible y explícita la información que la IA proporciona.

#### 4. La protección de los datos

Respecto al derecho a la intimidad y la privacidad de los datos es importante que los operadores de la administración de justicia velen porque las personas conserven los derechos sobre sus datos.<sup>64</sup> En ese sentido, el marco de protección debe dirigirse a que se garantice la transparencia frente al usuario del sistema judicial del uso que el operador judicial dio a sus datos y asegurar un tratamiento con salvaguardas para datos sensibles. Especialmente debe evitarse que el uso de algoritmos no licenciados permita que privados o terceros procesen datos personales y los reutilicen haciéndolos ampliamente accesibles.<sup>65</sup> Al mismo tiempo, una política que persiga la protección de la intimidad y de la privacidad de los datos debe estar conforme a la legislación ya existente sobre tratamiento de datos y, en ese sentido, atender a los principios de finalidad, seguridad, transparencia y demás establecidos en la Ley 1581 de 2012.

#### 5. La responsabilidad en el uso de las herramientas

La búsqueda de la eficiencia y la existencia del sesgo de automatización en los humanos propician escenarios en los que se opta de ceder tareas a herramientas de IA, y aún cuando debe ser posible acudir a estos sistemas en la ejecución de determinadas tareas no se puede reemplazar la responsabilidad final de los seres humanos. La administración de justicia toma decisiones trascendentales para la vida de las personas, que deben seguir

<sup>63</sup> European Commission for the Efficiency of Justice (CEPEJ). European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. Diciembre de 2018. Principio 4.

<sup>64</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 73.

<sup>65</sup> European Commission for the Efficiency of Justice (CEPEJ). European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment. Diciembre de 2018. Párr. 143.

siendo tomadas por seres humanos. En estos casos es imprescindible que la responsabilidad ética, disciplinaria y jurídica del uso de IA en la administración de justicia siempre sea atribuible a una persona, el funcionario judicial.<sup>66</sup> Por ello, la política pública que se desarrolle debería introducir marcos de responsabilidad o aclarar la interpretación de los marcos existentes para garantizar la atribución de la responsabilidad disciplinaria por el uso irresponsable y los resultados que generen los sistemas de IA en la administración de justicia.<sup>67</sup> A su vez, este marco debe acompañarse de medidas correctivas que busquen reparar los daños que se causen a usuarios del sistema judicial por el uso indebido de los sistemas de IA.<sup>68</sup>

Solicitudes:

En razón de lo expuesto antes, solicitamos a la Honorable Corte:

1. CONFIRMAR la decisión del juez de segunda instancia, toda vez que, garantiza los derechos de los accionantes y el debate en sede de revisión se basa no en el fondo del asunto sino en las herramientas utilizadas para la toma de la decisión.
2. ORDENAR al Consejo de Administración Judicial o la autoridad competente que en un término razonable desarrolle una política pública sobre las salvaguardias para el uso de IA en la administración de justicia que incluya como mínimo, pero no de forma excluyente, las recomendaciones tratadas en el apartado inmediatamente anterior.
3. ORDENAR al Ministerio de Justicia y del Derecho o la autoridad competente que, una vez cumplida la solicitud anterior, los lineamientos de la política pública sean divulgados y fomentados en los centros de conciliación y arbitraje.
4. INSTAR a la Escuela Judicial “Rodrigo Lara Bonilla” o a la autoridad competente a desarrollar un módulo que se incluya dentro del Programa de Formación para la Rama Judicial y tenga como principal contenido la necesidad de salvaguardas para el uso de herramientas de IA en la administración de justicia.

<sup>66</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 36.

<sup>67</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 68.

<sup>68</sup> UNESCO. Recomendación sobre la ética de la inteligencia artificial. 23 de noviembre de 2021. Párr. 55.



Fundación  
**Karisma**

Firman:

Ana Bejarano Ricaurte  
El Veinte  
C.C. 1.136.879.823

Juan Diego Castañeda  
Fundación Karisma  
C.C. 1.020.739.106

Susana Echavarría Medina  
El Veinte  
C.C. 1.152.225.226

Juan de Brigard  
Fundación Karisma  
C.C. 1.018.452.918

Laura Urrego Aguilera  
El Veinte  
C.C. 1.000.793.820

