

## OCTAVA MESA DE TRABAJO MULTI-ACTOR SOBRE REGULACIÓN DE IA EN COLOMBIA

### INFORMACIÓN GENERAL

<b>Organizadores:</b>	Defensoría Delegada para la Protección de Derechos en Ambientes Digitales y Libertad de Expresión de la Defensoría del Pueblo, Escuela de Gobierno Alberto Lleras Camargo de la Universidad de los Andes y Departamento de Derecho de las Telecomunicaciones de la Universidad Externado de Colombia.
<b>Objetivo:</b>	Construir una hoja de ruta sobre la regulación de la IA con enfoque de derechos humanos.
<b>Metodología:</b>	Diálogo y deliberación en un ambiente plural, inclusivo y respetuoso. Presencial.
<b>Participantes:</b>	60 participantes en el evento presencial: que provienen de las tres ramas del poder público, órganos de control, organizaciones de la sociedad civil, academia, empresas y organizaciones gremiales. Video disponible en <a href="#">YouTube</a> .
<b>Fecha:</b>	13 de marzo de 2025, 09:10 a.m. – 01:20 p.m.
<b>Lugar:</b>	Sede Nacional de la Defensoría del Pueblo – Auditorio Alfredo Molano (Calle 55 # 10-32).

### AGENDA

<b>9:10 am – 9:20 am</b>	Bienvenida y apertura del evento a cargo del Dr. Roberto Molina Palacios, Vicefensor del Pueblo.
<b>9:20 am – 09:50 am</b>	Intervención de Santiago Pardo Rodríguez, Defensor Delgado para la Protección de los Derechos Digitales y la Libertad de Expresión.
<b>09:50 am – 10:10 am</b>	Intervención de Carolina Botero, Asesora de la Dirección de la Fundación Karisma.
<b>10:00 am – 11:20 am</b>	Taller sobre los Sistemas Automatizados de toma de Decisiones en el sector público de Colombia y sus posibles implicaciones para los derechos fundamentales.
<b>11:20 am – 12:40 pm</b>	Intervención de Margaret Hagan, Directora del Laboratorio de Diseño Legal de la Universidad de Stanford.
<b>12:40 pm – 01:00 pm</b>	Presentación de las conclusiones del taller.
<b>01:00 pm - 01:05 pm</b>	Cierre y compromiso.

## RELATORÍA/MEMORIAS DE LA OCTAVA MESA DE TRABAJO MULTI-ACTOR SOBRE REGULACIÓN DE IA EN COLOMBIA<sup>1</sup>

### 1. BIENVENIDA Y APERTURA DEL EVENTO

#### 1.1. INSTALACIÓN DEL EVENTO

El evento inició con la intervención de Lina Berrocal, maestra de ceremonias, quien dio la bienvenida a los participantes y resaltó que esta era la primera ocasión en que se realizaba la Mesa de Trabajo Multi-actor sobre regulación de Inteligencia Artificial (IA) en un organismo de control dedicado a la protección de los derechos humanos.<sup>2</sup> También manifestó que esperaba que se generaran recomendaciones que orienten y contribuyan a los procesos regulatorios futuros, siempre bajo un enfoque que priorice la protección de los derechos humanos de todas las personas. La IA, en sus palabras, es una herramienta poderosa: plantea tanto oportunidades como riesgos. Por ello, enfatizó en que la regulación de la IA debe ir más allá de los aspectos técnicos y económicos, y basarse también en un enfoque de derechos humanos, garantizando que el uso de la IA proteja y promueva los derechos de todos los ciudadanos, especialmente los de los más vulnerables. A continuación, la maestra de ceremonias concedió el uso de la palabra al Dr. Roberto Molina Palacios, Vicedefensor del Pueblo.

El Vicedefensor del Pueblo saludó a todos los asistentes y reconoció que para la Defensoría del Pueblo era un honor darles la bienvenida a la octava sesión de la Mesa de Trabajo Multi-actor para la regulación de la IA en Colombia. El Dr. Roberto Molina Palacios destacó que se realizaba la Mesa en este escenario porque el desarrollo tecnológico de la inteligencia artificial necesita un enfoque de derechos humanos, basado en dos derechos clásicos y fundamentales: el derecho a la igualdad y a la libertad. En sus palabras, la IA podría implicar riesgos muy grandes en materia de discriminación, trato desigual a las personas, y podría afectar principios de libertad como la privacidad, la autonomía y, el principio de todos los principios, el de la dignidad humana. De allí que, es esencial hablar de cómo equilibramos los avances tecnológicos con la garantía de los derechos humanos. En este sentido, él se refirió a tres escenarios fundamentales: uno relacionado con la prevención del daño; un

---

<sup>1</sup> Este documento fue elaborado por Sarah Muñoz-Cadena, y luego fue complementado y revisado por Santiago Pardo Rodríguez, Defensor Delgado para la Protección de los Derechos Digitales y la Libertad de Expresión de la Defensoría del Pueblo; Juan David Gutiérrez profesor de la Escuela de Gobierno de la Universidad de los Andes; y Sandra Ortiz, profesora de la Universidad Externado de Colombia. Para más información sobre la Mesa de Trabajo Multi-actor contactar al profesor Juan David Gutiérrez al correo [juagutie@uniandes.edu.co](mailto:juagutie@uniandes.edu.co). Para conocer más sobre la Defensoría Delegada para la Protección de los Derechos Digitales y la Libertad de Expresión se puede escribir a [ambientes.digitales@defensoria.gov.co](mailto:ambientes.digitales@defensoria.gov.co).

<sup>2</sup> La información sobre las anteriores sesiones de la Mesa de Trabajo Multi-actor pueden consultarse en la siguiente página web: <https://sistemaspublicos.tech/mesa-de-trabajo-multiactor-sobre-regulacion-de-ia-en-colombia/>

segundo escenario asociado a la participación democrática; y un tercer escenario vinculado a la responsabilidad de todos los actores involucrados en el desarrollo de esta tecnología.

El Dr. Molina Palacios, en términos de prevención del daño, destacó la necesidad de recordar los principios, por ejemplo, el principio de prevención y el principio de precaución. Él destacó que es necesario evaluar los impactos y los riesgos que implica el desarrollo de la IA, en particular los riesgos en tres escenarios fundamentalmente. Primero, los riesgos en términos de decisión de justicia. Si bien está muy de moda hablar de que la IA puede tomar decisiones en asuntos referidos a la justicia, estas decisiones pueden ser claramente discriminadoras o afectar la autonomía de las personas o el acceso a servicios como la seguridad social y la salud. El Vicedefensor citó a la Dra. Wendy Wong, quien dice que nosotros no somos sujetos pasivos o fuentes de datos, sino titulares de derechos y de garantías, y que los datos vistos en sus términos básicos no son simple datos, sino que son expresión de nuestra identidad, nuestra autonomía y, en ese término, de identidad y autonomía, la afectación de derechos fundamentales puede implicar un riesgo grande.

En segundo lugar, está el escenario de la democracia. En este escenario, el Dr. Molina Palacios, resaltó que la IA no es un problema exclusivo de los estados o de las grandes empresas o de quienes diseñan los sistemas de programación. Desde su perspectiva, el problema de la inteligencia artificial es un problema de la democracia deliberativa. Él explicó que, por ello, deben intervenir todos los sujetos implicados y todos los puntos de vista frente a estas nuevas tecnologías. Dado que la IA impacta la vida diaria de todas las personas, esas personas deben ser escuchadas en este tipo de debates, en particular, porque el estado social de derecho está en juego. Él manifestó que hay sectores muy sensibles a la implementación de IA como, por ejemplo, la seguridad, el acceso a créditos y financiación, en términos de economía solidaria, de asignación de beneficios sociales y de acceso y permanencia en espacios educativos.

El tercer escenario al que se refirió el Vicedefensor es el de la responsabilidad. En sus palabras, en términos de responsabilidad, hay que deslindar muy bien la responsabilidad de quienes diseñan los programas, quienes los implementan y quienes hacen uso de esos programas. Él argumentó que, así como es de sofisticada la tecnología, también se deben crear escenarios de fiscalización y control a la IA para evitar errores, bajo un principio fundamental que es la transparencia algorítmica y el principio de la explicabilidad. Es decir, que las personas afectadas por decisiones tomadas por un algoritmo, puedan entender cómo y por qué se tomó una determinación, por ejemplo, tanto en materia de contratación como en acceso a crédito.

El Dr. Molina Palacios, enfatizó en que sin transparencia, la inteligencia artificial se convierte en una caja negra que erosiona la confianza de las instituciones. Las personas afectadas por decisiones deben exigir y pueden exigir correcciones, compensaciones y garantías de no repetición, es decir, es necesario diseñar mecanismos de pretensión y protección de derechos fundamentales en escenarios donde se utiliza la inteligencia artificial. El Vicedefensor argumentó que, por tanto, se debe generar regulación normativa, de

procedimiento administrativo, mecanismos de protección constitucional o de acciones constitucionales en esta materia, los cuales aún están por desarrollarse.

El Vicedefensor del Pueblo dijo que combinando el escenario de la responsabilidad con el de la participación, es necesario crear o habilitar escenarios de auditoría pública en la que participen la sociedad civil, la academia y otros actores. En particular, porque hay una oportunidad, en el escenario colombiano, de crear sujetos informados que puedan representar debidamente a las poblaciones afectadas. De allí que destacó la necesidad de crear escenarios de representación de sujetos informados para la defensa de posibles discriminaciones y desigualdades en materia de inteligencia artificial.

Para concluir, el Vicedefensor enfatizó en que la inteligencia artificial no puede desarrollarse ni regularse al margen de los derechos humanos y, por ello, hay que tener en cuenta los tres escenarios: la prevención del daño, la participación democrática y la responsabilidad efectiva de quienes participan en el diseño y uso de la IA. Desde la Defensoría del Pueblo, precisamente, se reiteró el compromiso de la entidad para acompañar este proceso con un enfoque garantista. El Dr. Molina Palacios manifestó que la labor de la Defensoría es asegurar que la regulación de la IA no quede en manos exclusiva de los desarrolladores tecnológicos o de las dinámicas del mercado, sino que tenga como eje central la protección de los derechos fundamentales de todas las personas. Este es un desafío que requiere compromiso, visión y voluntad política, por lo cual, agradeció a todas y todos los participantes en este espacio de diálogo y los invitó a seguir construyendo juntos un futuro donde la tecnología esté verdaderamente al servicio de las personas y de la sociedad.

Luego, la maestra de ceremonias, Lina Berrocal, agradeció la intervención del Dr. Molina Palacios y le dio la bienvenida al Delegado para la Protección de los Derechos en Ambientes digitales y libertad de expresión, Santiago Pardo Rodríguez.

## **1.2. INTERVENCIÓN DE SANTIAGO PARDO RODRÍGUEZ, DEFENSOR DELGADO PARA LA PROTECCIÓN DE LOS DERECHOS DIGITALES Y LA LIBERTAD DE EXPRESIÓN.**

El Defensor Delegado, Santiago Pardo Rodríguez, saludó a los asistentes, agradeció la intervención del Vicedefensor y dio las gracias a la profesora Sandra Ortiz de la Universidad de Externado y al profesor Juan David Gutiérrez de la Universidad de los Andes por coorganizador este espacio, por primera vez, en una institución de derechos humanos en el país. Él destacó que la Tercera Mesa se realizó exitosamente en el Consejo Superior de la Judicatura y que la idea de esta sesión era retomar estas discusiones dentro de espacios públicos y, particularmente, dentro de la institución líder de derechos humanos en el país en estos términos que es la Defensoría Delegada para la Protección de los Derechos Digitales y la Libertad de Expresión (DADLE).

El Defensor Delegado explicó que la Delegada se creó mediante la resolución 1589 del 2020 y que tiene, a grandes rasgos, una gran función que está relacionada con promover,

defender y divulgar la vigencia de los derechos humanos y los derechos fundamentales en escenarios emergentes digitales. Precisamente, la IA es uno de esos escenarios emergentes que sin duda merecen y requieren un enfoque de derechos humanos. Según manifestó el Sr. Pardo Rodríguez dentro del nuevo proyecto del plan estratégico institucional, liderado por la Defensora, la DADLE ocupa un lugar central para la protección de derechos innominados y emergentes. La IA plantea retos frente a la materialización de derechos ya existentes, pero también existe toda una nueva área sobre derechos emergentes. Por ejemplo, el derecho a la transparencia algorítmica, el derecho a la auditoría algorítmica, que hacen parte del derecho al acceso a la información pública. El Delegado explicó que estos temas hacen parte fundamental de las discusiones que la Defensoría del Pueblo (bajo el liderazgo de Defensora y el Vicedefensor) está teniendo en el marco de las tareas que realiza la DADLE.

En este sentido, el Delegado Pardo Rodríguez explicó que la estructura de la DADLE consta de tres equipos de trabajo que se articulan entre sí con el fin de incidir y participar en discusiones de política pública digital como estas. El primer equipo de trabajo es el de Política Pública Digital, el cual busca promover una agenda de derechos humanos con un enfoque en la mitigación de riesgos y el diseño centrado en las personas. En este punto, el Delegado recordó que en el [Documento CONPES 4144, Política Nacional de Inteligencia Artificial](#), por ejemplo, se incorporó un capítulo de mitigación de riesgos y reducción de daños. Precisamente, la Defensoría del Pueblo quiere ocupar un rol de liderazgo y de participación activa en estas discusiones.

El segundo equipo de la DADLE es el de alfabetización informacional y comunicacional, que se centra en generar modelos de curaduría especialmente para poblaciones vulnerables (como niños, niñas, adolescentes, minorías étnicas), que les permitan construir un modelo de pedagogía de derechos digitales propios y también reducir daños frente a, por ejemplo, fenómenos asociados al uso masivo de la información y de la inteligencia artificial como la desinformación.

El tercer equipo, según explicó el Delegado Pardo Rodríguez, es el de libertad de expresión y nuevas formas de expresión. La Defensoría del Pueblo desde la creación de esta Delegada tiene un mandato muy importante para la protección de la libertad de expresión y la libertad de prensa. Labor que se ha ampliado dada la incidencia en debates sobre nuevas expresiones como, por ejemplo, la regulación de contenidos y moderación de contenidos en nuevos medios digitales. Donde surgen preguntas como ¿qué tipo de autorregulación tienen que tener las plataformas como TikTok, Instagram, las plataformas de meta, o X frente a debates públicos? ¿Debería existir un modelo de autorregulación o debería ser un modelo más específico como de regulación de contenidos de un tercero, como en el modelo europeo? El Delegado explicó que desde la Defensoría buscan consolidarse como un espacio referente para las discusiones de política pública digital y, por ello, es tan importante que la Octava Mesa de Trabajo Multi-actor se realice en este espacio.

El Delegado destacó un libro de la profesora Wendy Wong, [We, the Data](#), en el que explica que los datos, en general, existen porque alguien los creó, no son el producto de una generación espontánea, por lo cual, los datos son humanos. El Delegado manifestó que es muy importante construir a partir de tres preguntas abiertas que, desde su perspectiva, se podrían abordar en un espacio como este: i) ¿Qué significa tener realmente un enfoque de derechos humanos en la inteligencia artificial?; ii) ¿Cómo se pueden establecer modelos para detectar realmente si se está violando un derecho humano o un derecho fundamental con la aplicación de la inteligencia artificial o en cualquier política pública digital?; y iii) ¿Cuál es el remedio ante estas situaciones?

Frente a estas preguntas, Santiago Pardo Rodríguez destacó que en la Defensoría del Pueblo se ha discutido mucho, no solo en incidencia en política pública digital sino, en términos generales, si la incidencia en estos temas debe ser judicial o política en términos de abogar por la protección de una agenda de derechos humanos, o si se deben utilizar otro tipo de estrategias. El Defensor Delegado recordó, por ejemplo, que estas discusiones toman relevancia en el marco de la [Sentencia T-067 del 2025](#), en la cual la Corte Constitucional se refiere específicamente a la protección del principio de transparencia algorítmica como parte, elemento y/o núcleo esencial del derecho al acceso a la información pública. En este sentido, el Defensor Delegado argumentó que esta Sentencia muestra que esos escenarios de judicialización son importantes y tienen que existir, pero no deben ser los únicos escenarios. De allí que invitó a que, en esta Mesa se reconozcan tanto los impactos positivos, que sin duda pueden tener los sistemas de decisión automatizada que está utilizando el Estado colombiano, pero también los riesgos que pueden acarrear.

El Delegado Pardo Rodríguez terminó su intervención destacando el [Decálogo](#) promovido por la Defensora del Pueblo, Dra. Iris Marín Ortiz, el cual es la guía institucional de la Defensoría y el compás hacia las acciones defensoriales e intervenciones defensoriales. En particular, él destacó tres de los 10 puntos. Primero, cualquier intervención defensorial tiene que promover un enfoque de género, lo cual cobra relevancia si se considera que la Dra. Iris Marín es la primera mujer Defensora en la historia del país. En particular, en el tema que atañe a las funciones de la Delegada, este primer punto se relaciona con cómo se pueden crear enfoques diferenciales para la protección de los derechos humanos en la regulación de la IA. Segundo, el decálogo incorpora un punto muy importante sobre el derecho a las víctimas. Asociado a este punto hay preguntas sobre ¿cómo se puede utilizar la IA para mejorar la atención del Estado en términos de memoria histórica, reparación integral y acceso a la justicia? Tercero, en el Decálogo también hay un enfoque hacia las regiones. La Defensoría del Pueblo es una institución con alcance nacional y, por ello, estas discusiones también se tienen que dar desde las regiones. El Delegado Pardo Rodríguez manifestó que desde la Defensoría del Pueblo se quiere acompañar a quienes estén impulsando iniciativas de trabajo legislativo o de trabajo de política pública digital desde las regiones, para apoyarlos y acompañarlos en estos procesos, con el fin de coliderar las discusiones regionales.

Luego de la instalación del evento y la intervención del Defensor Delegado para la Protección de Derechos en Ambientes Digitales y Libertad de Expresión, Santiago Pardo Rodríguez, la

señora Lina Berrocal, maestra de ceremonias del evento, cedió la palabra a Carolina Botero, abogada experta en derechos digitales, quien se desempeñó como Directora de la Fundación Karisma y quien continúa como consultora en dicha institución.

## **2. INTERVENCIÓN DE CAROLINA BOTERO, ASESORA DE LA DIRECCIÓN DE LA FUNDACIÓN KARISMA.**

La Dra. Botero saludó a los asistentes y agradeció a la Defensoría del Pueblo por la invitación a este espacio. Ella inició su intervención reconociendo que el [Documento CONPES 4144, Política Nacional de Inteligencia Artificial](#) aborda aspectos relacionados con los derechos humanos para el desarrollo y el uso de la IA en Colombia. Sin embargo, desde su perspectiva, lo hace de una forma “suave”. Ella explicó que este Documento no es una apuesta decidida para tener esa perspectiva de derechos humanos, aunque de una forma legítima se inserta dentro de una tendencia a hacerlo desde el marco ético. Lo cual se podría corregir en la implementación del Documento.

La Dra. Botero explicó que en la primera parte del Documento CONPES se recoge lo que ya venían diciendo la Organización para la Cooperación y el Desarrollo Económico (OCDE) y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), cuando reconocían la importancia de tener a los derechos humanos y los valores democráticos en el desarrollo y uso de la IA. Además, estas entidades enfatizan en transparencia y explicabilidad, en la importancia de desarrollar sistemas de IA robustos y seguros, minimizar riesgos, en la responsabilidad en el cumplimiento de principios éticos que aplican a desarrolladores y a los usuarios, en asegurar que la IA contribuya al crecimiento inclusivo y sostenible. Todo lo cual es loable. Sin embargo, la Dra. Botero explicó que tiene reservas con respecto a esto, porque esta tendencia no está apoyada en obligaciones de derechos humanos, que son las obligaciones que tienen los estados porque han firmado convenciones que les imponen cuidar sobre estos temas. En este sentido, ella argumentó que se debe replantear esta tendencia a considerar que la ética es suficiente para velar y cuidar que se cumplan los derechos humanos. Su propuesta radica en que se cambie esta tendencia por una en la cual, cumplir con los derechos humanos sea lo que garantiza que la IA sea ética. Según manifestó, eso parece un cambio menor, pero no lo es, es un cambio importante y esa es la ruta que le propone a la Defensoría liderar, por ejemplo, a través de la implementación del Documento CONPES antes mencionado.

La Dra. Botero realizó tres propuestas concretas. Primera, mapear y exigir mecanismos de transparencia y explicabilidad, incluyendo herramientas de rendición de cuentas para el desarrollo y uso de la IA que estén en línea con las obligaciones de derechos humanos y los estándares internacionales. Este ejercicio inicial de mapeo revelaría las exigencias mínimas para la industria que desarrolle IA y para quienes lo implementen en servicios a las personas. Una forma de materializar esto es con obligaciones de reparar cuando se produzcan graves violaciones a los derechos humanos. Por ejemplo, con la creación del sistema financiero abierto se manifiesta que se va a crear un mercado de datos que, entre otras, va a utilizar

IA, pero cuando se revisa el detalle de los mecanismos de seguimiento y control para ese sistema, aunque en el principio explican que los derechos humanos y la protección de datos hacen parte central del tema, en la realidad no existe ni un solo control o seguimiento que nos hable de cómo se va a hacer efectiva esa protección; de hecho, ni siquiera existe la obligación central de protección de datos. No se menciona a la Superintendencia de Industria y Comercio (SIC), ni tampoco se hace alusión a mecanismos que involucren a la academia o a la sociedad civil. Ella enfatizó en que se está confiando en que la ética proteja los derechos humanos, y no que los derechos humanos garanticen que este o aquel sea un sistema ético.

La segunda propuesta que hizo la Dra. Botero está relacionada con reconfigurar y reinterpretar conceptos como robustez y seguridad, para que los vean a la luz de estándares de derechos humanos que han sido ya muy desarrollados y que son obligatorios como los de la necesidad y proporcionalidad. Según ella explicó esto cambiaría totalmente la dinámica.

La tercera propuesta que realizó la Dra. Botero en su intervención fue que se diera alcance al concepto de responsabilidad en el cumplimiento de principios éticos y se exigiera que se pensara en que los usos que se daban a estas tecnologías fueran legales y fueran en línea con las normas del derecho internacional. Ella declaró que para el desarrollo y uso de la IA solo deberían ser lícitos los sistemas que se ajusten a normas de derechos humanos, de modo que no solo la industria piense en cómo evitar el daño y ser precavida, sino que también se llegue a acuerdos sobre algunos usos que son demasiado riesgosos para los derechos de las personas y que, por ende, deban regularse restrictivamente. Por ejemplo, se podría analizar en los casos de reconocimiento facial y en los mecanismos de identificación biométrica a través de IA que está usando la fuerza pública. En este sentido, ella argumentó que en el Documento CONPES 4144 la mayoría de las acciones relacionadas con este tipo de tecnologías están en manos del Ministerio de Defensa. Al respecto ella se preguntó ¿por qué no iniciar, entonces, con un análisis de impacto de derechos humanos a este tipo de procesos? Porque esto puede materializar lo que en ética pareciera ser tan etéreo. Un análisis de impacto de derechos humanos sobre un sistema de IA en desarrollo o en uso, le permitiría a la industria y a los usuarios entender cómo se concretan propuestas como las tres que mencionó. Este tipo de análisis, además, según ella explicó, tendría un beneficio adicional y es que se podría incorporar a dos sectores centrales, como lo son la sociedad civil y la academia, adicional a involucrar a los usuarios.

Finalmente, la Dra. Botero enfatizó en que el Documento CONPES no cubrió en la amplitud que requería temas como los de propiedad intelectual, a pesar de las advertencias de diferentes organizaciones sobre este aspecto. Desde su perspectiva, el Documento CONPES se concentró, legítimamente, en la protección de los derechos de propiedad intelectual de los titulares pero olvidó por completo proteger a la sociedad en general. Por ejemplo, el Documento CONPES no planteó ninguna ruta para incorporar excepciones necesarias para la minería de textos y datos que garanticen que se puedan usar métodos de investigación de IA para el desarrollo de derechos como los relacionados con salud, ciencia y educación.

### **3. TALLER SOBRE LOS SISTEMAS AUTOMATIZADOS DE TOMA DE DECISIONES EN EL SECTOR PÚBLICO DE COLOMBIA Y SUS POSIBLES IMPLICACIONES PARA LOS DERECHOS FUNDAMENTALES.**

La maestra de ceremonias, Lina Berrocal, retomó la palabra para agradecer la intervención de la Dra. Carolina Botero. A continuación, el profesor de la Escuela de Gobierno Alberto Lleras Camargo de la Universidad de los Andes, Juan David Gutiérrez, explicó cuál sería la metodología para que los asistentes al evento pudieran trabajar en mesas.

El profesor Juan David Gutiérrez saludó a los asistentes de la octava sesión de la Mesa de Trabajo Multi-actor y reflexionó acerca de cómo se está construyendo comunidad en torno a este espacio de trabajo que procura ser diverso e inclusivo. Precisamente, de allí que en este espacio no solo se busque escuchar a personas que ilustren y generen unas primeras pinceladas de la discusión que debería darse en torno a cómo el Estado debería proteger, promover y respetar los derechos fundamentales en el uso de este tipo de sistemas de IA. El profesor Gutiérrez manifestó que en estas mesas de trabajo el protagonismo está en sus integrantes, y en la pluralidad de estos, porque no solo participan personas que provienen de las tres ramas del poder público, sino también de la sociedad civil, de las empresas. Él también agregó que el ejercicio que se iba a realizar buscaba poner a prueba qué tanto los sistemas de toma automatizada de decisiones que está, en este momento, desplegando el Estado colombiano cumplen o no con los principios básicos de protección y respeto de derechos fundamentales. Para este ejercicio se utilizó el [Repositorio de Sistemas automatizados de toma de decisiones en el sector público de Colombia](#), repositorio creado por la Universidad de los Andes, con información para cerca de 400 sistemas automatizados de toma de decisiones que han sido desplegados en el país.

Según explicó el profesor Gutiérrez, los asistentes al evento podían seleccionar uno de los nueve sectores en los cuales estaban clasificados los sistemas y, el ejercicio consistía en que en cada uno de los grupos se identificara al menos un sistema automatizado que represente un alto riesgo para los derechos fundamentales y otro que pueda promover su garantía. Después de seleccionar los dos casos, en el grupo se debía responder a tres preguntas guía:

- 1) ¿Qué derechos fundamentales se ven afectados?
- 2) ¿El sistema automatizado contribuye al goce efectivo de un derecho fundamental o, por el contrario, lo restringe?
- 3) ¿Qué acciones debería emprender el Estado para mitigar los riesgos o potenciar los impactos positivos en estos derechos fundamentales?

Para responder a las preguntas propuestas los grupos contaron con más de una hora. Al final del ejercicio uno o un par de relatores por grupo le contaron al resto de asistentes cuales fueron las conclusiones a las que llegaron.

A continuación, se presentan sus conclusiones.

## Grupo No. 1: Justicia

En este grupo, el primer sistema que analizaron fue el de “[Sentencias para la gente a través de IA](#)” que, según explicaron, es un aplicativo o una plataforma que permite llevar algunas sentencias a la gente de una manera lúdica o dinámica que les permitan entenderlas. En este grupo encontraron que los derechos que se pueden ver afectados por este sistema son el derecho a la información y a la justicia, el derecho a la igualdad y a la no discriminación, el derecho a la privacidad y a la protección de datos, y al debido proceso. Este sistema tiene pros y contras. En particular, destacaron que es positivo llevarle de forma clara y concisa esta información a las personas. Pero, reconocieron que es fundamental evitar ciertos sesgos, para impedir la vulneración a ciertos derechos. Por ejemplo, si la información se presenta por medio de avatares, hay que cuestionarse acerca de ¿qué tipo de avatares se van a manejar? ¿si los avatares representarían a algún tipo de etnia o tendría algún tipo de acento? En lo que se refiere a si ¿el sistema automatizado contribuye al goce efectivo de los derechos? La conclusión a la que llegaron en este grupo es que esto va a depender del diseño y la implementación del sistema. En este sentido, reflexionaron acerca de la importancia de la transparencia, no solo para que una persona conozca sobre un sistema que se está usando, sino también para que haya información acerca de los criterios de selección que se utilizaron para garantizar que la información que alimenta el sistema no fue manipulada, alterada, o cambiada. En el grupo identificaron que este sistema puede efectivamente contribuir al acceso de la información, simplificando procesos y proporcionando una información más accesible a los usuarios. Aunque manifestaron que se debe fortalecer la supervisión y auditoría de los procesos.

El segundo sistema que analizaron es el de [Watson](#), sistema de la Fiscalía General de la Nación. Frente a este sistema, el grupo concluyó que quizás podría ser aplicado primero para sistemas laborales o civil, o para casos que tengan menos impacto, toda vez que Watson analiza patrones “con el propósito de asociar los casos, las denuncias y los hechos criminales” y pueden ocurrir sesgos que afecten el análisis.

## Grupo No. 2: Educación

En este grupo analizaron diferentes sistemas, como chatbots, el [Modelo Instituciones de Educación Superior \(IES\)](#) de la Contraloría General de la República, y el sistema de [Integración de Plexi con plataformas basadas en inteligencia artificial para supervisión y vigilancia de exámenes por computador](#). En el grupo identificaron cuatro dificultades frente a estos sistemas. Primero, un problema asociado al alcance de la información, es decir, la profundidad que se puede presentar, por ejemplo, frente a los chatbots debido a que muchas de estas herramientas tienden a ser sobre todo enunciativas. En segundo lugar, el acceso a internet, asociado también acceso al servicio de la luz, y a servicios públicos en general. Tercero, un reto frente a la participación, en otras palabras, quiénes tienen las posibilidades de participar frente a estas herramientas. Cuarto, la privacidad frente a la evaluación. En particular, en el grupo se refirieron a la biometría, los sesgos y el tratamiento

de datos. En el tema de la biometría, por ejemplo, mencionaron que durante la pandemia de COVID-19 se presentaron dificultades para que adultos mayores tuviesen acceso a ciertos servicios, por ejemplo, los bancarios, por problemas para identificar sus huellas. También se refirieron a neuro-derechos frente a estos exámenes ¿a dónde van los resultados? Y asociaron esta pregunta a la protección de datos y la necesidad de conocer que se hace con la información de las personas. Finalmente, reconocieron la importancia de realizar acciones para mitigar posibles riesgos y realizar acciones para reducir la brecha digital.

### **Grupo No. 3: Salud y servicios públicos**

La relatora del grupo inició su intervención manifestando su asombro frente a la cantidad de sistemas que se están desplegando. A continuación, manifestó que uno de los derechos posiblemente vulnerados es el derecho a una explicación, en sus palabras, no solamente existe un derecho a la información sino también un derecho a que nos expliquen qué pasa. Ella expresó que se deben reconocer los contextos donde hay población vulnerable y que tiene una especial protección constitucional para asegurar que se protejan en el marco del uso de estas herramientas o tecnologías. Además, mencionó la necesidad de tener un lenguaje claro e incluyente. En este sentido, enfatizó en la importancia de contar con grupos interdisciplinarios para el diseño de estas herramientas, de tal forma que se reconozca la diversidad de contextos en el país. Ella manifestó que encontraron más sistemas en la categoría de salud, que en la de servicios públicos. Algunas de estas herramientas fueron desplegadas por entidades del orden nacional y otras por entidades del orden local, algunas recopilan datos de los usuarios para controlar el sistema de salud pública. Sin embargo, en la categoría de servicios públicos solo encontraron tres herramientas. Frente a este hallazgo, la relatora del grupo expresó su preocupación acerca de ¿por qué las empresas de servicios públicos y, en particular, las que están concesionadas no tienen una IA para responderle al usuario?

### **Grupo No. 4: Medioambiente y movilidad**

En este grupo se concentraron en los sistemas relacionados con medioambiente y, en medio de la discusión, reiteradamente se habló, según el relator del grupo, acerca de los problemas centrales del sistema nacional ambiental y los problemas ambientales más serios del país. En lo que respecta a los sistemas, identificaron que los derechos que se ven más seriamente impactados son el derecho al medio ambiente sano y, en un primer círculo de influencia, los [derechos del tratado Escazú](#), que son los derechos a la participación, los derechos a acceso a la justicia y a la información en asuntos ambientales. También, el relator del grupo mencionó que, en el análisis se debe incluir el tema de la intimidad y el manejo de datos personales.

En este grupo identificaron que un potencial de estos sistemas está en cómo se hace más eficiente producir alertas tempranas para que la autoridad ambiental pueda actuar de forma preventiva o sancionatoria, según corresponda. Ellos también reconocieron que estos sistemas pueden aportar a la trazabilidad. Sin embargo, reconocieron unos riesgos. Por

ejemplo, en lo referente a la trazabilidad de esa información, se cuestionaron acerca de cómo la ciudadanía puede participar, no solamente desde el punto de vista de dónde salieron esos datos para que se tomara una decisión sancionatoria o de licenciamiento, sino también previamente cómo se puede lograr que los datos con los que se alimentan estos sistemas, en un país mega biodiverso y complejo como Colombia, permitan que haya una participación efectiva de una ciudadanía informada.

Otro de los retos que identificaron es que en Colombia hay muchas áreas del territorio a las que simplemente no es posible acceder por razones de violencia armada o por dificultades geográficas, frente a lo cual un sistema de monitoreo a través de IA puede ofrecer ciertos beneficios. Pero, frente a este tipo de tecnologías reconocieron que es importante: i) garantizar que haya una posibilidad de supervisión, lo más generalizada posible, institucional y ciudadana; y ii) comprender que no todo se resuelve con IA, que la IA es una herramienta más entre muchas otras y que aún persisten los problemas de articulación institucional, de modernización del sector. El relator de este grupo finalizó su intervención expresando que hay un reto para regular la inteligencia artificial, reto conectado con los derechos humanos, porque se está hablando de obligaciones de protección y de respeto como deberes generales, pero también de adecuación del derecho interno. En otras palabras, se debe responder a estos derechos consignado en las convenciones e instrumentos internacionales, pero adecuando esa respuesta al derecho interno.

#### **Grupo No. 5: Orden público, seguridad y defensa**

En este grupo se identificaron dos sistemas, pero el análisis se concentró en el [Sistema Predictivo de Seguridad](#) de la Secretaría Distrital de Seguridad, Convivencia y Justicia de la Alcaldía Mayor de Bogotá. Según explicó la relatora del grupo, este sistema predice a partir de los datos históricos, como el registro de denuncias por delitos como homicidios, riñas, y lesiones personales, delitos contra el patrimonio, con uso de la violencia, entre otros, cuáles son las zonas de la ciudad de Bogotá y los horarios en los cuales suelen ocurrir más crímenes, para anticiparse a estos hechos y optimizar los recursos policiales. La relatora del grupo enfatizó en dos aspectos: i) el sistema realiza predicciones y ii) la pluralidad de personas que integraron el grupo, en el grupo se contó con la participación de delegados de Ministerio de Defensa, de la Escuela Militar de Cadetes, la Corte Constitucional, un ingeniero electrónico que estudia derecho y un politólogo que es especialista en análisis de datos, un programador, funcionarios de la Defensoría del Pueblo, y de la Fundación Karisma.

En el grupo concluyeron que los derechos fundamentales que se pueden ver afectados por el uso de este sistema son: el derecho a la dignidad humana, el derecho a la privacidad y el derecho a la protesta. Con respecto a si el sistema automatizado contribuye al goce efectivo de un derecho fundamental o por el contrario lo restringe, la relatora del grupo explicó que este sistema puede contribuir al derecho a la paz y la seguridad. Finalmente, frente a la pregunta de ¿qué acciones debería emprender el Estado para mitigar los riesgos o potenciar los impactos positivos en estos derechos? Se mencionaron tres: i) la delimitación de soluciones; ii) establecer que el sistema genera un insumo para la toma de decisiones, pero

la toma de decisiones debe realizarla un ser humano, en particular, porque este sistema realiza predicciones; y iii) la responsabilidad debe recaer en la persona, es decir, la persona que pertenezca al Estado o al sector privado debe asumir una responsabilidad frente al uso de estos sistemas para la toma de decisiones.

## **Grupo No. 6: Protección social**

En este grupo se analizó el sistema del Instituto Colombiano de Bienestar Familiar (ICBF) que se denomina [Modelo probabilístico de vulneración de derechos de niñas, niños y adolescentes](#). Una de las primeras preguntas que se realizaron en el grupo fue si este es un sistema que realiza predicciones a partir de datos pasados. Siguiendo con este análisis, los integrantes del grupo se cuestionaron acerca de dónde vienen los datos que alimentan el sistema, cuáles son las bases de datos que se utilizan para realizar cruces de información para determinar si hay un potencial riesgo de que se vulneren los derechos de los menores de edad. También, según el relator, debatieron acerca de la interoperabilidad entre instituciones para poder llegar a hacer un pronóstico.

En lo que respecta a los derechos que se pueden ver afectados por el sistema mencionaron el derecho a la presunción de inocencia, el derecho al buen nombre, a la transparencia y a la explicabilidad. En lo que se refiere a si el sistema automatizado contribuye al goce efectivo de un derecho fundamental o por el contrario lo restringe, expresaron que depende de la calidad del algoritmo y la calidad de los datos. Frente a la pregunta de ¿qué acciones debería emprender el Estado para mitigar los riesgos o potenciar los impactos positivos en estos derechos? El relator del grupo expresó que ellos consideran que el resultado del algoritmo se debe tomar como un insumo más para la toma de decisiones. De tal forma que se garantice la transparencia, trazabilidad, y se contraste dicha predicción con otra información. Finalmente, el relator mencionó que se debe establecer la robustez del sistema, más aún, considerando que este sistema toma datos de menores de edad.

### **4. INTERVENCIÓN DE MARGARET HAGAN, DIRECTORA DEL LABORATORIO DE DISEÑO LEGAL DE LA UNIVERSIDAD DE STANFORD<sup>3</sup>**

Continuando con la agenda el Defensor Delgado para la Protección de los Derechos Digitales y la Libertad de Expresión, Santiago Pardo Rodríguez, saludó, presentó y agradeció a la profesora [Margaret Hagan](#) por acompañar el evento.

La profesora Margaret Hagan agradeció el espacio y manifestó estar feliz de hablar sobre la IA, los derechos humanos, el acceso a la justicia; en particular, a partir de algunas investigaciones que están realizando en la Universidad de Stanford sobre el tema.

---

<sup>3</sup> La intervención de la profesora Margaret Hagan fue en inglés.

La Dra. Hagan empezó su intervención contando que lo que realmente motiva su trabajo y el trabajo desde la Universidad de Stanford es investigar si ¿puede la IA realmente ayudar a cumplir la promesa de cerrar la brecha de la justicia? ¿Si puede la IA a las personas de la comunidad que necesitan ayuda, necesitan proteger sus derechos, necesitan una vida mejor, ayudarles a obtener justicia? Frente a estas preguntas la Dra. Hagan expresó que en su comunidad de abogados, investigadores, jueces y otras personas que trabajan en el sector de la justicia, ella ha observado dos bandos. Uno son los optimistas sobre la tecnología, que están muy entusiasmados con el poder que la IA podría aportar para hacer el trabajo legal, como escribir cartas, rellenar formularios, hacer investigaciones, hacer todas esas cosas que los abogados hacen para ayudar a la gente a hacer valer sus derechos. Por otro lado, están los pesimistas, los que piensan que los sistemas de IA están sobrevalorados, que no son más que máquinas incoherentes, y que no deberíamos confiar en estas herramientas de IA para realizar trabajos jurídicos de alto riesgo. Estas son las dos vertientes de la polémica que se han venido desarrollando en los últimos años.

Frente a ello, la Dra. Hagan y su grupo de investigación está tratando de tomar un camino intermedio, en sus palabras, algo más realista y también empírico. Por lo cual, en lugar de limitarse a decir “¿Esto va a ser bueno o va a ser malo?” están intentando hablar con la gente sobre el terreno y en la comunidad para conocer si quieren utilizar la IA para sus problemas jurídicos. Por ello, se han preguntado ¿hasta qué punto los distintos modelos de lenguaje a gran escala cumplen realmente tareas jurídicas importantes, como responder a las preguntas de los ciudadanos o informarles sobre sus derechos? ¿Se puede hacer una investigación empírica para ver si estos sistemas son buenos o malos haciendo estas tareas? ¿Cómo identificar las oportunidades más maduras para que la IA mejore realmente la eficiencia en temas legales?

Bajo este enfoque, la Dra. Hagan y su equipo ha estado desarrollando una variedad de diferentes tipos de productos de investigación. En esta sesión, la Dra. se refirió específicamente a una investigación con los usuarios, con los miembros de la comunidad; aunque aclaró que si deseaban información sobre otras investigaciones podrían encontrarlas en su [sitio web](#). Refiriéndose a los usuarios, las preguntas de investigación que motivaron a la profesora Hagan y su equipo fueron ¿quieren los ciudadanos de los Estados Unidos (EE. UU.) utilizar la IA para resolver sus problemas judiciales? ¿Cómo utilizan realmente la IA? ¿Encuentran valor en ella o no?

En esta investigación la Dra. Hagan y su equipo encontró que los adultos en los EE. UU. están muy interesados y son mucho más propensos a utilizar la IA para la resolución de sus problemas jurídicos y, sinceramente, para todos los demás tipos de problemas de la vida en los próximos años. Según ella comentó, cada vez más esperamos que acudan a lugares como ChatGPT, Gemini u otras plataformas de IA de grandes empresas, como uno de los primeros lugares a los que acudirán cuando les ocurra algo importante, por ejemplo, cuando tengan un problema jurídico. En este sentido, ella reflexionó que aunque en el sector de la justicia sean escépticos respecto a la IA, tenemos que reconocer que cada vez más gente va a utilizar la IA para averiguar sobre sus derechos legales y emprender acciones. Así que invitó a

trabajar ahora mismo para ver cómo se puede lograr que ese mundo futuro sea lo más seguro y eficaz posible.

La investigación de la Dra. Hagan y su equipo inició alrededor de seis meses después de que ChatGPT fue lanzado y se presentó todo ese entusiasmo en torno a la IA. En ese momento, ellos empezaron a entrevistar a adultos en los EE.UU y querían entender, desde su punto de vista, si estaban entusiasmados con la IA para la resolución de problemas legales y cómo la usarían. Entrevistaron alrededor de 50 personas. A algunas de estas personas las reclutaron online y a otras las encontraron en los pasillos de los juzgados, cuando los tribunales les permitían investigar y podían invitar a personas que estaban esperando en el juzgado a ser entrevistadas por el equipo. Buscaron variedad entre los entrevistados y hablaron con ellos durante unos 30 a 60 minutos. A los entrevistados les planteaban un escenario en el cual debían imaginar que acababan de llegar a casa y en la puerta encontraban un documento o carta donde se les expresaba que el casero o arrendador los iba a expulsar de la propiedad en una semana. Frente a ese escenario les preguntaban si, después de recibir esa comunicación, al abrir su computador y/o teléfono ¿a dónde se dirigirían para saber qué hacer en esas circunstancias? ¿Llamarían a alguien? ¿Buscarían por internet? ¿Qué harían? Luego, a los entrevistados se les preguntaba si estarían interesados en utilizar una herramienta de IA como ChatGPT o Bard (hoy en día Gemini) para solucionarlo.

La gran mayoría de los entrevistados manifestaron que utilizarían la IA. Según contó la Dra. Hagan fue mínimo el número de personas que estaban en contra de usar la IA. De hecho, después de presentarles la herramienta, sólo dos personas manifestaron que desconfiaban de la IA y que no querían usarla en absoluto para sus problemas legales. Una de las personas que no quería utilizar la IA expresó que tenía reservas acerca de sus datos y la vigilancia, porque no sabía quién podría tener acceso a sus datos si introducía información a la IA o quién pudiera usar eso en su conta. La otra persona manifestó su paranoia y desconfianza frente a las empresas tecnológicas. Con lo cual, solo dos personas estaban preocupadas o expresaban reservas frente a la IA, mientras el resto de participantes aceptó utilizar la IA.

Luego, los participantes utilizaron la IA y, al observar por encima de su hombro, que palabras utilizaban para consultar el tema, la Dra. Hagan y su equipo encontró que la mayoría de los participantes utilizaban la IA de la misma forma como realizarían una búsqueda en Google u otro motor de búsqueda. Es decir, los usuarios no escribían un párrafo largo sobre su problema, sino frases cortas como las de una búsqueda en Google, con palabras clave como: derechos de los inquilinos, problemas con el propietario, desalojado por el propietario. Sin embargo, algunos usuarios que tenían más experiencia con la tecnología y sabían un poco más sobre cómo “comunicarse” con la IA para obtener mejor información, optaron por escribir frases completas o explicar en detalle lo que les estaba ocurriendo para pedir ayuda específica.

La Dra. Hagan enfatizó en que este ejercicio se realizó en 2023, momento en el cual pudieron observar que, en general, la IA seguía el mismo patrón de respuesta. Cuando una persona preguntaba sobre un escenario legal o sus derechos, ChatGPT o Gemini normalmente

empezaba con una primera frase muy empática, como “siento mucho oír que estás lidiando con esto”. A continuación, en la respuesta había uno o dos párrafos sobre lo que dice la ley, diciendo: “Aquí están las normas en tu Estado, aquí están tus derechos, aquí están los derechos de la otra persona”. Luego, la IA en sus respuesta animaba a la persona a buscar ayuda de expertos, a acudir a su tribunal, a su organización de ayuda legal, e incluso daba un número de teléfono al que llamar. Y la respuesta terminaba con una lista de puntos de contacto de organizaciones jurídicas. Según comentó la Dra. Hagan, este tipo de respuesta le encantaba a los participantes, porque eran cuatro o cinco párrafos, que los participantes encontraron increíblemente valiosos por su simplicidad, porque eran directos, claros, estaban orientados a la acción, y además, contaban con un formato agradable (frases cortas y listas de viñetas). En opinión de los participantes era mejor que buscar en Google y obtener una larga lista de páginas entre las que elegir.

Antes de que la gente utilizara la IA, la Dra. Hagan y su equipo les preguntó: “¿Cuánto crees que te fiarás de lo que te diga la IA?” utilizando una escala de 0 a 5, la mayoría de personas estaba en 2 o 3 de 5. Sin embargo, después de la interacción con la IA, les hicieron la misma pregunta y los participantes tendían a puntuar más alto su confianza en la respuesta obtenida de la IA, pasaron de un nivel medio a alto. Otra pregunta que se les realizó a los participantes era “¿en qué medida te ha ayudado la IA?”. A lo cual, la mayoría de la gente puntuó 5 sobre 5, es decir, que era extremadamente útil. También les preguntaron después de usar la IA, “¿qué probabilidades hay de que vuelvas a usarla si tienes un problema similar en el futuro?” y los participantes en su mayoría puntuaron 5 sobre 5, es decir, que era muy probable que volvieran a utilizar esta herramienta si tenían un problema legal similar. En suma, los participantes que antes de usar la IA eran un poco escépticos frente a sus resultados, después de usarla se mostraron más confiados.

¿Por qué confía la gente en este tipo de IA? Fue otra pregunta que les realizaron a los participantes, ante lo cual se expresaron varias razones. Por un lado, la información que recibieron les suena a cosas que ya saben, quizás ya conocen un poco sus derechos y la IA les está diciendo algo que parece cierto. Otras veces, según contó la Dra. Hagan, los participantes expresaron que confiaban en la empresa tecnológica porque respetan a Google y Google parece saberlo todo en Internet; bajo está lógica, si Google publica esta IA, el usuario cree que si Google lo dice tiene que ser verdad. Por ejemplo, un joven de unos 20 años con estudios universitarios, que trabaja en la gestión de la construcción, encontró realmente valiosa la IA porque le facilitaba mucho la desalentadora tarea de hacer investigaciones jurídicas por su cuenta. Desde su perspectiva, era una herramienta potente, que facilitaba las cosas y le daba la sensación de estar tratando con un ser humano. A otra usuaria, una mujer negra de unos 30 años, con estudios universitarios, que trabaja en recursos humanos, le gustó la IA porque lo explicaba todo paso a paso, no utilizaba términos jurídicos, lo explicaba todo de forma que ella pudiera entenderlo, lo dejaba muy claro y parecía muy fiable. Otra de las personas que se mostró un poco más escéptica, una mujer blanca de unos 50 años, con estudios universitarios, expresó que la IA le había aportado algo de valor, que le había dado una idea de los recursos que hay, que hay formularios, que hay

agencias, en otras palabras, que le había dado un punto de partida, pero dijo que no era extremadamente útil porque todavía tenía muchas preguntas.

La Dra. Hagan y su equipo encontró que entre las 50 personas que entrevistaron había una gran variedad de formas en las cuales los distintos usuarios utilizarían la IA. Algunas personas, a las que llamaron “Primer grupo” confiaban mucho en lo que les dice la IA. De hecho, cuando les preguntaron “¿Qué harías en la vida real con tu casero?”, respondieron que “Sacarían su teléfono, tomarían una captura de pantalla de lo que dice la IA, y se la enviarían al casero, con un mensaje en el que expresaran que esto es lo que dice la ley y, por ende, deben darles más tiempo o devolverles su dinero”. Este grupo, creía tanto en la respuesta de la IA que pensaban que podían utilizarla casi como prueba de la verdad.

Según describió la Dra. Hagan, en el segundo grupo estaban los participantes que se fijaban un poco en los detalles de la respuesta de la IA y prestaban atención, pero estaban muy sesgados hacia su propia postura. Así que leían lo que la IA les decía, incluso lo que podía ir en contra de su caso, pero solo se fijaban en lo positivo. Así que, en todos los detalles, en todo el contexto, solo buscaban algo que estuviera de su parte, que apoyara su versión de los hechos: seleccionaron eso e ignoraron todo lo demás. Así, si veían un detalle en el que la IA les decía que tenían 30 días más para quedarse en su casa, se centraban en eso, lo veían como la verdad, e ignoraban todo lo demás. La Dra. Hagan explicó que este grupo muestra que los participantes no siempre leen todo lo que la IA les está diciendo y podrían cometer un error por centrarse demasiado en los aspectos positivos y no mirar los negativos.

En un tercer grupo, se clasificaron a los participantes que tenían un grado muy alto de capacidad de pensamiento crítico. En otras palabras, asimilaban todo lo que IA les decía para entender sus derechos, para entender la ley, pero decían: “Este es un buen punto de partida, pero tengo que investigar más, tengo que consultar fuentes más autorizadas y directas de la ley”. Aunque ellos agradecían que el chatbot les ofreciera un panorama básico, no confiaban totalmente en la respuesta generada por este y preferían investigar un poco más.

La Dra. Hagan finalizó su intervención contando que junto con su equipo hacían un seguimiento del rendimiento de la IA. Ellos encontraron que la IA daba mucha información buena, pero también cometía muchos errores. En este punto, la Dra. Hagan enfatizó en que era necesario recordar que esta investigación se realizó hace dos años y que, desde entonces, los chatbot han cambiado mucho. El equipo de investigadores observaron que la IA inventaba organizaciones de asistencia jurídica, organizaciones que no existían. También, la IA alucinaba con casos legales que no existían o utilizaba las normas de una jurisdicción en otra y, por tanto, comunicaba a los usuarios información errónea debido a una confusión geográfica. Ella también resaltó que los usuarios cometían errores, así que, a veces, aunque la IA ofrecía un buen resumen de la ley, los usuarios la malinterpretaban o seleccionaban solo los detalles favorables.

La Dra. Hagan expresó su agradecimiento por el espacio para hablar sobre su investigación y destacó que esta puede dar una idea general de hacia dónde vamos en los servicios jurídicos y la IA, y también sobre cómo es más probable que cada vez las personas del común utilicen la IA para las tareas jurídicas.

El Defensor Delgado para la Protección de los Derechos Digitales y la Libertad de Expresión, Santiago Pardo Rodríguez, agradeció a la Dra. Hagan por su intervención y abrió el espacio para que personas del público le pudieran realizar preguntas a la invitada. A continuación, se presentan las preguntas y respuestas.

**¿Usted considera que los altos índices de desaprobación que tienen las instituciones públicas tanto en los Estados Unidos como en Colombia podrían explicar, de alguna manera, el uso masivo de inteligencia artificial?**

La Dra. Hagan comentó que, en un taller con jueces estatales y con representantes de Google, un juez les preguntó ¿por qué cuando alguien busca sobre temas legales no los enviaban a su sitio oficial? A lo cual, los representantes de Google manifestaron que “los usuarios no quieren ir a ese sitio y ellos los envían a donde los usuarios quieren ir”. Esto refleja que las personas confían en el tribunal si el tribunal les da la información autorizada en la que pueden confiar, pero lo quieren de una manera fácil de usar, lo quieren en un lenguaje sencillo, procesable, de manera comprensible. Y, sinceramente, confían mucho en las empresas tecnológicas, especialmente en empresas como Google, que, ya sabes, si lo buscas todo en Google, Google lo sabe todo. Los usuarios no saben exactamente cómo funciona la búsqueda, pero realmente confían en que estas empresas están más informadas, actualizadas, y también que tienen el conocimiento de otras personas, lo que significa que pueden ver cómo la gente está hablando en Reddit u otros lugares donde la gente real está hablando de sus experiencias y cómo funciona. Por eso mucha gente empieza en una empresa tecnológica en lugar de ir directamente al juzgado o al sitio web del tribunal para informarse.

**Teniendo en consideración que los datos de esta investigación corresponden a 2023, ¿ustedes cuentan con nuevos datos que indiquen si las alucinaciones y otro tipo de problemas asociados al uso de la IA se han corregido o están siendo abordados por las empresas tecnológicas?**

La Dra. Margaret Hagan comentó que este es un proyecto de investigación en curso, porque los modelos de lenguaje a gran escala se actualizan todo el tiempo. Ella agregó que tienen un grupo de expertos revisando la precisión de las leyes, organizaciones, reglas y todos los demás detalles legales clave que suelen proporcionar los chatbot en sus respuestas. Con información obtenida en el verano de 2024, encontraron que estos modelos han mejorado reduciendo la cantidad de alucinaciones. Antes, en 2023, simplemente inventaban números de teléfono, inventaban organizaciones legales, e inventaban todo tipo de cosas, pero a partir de 2024 han observado mejoras en los datos. Aunque, ella reconoció que todavía estaban observando problemas en temas clave. Un primer problema asociado a la jurisdicción, lo cual implica que la información brindada a los usuarios de un Estado no

necesariamente corresponde con la realidad jurídica de ese Estado. Por ejemplo, cuando se refieren al número de días que tienen los usuarios para presentar un documento, cuáles son sus derechos o el mejor curso de acción. La IA decía cosas con mucha confianza a un usuario, pero cuando un experto revisaba si esa información correspondía con la jurisdicción de ese Estado encontraban errores e imprecisiones. Al respecto, la Dra. Hagan expresó que podría ocurrir que la IA esté tomando un promedio de todos los estados y, luego, produciendo una especie de mezcla de cosas generales, pero sin dar una respuesta precisa para el contexto local del usuario.

Un segundo problema que identificaron es que las respuestas generadas por la IA tienden a dar demasiada confianza a las acciones emprendidas por el usuario y las hacen ver fáciles, cuando en la realidad son tareas muy arduas. Por ejemplo, en algunos lugares de los Estados Unidos, se expresó que “si tu arrendador no está reparando tu casa, puedes dejar de pagar el alquiler, retener tu alquiler”, aunque en la realidad, si la persona hace eso y no sigue las reglas verdaderas, se equivocará y su arrendador podrá desalojarlo legalmente. Los expertos que revisaron las respuestas, manifestaron que la IA no está citando la ley incorrectamente, pero está haciendo que parezca fácil y es demasiado alentadora para el usuario. En este punto la Dra. Hagan reconoció que, en la práctica si se animan a los usuarios a que lo hagan ellos mismos, muchas veces cometerán un error y se meterán en una situación mucho peor, y aún no se está calibrado el nivel de riesgo al que se exponen.

**¿Qué tan receptivas son estas empresas de IA, por ejemplo, OpenAI o Google, a los resultados de su investigación? ¿Todavía hay una brecha de comunicación con estos desarrolladores?**

La Dra. Hagan manifestó que el principal problema es que no está claro cómo plantear problemas, no hay un defensor del pueblo, no hay una vía clara para la formulación de políticas, así que se deben probar muchas estrategias diferentes. En su experiencia, ellos realizan muchos talleres y espacios en los cuales se sientan diferentes actores, desde jueces hasta ingenieros, con el propósito de entender al sector jurídico. Desde su perspectiva, no se entiende muy bien al sector jurídico y dónde encontrar la información autorizada. Lo cual puede conducir a que, por defecto, simplemente “raspando” Internet y utilizando sus modelos se considere que está realizada la tarea, cuando se pueden cometer muchos errores. Ella expresó que gran parte de lo que tenemos que hacer en el sector de la justicia es hacer una mezcla de educación a los tecnólogos sobre lo que hace que este dominio sea especial, y que lo que funciona en otros sectores, no necesariamente funciona en el sector jurídico. Por ejemplo, expresarles por qué las regiones geográficas realmente importan, y lo devastadoras que las consecuencias pueden ser para una persona si se les da información errónea.

La Dra. Hagan argumentó que realizar informes académicos es muy valioso para mostrar con datos que algo está ocurriendo y para que la información no se quede simplemente en lo anecdótico. Ella resaltó la importancia de realizar evaluaciones comparativas sistemáticas, puntuaciones y clasificaciones. Por ejemplo, ellos están haciendo auditorías con el fin de

generar una puntuación de referencia cuantitativa que diga, a partir de 2024, cómo iban los diferentes modelos: ChatGPT 3.5 frente a Gemini, frente a Meta, frente a Antrópico, entre otros. Esta calificación podría llevarlos a prestar atención al sector legal y tratar de trabajar para mejorar sus modelos.

Desde su perspectiva, ella considera que las empresas sí están abiertas a estas investigaciones, porque están en una carrera entre ellos por la calidad, por la financiación, para que los consumidores los usen y confíen en ellos. Esa competencia puede contribuir a hacerlos más sensibles frente a este tipo de trabajos.

**Frente a los cambios que se están presentando rápidamente en el mercado de este tipo de tecnologías, ¿qué tipo de escenarios se podrían presentar? Por ejemplo, ¿es posible que, en el futuro cercano, haya una IA de alta calidad, muy fiable y asombrosa disponible solo para aquellos que puedan pagarla, mientras las gratuitas serán de menor calidad y con más problemas?**

La Dra. Hagan manifestó que este es el momento para presionar a las empresas. Desde su perspectiva, este año, en particular, es el momento de tratar de llamar la atención de las empresas de tecnología para asegurarse de que, a medida que, desarrollan sus modelos, están dando un trato especial al sector legal, porque existe esa sensación por parte de muchos ingenieros de que los modelos de IA por su cuenta y las técnicas se encargarán de ello. En otras palabras, existe la idea de que no es necesario prestar ninguna atención especial a sectores concretos como el jurídico, el financiero o el médico, que los modelos lo resolverán por sí solos. No obstante, ante esta perspectiva, la Dra. Hagan enfatizó en que debemos prestar especial atención a la información jurídica, porque hay mucho en juego y muchos matices de un lugar a otro. Ella agregó que no quiere dejar que los tecnólogos piensen que los modelos serán lo bastante inteligentes como para saber todas estas cosas por sí solos. Por lo cual, tanto para la asequibilidad y que la IA de alta calidad pueda estar abierta incluso para aquellos que no pueden costearlas, es fundamental constituir grupos interdisciplinarios que trabajen en ello.

El Defensor Delegado, Santiago Pardo Rodríguez, agradeció la intervención de la Dra. Hagan e invitó a los participantes a darle un fuerte aplauso.

## **5. CIERRE Y COMPROMISO**

Al final del evento el Defensor Delegado para la Protección de los Derechos Digitales y la Libertad de Expresión, Santiago Pardo Rodríguez agradeció a los asistentes por su activa participación en el taller y cedió la palabra a la profesora Sandra Ortiz de la Universidad Externado de Colombia para que concluyera el evento.

La profesora Sandra Ortiz expresó su agradecimiento a los asistentes, reconoció que estos espacios son particularmente productivos, en un espacio que se viene desarrollando desde 2023. La profesora invitó a que se analice el Documento CONPES 4144 de Inteligencia

Artificial para que se busque, en las más de 100 acciones que plantea, cómo se puede materializar ese modelo de gobernanza abierta. También manifestó que la próxima Mesa está prevista para realizarse en el mes de mayo y que la discusión de esa Mesa girará en torno a los modelos de gobernanza abierta que se necesitan y que propone el Documento CONPES 4144. La profesora Ortiz reflexionó acerca de que el objetivo general de estas Mesas es generar una propuesta de instrumento normativo en el marco de un modelo de la gobernanza. De allí que invitó a acompañar este proceso y agradeció a la Defensoría del Pueblo por haber tomado la iniciativa de liderar y realizar la Octava Mesa en sus instalaciones.

\*\*\*