



Asamblea General

Distr. general
29 de agosto de 2018
Español
Original: inglés

Septuagésimo tercer período de sesiones

Tema 74 b) del programa provisional**

**Promoción y protección de los derechos humanos:
cuestiones de derechos humanos, incluidos otros
medios de mejorar el goce efectivo de los derechos
humanos y las libertades fundamentales**

Promoción y protección del derecho a la libertad de opinión y expresión***

Nota del Secretario General

El Secretario General tiene el honor de transmitir a la Asamblea General el informe preparado por el Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Sr. David Kaye, presentado de conformidad con la resolución [34/18](#) del Consejo de Derechos Humanos. En el informe el Relator Especial estudia las consecuencias de las tecnologías de inteligencia artificial para los derechos humanos en el entorno de información, centrándose en particular en los derechos a la libertad de opinión y expresión, a la privacidad y a la no discriminación.

* Publicado nuevamente por razones técnicas el 26 de octubre de 2018.

** [A/73/150](#).

*** Este informe se presenta después del plazo establecido para incluir la información más reciente.



Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión

Índice

	<i>Página</i>
I. Introducción	3
II. Entender la inteligencia artificial	3
A. ¿Qué es la inteligencia artificial?	3
B. La inteligencia artificial y el entorno de información	6
III. Marco jurídico de derechos humanos para la inteligencia artificial	10
A. Alcance de las obligaciones de derechos humanos en el contexto de la inteligencia artificial	10
B. Derecho a la libertad de opinión	11
C. Derecho a la libertad de expresión	12
D. Derecho a la privacidad	14
E. Obligación de no discriminar	15
F. Derecho a un recurso efectivo	16
G. Respuestas legislativas, reglamentarias y normativas a la inteligencia artificial	16
IV. Enfoque de la inteligencia artificial basado en los derechos humanos	18
A. Normas sustantivas para los sistemas de inteligencia artificial	19
B. Procesos para los sistemas de inteligencia artificial	20
V. Conclusiones y recomendaciones	22

I. Introducción

1. La inteligencia artificial (IA) influye cada vez más en el entorno de información en todo el mundo. Permite a las empresas administrar resultados de búsquedas y transmisiones de noticias además de la colocación de anuncios publicitarios, organizando lo que ven los usuarios y cuándo lo ven. Las compañías de medios sociales usan tecnologías de IA como instrumento para moderar el contenido de sus plataformas, que a menudo sirve de primera línea de defensa contra contenidos que pueden infringir sus reglas. La IA recomienda personas como amigos o seguidores, artículos de noticias que leer y lugares que visitar o en los cuales comer, hacer compras o dormir. Ofrece velocidad, eficiencia y magnitud y ayuda a las empresas más grandes del sector de la tecnología de la información y las comunicaciones a administrar la gran cantidad de contenido que se carga cada día en sus plataformas. Las tecnologías de IA pueden permitir un intercambio más amplio y rápido de información e ideas a nivel mundial, que es una gran oportunidad para la libertad de expresión y el acceso a la información. Al mismo tiempo, la opacidad de la IA también crea el riesgo de injerencia en la autodeterminación individual o lo que se llama en el presente informe “autonomía y acción individuales”¹. Una gran cuestión mundial se plantea a todos los que promueven los derechos humanos y el estado de derecho: ¿cómo pueden los Estados, las empresas y la sociedad civil asegurar que las tecnologías de IA refuercen y respeten los derechos humanos, en vez de socavarlos y ponerlos en peligro?

2. El presente informe no pretende ser la última palabra en IA y derechos humanos. Trata más bien de hacer tres cosas: definir los términos clave esenciales para un debate sobre la IA en relación con los derechos humanos, identificar el marco jurídico de derechos humanos pertinente a la IA y presentar algunas recomendaciones preliminares para asegurar que, a medida que evolucionan las tecnologías que forman la IA, se incorporen en ese proceso consideraciones de derechos humanos. El informe debe leerse como complemento de mi informe más reciente al Consejo de Derechos Humanos (A/HRC/38/35), que presenta un enfoque de la moderación del contenido en línea basado en los derechos humanos².

II. Entender la inteligencia artificial

A. ¿Qué es la inteligencia artificial?

3. Con IA se suele hacer referencia al aumento de la independencia, la velocidad y la magnitud relacionado con la adopción informática y automatizada de decisiones. No es una sola cosa, sino más bien una “constelación” de procesos y tecnologías que permiten que las computadoras complementen o reemplacen tareas específicas que de otro modo serían ejecutadas por seres humanos, como tomar decisiones y resolver problemas³. “IA” puede ser un término problemático, porque sugiere que las

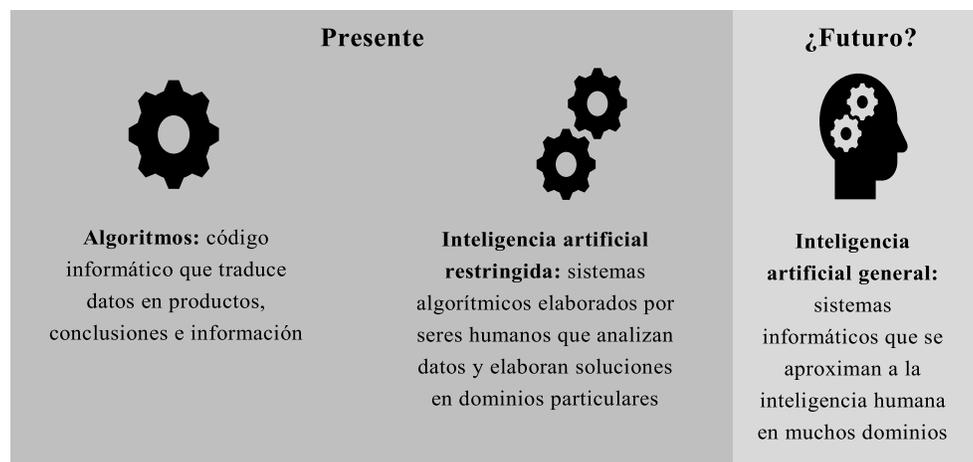
¹ Véase Mariarosaria Taddeo y Luciano Floridi, “How AI can be a force for good”, *Science*, vol. 361, núm. 6404 (24 de agosto de 2018). Disponible en <http://sciencemag.org/content/361/6404/751.full>.

² El presente informe se benefició de una consulta de expertos celebrada en Ginebra en junio de 2018, financiada por una donación de la Unión Europea, y de las contribuciones de expertos a la elaboración del documento A/HRC/35/38 en 2017 y 2018. El Relator Especial desea dar las gracias especialmente a Carly Nyst y a Amos Toh, que aportaron trabajos esenciales de investigación y redacción a este proyecto.

³ Véanse AI Now, “The AI now report: the social and economic implications of artificial intelligence technologies in the near term”, 2016, disponible en https://ainowinstitute.org/AI_Now_2016_Report.pdf, y Comité Especial sobre Inteligencia

máquinas pueden funcionar según los mismos conceptos y reglas que la inteligencia humana. No es así. La IA generalmente optimiza la ejecución de tareas computarizadas asignadas por seres humanos mediante repetición e intentos iterativos. Sin embargo, como este es el término utilizado en la cultura, las empresas y los Gobiernos, es el que se usa aquí.

4. La cultura popular a menudo sugiere que la sociedad avanza hacia la IA general, que es la capacidad todavía distante (la “singularidad”) de un sistema informático que se aproxime a la inteligencia humana o la supere en muchos dominios⁴. En el futuro previsible seguirá habiendo progresos en la IA restringida, con la cual los sistemas informáticos ejecutan tareas programadas (algoritmos elaborados por seres humanos) en dominios particulares. Por ejemplo, la IA restringida sirve de base a la asistencia vocal en dispositivos móviles y *chatbots* de servicio al cliente, a los instrumentos de traducción en línea y a los automóviles autoconducidos, a los resultados de motores de búsqueda y a los servicios de mapeo. El aprendizaje automático es una clase de técnicas de IA restringida que se usan en la capacitación de algoritmos para que usen conjuntos de datos a fin de reconocer y ayudar a resolver problemas. Por ejemplo, los dispositivos domésticos inteligentes dotados de IA “aprenden” continuamente, de los datos reunidos sobre el lenguaje cotidiano y las pautas lingüísticas, a procesar y responder más exactamente a las preguntas de sus usuarios. En todos los casos los seres humanos desempeñan un papel fundamental en el diseño y la difusión de sistemas de IA, definiendo los objetivos de una aplicación de IA y, según el tipo de aplicación, eligiendo y etiquetando conjuntos de datos y clasificando productos. Los seres humanos siempre determinan la aplicación y el uso de los productos de la IA, incluido el grado en que complementan o reemplazan la adopción de decisiones humana.



5. En la base de la IA están los algoritmos, códigos informáticos diseñados y escritos por seres humanos que ejecutan instrucciones para traducir datos en conclusiones, información o productos. Hace tiempo que los algoritmos son esenciales para el funcionamiento de los sistemas cotidianos de comunicación e infraestructura. El enorme volumen de datos de la vida moderna y la capacidad de analizarlos impulsan la IA. El sector privado ciertamente ve los datos de esa manera: cuantos más datos estén disponibles para alimentar los algoritmos y cuanto mejor sea la calidad de esos datos, más poderosos y precisos pueden llegar a ser los algoritmos.

Artificial de la Cámara de los Lores del Reino Unido de Gran Bretaña e Irlanda del Norte, “AI in the United Kingdom: ready, willing and able?”, 2018, pág. 13.

⁴ Artículo 19 y Privacy International, “Privacy and freedom of expression in an age of artificial intelligence”, Londres, 2018, pág. 8.

Los sistemas algorítmicos pueden analizar rápidamente grandes volúmenes de datos y así hacer que los programas de IA puedan ejecutar funciones de adopción de decisiones que antes eran dominio de seres humanos que actuaban sin instrumentos informáticos.



La acción humana es parte integrante de la IA, pero las características distintivas de la IA merecen escrutinio desde el punto de vista de los derechos humanos con respecto a por lo menos tres de sus aspectos: la automatización, el análisis de datos y la adaptabilidad⁵.

6. **Automatización.** La automatización elimina la intervención humana en partes de un proceso de adopción de decisiones cumpliendo tareas determinadas con instrumentos informáticos. Esto puede tener consecuencias positivas desde el punto de vista de los derechos humanos si el diseño limita el sesgo humano. Por ejemplo, un sistema automatizado de entrada en la frontera puede señalar a una persona para que sea objeto de un examen detallado sobre la base de características objetivas como los antecedentes penales o la situación de visado y limitar el uso de evaluaciones subjetivas (y tendenciosas) de la presentación física, el grupo étnico, la edad o la religión. La automatización también permite procesar grandes cantidades de datos a una velocidad y en una cantidad que los seres humanos no pueden alcanzar,

⁵ Consejo de Europa, *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, estudio del Consejo de Europa, núm. DGI (2017) 12, 2018. Disponible en <https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published>, pág. 5.

potencialmente en beneficio de la seguridad pública, la salud y la seguridad nacional. Pero los sistemas automatizados se basan en conjuntos de datos que, en su diseño o aplicación, pueden incluir sesgos y por tanto producir efectos discriminatorios. Por ejemplo, los antecedentes penales o los datos relativos al visado mencionados pueden ellos mismos incorporar sesgos. Basarse y confiar demasiado en decisiones automatizadas y no reconocer esta cuestión fundamental puede a su vez socavar el examen crítico de los resultados de la IA e impedir a las personas el acceso a recursos contra decisiones adversas tomadas con IA. La automatización puede dificultar la transparencia y la escrutabilidad de un proceso e impedir que incluso autoridades bien intencionadas den una explicación de los resultados⁶.

7. **Análisis de datos.** Grandes conjuntos de datos forman la base de la mayoría de las aplicaciones de IA. Cualquier conjunto de datos puede formar la base de un sistema de IA, desde los hábitos de navegación en Internet hasta los datos sobre corrientes de tráfico en las carreteras. Algunos conjuntos de datos contienen datos personales, otros datos anonimizados. El uso de esos conjuntos de datos por la IA plantea preocupaciones graves, en particular con respecto a sus orígenes, su exactitud y los derechos de las personas sobre ellos; la capacidad de los sistemas de IA de desanonimizar los datos anonimizados; y los sesgos que pueden estar establecidos en los conjuntos de datos o ser infundidos por capacitación o etiquetado humano de los datos. La evaluación de datos por IA puede determinar correlaciones, pero no necesariamente causalidad, lo que puede producir resultados sesgados y defectuosos que son difíciles de escrutar.

8. **Adaptabilidad.** Los sistemas de IA de aprendizaje automático son adaptables, porque los algoritmos en que se basan son capaces de determinar progresivamente nuevos problemas y elaborar nuevas respuestas. Según el grado de supervisión, los sistemas pueden determinar pautas y elaborar conclusiones imprevistas por las personas que los programaron o les asignaron tareas. Esta falta de previsibilidad es la verdadera promesa de la IA como tecnología transformativa, pero también revela sus riesgos: a medida que los seres humanos son excluidos progresivamente de la definición de los objetivos y productos de un sistema de IA, se hace más difícil asegurar la transparencia, la rendición de cuentas y el acceso a un recurso efectivo, al igual que prever y mitigar los efectos adversos en los derechos humanos.

B. La inteligencia artificial y el entorno de información

9. La IA tiene consecuencias particularmente importantes, y a veces problemáticas, para el entorno de información, el ecosistema complejo de tecnologías, plataformas y actores privados y públicos que facilitan el acceso a la información y la difusión de información por medios digitales. Hay algoritmos y aplicaciones de IA en todos los rincones de Internet, en dispositivos digitales y sistemas técnicos y en motores de búsqueda, plataformas de redes sociales, aplicaciones de mensajería y mecanismos de información pública. Conforme al tema principal del mandato, el Relator Especial indica a continuación tres aplicaciones de la IA en el entorno de información que suscitan preocupación.

10. **Presentación y personalización del contenido.** Los medios sociales y las plataformas de búsqueda dominan cada vez más la forma en que las personas obtienen e intercambian información e ideas y la forma en que se difunden las noticias. Los algoritmos y las aplicaciones de IA determinan cuánto, cuándo y con qué públicos e individuos se comparte el contenido. Conjuntos de datos masivos que combinan historiales de navegación, datos demográficos del usuario, análisis semánticos y de

⁶ Consejo de Europa, *Algorithms and Human Rights*, pág. 8.

sentimientos y muchos otros factores se incorporan en modelos algorítmicos cada vez más personalizados para clasificar y administrar información, es decir, para presentarla a las personas o excluirla implícitamente. El contenido pagado, patrocinado o etiquetado puede ser promovido con exclusión o degradación de otro contenido. Las fuentes de noticias de las redes sociales presentan contenido según evaluaciones subjetivas del grado de interés o atracción que puede tener para el usuario, con lo cual las personas pueden tener poca o ninguna exposición a ciertos tipos de materiales o contenidos sociales o políticos críticos publicados en sus plataformas⁷. La IA configura el mundo de la información de una manera que es opaca para el usuario y a menudo incluso para la plataforma administradora.

11. La búsqueda en línea es una de las formas más generalizadas de presentación y personalización de contenidos mediante IA. Los motores de búsqueda presentan resultados de consultas (y completan o predicen consultas) usando sistemas de IA que procesan datos extensos sobre el usuario o un grupo de usuarios. Como es poco probable que se vean los contenidos que tienen rango bajo o que han sido completamente excluidos de los resultados de búsqueda, las aplicaciones de IA para la búsqueda tienen una enorme influencia sobre la difusión del conocimiento⁸. Los agregadores de contenido y los sitios de noticias⁹ análogamente eligen la información que presentan al usuario no sobre la base de acontecimientos recientes o importantes, sino según aplicaciones de IA que predicen el interés del usuario y las pautas de noticias sobre la base de conjuntos de datos extensos. En consecuencia, la IA desempeña un papel importante pero generalmente oculto en la determinación de la información que las personas consumen o incluso son conscientes de consumir.

12. La IA en el campo de la presentación de contenido tiende a aumentar la personalización de la experiencia en línea de cada persona; en esta era de abundancia de información¹⁰, la personalización promete ordenar el caos de Internet, permitiendo a las personas encontrar la información solicitada. Entre los beneficios pueden estar la capacidad de obtener información y servicios en más idiomas¹¹ o información que sea más oportuna y pertinente a la experiencia o las preferencias personales. La personalización impulsada por la IA también puede minimizar la exposición a puntos de vista diversos, lo que interfiere con la iniciativa individual de buscar e intercambiar ideas y opiniones entre grupos ideológicos, políticos o sociales. Esa personalización puede reforzar los sesgos e incentivar la promoción y recomendación de contenido incendiario o desinformación para mantener la participación en línea del usuario¹².

⁷ World Wide Web Foundation, “The invisible curation of content: Facebook’s News Feed and our information diets”, 22 de abril de 2018. Disponible en <https://webfoundation.org/research/the-invisible-curation-of-content-facebooks-news-feed-and-our-information-diets/>.

⁸ Consejo de Europa, *Algorithms and Human Rights*, pág. 17.

⁹ Por ejemplo, véanse “How Reuters’s revolutionary AI system gathers global news,” MIT Technology Review, 27 de noviembre de 2017, disponible en <https://www.technologyreview.com/s/609558/how-reuterss-revolutionary-ai-system-gathers-global-news/>, y Paul Armstrong y Yue Wang, “China’s \$11 billion news aggregator Jinri Toutiao is no fake,” *Forbes*, 26 de mayo de 2017, disponible en <https://www.forbes.com/sites/ywang/2017/05/26/jinri-toutiao-how-chinas-11-billion-news-aggregator-is-no-fake/#1d8b97804d8a>.

¹⁰ Carly Nyst y Nick Monaco, *State-Sponsored Trolling: How Governments are Deploying Disinformation as Part of Broader Digital Harassment Campaigns* (Palo Alto, Institute for the Future, 2018).

¹¹ World Wide Web Foundation, “Artificial intelligence: the road ahead in low- and middle-income countries”, Washington D.C., 2017.

¹² Zeynep Tufekci, “YouTube, the great radicaliser”, *New York Times*, 10 de marzo de 2018. Disponible en <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>; James Williams, *Stand Out of our Light: Freedom and Resistance in the Attention Economy* (Cambridge, Cambridge University Press, 2018).

Sin duda, todo tipo de situaciones sociales y culturales pueden limitar la exposición de la persona a la información. Pero al optimizar los sistemas para estimular la participación y la viralidad en gran escala, la personalización asistida por IA puede socavar la elección de la persona de encontrar ciertos tipos de contenido. Esto se debe especialmente a que los algoritmos típicamente reducen la prioridad del contenido que tiene niveles más bajos de participación y relegan a la oscuridad el contenido independiente o generado por usuarios¹³. Los actores informados pueden explotar los sistemas de IA basados en reglas optimizados para estimular la participación a fin de conseguir un grado más alto de exposición, y apropiándose de etiquetas populares o usando *bots* pueden lograr un alcance en línea muy grande en detrimento de la diversidad de la información.

13. **Moderación y eliminación de contenido.** La IA ayuda a las empresas de medios sociales a moderar el contenido según las normas y reglas de la plataforma, en particular la detección de correo basura, la tecnología de comparación de hash (uso de huellas digitales para detectar, por ejemplo, contenido terrorista o de explotación de niños), filtros de palabras clave, procesamiento de lenguas naturales (por el cual la naturaleza del contenido se juzga según palabras o imágenes prohibidas) y otros algoritmos de detección. La IA se puede usar para imponer una advertencia, suspensión o desactivación a la cuenta del usuario por infracción de las condiciones de servicio o para bloquear o filtrar sitios web según datos o contenidos de dominios prohibidos. Las compañías de medios sociales usan la IA para filtrar contenido en toda la gama de sus reglas (de la desnudez al acoso y al discurso de odio, etc.), aunque no se sabe hasta qué punto dichas empresas usan la automatización sin intervención humana en casos determinados¹⁴.

14. El apoyo y la presión a favor del aumento del papel de la IA proceden tanto del sector privado como del sector público. Las empresas afirman que el volumen de contenido ilegal, inadecuado y nocivo en línea supera con mucho la capacidad de moderación humana y argumentan que la IA es uno de los instrumentos que pueden ayudar a atacar mejor este problema. Según algunas plataformas, la IA no solo es más eficiente en la detección de contenido inadecuado (según sus reglas) y contenido ilegal para su eliminación (generalmente por un moderador humano) sino que también tiene un grado de exactitud más alto que la adopción humana de decisiones. Los Estados, por su parte, ejercen presión a favor de una moderación automatizada eficiente y rápida en relación con una serie de problemas, desde el abuso sexual de niños y el contenido terrorista, donde la IA ya se despliega ampliamente, hasta los derechos de autor y la eliminación de contenido “extremista” y “de odio”¹⁵. La

¹³ Recientemente algunas plataformas tecnológicas han indicado su intención de pasar de la personalización basada en la “participación” a una personalización que prioriza la calidad de la experiencia en línea del usuario; véase Julia Carrie Wong, “Facebook overhauls News Feed in favour of ‘meaningful social interactions’”, *The Guardian*, 11 de enero de 2018. Disponible en <https://www.theguardian.com/technology/2018/jan/11/facebook-news-feed-algorithm-overhaul-mark-zuckerberg>. Sin embargo, sin un alto grado de transparencia, notificación y métrica de la forma en que los sistemas de inteligencia artificial hacen y aplican esas evaluaciones, es difícil estimar si este cambio tiene un efecto demostrable en la experiencia de los usuarios de Internet.

¹⁴ Un instrumento de Instagram, Deep Text, intenta juzgar la “toxicidad” del contexto además de permitir al usuario adaptar sus filtros propios de palabras y emojis, y también evalúa la relación del usuario a fin de establecer el contexto (por ejemplo, si un comentario es solo una broma entre amigos). Andrew Hutchison, “Instagram’s rolling out new tools to remove ‘toxic comments’”, *Social Media Today*, 30 de junio de 2017. Disponible en <https://www.socialmediatoday.com/social-networks/instagrams-rolling-out-new-tools-remove-toxic-comments>.

¹⁵ Según informes el Reino Unido ha elaborado un instrumento para detectar y eliminar automáticamente el contenido terrorista en el punto de carga. Por ejemplo, véase Home Office, “New technology revealed to help fight terrorist content online”, comunicado de prensa, 13 de febrero de 2018. Véanse Comisión Europea, propuesta de directiva del Parlamento Europeo y

Recomendación de la Comisión Europea sobre medidas para seguir aumentando la eficacia de la lucha contra el contenido ilícito en línea de marzo de 2018 insta a las plataformas de Internet a que usen filtros automáticos para detectar y eliminar el contenido terrorista y sugiere que en algunos casos se use la revisión humana como contrapeso necesario de los errores inevitables causados por los sistemas automatizados¹⁶.

15. Los esfuerzos de automatizar la moderación del contenido pueden tener un costo para los derechos humanos (véase [A/HRC/38/35](#), párr. 56). La moderación del contenido por IA tiene varias limitaciones, en particular el problema de evaluar el contexto y tener en cuenta la variación generalizada de las claves lingüísticas, los significados y las particularidades lingüísticas y culturales. Como las aplicaciones de IA a menudo se basan en conjuntos de datos que incorporan supuestos discriminatorios¹⁷, y en condiciones en que el costo de la moderación excesiva es bajo, hay mucho riesgo de que esos sistemas supriman contenido en línea o suspendan cuentas que no son problemáticas y de que el contenido se suprima según conceptos sesgados o discriminatorios. Por consiguiente, los grupos vulnerables son los que más tienden a ser desfavorecidos por los sistemas de moderación de contenido por IA. Por ejemplo, la aplicación DeepText de Instagram detectó “mexicano” como injurioso porque sus conjuntos de datos estaban poblados con datos en que “mexicano” estaba asociado con “ilegal”, término codificado negativamente en el algoritmo¹⁸.

16. La IA hace difícil escrutar la lógica en que se basan las medidas relativas al contenido. Incluso cuando la moderación del contenido algorítmico se complementa con la revisión humana —disposición que según las grandes plataformas de medios sociales es cada vez más inviable en la escala en que operan— la tendencia a respetar las decisiones automáticas (según los supuestos de objetividad indicados) dificulta el examen de los resultados de la moderación de contenido, especialmente cuando el diseño técnico del sistema impide ese tipo de transparencia.

17. **Perfilado, publicidad y selección de destinatarios.** Los progresos de la IA se han beneficiado del modelo de negocios basado en datos de Internet y han estimulado este modelo, que consiste en que las personas pagan el contenido y los servicios gratuitos con sus datos personales. Con los enormes recursos de datos acumulados durante años de seguimiento y perfilado en línea, las empresas pueden equipar los sistemas de IA con conjuntos de datos ricos para elaborar modelos de predicción y selección cada vez más precisos. Hoy la publicidad de actores privados y públicos se puede lograr a nivel individual; los consumidores y los votantes son objeto de una

del Consejo sobre los derechos de autor en el mercado único digital, COM(2016) 593 final, art. 13, y la carta del Relator Especial al Presidente de la Comisión Europea, referencia núm. OL OTH 41/2018, 13 de junio de 2018. Disponible en <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-OTH-41-2018.pdf>.

¹⁶ Recomendación de la Comisión de 1 de marzo de 2018 sobre las medidas para luchar eficazmente contra los contenidos ilícitos en línea (C(2018) 1177 final). Disponible en <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online>; véase también Daphne Keller, “Comment in response to European Commission’s March 2018 recommendation on measures to further improve the effectiveness of the fight against illegal content online”, Stanford Law School, Center for Internet and Society, 29 de marzo de 2018. Disponible en <http://cyberlaw.stanford.edu/publications/comment-response-european-commissions-march-2018-recommendation-measures-further>.

¹⁷ Véanse Aylin Caliskan, Joanna Bryson y Arvind Narayanan, “Semantics derived automatically from language corpora contain human-like biases”, *Science*, vol. 356, núm. 6334 (14 de abril de 2017), y Solon Barocas y Andrew Selbst, “Big data’s disparate impact”, *California Law Review*, vol. 104, núm. 671 (2016).

¹⁸ Nicholas Thompson, “Instagram’s Kevin Systrom wants to clean up the &#%@! Internet”, *Wired*, 14 August 2017. Disponible en <https://www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet/>.

“microselección” destinada a responder a las idiosincrasias individuales y a explotarlas.

18. La selección de destinatarios basada en la IA incentiva la reunión y explotación generalizadas de datos personales y aumenta el riesgo de manipulación del usuario mediante la difusión de desinformación. La selección puede perpetuar la discriminación lo mismo que privar de información o de oportunidades al usuario, por ejemplo permitiendo dirigir anuncios especiales de empleo y vivienda que excluyan a los trabajadores mayores, a las mujeres o a minorías étnicas¹⁹. Por ejemplo, en vez de exponer a las personas a la paridad y a la diversidad en los mensajes políticos, el uso de la microselección por las plataformas de redes sociales está creando una visión del mundo administrada desfavorable al discurso político pluralista.

III. Marco jurídico de derechos humanos para la inteligencia artificial

A. Alcance de las obligaciones de derechos humanos en el contexto de la inteligencia artificial

19. Los instrumentos de IA, como todas las tecnologías, deben diseñarse, elaborarse y desplegarse de modo que sean compatibles con las obligaciones de los Estados y las responsabilidades de los actores privados conforme al derecho internacional de los derechos humanos. El derecho de los derechos humanos impone a los Estados a la vez obligaciones negativas de abstenerse de aplicar medidas que afecten el ejercicio de la libertad de opinión y expresión y obligaciones positivas de promover el derecho a la libertad de opinión y expresión y de proteger su ejercicio.

20. Con respecto al sector privado, los Estados están obligados a garantizar el respeto de los derechos individuales²⁰, especialmente el derecho a la libertad de opinión y expresión, en particular protegiendo a las personas contra infracciones cometidas por particulares (artículo 2 1) del Pacto Internacional de Derechos Civiles y Políticos). Los Estados pueden cumplir esta obligación con medidas legislativas para restringir o influir en la elaboración y la implementación de aplicaciones de IA, con políticas relativas a la adquisición de aplicaciones de IA de empresas privadas por actores del sector público, con planes de autorregulación y corregulación y aumentando la capacidad de las empresas privadas de reconocer y priorizar el derecho a la libertad de opinión y expresión en sus actividades.

21. Las empresas también tienen responsabilidades en virtud del derecho de los derechos humanos que deben guiar la creación, adopción y movilización de las tecnologías de IA (A/HRC/38/35, párr. 10). Los Principios Rectores sobre las Empresas y los Derechos Humanos: Puesta en Práctica del Marco de las Naciones Unidas para “Proteger, Respetar y Remediar” establecen “normas de conducta mundial aplicables a todas las empresas que deben regir todas las operaciones de las

¹⁹ Julia Angwin, Noam Scheiber y Ariana Tobin, “Dozens of companies are using Facebook to exclude older workers from job ads”, ProPublica, 20 de diciembre de 2017. Disponible en <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>; Julia Angwin, Ariana Tobin y Madeleine Varner, “Facebook (still) letting housing advertisers exclude users by race”, ProPublica, 21 de noviembre de 2017. Disponible en <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>.

²⁰ Véase el principio 3 de los Principios Rectores sobre las Empresas y los Derechos Humanos: Puesta en Práctica del Marco de las Naciones Unidas para “Proteger, Respetar y Remediar” (A/HRC/17/31) y A/HRC/38/35, párrs. 6 a 8.

empresas cualquiera que sea el lugar en el que operen” (principio 11), incluidas las redes sociales y las empresas de búsqueda. Para adaptar las conclusiones de los Principios Rectores al dominio de la IA (*ibid.*, párr. 11), los Principios Rectores requieren que las empresas, como mínimo, contraigan compromisos de política de alto nivel orientados a respetar los derechos humanos de sus usuarios en todas las aplicaciones de IA (principio 16); se abstengan de causar o contribuir a cualquier efecto negativo para los derechos humanos de la tecnología de IA y prevenir o mitigar todo efecto negativo relacionado con sus operaciones (principio 13); hagan operaciones de diligencia debida sobre los sistemas de IA para determinar y resolver las posibles repercusiones en los derechos humanos (principios 17 a 19); apliquen estrategias de prevención y mitigación (principio 24); hagan un seguimiento continuo de las actividades relacionadas con la IA, en particular mediante consultas con los interesados y consultas públicas (principios 20 y 21), y ofrezcan mecanismos de reclamación accesibles para remediar los efectos negativos en los derechos humanos de los sistemas de IA (principios 22, 29 y 31).

B. Derecho a la libertad de opinión

22. La libertad de tener una opinión sin injerencia es un derecho absoluto, consagrado en el artículo 19 1) del Pacto Internacional de Derechos Civiles y Políticos y en el artículo 19 de la Declaración Universal de Derechos Humanos. El Pacto “no autoriza excepción ni restricción alguna”, sea “por ley u otro poder”²¹. En el informe que presentó en 2015 al Consejo de Derechos Humanos sobre el cifrado y el anonimato en las comunicaciones digitales ([A/HRC/29/32](#)), el Relator Especial observó que las formas en que se almacena, transmite y asegura la información en la era digital afectan singularmente el ejercicio del derecho a opinar. Las consultas de búsqueda, las actividades de navegación, las comunicaciones de correo electrónico y de textos y los documentos y recuerdos guardados en la nube: juntas, estas actividades y registros digitales forman la estructura de las opiniones que tienen los usuarios (*ibid.*, párr. 12). Tanto los actores estatales como los no estatales pueden intervenir en estos mecanismos y procesos de formación y sostenimiento de opiniones.

23. Un elemento esencial del derecho a sostener una opinión es el “derecho a formarse una opinión y a desarrollarla mediante el razonamiento”²². El Comité de Derechos Humanos ha llegado a la conclusión de que este derecho requiere que la persona esté libre de coacción indebida en el desarrollo de sus creencias, ideologías, reacciones y posiciones²³. En consecuencia, las intervenciones neurológicas forzadas, los programas de adoctrinamiento (como los “campamentos de reeducación”) y las amenazas de violencia destinadas a obligar a las personas a formarse opiniones determinadas o a cambiar de opinión infringen el artículo 19 1) del Pacto. El Comité también concluyó que el “aliciente de recibir trato preferente” coactivo puede llegar a un grado de persuasión que afecta el derecho a formarse y sostener opiniones (véase [CCPR/C/78/D/878/1999](#)).

24. La intersección de la tecnología y la administración de contenido plantea nuevas cuestiones sobre los tipos de coacción o incentivo que pueden considerarse injerencia en el derecho a formarse una opinión. La administración de contenido ha informado durante mucho tiempo la capacidad de la persona de formarse opiniones: por ejemplo, los medios de información elevan noticias determinadas a la primera página con la

²¹ Comité de Derechos Humanos, Observación general núm. 34 (2011) sobre la libertad de opinión y de expresión, párr. 9; disponible en www2.ohchr.org/english/bodies/hrc/docs/GC34.pdf; Manfred Nowak, *U.N. Covenant on Civil and Political Rights: CCPR Commentary* (1993).

²² Nowak, *U.N. Covenant on Civil and Political Rights*.

²³ *Yong Joo-Kang c. República de Corea*, Comité de Derechos Humanos, comunicación núm. 878/1999, 16 de julio de 2003 ([CCPR/C/78/D/878/1999](#)).

intención de modificar e influir en el conocimiento que la persona tiene de las noticias importantes del día. La publicidad comercial también ha buscado producir opiniones favorables a productos y servicios determinados y estimular el deseo de adquirirlos.

25. El uso de la IA extiende y aumenta la tradición de administración de contenido en Internet, ofreciendo medios más refinados y eficientes de personalizar y administrar el contenido para el usuario en una escala que está fuera del alcance de los medios de información tradicionales. El predominio de modos particulares de administración asistida por la IA es motivo de preocupación por su efecto en la capacidad de la persona de formar y desarrollar opiniones. Por ejemplo, un puñado de compañías tecnológicas recibe la gran mayoría de las consultas de búsqueda que se hacen en línea. El monopolio de unas pocas empresas sobre el mercado de búsquedas hace extremadamente difícil para el usuario salir de la clasificación algorítmica y la administración de los resultados de búsquedas y también puede inducirlo a creer (como desean las empresas) que los resultados producidos son la información más pertinente u objetiva disponible sobre un tema determinado. La falta de transparencia sobre la forma en que los criterios de búsqueda se elaboran y aplican mediante la IA también puede reforzar la suposición de que los resultados generados en una plataforma determinada son una presentación objetiva de información sobre los hechos.

26. En consecuencia, las cuestiones que plantea la dominación del mercado en el campo de la administración asistida con IA ponen a prueba las interpretaciones históricas de la forma en que la administración de contenido afecta o no la capacidad de formarse una opinión. La novedad de las cuestiones planteadas, junto con la falta general de jurisprudencia sobre la injerencia en el derecho de opinión, producen más cuestiones que respuestas sobre el efecto en los derechos humanos de la administración asistida con IA en el entorno digital actual. Sin embargo, estas cuestiones deben impulsar la investigación inspirada en los derechos sobre los efectos sociales, económicos y políticos de la administración asistida con IA. Las empresas deben, como mínimo, dar información pertinente sobre la forma en que se elaboran y aplican los criterios de administración y personalización de contenido en sus plataformas, en particular las políticas y procesos para detectar sesgos sociales, culturales o políticos en el diseño y elaboración de los sistemas de IA pertinentes.

C. Derecho a la libertad de expresión

27. El artículo 19 2) del Pacto garantiza un derecho amplio de “buscar, recibir y difundir informaciones e ideas de toda índole”, que debe ser protegido y respetado sin consideración de fronteras o tipo de medio de comunicación. El disfrute del derecho a la libertad de expresión está íntimamente relacionado con el ejercicio de otros derechos y es fundamental para el funcionamiento efectivo de las instituciones democráticas y, en consecuencia, la protección, el respeto y la promoción del derecho a la libertad de expresión implica el deber de incluir la promoción de la diversidad e independencia de los medios de información y la protección del acceso a la información²⁴.

²⁴ Relator Especial de las Naciones Unidas sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa, Relator Especial para la Libertad de Expresión de la Organización de los Estados Americanos y Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos, “Declaración conjunta sobre libertad de expresión y ‘noticias falsas’, desinformación y propaganda”, 3 de marzo de 2017, disponible en <https://www.osce.org/fom/302796>; véase también Comité de Derechos Humanos, Observación

28. A diferencia del derecho a formarse y sostener opiniones, los derechos a expresar y acceder a información e ideas pueden estar sujetos a restricciones en circunstancias limitadas (artículo 19 3) del Pacto). Las restricciones deben ajustarse a las normas de legalidad, lo que significa que deben ser dispuestas públicamente por una ley que cumpla normas de claridad y precisión y ser interpretadas por autoridades judiciales independientes; necesidad y proporcionalidad, lo que significa que deben ser la injerencia mínima necesaria para asegurar el interés legítimo en cuestión y no deben afectar la esencia del derecho; y legitimidad, lo que significa que deben proteger un interés legítimo enumerado, a saber, la protección de los derechos o la reputación de los demás, la seguridad nacional o el orden público o la salud o la moral públicas (A/HRC/38/35, párr. 7). Dentro de este marco, los derechos de expresión también pueden restringirse de conformidad con el artículo 20 2) del Pacto, que requiere a los Estados prohibir “toda apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia”, pero las restricciones deben cumplir las condiciones acumulativas de legalidad, necesidad y legitimidad²⁵.

29. La complejidad de la adopción de decisiones inherente a la moderación de contenido puede agravarse con la introducción de procesos automatizados. A diferencia de los seres humanos, los algoritmos no son hoy capaces de evaluar el contexto cultural, detectar la ironía o hacer el análisis crítico necesario para detectar con exactitud, por ejemplo, el contenido “extremista” o el discurso de odio²⁶, y por tanto tienden más a bloquear y restringir el contenido y así a menoscabar los derechos de los usuarios a ser escuchados y el derecho a obtener información sin restricción ni censura.

30. En un sistema gobernado por la IA, la difusión de información e ideas se rige por fuerzas opacas con prioridades que pueden no ser compatibles con un entorno propicio a la diversidad de medios y a las voces independientes. El Comité de Derechos Humanos ha observado de manera pertinente que los Estados deben “adoptar medidas adecuadas ... para impedir un excesivo predominio o concentración de los medios de comunicación por grupos mediáticos bajo control privado, en situaciones monopolísticas que pueden menoscabar la diversidad de fuentes y opiniones”²⁷.

31. Los usuarios también carecen de acceso a las reglas del juego cuando se trata de plataformas y sitios web regidos por IA. La falta de claridad sobre el grado y alcance de la IA y las aplicaciones algorítmicas en línea impiden a las personas entender cuándo y según qué métrica se difunde, restringe o dirige la información. Pequeñas concesiones para hacer frente a este problema, como la identificación selectiva de resultados de búsqueda patrocinados²⁸ o las plataformas de redes sociales que destacan cuándo la publicidad ha sido pagada por actores políticos, pueden ayudar ligeramente a los usuarios a entender las reglas del entorno de información, pero esas concesiones no captan ni resuelven las preocupaciones acerca del grado en que los procesos algorítmicos configuran ese entorno.

32. Incluso cuando las personas son informadas de la existencia, el alcance y el funcionamiento de sistemas de IA, estos sistemas pueden frustrar los esfuerzos de

general núm. 34 (2011) sobre las libertades de opinión y de expresión; A/HRC/29/32, párr. 61, y A/HRC/32/38, párr. 86.

²⁵ Comité de Derechos Humanos, Observación general núm. 34 (2011) sobre las libertades de opinión y de expresión, párr. 50.

²⁶ Consejo de Europa, *Algorithms and Human Rights*, pág. 21.

²⁷ Comité de Derechos Humanos, Observación general núm. 34 (2011) sobre las libertades de opinión y de expresión, párr. 40.

²⁸ Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (Nueva York, New York University Press, 2018).

transparencia e idoneidad. Hasta la fecha no se han elaborado medios refinados y escalables de escrutar y hacer transparentes los fundamentos técnicos de las decisiones automatizadas en el ciberespacio²⁹. Esto significa que a menudo se afectan los derechos de expresión de las personas sin que estas puedan investigar o entender por qué, cómo o sobre qué base.

D. Derecho a la privacidad

33. El derecho a la privacidad a menudo actúa como puerta de acceso al disfrute de la libertad de opinión y expresión³⁰. El artículo 17 del Pacto protege al individuo contra “injerencias arbitrarias o ilegales en su vida privada, su familia, su domicilio o su correspondencia” y “ataques ilegales a su honra y reputación” y establece que “toda persona tiene derecho a la protección de la ley contra esas injerencias o esos ataques”. La Oficina del Alto Comisionado para los Derechos Humanos y el Consejo de Derechos Humanos han subrayado que toda injerencia en la vida privada debe cumplir normas de legalidad, necesidad y proporcionalidad (*A/HRC/27/37*, párr. 23, y resolución *34/7* del Consejo de Derechos Humanos, párr. 2).

34. Los sistemas de adopción de decisiones basados en IA dependen de la reunión y explotación de datos, que van de los datos ambientales y no personales a la información que permite identificar a una persona, y la gran mayoría de los datos usados para alimentar los sistemas de IA están en el medio del espectro, es decir, son datos que se deducen o extraen de datos personales o son datos personales que se han anonimizado (a menudo imperfectamente). Las empresas usan datos derivados del perfilado en línea y de las huellas digitales, obtienen conjuntos de datos de terceros, como los intermediarios de datos, y sacan nuevos datos de grandes conjuntos de datos agregados para alimentar los sistemas de IA. Los productos para el consumidor y los sistemas autónomos basados en IA a menudo tienen sensores que producen y reúnen grandes cantidades de datos sobre las personas en su proximidad³¹ y en las plataformas de las redes sociales se usan métodos de IA para inferir y producir información delicada sobre las personas que estas no han dado o confirmado, como la orientación sexual, las relaciones familiares, las opiniones religiosas, el estado de salud o la afiliación política.

35. La IA desafía las nociones tradicionales de consentimiento, objeto y limitación del uso, transparencia y rendición de cuentas, los pilares en que se basan las normas internacionales de protección de datos³². Como los sistemas de IA funcionan explotando los conjuntos de datos que ya existen y creando otros nuevos, la capacidad de las personas de saber, entender y ejercer control sobre la forma en que se usan sus datos pierde sentido práctico en el contexto de la IA. Una vez que se usan para otro fin en un sistema de IA, los datos pierden su contexto original, aumenta el riesgo de que los datos sobre las personas se vuelvan inexactos o anticuados y se priva a las personas de la capacidad de rectificarlos o suprimirlos. Los sistemas basados en la IA se usan para tomar decisiones importantes sobre la base de esos datos, algunas de las cuales afectan profundamente la vida de las personas³³, pero estas tienen pocos

²⁹ Mike Ananny y Kate Crawford, “Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability”, *New Media and Society*, vol. 20, núm. 3 (13 de diciembre de 2016). Disponible en <http://journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsa>.

³⁰ Véanse *A/HRC/29/32*, párr. 16, la resolución 68/167 de la Asamblea General y la resolución 20/8 del Consejo de Derechos Humanos.

³¹ Artículo 19 y Privacy International, “Privacy and freedom of expression”.

³² Comité de Derechos Humanos, Observación general núm. 16: artículo 17 (1988) sobre el derecho a la intimidad, párr. 10.

³³ Artículo 19 y Privacy International, “Privacy and freedom of expression”, pág. 19.

medios de ejercer control sobre los datos que se han derivado de sus datos personales, incluso cuando las técnicas de anonimización siguen teniendo deficiencias.

E. Obligación de no discriminar

36. La no discriminación es un principio intrínseco del derecho de los derechos humanos, que no solo califica las obligaciones de los Estados de garantizar el disfrute de todos los demás derechos humanos sin discriminación, sino también, como se establece en el artículo 26 del Pacto, como garantía independiente de igualdad ante la ley e igual protección de la ley. Los Estados tienen la clara obligación de prohibir “toda discriminación” y garantizar “a todas las personas protección igual y efectiva contra cualquier discriminación por motivos de raza, color, sexo, idioma, religión, opiniones políticas o de cualquier índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social”. De esta manera los artículos 17 y 19 incorporan derechos individuales a la no discriminación en el sostenimiento y formación de opiniones, la expresión de ideas e información y el acceso a ellas y el ejercicio del derecho a la privacidad y a la protección de los datos personales.

37. El potencial de la IA de establecer y perpetuar sesgos y discriminación se extiende a la discriminación en el ejercicio de la libertad de opinión y expresión. Los algoritmos de moderación pueden no tener en cuenta contextos y sensibilidades culturales, lingüísticos o de género, o el interés público en el contenido³⁴. Las transmisiones de noticias basadas en la IA pueden perpetuar y reforzar actitudes discriminatorias, y los sistemas de perfilado y publicidad basados en la IA han facilitado demostrablemente la discriminación por raza, religión y género³⁵. Las funciones de autocompletar de la IA también han producido resultados racialmente discriminatorios³⁶.

38. Una serie de factores incorporan sesgo en los sistemas de IA y aumentan su potencial discriminatorio. Entre ellos están la forma en que se diseñan los sistemas, las decisiones sobre el origen y el alcance de los conjuntos de datos con que se entrenan, los sesgos sociales y culturales que los creadores de aplicaciones pueden incorporar en esos conjuntos de datos, los modelos mismos de IA y la forma en que los productos del modelo de IA se ejecutan en la práctica. Por ejemplo, las aplicaciones de reconocimiento facial son imperfectas porque se basan en conjuntos de datos en que predominan los hombres blancos y tienen errores de hasta 20% para las mujeres y las personas con color de la piel más oscuro³⁷. Cuando esos sistemas se

³⁴ Esto ha conducido, por ejemplo, a la supresión de fotografías históricas de significado cultural especial. Véase Julia Carrie Wong, “Mark Zuckerberg accused of abusing power after Facebook deletes ‘napalm girl’ post”, *The Guardian*, 9 de septiembre de 2016. Disponible en <https://www.theguardian.com/technology/2016/sep/08/facebook-mark-zuckerberg-napalm-girl-photo-vietnam-war>; véase también A/HRC/38/35, párr. 29.

³⁵ Julia Angwin, Madeleine Varner y Ariana Tobin, “Facebook enabled advertisers to reach ‘Jew haters’”, ProPublica, 14 de septiembre de 2017, disponible en <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>; Ariana Tobin, “Why we had to buy racist, sexist, xenophobic, ableist and otherwise awful Facebook ads”, ProPublica, 27 de noviembre de 2017, disponible en <https://www.propublica.org/article/why-we-had-to-buy-racist-sexist-xenophobic-ableist-and-otherwise-awful-facebook-ads>.

³⁶ Paris Martineau, “YouTube’s search suggests racist autocompletes”, *The Outline*, 13 de mayo de 2018. Disponible en <https://theoutline.com/post/4536/youtube-s-search-autofill-suggests-racist-results?zd=1&zi=3ygt6hw>.

³⁷ Joy Buolamwini, “The dangers of supremely white data and the coded gaze”, presentado en Wikimania 2018, Ciudad del Cabo. Disponible en <https://www.youtube.com/watch?v=ZSJXKoD6mA8&feature=youtu.be>.

usan, por ejemplo, para clasificar imágenes obtenidas con un motor de búsqueda, su potencial de discriminación puede traducirse en injerencias concretas en el ejercicio de los derechos de la persona de buscar, recibir e impartir información y de reunirse o asociarse libremente.

F. Derecho a un recurso efectivo

39. El derecho de los derechos humanos garantiza a las personas un recurso determinado por la autoridad judicial, administrativa o legislativa competente (artículo 23) del Pacto). Los recursos deben ser conocidos por toda persona cuyos derechos hayan sido infringidos y accesibles a ella; deben incluir una investigación inmediata, exhaustiva e imparcial de las presuntas violaciones³⁸ y deben poder poner fin a las violaciones en curso (A/HRC/27/37, párrs. 39 a 41).

40. Los sistemas de IA a menudo dificultan el ejercicio del derecho a un recurso. En primer lugar, la notificación individual casi está excluida por la naturaleza del sistema. En casi todas las aplicaciones de la tecnología de IA en el entorno de información, los individuos desconocen el alcance, la extensión y hasta la existencia de los procesos decisorios algorítmicos que pueden afectar el disfrute de los derechos de opinión y expresión. El segundo aspecto, más complejo, es la escrutabilidad del sistema de IA en sí mismo. La lógica en que se basa una decisión algorítmica puede no ser evidente ni siquiera para un experto capacitado en la mecánica del sistema. Es lógico suponer que el aumento de la transparencia de los sistemas de IA permitiría más escrutinio, pero la transparencia algorítmica no equivale necesariamente a explicaciones inteligibles de los procesos de adopción de decisiones. Los algoritmos pueden ocultar la adopción de una decisión importante o ser tan complejos y depender tanto del contexto que impidan la explicación. La situación se complica porque las empresas que operan en el entorno de información actualizan a menudo sus algoritmos³⁹; asimismo, las aplicaciones de aprendizaje automático pueden cambiar con el tiempo sus reglas y algoritmos.

41. A estas preocupaciones se suma el paso a la automatización de los propios sistemas de recursos, de manera que las quejas de los usuarios, sea sobre decisiones de moderación de contenido o sobre los efectos adversos de las tecnologías de IA en los derechos humanos, son consideradas y resueltas por tecnologías de IA⁴⁰. Los procesos automáticos de respuesta suscitan dudas de que los mecanismos de atención de quejas constituyan un recurso efectivo, en vista de la falta de discreción, análisis contextual y determinación independiente incorporada en esos procesos⁴¹.

G. Respuestas legislativas, reglamentarias y normativas a la inteligencia artificial

42. Muchos Estados están diseñando estrategias nacionales de IA para estudiar y elaborar políticas e iniciativas destinadas a maximizar los posibles beneficios de la

³⁸ Comité de Derechos Humanos, Observación general núm. 31 (2004) sobre la naturaleza de la obligación jurídica general impuesta a los Estados Partes en el Pacto, párr. 15.

³⁹ Barry Schwartz, “Google: we make thousands of updates to search algorithms each year”, Search Engine Roundtable, 5 de junio de 2015. Disponible en <https://www.seroundtable.com/google-updates-thousands-20403.html>.

⁴⁰ Consejo de Europa, *Algorithms and Human Rights*, pág. 24.

⁴¹ Pei Zhang, Sophie Stalla-Bourdillon y Lester Gilbert, “A content-linking-context model for ‘notice-and-take-down’ procedures”, *WebSci '16*, mayo de 2016. Disponible en <http://takedownproject.org/wp-content/uploads/2016/04/ContentLinkingModelZhangStallaGilbert.pdf>.

IA para sus ciudadanos⁴². Ningún Estado ha propuesto una ley o reglamento general de la IA, pero hay razones para proceder con cautela en este enfoque, que puede ser inadecuado para un campo tan innovador y puede compensar la falta de detalle con disposiciones demasiado restrictivas o demasiado permisivas. La reglamentación sectorial puede ser preferible, aunque se puede alegar que la legislación y reglamentación vigente, por ejemplo en materia de protección de datos, podría ser flexible y estar disponible sin necesidad de más legislación.

43. Al mismo tiempo, los Estados deben asegurar que la IA se desarrolle de conformidad con las normas de derechos humanos. Todo esfuerzo de elaborar una política o reglamentación del Estado en el campo de la IA debe asegurar que se tengan en cuenta cuestiones de derechos humanos⁴³. Los derechos a la libertad de opinión y expresión, en particular, a menudo se excluyen de los debates públicos y políticos sobre la IA, que, en la medida en que tratan cuestiones de derechos humanos, tienden a centrarse en el sesgo y la discriminación en la prestación de servicios.

44. Dado que la elaboración de sistemas eficaces de IA depende de la adquisición de grandes conjuntos de datos además de inversión a largo plazo en capacidad técnica, es probable que las entidades del sector privado dominen en elaboración, producción y capacidad, lo cual aumentará el uso que el sector público hace de empresas privadas para el acceso a sistemas de IA. Esto a su vez aumenta la probabilidad de que los intereses privados y estatales estén cada vez más entrelazados y la percepción pública de este entrelazamiento. Esto vale especialmente para el entorno de información, en el cual los Gobiernos a menudo son usuarios – de plataformas de redes sociales, motores de búsqueda y otras tecnologías – y no proveedores. La alineación de los intereses públicos y privados no afecta en sí misma los derechos humanos, pero suscita preocupación por la transparencia y la rendición de cuentas. A medida que avanza el desarrollo privado de la IA, surge un riesgo muy real de que los Estados deleguen en las empresas mandatos de censura y vigilancia cada vez más complejos y onerosos.

45. Todo intento del Estado de articular leyes o políticas en el campo de la IA debe tratar tanto las aplicaciones de IA del sector privado como las del sector público, y no limitarse a la reglamentación de la IA del sector público. Como concluyó el Consejo de Europa, “las cuestiones relativas a la gobernanza o la reglamentación algorítmicas son prerrogativas de política pública y no deben dejarse solo en manos de los actores privados”⁴⁴. Los enfoques estatales pueden incluir más requisitos de transparencia y más obligaciones de divulgación para las empresas y una legislación robusta de protección de datos que responda a las preocupaciones relativas a la IA.

46. Están proliferando las iniciativas del sector público y del sector privado encaminadas a estudiar e integrar la ética en la adquisición, el diseño, el despliegue y la aplicación de sistemas de IA. El Relator Especial recomienda encarecidamente la integración de las cuestiones de derechos humanos en esas actividades. La concentración del sector privado en la ética y la presión del sector público en ese sentido a menudo implican resistencia a la reglamentación basada en los derechos

⁴² Tim Dutton, “An overview of national AI strategies”, Medium, 28 de junio de 2018. Disponible en <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>.

⁴³ Por ejemplo, es motivo de preocupación que una comisión parlamentaria del Reino Unido de Gran Bretaña e Irlanda del Norte haya publicado un informe de 200 páginas en que los derechos humanos no se mencionan siquiera una vez. Véase Comité Especial sobre Inteligencia Artificial de la Cámara de los Lores del Reino Unido de Gran Bretaña e Irlanda del Norte, “AI in the United Kingdom”.

⁴⁴ Consejo de Europa, *Algorithms and Human Rights*, pág. 44.

humanos⁴⁵. La ética ofrece un marco crítico para resolver problemas particulares en el campo de la IA, pero no reemplaza el respeto de los derechos humanos, al cual cada Estado está obligado por la ley. Las empresas y los Gobiernos deben asegurar que las consideraciones y responsabilidades de derechos humanos estén firmemente integradas en todos los aspectos de sus operaciones de IA, incluso cuando elaboran códigos y orientación éticos⁴⁶.

IV. Enfoque de la inteligencia artificial basado en los derechos humanos

47. En informes recientes el titular del mandato ha enunciado consideraciones jurídicas y prácticas para las empresas a fin de poner los principios de derechos humanos en el centro de sus políticas de regulación de contenido y ha detallado normas y procesos sustantivos que aseguren que las empresas puedan cumplir sus obligaciones de derechos humanos conforme a los Principios Rectores sobre las Empresas y los Derechos Humanos en todos los aspectos de sus operaciones. El mismo marco rige el enfoque que se propone aquí con respecto a las tecnologías de IA. Las normas y procesos sustantivos que se proponen a continuación son aplicables a las empresas, en su calidad de actores que diseñan, despliegan y aplican sistemas de IA, y a los Estados, que tienen la obligación de abstenerse de infringir los derechos humanos al adoptar y usar sistemas de IA. Estas normas y procesos tienen por objeto asegurar que el derecho de los derechos humanos se ponga en el centro de los progresos en el campo de la IA. Dos principios fundamentales inspiran las normas y procesos propuestos: la necesidad de proteger y respetar la acción y la autonomía individuales, que es un requisito fundamental para el ejercicio del derecho a la libertad de opinión y expresión; y la importancia de que los actores públicos y sectoriales hagan una divulgación significativa, caracterizada por esfuerzos abiertos e innovadores de explicar al público las tecnologías de IA y facilitar su escrutinio.

A. Normas sustantivas para los sistemas de inteligencia artificial

48. Las empresas deben orientar sus normas, reglas y diseños de sistemas en torno a los principios universales de derechos humanos (A/HRC/38/35, párrs. 41 a 43). Las condiciones y directrices que se presentan al público deben complementarse con compromisos de política interna de incorporar consideraciones de derechos humanos en todas las operaciones de la empresa, especialmente en relación con la elaboración y el despliegue de sistemas de IA y algorítmicos. Las empresas deben considerar cómo elaborar normas profesionales para los ingenieros de IA, traduciendo las obligaciones de derechos humanos en orientación para el diseño técnico y las opciones operativas. La elaboración de códigos de ética y las estructuras institucionales concomitantes puede ser un complemento importante de los compromisos de derechos humanos, pero no puede reemplazarlos. Los códigos y directrices emitidos por órganos del sector público y del sector privado deben subrayar que el derecho de los derechos humanos aporta las reglas fundamentales para la protección de las personas en el contexto de la IA, mientras que los marcos éticos pueden ayudar a seguir desarrollando el contenido y la aplicación de los derechos humanos en circunstancias determinadas.

⁴⁵ Ben Wagner, "Ethics as an escape from regulation: from ethics-washing to ethics-shopping?", en *Being Profiling: Cogitas Ergo Sum*, Mireille Hildebrandt, ed. (Amsterdam University Press (de próxima publicación)).

⁴⁶ Artículo 19 y Privacy International, "Privacy and freedom of expression", pág. 13.

49. Las empresas y los Gobiernos deben indicar explícitamente a las personas las decisiones en el entorno de información que se toman con sistemas automáticos y las que van acompañadas de revisión humana, lo mismo que los elementos generales de la lógica que usan esos sistemas. Las personas también deben ser informadas cuando los datos personales que dan a un actor del sector privado (sea explícitamente o mediante el uso de un servicio o sitio) se incluirán en un conjunto de datos usado por un sistema de IA, para que puedan tener en cuenta ese conocimiento al decidir si darán o no su consentimiento a la reunión de datos y qué tipos de datos desean revelar⁴⁷. Al igual que los avisos públicos requeridos para el uso de cámaras de televisión de circuito cerrado, los sistemas de IA deben revelar activamente a las personas (por medios innovadores como las ventanas de diálogo) de manera clara y comprensible que están sujetas o aportan datos a un proceso de adopción de decisiones basado en IA, y también proporcionarles información significativa sobre la lógica empleada en el proceso y la importancia de las consecuencias para la persona.

50. La transparencia no se limita a informar al usuario de la existencia de tecnologías de IA en las plataformas y servicios en línea que usa. Las empresas y los Gobiernos deben adoptar la transparencia en cada aspecto de la cadena de valor de la IA. La transparencia no necesita ser compleja para ser eficaz; incluso las explicaciones simplificadas del objeto, las políticas, las entradas y los productos de un sistema de IA pueden contribuir a la educación y al debate públicos⁴⁸. En vez de lidiar con el problema de hacer legibles para el público procesos técnicos complicados, las empresas deben procurar lograr la transparencia dando indicaciones no técnicas sobre el sistema. Para ello, la atención debe centrarse en educar al usuario sobre la existencia, el objeto, la constitución y el efecto del sistema de IA, en lugar de sobre el código fuente, los datos de capacitación y las entradas y productos⁴⁹.

51. La transparencia radical sobre el efecto de un sistema de IA en el entorno de información requiere revelar datos, por ejemplo, sobre cuánto contenido suprimen los sistemas de IA, con qué frecuencia un moderador humano aprueba la supresión de contenido sugerida por la IA, con qué frecuencia se impugnan las supresiones y con qué frecuencia se aprueban las impugnaciones. Los datos agregados que ilustran las tendencias de la presentación de contenido deben poder ser examinados por los usuarios, junto con estudios de casos que ilustren por qué se da más prioridad a un contenido que a otro. La divulgación de las fuentes y los beneficiarios de la publicidad política y comercial es un elemento esencial de la transparencia radical. Los actores del sector público y del sector privado que establecen sistemas basados en IA también deben ser transparentes respecto de los límites del sistema de IA, por ejemplo mediante medidas de confianza, escenarios de falla conocidos y limitaciones convenientes del uso⁵⁰.

52. La lucha contra la prevalencia de la discriminación en los sistemas de IA es un reto existencial para las empresas y los Gobiernos; si no se tratan y resuelven los elementos y efectos discriminatorios, la tecnología será no solo ineficaz sino también peligrosa. Hay amplio liderazgo intelectual y muchos recursos que las empresas y los Gobiernos pueden aprovechar para considerar cómo combatir el sesgo y la discriminación en los sistemas de IA; en términos generales, hay que aislar y dar

⁴⁷ Comité de Derechos Humanos, Observación general núm. 16 (1988) sobre el derecho a la intimidad.

⁴⁸ Aaron Rieke, Miranda Bogen y David Robinson, “Public scrutiny of automated decisions: early lessons and emerging methods” (Omidyar and Upturn, 2018), pág. 5.

⁴⁹ Rieke, Bogen y Robinson, “Public scrutiny of automated decisions”, pág. 8.

⁵⁰ Amnistía Internacional y Access Now, “Toronto declaration: protecting the right to equality and non-discrimination in machine learning systems”, art. 27 d), 2018. Disponible en <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/>.

cuenta de la discriminación a nivel de entrada y de producto. Esto implica, como mínimo, suprimir errores de muestreo (cuando los conjuntos de datos no son representativos de la sociedad), depurar conjuntos de datos para eliminar datos discriminatorios y tomar medidas para compensar los datos que “contienen la impronta de pautas históricas y estructurales de discriminación”⁵¹ y de los cuales los sistemas de IA tienden a derivar representantes discriminatorios. El seguimiento activo de los resultados discriminatorios de los sistemas de IA también es esencial para evitar y mitigar los efectos adversos en los derechos humanos de las personas.

B. Procesos para los sistemas de inteligencia artificial

53. **Evaluaciones del efecto en los derechos humanos.** Para abrazar la transparencia radical en todo el ciclo vital de la IA, las empresas y los Gobiernos deben tomar medidas para que los sistemas puedan ser escrutados y criticados desde la concepción hasta la ejecución. La evaluación del efecto en los derechos humanos es un instrumento que puede demostrar la dedicación a hacer frente a las consecuencias de los sistemas de IA para los derechos humanos y debe hacerse antes de la adquisición, la elaboración o el uso y requiere autoevaluación y examen externo. El centro de estudio AI Now ha propuesto una evaluación algorítmica del efecto a cargo de un organismo público que requiera que los Gobiernos hagan un examen interno de los sistemas de IA, además de facilitar los procesos de examen por investigadores externos para probar y verificar supuestos y conclusiones⁵². Las empresas también deben hacer evaluaciones de ese tipo.

54. La adquisición de tecnologías de IA de proveedores privados por el sector público debe ir acompañada de una consulta pública para obtener opiniones y aportes de la sociedad sobre el diseño y la ejecución del sistema de IA antes de la adquisición. Tanto las empresas como los Gobiernos deben hacer consultas serias y sostenidas con la sociedad civil, grupos de derechos humanos, las comunidades locales pertinentes y representantes de poblaciones históricamente marginadas o subrepresentadas antes de elaborar, adquirir o usar sistemas y tecnologías de IA.

55. **Auditorías.** La facilitación del examen externo de los sistemas de IA aporta una garantía esencial de rigor e independencia en la transparencia. Por esta razón, las auditorías independientes en curso deben complementar las evaluaciones del efecto en los derechos humanos anteriores a la adquisición como mecanismo importante de transparencia y rendición de cuentas en los sistemas de IA. Los actores del sector privado han formulado objeciones a la viabilidad de las auditorías en el espacio de la IA, dado el imperativo de proteger la tecnología patentada. Estas preocupaciones pueden ser fundadas, pero el Relator Especial está de acuerdo con AI Now en que, especialmente cuando una aplicación de IA es usada por un organismo del sector público, la negativa del proveedor a ser transparente con respecto al funcionamiento del sistema sería incompatible con las obligaciones de rendición de cuentas del organismo público.

56. En todo caso, abundan las sugerencias innovadoras sobre las auditorías de la tecnología de IA que permiten mantener el secreto de la propiedad intelectual: sería posible producir, mediante algoritmos, pruebas de conocimiento cero para demostrar que se ajusta a ciertos requisitos, obviando la necesidad de escrutar el algoritmo

⁵¹ Jason Gabriel, “The case for fairer algorithms”, Medium, 14 de marzo de 2018. Disponible en https://medium.com/@Ethics_Society/the-case-for-fairer-algorithms-c008a12126f8.

⁵² Dillon Reisman y otros, “Algorithmic impact assessments: a practical framework for public agency accountability” (AI Now, 2018). Disponible en <https://ainowinstitute.org/aiareport2018.pdf>.

subyacente⁵³, o los algoritmos podrían revelarse a terceros expertos que los mantuvieran en custodia bajo condición de confidencialidad, lo cual permitiría el escrutinio que exige el interés público pero no la publicidad del algoritmo⁵⁴. Los reguladores estatales de las telecomunicaciones o la competencia podrían tener acceso confidencial a los sistemas de IA, como ya se hace, por ejemplo, en la reglamentación de las máquinas de juego de Australia y Nueva Zelanda, por la cual las empresas deben someter sus sistemas algorítmicos a una auditoría del regulador⁵⁵. La literatura académica contiene otras sugerencias sobre formas innovadoras de auditoría de la IA⁵⁶.

57. Cada uno de estos mecanismos puede encontrar problemas de ejecución, especialmente en el entorno de información, pero las empresas deben procurar hacer factible la auditoría de los sistemas de IA. Los Gobiernos deben contribuir a la eficacia de las auditorías considerando la posibilidad de hacer intervenciones normativas o legislativas que requieran a las empresas hacer auditable el código de IA, garantizando la existencia de registros de auditoría y así creando más oportunidades de transparencia para las personas afectadas.

58. **Autonomía individual.** La IA no debe suplantar, manipular ni perturbar invisiblemente la capacidad de las personas de formarse y mantener opiniones o de recibir y expresar ideas en el entorno de información. Respetar la autonomía individual significa, como mínimo, asegurar que los usuarios tengan conocimiento, elección y control. Las aplicaciones de IA omnipresentes y ocultas que oscurecen los procesos de presentación, personalización y moderación de contenido, perfilado y selección socavan la capacidad de las personas de ejercer los derechos a la libertad de opinión y expresión y a la privacidad. Las empresas deben tener presentes los efectos adversos en los derechos humanos que resultan de las aplicaciones de IA que priorizan los intereses comerciales o políticos sobre la transparencia y la elección individual.

59. **Aviso y consentimiento.** Las empresas deben asegurar que los usuarios estén plenamente informados de la forma en que la adopción algorítmica de decisiones configura su uso de una plataforma, sitio o servicio. Esto se puede lograr mediante campañas educativas, ventanas de diálogo, anuncios intersticiales y otros medios de señalización cuando un sistema de IA determina la experiencia del usuario de un motor de búsqueda, sitio de noticias o plataforma de una red social. La imposición por el Estado de requisitos de divulgación puede ser un medio conveniente de proteger el aviso y el consentimiento. Las personas también tienen derecho a saber cuándo una aplicación de IA reúne sus datos y si los datos formarán parte de un conjunto de datos que después informará una aplicación de IA, al igual que las condiciones en que se usarán, almacenarán y suprimirán esos datos.

60. **Medidas correctivas.** Los efectos adversos de los sistemas de IA en los derechos humanos deben ser corregibles y corregidos por las empresas responsables. La condición previa del establecimiento de procesos correctivos eficaces es asegurar que las personas sepan que han sido objeto de una decisión algorítmica (incluida una decisión sugerida por un sistema de IA y aprobada por un interlocutor humano) y estén provistas de información sobre la lógica en que se basa la decisión. Por otra parte, las empresas deben asegurar la revisión humana de las solicitudes de corrección,

⁵³ Consejo de Europa, *Algorithms and Human Rights*, pág. 36.

⁵⁴ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, Harvard University Press, 2015).

⁵⁵ Consejo de Europa, *Algorithms and Human Rights*, pág. 34.

⁵⁶ Christian Sandvig y otros, "Auditing algorithms: research methods for detecting discrimination on Internet platforms", ponencia presentada en *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, sesión preliminar de la 64ª reunión anual de la International Communication Association, Seattle, 22 de mayo de 2014.

a fin de ofrecer un control adecuado de los sistemas y garantizar la rendición de cuentas. Se deben publicar datos sobre la frecuencia con que se activan los mecanismos correctivos de las decisiones tomadas por las tecnologías de IA.

V. Conclusiones y recomendaciones

61. En este informe el Relator Especial ha estudiado los efectos actuales y potenciales de la IA en los derechos a la libertad de opinión y expresión y postulado que la IA es ahora una parte esencial del entorno de información, que presenta beneficios y riesgos para el disfrute de los derechos de las personas. El Relator Especial ha propuesto un marco conceptual para reflexionar sobre las obligaciones de los Estados y las responsabilidades de las empresas respecto de la defensa de estos derechos ante la expansión de las capacidades tecnológicas y sugerido medidas concretas que podrían aplicar tanto los Gobiernos como las empresas para asegurar el respeto de los derechos humanos mientras crecen el poder, el alcance y la extensión de la tecnología de IA.

Recomendaciones para los Estados

62. Cuando adquieren o despliegan sistemas o aplicaciones de IA, los Estados deben asegurar que los órganos del sector público actúen de conformidad con los principios de derechos humanos. Esto incluye, entre otras cosas, hacer consultas públicas y evaluaciones del efecto en los derechos humanos o evaluaciones algorítmicas del efecto hechas por organismos públicos antes de adquirir o desplegar sistemas de IA. Se debe prestar especial atención al efecto desigual de esas tecnologías en las minorías raciales y religiosas, la oposición política y los activistas. El despliegue por el Estado de sistemas de IA debe ser objeto de auditorías periódicas a cargo de expertos externos independientes.

63. Los Estados deben asegurar que los derechos humanos tengan un lugar central en el diseño, el despliegue y la ejecución de los sistemas de IA del sector privado. Esto incluye actualizar y aplicar la reglamentación vigente, particularmente la reglamentación de protección de datos, al campo de la IA, según planes regulatorios o corregulatorios destinados a requerir que las empresas hagan evaluaciones del efecto y auditorías de las tecnologías de IA y a asegurar la existencia de mecanismos externos eficaces de rendición de cuentas⁵⁷. Cuando corresponda, la reglamentación sectorial de determinadas aplicaciones de la IA puede ser necesaria y eficaz para la protección de los derechos humanos. En la medida en que esas restricciones introduzcan o faciliten injerencias en la libertad de expresión, los Estados deben asegurar que sean necesarias y proporcionadas para alcanzar un objetivo legítimo de conformidad con el artículo 19 3) del Pacto. La reglamentación relativa a la IA también debe elaborarse mediante una amplia consulta pública que incluya el diálogo con la sociedad civil, grupos de derechos humanos y representantes de usuarios finales marginados o subrepresentados.

64. Los Estados deben crear un marco normativo y legislativo propicio para un entorno de información diverso y pluralista. Esto incluye tomar medidas para garantizar la competencia en el campo de la IA. Estas medidas pueden incluir la regulación de los monopolios tecnológicos para prevenir la concentración de pericia y poder en materia de IA en manos de unas pocas empresas dominantes, la regulación encaminada a aumentar la interoperabilidad de servicios y

⁵⁷ Wagner, “Ethics as an escape from regulation”.

tecnologías y la adopción de políticas que apoyen la neutralidad de la red y del dispositivo⁵⁸.

Recomendaciones para las empresas

65. Todos los esfuerzos de elaborar directrices o códigos sobre las consecuencias éticas de las tecnologías de IA deben basarse en los principios de derechos humanos. Toda elaboración y todo despliegue de IA por el sector público y por el sector privado deben dar oportunidades a la sociedad civil de hacer observaciones. Las empresas deben reiterar en sus políticas y su orientación técnica para ingenieros, creadores de aplicaciones, técnicos de datos, depuradores de datos, programadores y otros participantes en el ciclo vital de la IA que las responsabilidades en el ámbito de los derechos humanos guían todas sus operaciones comerciales y que los principios éticos pueden ayudar facilitando la aplicación de principios de derechos humanos a situaciones determinadas de diseño, despliegue y ejecución de la IA. En particular, las condiciones de servicio de las plataformas deben basarse en los principios universales de derechos humanos.

66. Las empresas deben hacer explícitos dónde y cómo se usan en sus plataformas, servicios y aplicaciones las tecnologías de inteligencia artificial y las técnicas automatizadas. El uso de medios innovadores para indicar a las personas cuándo son objeto de un proceso de adopción de decisiones basado en IA, cuándo esta desempeña un papel en la presentación o moderación de contenido o cuándo los datos personales de las personas pueden integrarse en un conjunto de datos que se usará para informar sistemas de IA es esencial para dar a los usuarios el aviso necesario para entender y hacer frente al efecto de los sistemas de IA en el disfrute de los derechos humanos. Las empresas también deben publicar datos sobre la supresión de contenido, incluida la frecuencia con que se impugnan las supresiones y se aprueban las impugnaciones, y también datos sobre las tendencias de la presentación de contenido, junto con estudios de casos y educación sobre el perfilado comercial y político.

67. Las empresas deben prevenir y dar cuenta de la discriminación a nivel de las entradas y los productos de los sistemas de IA. Esto implica asegurar que los equipos que diseñan y despliegan sistemas de IA reflejen actitudes diversas y no discriminatorias y priorizar la prevención de sesgos y discriminación en la elección de conjuntos de datos y el diseño del sistema, en particular la corrección de errores de muestreo, la depuración de conjuntos de datos para eliminar datos discriminatorios y la adopción de medidas para compensar esos datos. También es esencial el seguimiento activo de los resultados discriminatorios de los sistemas de IA.

68. Las evaluaciones del efecto en los derechos humanos y las consultas públicas deben hacerse durante el diseño y el despliegue de nuevos sistemas de IA, incluido el despliegue de sistemas que ya existen en nuevos mercados mundiales. Las consultas públicas y el diálogo con el público deben ocurrir antes de la finalización o implantación de un producto o servicio, para asegurar que sean significativos, y deben incluir el diálogo con la sociedad civil, defensores de los derechos humanos y representantes de usuarios finales marginados o subrepresentados. Los resultados de las evaluaciones del efecto en los derechos humanos y de las consultas públicas deben hacerse públicos.

⁵⁸ Autorité de régulation des communications électroniques et des postes, *Devices, the Weak Link in Achieving an Open Internet* (2018). Disponible en https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf.

69. **Las empresas deben hacer que todos los códigos de IA sean plenamente auditables y deben buscar medios innovadores para permitir la auditoría externa e independiente de los sistemas de IA, independientemente de los requisitos reglamentarios. Los resultados de las auditorías de la IA deben hacerse públicos.**

70. **Los usuarios deben tener acceso a medidas correctivas de los efectos adversos de los sistemas de IA en los derechos humanos. Las empresas deben establecer sistemas de revisión y corrección humanas para responder oportunamente a las quejas de todos los usuarios y a las apelaciones contra sistemas basados en IA. Deben publicarse periódicamente datos sobre la frecuencia con que los sistemas de IA son objeto de quejas y solicitudes de corrección y sobre los tipos y la eficacia de las medidas correctivas disponibles.**
