La Alianza Global Jus Semper

*En Pos del Paradigma de la Gente y el Planeta*

Desarrollo Humano Sostenible

Septiembre 2021 ENSAYOS SOBRE DEMOCRACIA REAL Y CAPITALISMO

# Derechos Humanos en la Era de la Inteligencia Artificial

## Access Now

### Resumen Ejecutivo

A medida que la inteligencia artificial sigue abriéndose paso en nuestra vida cotidiana, su propensión a interferir con los derechos humanos no hace más que agravarse. Teniendo esto en cuenta, y observando que la tecnología se encuentra todavía en sus fases iniciales, Access Now realiza este estudio preliminar para examinar la gama potencial de cuestiones de derechos humanos que pueden plantearse hoy o en un futuro próximo.

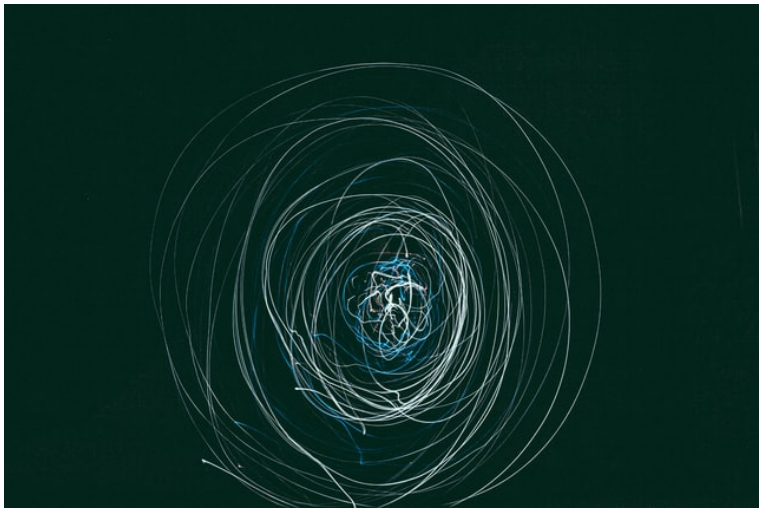

Foto por Gertrūda Valasevičiūtė en Unsplash

Muchas de las cuestiones que se plantean al examinar este ámbito no son nuevas, pero se ven muy exacerbadas por la escala, la proliferación y el impacto en la vida real que propicia la inteligencia artificial. Por ello, el potencial de la inteligencia artificial tanto para ayudar como para perjudicar a las personas es mucho mayor que el de las tecnologías anteriores. Aunque ya hemos visto algunas de estas consecuencias, los impactos no harán más que crecer en gravedad y alcance. Sin embargo, si empezamos ahora a examinar qué salvaguardias y estructuras son necesarias para abordar los problemas y los abusos, los peores daños -incluidos los que afectan de forma desproporcionada a las personas marginadas- pueden prevenirse y mitigarse.

*El potencial de la inteligencia artificial tanto para ayudar como para perjudicar a las personas es mucho mayor que el de las tecnologías anteriores. Aunque ya hemos visto algunas de estas consecuencias, los impactos no harán más que crecer en gravedad y alcance.*

Hay varias ópticas a través de las cuales los expertos examinan la inteligencia artificial. El uso de la legislación internacional sobre derechos humanos y sus normas e instituciones bien desarrolladas para examinar los sistemas de

inteligencia artificial puede contribuir a las discusiones que ya están teniendo lugar, y proporcionar un vocabulario universal y foros establecidos para abordar las diferencias de poder.

Además, las leyes de derechos humanos aportan un marco de soluciones, que ofrecemos aquí en forma de recomendaciones. Nuestras recomendaciones se enmarcan en cuatro categorías generales: normas de protección de datos para proteger los derechos en los conjuntos de datos utilizados para desarrollar y alimentar los sistemas de inteligencia artificial; salvaguardias especiales para los usos de la inteligencia artificial por parte de los gobiernos; salvaguardias para los usos de los sistemas de inteligencia artificial por parte del sector privado; e inversión en más investigación para seguir examinando el futuro de la inteligencia artificial y sus posibles interferencias con los derechos humanos.

Esperamos que este informe proporcione un punto de partida para nuevas diálogos e investigaciones en este espacio en desarrollo. Todavía no sabemos qué significará la inteligencia artificial para el futuro de la sociedad, pero podemos actuar ahora para construir las herramientas que necesitamos para proteger a las personas de sus aplicaciones más peligrosas. Esperamos seguir explorando las cuestiones planteadas en este informe, incluso a través del trabajo con nuestros socios, así como con las principales instituciones corporativas y gubernamentales.



## Introducción

El concepto de inteligencia artificial ha pasado del ámbito de la ciencia ficción a los debates en los círculos más altos del mundo académico, la industria y el gobierno. Sin embargo, los expertos acaban de empezar a estudiar el impacto de la inteligencia artificial en los derechos humanos, y hasta ahora no parecen ponerse de acuerdo en lo que significa el término.

Es evidente que el uso de la inteligencia artificial y la tecnología de aprendizaje de máquina tiene el potencial de efectuar cambios revolucionarios en el mundo. En 2018, fue un tema clave en RightsCon, la conferencia anual de Access Now sobre la intersección de los derechos humanos y la tecnología. En vísperas de RightsCon, trabajamos con socios cercanos para redactar y publicar la Declaración de Toronto sobre la protección de los derechos a la igualdad y la no discriminación en los sistemas de aprendizaje de máquina.[1] También participamos en un taller sobre inteligencia artificial y derechos humanos organizado por el Data & Society Research Institute en Nueva York, cuyo objetivo era "considerar el valor de los derechos humanos en el espacio de la IA, fomentar el compromiso y la colaboración entre

[1] The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems

sectores, y desarrollar ideas y resultados para beneficiar a las partes interesadas que trabajan en esta cuestión de cara al futuro".[2]

Este informe es un análisis preliminar de la intersección entre la inteligencia artificial y los derechos humanos. En la primera sección se proponen definiciones de términos y conceptos clave, como "inteligencia artificial" y "aprendizaje de máquina". A continuación se examina cómo se utilizan los distintos sistemas de inteligencia artificial en el mundo actual y cómo pueden ayudar o perjudicar a la sociedad. En cuanto a los derechos humanos, examinamos el papel que puede desempeñar la legislación sobre derechos humanos en el desarrollo de la inteligencAia artificial, incluida la interacción entre estos derechos fundamentales y la ética. A continuación, analizando los instrumentos de derechos humanos ampliamente adoptados, destacamos las formas en que los usos actuales y previsibles de la inteligencia artificial pueden interferir con una amplia gama de derechos humanos. Por último, ofrecemos una lista de recomendaciones para que las partes interesadas protejan esos derechos.

Reconocemos que estamos ofreciendo recomendaciones en las primeras etapas del desarrollo y el uso de la inteligencia artificial, y que sólo estamos empezando a lidiar con sus posibles consecuencias. Por ello, una de nuestras recomendaciones es destinar fondos y recursos adicionales a investigar más a fondo las cuestiones planteadas en este informe para determinar cuáles deben ser las salvaguardias y estructuras para prevenir o mitigar futuros abusos de los derechos humanos.

## Definiciones

1. INTELIGENCIA ARTIFICIAL O IA: No existe una definición consensuada de inteligencia artificial. Marvin Minsky, uno de los fundadores de la IA, la define como "la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran los hombres".[3] Otro de los fundadores, John McCarthy, la define como "la ciencia y la ingeniería de hacer máquinas inteligentes".[4] Un reciente informe de la Universidad de Stanford define la IA como "una ciencia y un conjunto de tecnologías computacionales que se inspiran en -pero que suelen funcionar de forma muy diferente- las formas en que las personas utilizan sus sistemas nerviosos y sus cuerpos para sentir, aprender, razonar y actuar".[5]

Stuart Russell y Peter Norving, autores de un popular libro de texto sobre IA, sugieren que ésta puede dividirse en las siguientes categorías 1) sistemas que piensan como los humanos; 2) sistemas que actúan como los humanos; 3) sistemas que piensan racionalmente; y 4) sistemas que actúan racionalmente.[6]

En realidad, la IA se considera más un campo que una "cosa" fácilmente definible, y puede dividirse en muchos subcampos, como el aprendizaje de máquina, la robótica, las redes neuronales, la visión, el procesamiento del lenguaje natural y el procesamiento del habla. Estos subcampos se entrecruzan de forma significativa. La IA también se nutre de otros campos además de la informática, como la psicología, la neurociencia, la ciencia cognitiva, la filosofía, la lingüística, la probabilidad y la lógica.

---

2 ↩ Mark Latonero: Artificial Intelligence & Human Rights: A Workshop at Data & Society — Points | Data & Society, 11 May 2018

3 ↩ "Report of COMEST on Robotics Ethics; 2017," n.d., 17.

4 ↩ McCarthy, John. 2018. " What Is AI? / Basic Questions "Jmc.Stanford.Edu. Accessed 15 June 2018.

5 ↩ 2018. ai100.Stanford.Edu. Accessed June 15 2018. https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf.

6 ↩ Qtd. in Committee on Technology, National Science and Technology Council, "Preparing for the Future of Artificial Intelligence" (Executive Office of the President of the United States, October 2016), 5,

La "IA estrecha" -la que se utiliza actualmente- es la aplicación de la inteligencia artificial a una sola tarea para usos como el reconocimiento de imágenes, la traducción de idiomas y los vehículos autónomos. En la actualidad, las máquinas realizan este tipo de tareas con mayor precisión que los humanos. En el futuro, los investigadores esperan conseguir una "inteligencia general artificial" (AGI). Esto supondría sistemas que muestran un comportamiento inteligente en toda una serie de tareas cognitivas. Sin embargo, los investigadores estiman que estas capacidades no se alcanzarán hasta dentro de unas décadas.[7]

2. GRANDES DATOS: Conjuntos de datos que son demasiado grandes o complejos para que los programas tradicionales de procesamiento de datos puedan analizarlos. La creciente disponibilidad de grandes datos, gracias al uso cada vez mayor de internet por parte de la sociedad, y unido a las rápidas mejoras en la potencia de cálculo, ha permitido los importantes avances en la IA en los últimos 10 años.

3. MINADO DE DATOS: Proceso de descubrimiento de patrones y extracción de información de grandes conjuntos de datos. En la era de los grandes datos, la minería de datos suele ser facilitada por el aprendizaje de máquina.

4. APRENDIZAJE DE MÁQUINA (AM): El aprendizaje de máquina es un subcampo de la IA. Harry Surden define el aprendizaje de máquina como "algoritmos informáticos que tienen la capacidad de "aprender" o mejorar su rendimiento a lo largo del tiempo en alguna tarea".[8] Esencialmente, es una máquina que aprende de los datos a lo largo del tiempo. Este aprendizaje se realiza a través de "un proceso estadístico que comienza con un conjunto de datos y trata de derivar una regla o procedimiento que explique los datos o pueda predecir datos futuros".[9] El resultado se llama modelo. Esto es diferente del enfoque tradicional de la inteligencia artificial, que implicaba que un programador intentara traducir la forma en que los humanos toman decisiones en código de software. La gran mayoría de la inteligencia artificial del mundo actual se basa en el aprendizaje de máquina. En la actualidad, muchos sistemas de AM son mucho más precisos que los humanos en una serie de tareas, desde la conducción hasta el diagnóstico de ciertas enfermedades.[10]

El aprendizaje de máquina funciona así:[11]

(1) Los programadores comienzan con un conjunto de datos históricos, que se dividen en un conjunto de entrenamiento y un conjunto de prueba.
(2) A continuación, eligen un modelo, una estructura matemática que caracteriza una serie de posibles reglas de decisión. Este modelo incluye parámetros ajustables. El modelo es como una caja, y los parámetros son los botones ajustables de la caja.
(3) Definen una función objetivo que sirve para evaluar la conveniencia del resultado.
(4) Entrenan el modelo, que es el proceso de ajuste de los parámetros para maximizar la función objetivo.
(5) Una vez entrenado, utilizan el conjunto de datos de prueba para evaluar la precisión y eficacia del modelo. Lo ideal es que el modelo tenga un rendimiento similar en los datos de prueba. El objetivo es poder generalizar el modelo, para que sea preciso en respuesta a casos que nunca ha visto antes.

---

7 ↩ "Preparing for the Future of Artificial Intelligence," 7.

8 ↩ Surden, Harry. 2014. "Machine Learning And Law". Papers.Ssrn.Com. Accessed 15 June 2018.

9 ↩ "Preparing for the Future of Artificial Intelligence", 5.

10 ↩ Para una explicación visual de cómo funciona el aprendizaje de máquina, véase, A visual introduction to machine learning

11 ↩ "Preparing for the Future of Artificial Intelligence", 9.

5. APRENDIZAJE PROFUNDO: Técnica de aprendizaje de máquina que utiliza estructuras denominadas "redes neuronales" que se inspiran en el cerebro humano. Están formadas por un conjunto de unidades en capas, modeladas como neuronas. Cada capa de unidades procesa un conjunto de valores de entrada y produce valores de salida que se pasan a la siguiente capa de unidades. Las redes neuronales suelen constar de más de 100 capas, con un gran número de unidades en cada una de ellas para permitir el reconocimiento de patrones extremadamente complejos y precisos en los datos.

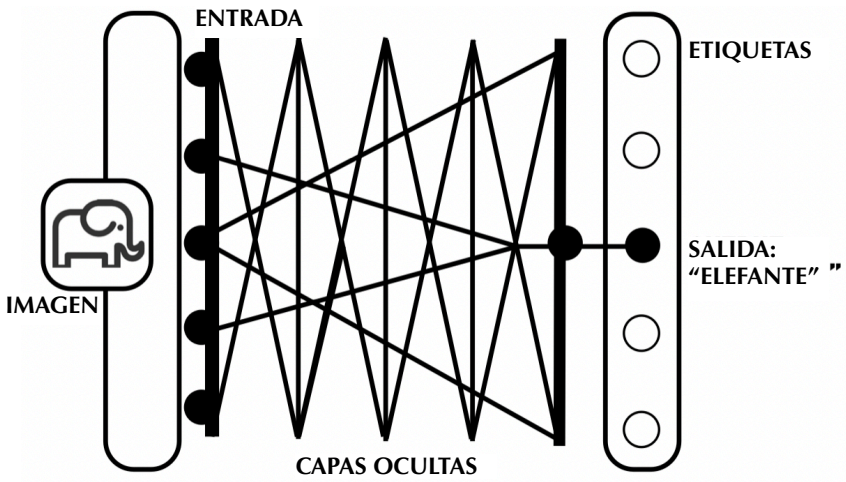


Fig 1. La Estructura de "redes neuronales"

Para profundizar en esta cuestión, pensemos en un programa de reconocimiento de imágenes que utiliza redes neuronales para identificar la imagen de un elefante. La primera capa de unidades podría buscar en los datos de la imagen en bruto los patrones más básicos, tal vez que hay una cosa con lo que parece ser cuatro patas. A continuación, la siguiente capa buscaría patrones dentro de estos patrones, tal vez que se trata de un animal. Entonces, tal vez la siguiente capa identifique el tronco. Este proceso continuaría a lo largo de muchas capas, reconociendo patrones cada vez más precisos en la imagen, hasta que la red es capaz de identificar que es en efecto una imagen de un elefante.

Los avances en el aprendizaje profundo han sido la causa de gran parte del optimismo sobre la IA debido a su capacidad para procesar y encontrar patrones en cantidades masivas de datos con precisión.[12] Mientras que el AM inicial suele utilizar una estructura de árbol de decisiones, el aprendizaje profundo se ha convertido en la técnica dominante. A menudo se utiliza para potenciar enfoques específicos de AM, como la visión artificial y el procesamiento del lenguaje natural.[13]

[12] ↩ Preparing for the Future of Artificial Intelligence", 9–10.

[13] ↩ Los asistentes virtuales Siri, Cortana y Alexa utilizan redes neuronales para reconocer el habla e imitar la conversación humana. En este caso, el aprendizaje profundo permite a estos asistentes virtuales detectar y comprender los matices del habla y producir una respuesta que parezca una conversación. Véase: Apple Machine Learning Research - Siri Team: Deep Learning for Siri's Voice: On-device Deep Mixture Density Networks for Hybrid Unit Selection Synthesis para más información. Watson, de IBM, utiliza técnicas de aprendizaje profundo para analizar la visión de las máquinas con el fin de interpretar rápidamente y con precisión la información médica y ayudar a los médicos a diagnosticar enfermedades. Véase: IBM: What is IBM Watson Health? para mayor información.

5.1. VISIÓN DE MÁQUINA: Un enfoque específico de AM que permite a los ordenadores reconocer y evaluar imágenes.[14] Es utilizado por Google para ayudar a buscar imágenes y por Facebook para etiquetar automáticamente a las personas en las fotos.

5.2. PROCESAMIENTO DEL LENGUAJE NATURAL: Un enfoque específico de AM que ayuda a los ordenadores a entender, interpretar y manipular el lenguaje humano. Lo hace descomponiendo el lenguaje en piezas más cortas y descubriendo cómo las piezas encajan para crear un significado. El procesamiento del lenguaje natural permite servicios de uso común como Google Translate y los chatbots.[15]

5.3 RECONOCIMIENTO DEL HABLA: Un enfoque específico de AM permite a los ordenadores traducir el lenguaje hablado en texto.[16] Permite utilizar el talk-to-text en el smartphone. A menudo se combina con el procesamiento del lenguaje natural y se utiliza para impulsar asistentes virtuales como Siri y Alexa.

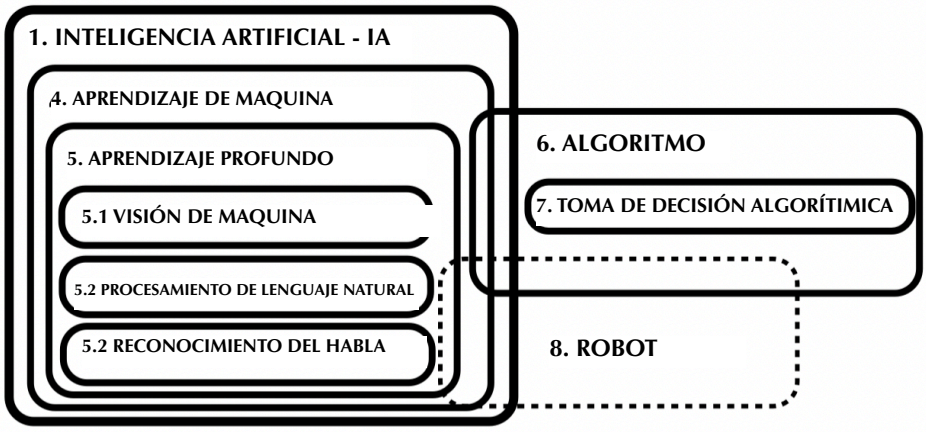


Fig 2. Las relaciones de los conceptos en la inteligencia artificial

6. ALGORITMO: En su forma más simple, un algoritmo es "un conjunto de directrices que describen cómo realizar una tarea".[17] Dentro de la informática, un algoritmo es una secuencia de instrucciones que indican a un ordenador lo que debe hacer.[18] La IA funciona mediante algoritmos (las redes neuronales son un tipo de algoritmo), pero no todos los algoritmos implican inteligencia artificial.

7. TOMA DE DECISIONES ALGORÍTMICA: Utilizar los resultados producidos por los algoritmos para tomar decisiones. Una de las primeras formas de toma de decisiones algorítmicas que todavía se utiliza en Estados Unidos son las directrices federales de sentencia para los jueces. No se trata más que de una ecuación matemática ponderada, extraída de las estadísticas, que recomienda una duración de la pena en función de los atributos del delito.[19]

---

14 ↩ "What Is a Machine Vision System (MVS)? - Definition from Techopedia," Techopedia.com, accessed 12 May 2018.

15 ↩ SAS "What Is Natural Language Processing?," accessed 12 May 2018.

16 ↩ "Speech Recognition," Wikipedia, 1 May 2018.

17 ↩ Jack Brogan: What's the Deal With Algorithms?," — Slate, accessed 12 May 2018.

18 ↩ ibid.

19 ↩ See: 2016 CHAPTER 5 CHAPTER FIVE - DETERMINING THE SENTENCE

8. ROBOT: Los robots suelen utilizar muchas de las formas de inteligencia artificial descritas anteriormente. Sin embargo, por definición, los robots tienen un cuerpo físico y movilidad. Los robots que utilizan la IA son capaces de percibir los cambios en su entorno y funcionar en consecuencia.[20] Aunque los robots suelen venir a la mente cuando se piensa en la inteligencia artificial, actualmente constituyen una cantidad muy pequeña de nuestras interacciones con la IA. La IA en el campo de la robótica es un área de investigación y desarrollo cada vez mayor, pero todavía no ha hecho tantos avances ni se ha vuelto tan omnipresente como las formas no robóticas de aprendizaje de máquinas.[21]

9. BOTS: Aplicaciones de software que ejecutan tareas automatizadas. Los bots se basan cada vez más en el AM, especialmente los chatbots, que utilizan el procesamiento del lenguaje natural para mantener conversaciones similares a las de los humanos con los usuarios.

10. DATOS ABIERTOS: Datos que están disponibles de forma gratuita para que todo el mundo pueda verlos, utilizarlos, compartirlos y volver a publicarlos sin restricciones. Existe un amplio movimiento de datos abiertos que defiende que los datos deben tratarse generalmente de esta manera.[22] En el contexto de la IA, muchos defensores sugieren que los datos de entrenamiento de los sistemas de AM sean abiertos para sacar a la luz los sesgos y errores, así como para arrojar luz sobre los resultados que producen los sistemas de AM. Existe una controversia sobre el mejor método para hacer esto respetando los intereses de privacidad de los sujetos de los datos.

11. INFORMACIÓN PROTEGIDA: Información que incluye, refleja, surge de, o es sobre las comunicaciones de una persona, y que no está fácilmente disponible y accesible al público en general. Aunque desde hace tiempo se ha acordado que el contenido de las comunicaciones merece una protección significativa en la ley debido a su capacidad para revelar información sensible, ahora está claro que otra información que surge de las comunicaciones -metadatos y otras formas de datos no relacionados con el contenido- puede revelar incluso más sobre un individuo que el propio contenido, y por lo tanto merece una protección equivalente. En la actualidad, cada uno de estos tipos de información puede, por sí solo o analizado colectivamente, revelar la identidad, el comportamiento, las asociaciones, las condiciones físicas o médicas, la raza, el color, la orientación sexual, los orígenes nacionales o los puntos de vista de una persona; o permitir el mapeo de la ubicación, los movimientos o las interacciones de la persona a lo largo del tiempo, o de todas las personas en un lugar determinado, incluso en torno a una manifestación pública u otro evento político.[23]

12. SESGO: Hay definiciones sociales y estadísticas de sesgo que entran en juego en la IA. La definición social de sesgo es "una inclinación o prejuicio a favor o en contra de una persona o grupo, especialmente de una manera que se considera injusta".[24] La definición estadística de sesgo es la diferencia entre el valor estimado -o predicho- y el valor real. En otras palabras, la diferencia entre lo que un sistema predice y lo que realmente ocurre.[25] En muchos casos, el sesgo estadístico presente en un determinado sistema de IA da lugar a resultados socialmente sesgados.

---

20 ↩ "Report of COMEST on Robotics Ethics; 2017."

21 ↩ Raghav Bharadwaj, "Artificial Intelligence in Home Robots – Current and Future Use-Cases," TechEmergence, 5 February 5 2018.

22 ↩ Véase: Wikipedia: Open Data

23 ↩ Véase: Necessary and Proportionate: International Principles on the Application of Human Rights to Communications Surveillance, accessed June 15 2018, available at. https://necessaryandproportionate.org/.

24 ↩ See Oxford Lexico: Bias

25 ↩ For a deeper discussion on statistical bias and fairness issues in AI, see talk by Princeton Computer Scientist Arving Narayanan

## ¿CÓMO ACTÚA EL SESGO EN LA IA?

*La IA puede estar sesgada tanto a nivel de sistema como de datos o de entrada. El sesgo a nivel del sistema implica que los desarrolladores incorporan sus propios prejuicios personales en los parámetros que consideran o en las etiquetas que definen. Aunque esto rara vez ocurre de forma intencionada, el sesgo involuntario a nivel del sistema es común. Esto suele ocurrir de dos maneras:*

- *Cuando los desarrolladores permiten que los sistemas confundan correlación con causalidad. Tomemos como ejemplo las puntuaciones de crédito. Las personas con bajos ingresos tienden a tener puntuaciones de crédito más bajas, por diversas razones. Si un sistema de AM utilizado para construir puntuaciones de crédito incluye las puntuaciones de crédito de tus amigos de Facebook como un parámetro, dará lugar a puntuaciones más bajas entre las personas con bajos ingresos, incluso si tienen otros indicadores financieros fuertes, simplemente por las puntuaciones de crédito de sus amigos.*
- *Cuando los desarrolladores deciden incluir parámetros que son sustitutos de sesgos conocidos. Por ejemplo, aunque los desarrolladores de un algoritmo intenten intencionadamente evitar el prejuicio racial al no incluir la raza como parámetro, el algoritmo seguirá teniendo resultados racialmente sesgados si incluye indicadores comunes de raza, como los ingresos, la educación o el código postal.*[26]

*El sesgo a nivel de datos o de entrada se produce de varias maneras:*[27]

- *El uso de datos históricos que están sesgados. Dado que los sistemas de AM utilizan un conjunto de datos existentes para identificar patrones, cualquier sesgo en esos datos se reproduce de forma natural. Por ejemplo, un sistema utilizado para recomendar admisiones en una universidad de alto nivel que utiliza los datos de los estudiantes previamente admitidos para entrenar el modelo es probable que recomiende a los hombres de clase alta por encima de las mujeres y de los grupos tradicionalmente infrarrepresentados.*
- *Cuando los datos de entrada no son representativos de la población objetivo. Esto se denomina sesgo de selección y da lugar a recomendaciones que favorecen a determinados grupos en detrimento de otros. Por ejemplo, si una aplicación de mapeo por GPS utilizara sólo los datos de entrada de los usuarios de teléfonos inteligentes para estimar los tiempos y las distancias de los viajes, podría ser más precisa en las zonas más ricas de las ciudades que tienen una mayor concentración de usuarios de teléfonos inteligentes, y menos precisa en las zonas más pobres o en los asentamientos informales, donde la penetración de los teléfonos inteligentes es menor y a veces no hay cartografía oficial.*
- *Cuando los datos de entrada están mal seleccionados. En el ejemplo de la aplicación de mapas GPS, esto podría implicar incluir sólo información relacionada con los coches, pero no los horarios del transporte público o los carriles bici, lo que daría lugar a un sistema que favorece a los coches y es inútil para los autobuses o la bicicleta.*
- *Cuando los datos son incompletos, incorrectos u obsoletos. Si no hay datos suficientes para sacar ciertas conclusiones, o los datos están desfasados, los resultados serán naturalmente inexactos. Y si un modelo de aprendizaje de máquina no se actualiza continuamente con nuevos datos que reflejen la realidad actual, será naturalmente menos preciso con el tiempo.*

*Desgraciadamente, los datos y los parámetros sesgados son la norma más que la excepción. Dado que los datos son producidos por seres humanos, la información lleva todo el sesgo humano natural dentro de ella. Los investigadores han empezado a tratar de averiguar la mejor manera de tratar y mitigar el sesgo, incluyendo si es posible enseñar a los sistemas de AM a aprender sin sesgo;*[28] *sin embargo, esta investigación está todavía en sus etapas iniciales. Por el momento, no hay cura para el sesgo en los sistemas de IA.*

---

26 ↩ "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy: Cathy O'Neil: 155–60, accessed 13 May 13 2018

27 ↩ Executive Office of the President of the United States, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," May 2016, 7–8,

28 ↩ Esto es ampliamente conocido en la comunidad FATML, "Fairness, Accountability and Transparency for Machine Learning."

## ¿Qué Hace que los Riesgos de la IA Sean diferentes?

Muchos de los problemas y riesgos analizados en este informe no son nuevos. Entonces, ¿en qué se diferencia la IA de las tecnologías que la han precedido? Debido a la forma en que la IA ha evolucionado con respecto a las tecnologías existentes, tanto en términos de sofisticación como de escala, la IA puede exacerbar las cuestiones existentes e introducir nuevos problemas a tener en cuenta, con enormes impactos para la responsabilidad y la fiabilidad. Para ilustrar esto, consideremos dos tendencias tecnológicas recientes: los grandes datos y el aumento de la toma de decisiones algorítmica.

En la actualidad, la toma de decisiones algorítmica es en gran medida digital. En muchos casos, emplea métodos estadísticos similares a los utilizados para crear el algoritmo de sentencia de papel y lápiz que hemos comentado anteriormente. Antes de la IA, los algoritmos eran deterministas, es decir, preprogramados e inmutables. Al estar basados en la modelización estadística, estos algoritmos sufren los mismos problemas que la estadística tradicional, como los datos mal muestreados, los datos sesgados y los errores de medición. Pero como están preprogramados, las recomendaciones que hacen pueden ser rastreadas.

El uso de la IA en la toma de decisiones algorítmicas ha introducido un nuevo conjunto de retos. Como los algoritmos de aprendizaje de máquina utilizan la estadística, también tienen los mismos problemas con los datos sesgados y los errores de medición que sus predecesores deterministas. Sin embargo, los sistemas de AM difieren en algunos aspectos clave. En primer lugar, mientras que la modelización estadística tradicional consiste en crear un modelo sencillo en forma de ecuación, el aprendizaje de máquina es mucho más preciso. Capta una multitud de patrones que no pueden expresarse en una sola ecuación. En segundo lugar, a diferencia de los algoritmos deterministas, los algoritmos de aprendizaje de máquina se calibran a sí mismos. Como identifican tantos patrones, son demasiado complejos para que los humanos los entiendan y, por tanto, no es posible rastrear las decisiones o recomendaciones que hacen. Además, muchos algoritmos de aprendizaje de máquina se recalibran constantemente a través de la retroalimentación.[29] Un ejemplo de ello son los filtros de correo electrónico, que aprenden y mejoran continuamente sus capacidades de detección de spam a medida que los usuarios marcan el correo electrónico como spam.

Otro problema es el impacto de las tasas de error. Debido a su base estadística, todos los sistemas de AM tienen tasas de error. Aunque en muchos casos los sistemas de AM son mucho más precisos que los seres humanos, existe el peligro de asumir que simplemente porque las predicciones de un sistema son más precisas que las de un ser humano, el resultado es necesariamente mejor. Incluso si la tasa de error es cercana a cero, en una herramienta con millones de usuarios, miles podrían verse afectados por las tasas de error. Consideremos el ejemplo de Google Photos. En 2015 se descubrió que el programa de reconocimiento de imágenes de Google Photos tenía un error terriblemente prejuicioso y ofensivo: ocasionalmente etiquetaba fotos de personas negras como gorilas. Debido a que el sistema utilizaba un complejo modelo AM, los ingenieros fueron incapaces de averiguar por qué ocurría esto. La única "solución" que pudieron encontrar para este AM "racista" fue simplemente una tirita: eliminaron cualquier palabra relacionada con los monos de la lista de etiquetas de las imágenes.[30]

Ahora imagina un programa similar utilizado por la Patrulla Fronteriza y de Aduanas de EUA que fotografía a cada persona que entra y sale de EUA y lo cruza con una base de datos de fotos de criminales y terroristas conocidos o

---

29 ↩ TAVISH SRIVASTAVA: "What Is The Difference Between Machine Learning & Statistical Modeling," — Analytics Vidhya accessed 12 May 2018,

30 ↩ "When It Comes to Gorillas, Google Photos Remains Blind | WIRED," accessed 13 May 13 2018.

sospechosos. En 2016, se calcula que llegaron a Estados Unidos 75,9 millones de personas.[31] Incluso si el sistema de reconocimiento facial tuviera una precisión del 99,9%, la tasa de error del 0,1% daría lugar a 75.900 personas mal identificadas. ¿Cuántas de estas personas serían falsamente identificadas como delincuentes buscados y detenidas? ¿Y cuál sería el impacto en sus vidas? A la inversa, ¿cuántos delincuentes conocidos se escaparían? Incluso los porcentajes de error relativamente reducidos en casos como éste pueden tener graves consecuencias.

**El quid de la cuestión: La escala, la proliferación y el impacto en la vida real de la IA exigen atención**

La proliferación de la IA en el análisis de datos ha llegado con el auge de los grandes datos. En su libro de 2015 Weapons of Math Destruction (Armas de destrucción matemática), la científica de datos Cathy O'Neil documentó cómo la toma de decisiones mediante algoritmos es ahora omnipresente en Occidente, desde la asignación de puntuaciones de crédito hasta la identificación de los mejores candidatos para un puesto de trabajo, pasando por la clasificación de los estudiantes para su admisión en la universidad. Hoy en día, estos sistemas algorítmicos de toma de decisiones emplean cada vez más el aprendizaje de máquina, y se están extendiendo rápidamente. Tienen muchos de los mismos problemas que el análisis estadístico tradicional. Sin embargo, la escala y el alcance de los sistemas de IA, la tendencia a un despliegue rápido y descuidado, el impacto inmediato que tienen en la vida de muchas personas y el peligro de que las sociedades consideren sus resultados como imparciales, plantean una serie de nuevos problemas.

## IA Benéfica y Dañina

Todas las grandes innovaciones tecnológicas tienen el potencial de hacer avanzar o perjudicar a la sociedad. Las capacidades de procesamiento y análisis de datos de la IA pueden ayudar a aliviar algunos de los problemas más acuciantes del mundo, desde permitir avances en el diagnóstico y el tratamiento de enfermedades, hasta revolucionar el transporte y la vida urbana, pasando por mitigar los efectos del cambio climático. Empero, estas mismas capacidades también pueden permitir la vigilancia a una escala nunca vista, pueden identificar y discriminar a los más vulnerables, y pueden revolucionar la economía tan rápidamente que ningún programa de reconversión laboral pueda seguir el ritmo. Y a pesar de los grandes avances en el desarrollo de la IA, la llamada "revolución de la inteligencia artificial" sólo tiene una década, lo que significa que hay muchas posibilidades desconocidas en lo que está por venir.

A continuación identificamos algunas de las formas en que la IA se está utilizando para ayudar o perjudicar a las sociedades. Es importante señalar que incluso los usos "útiles" de la IA tienen implicaciones potencialmente negativas. Por ejemplo, muchas aplicaciones de la IA en la atención sanitaria plantean graves amenazas a la privacidad y corren el riesgo de discriminar a las comunidades desatendidas y de concentrar la propiedad de los datos en las grandes empresas. Al mismo tiempo, el uso de la IA para mitigar los daños puede no resolver los problemas subyacentes y no debe tratarse como una cura para los males de la sociedad. Por ejemplo, aunque la IA puede aliviar la necesidad de profesionales de la medicina en zonas desatendidas, no proporciona los recursos o incentivos que esos profesionales necesitarían para trasladarse. Del mismo modo, algunos de los casos de uso clasificados como "perjudiciales" surgieron como resultado de buenas intenciones, pero están causando un daño significativo.

### INTELIGENCIA ARTIFICIAL BENÉFICA

Mejorar el acceso a la atención sanitaria y predecir los brotes de enfermedades: Ya se han producido avances significativos mediante el uso de la IA en el diagnóstico y la prevención de enfermedades. La IA también se utiliza para

[31] ↩ U.S. Travel Answer Sheet

mejorar el acceso a la atención sanitaria en regiones donde no hay acceso.[32] Las víctimas de los brotes de enfermedades también se benefician del uso de la IA para que los funcionarios sanitarios puedan intervenir de forma temprana para contener un brote antes de que se inicie.[33]

Facilitar la vida a los discapacitados visuales: Las herramientas de reconocimiento de imágenes ayudan a los discapacitados visuales a navegar mejor tanto por Internet como por el mundo real.[34]

Optimizar la agricultura y ayudar a los agricultores a adaptarse al cambio: La IA está combinando la información de las imágenes de satélite globales con los datos meteorológicos y agronómicos para ayudar a los agricultores a mejorar el rendimiento de las cosechas, diagnosticar y tratar las enfermedades de los cultivos y adaptarse a los cambios del entorno. Este enfoque de la agricultura se conoce como agricultura de precisión y puede ayudar a aumentar la productividad de las explotaciones para alimentar a una mayor parte de la creciente población mundial.

Mitigar el cambio climático, predecir las catástrofes naturales y conservar la fauna: Ante la aparición de los efectos del cambio climático en todo el mundo, el aprendizaje de máquina se está utilizando para elaborar modelos climáticos más precisos para los científicos. Ya se utiliza la IA para clasificar los modelos climáticos y predecir los fenómenos meteorológicos extremos,[35] así como para predecir mejor los fenómenos meteorológicos extremos y responder a los desastres naturales.[36] La IA también es útil para identificar y detener a los cazadores furtivos y localizar y capturar a los animales que propagan enfermedades.[37]

Hacer que los servicios gubernamentales sean más eficientes y accesibles: A pesar de ser a menudo lentos en la adopción de nuevas tecnologías, los gobiernos de todo el mundo están utilizando la IA, desde el nivel local hasta el nacional, para hacer que los servicios públicos sean más eficientes y accesibles, haciendo hincapié en el desarrollo de "ciudades inteligentes". La IA también se está utilizando para asignar recursos gubernamentales y optimizar los presupuestos.

## INTELIGENCIA ARTIFICIAL DAÑINA

Perpetuación de los prejuicios en la justicia penal: Hay muchos casos documentados de IA que han salido mal en el sistema de justicia penal. El uso de la IA en este contexto se produce a menudo en dos ámbitos diferentes: la calificación del riesgo -evaluar si un acusado tiene probabilidades de reincidir con el fin de recomendar la sentencia y fijar la fianza- o la llamada "policía predictiva", que utiliza información de varios puntos de datos para predecir dónde o

---

32 ↩ Watson de IBM se utiliza en hospitales de todo el mundo para ayudar a los médicos a diagnosticar y tratar enfermedades. Para más información, consulte https://www.ibm.com/watson/health/. Otro ejemplo es Aajoh, una empresa nigeriana que está desarrollando un sistema de IA para el diagnóstico médico a distancia. Los usuarios comparten sus síntomas mediante texto, audio y fotografías, y Aajoh utiliza la IA para ofrecer posibles diagnósticos. Véase Stephen Timm, "6 Artificial Intelligence Startups in Africa to Look out For," Venture Burn, 24 April 24 2017.

33 ↩ Zeena Saifi, Victoria Brown and Tom Page: AI and big data joins effort to predict deadly disease outbreaks — CNN Health, 6 March 2018.

34 ↩ Para ver algunos ejemplos, vea los esfuerzos de Facebook para ayudar a los ciegos a "ver" Facebook: "Using Artificial Intelligence to Help Blind People 'See' Facebook," Facebook Newsroom, April 4, 2016. Véase también Microsoft's work: "Seeing AI: An app for visually impaired people that narrates the world around you," Microsoft.

35 ↩ Nicola Jones: How Machine Learning Could Help to Improve Climate Forecasts – Mixing artificial intelligence with climate science helps researchers to identify previously unknown atmospheric processes and rank climate models — Scientific American, 23 August 2017.

36 ↩ Aili McConnon, "AI Helps Cities Predict Natural Disasters," The Wall Street Journal, June 26, 2018.

37 ↩ Véase Hila Mehr, "Artificial Intelligence for Citizen Services and Government," Ash Center for Democratic Governance and Innovation, Harvard Kennedy School, August 2017, and Daryl Pereira: "Watson helps cities help citizens: the 411 on how artificial intelligence transforms 311 services," Medium, 31 January 31 2017.

cuándo se producirá el delito y dirigir la acción policial en consecuencia.[38] En muchos casos, estos esfuerzos son probablemente bien intencionados. El uso del aprendizaje de máquina para calificar el riesgo de los acusados se anuncia como una forma de eliminar el conocido sesgo humano de los jueces en sus decisiones sobre sentencias y fianzas.[39] Y los esfuerzos de predicción policial tratan de asignar de la mejor manera posible los recursos policiales, a menudo limitados, para prevenir la delincuencia, aunque siempre existe un alto riesgo de desviación de la misión.[40] Sin embargo, las recomendaciones de estos sistemas de IA a menudo exacerban el mismo sesgo que intentan mitigar, ya sea directamente o incorporando factores que son sustitutos del sesgo.

Facilitar la vigilancia masiva: Dado que la IA proporciona la capacidad de procesar y analizar múltiples flujos de datos en tiempo real, no es de extrañar que ya se esté utilizando para facilitar la vigilancia masiva en todo el mundo.[41] El ejemplo más generalizado y peligroso es el uso de la IA en programas de reconocimiento facial.[42] Aunque la tecnología es todavía imperfecta, los gobiernos están buscando la tecnología de reconocimiento facial como una herramienta para vigilar a sus ciudadanos, facilitar la elaboración de perfiles de ciertos grupos, e incluso identificar y localizar a individuos.[43]

Permitir la elaboración de perfiles discriminatorios: Los programas de reconocimiento facial no sólo se utilizan para vigilar e identificar, sino también para seleccionar y discriminar.[44]

Ayudar a la difusión de desinformación: La IA puede utilizarse para crear y difundir propaganda selectiva, y este problema se agrava con los algoritmos de las redes sociales impulsados por la IA y basados en el "compromiso", que promueven los contenidos con más probabilidades de que se haga clic en ellos. El aprendizaje de máquina impulsa el análisis de datos que las empresas de redes sociales utilizan para crear perfiles de usuarios para la publicidad dirigida. Además, los bots disfrazados de usuarios reales difunden más contenidos fuera de los círculos de las redes sociales,

---

38 ↩Una investigación de ProPublica de 2016 reveló que COMPAS, un programa de inteligencia artificial ampliamente utilizado en el sistema de justicia penal de EUA, no solo era inexacto a la hora de pronosticar futuros delitos, sino que tenía un fuerte sesgo contra los acusados de raza negra. Los investigadores examinaron las puntuaciones de riesgo de más de 7.000 personas detenidas en el condado de Broward (Florida) y las compararon con los antecedentes penales posteriores. Descubrieron que sólo el 20% de las personas que se preveía que iban a cometer delitos violentos acabaron haciéndolo. Y al examinar toda la gama de delitos, sólo el 61% de los acusados considerados propensos a reincidir fueron realmente detenidos por un delito futuro. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner: "Machine Bias," — ProPublica, 23 May 23 2016.

39 ↩ Big Brother Watch: A CLOSER LOOK AT EXPERIAN BIG DATA AND ARTIFICIAL INTELLIGENCE IN DURHAM POLICE; Una investigación de la Comisión de Ciencia y Tecnología del Parlamento de HART, Un programa informático con tecnología AM utilizado por la policía de Durham (Inglaterra) para evaluar el riesgo de reincidencia, reveló que estaba calibrado para evitar falsos negativos, clasificando incorrectamente a una persona como de bajo riesgo cuando en realidad llega a cometer delitos graves; y Durham police criticised over 'crude' profiling.

40 ↩ Los registros públicos sugieren que el programa desarrollado por Palantir y utilizado por la policía en las investigaciones criminales en Nueva Orleans se utilizó más allá de su alcance original. Ali Winston: PALANTIR HAS SECRETLY BEEN USING NEW ORLEANS TO TEST ITS PREDICTIVE POLICING TECHNOLOGY — The Verge, 27 February 2018. Tras una serie de informes de investigación y una importante protesta pública, la ciudad puso fin a su contrato de seis años con Palantir en marzo de 2018.

41 ↩China, en particular, está persiguiendo agresivamente un estado de vigilancia basado en la IA. Véase Paul Mozur, "Inside China's Dystopian Dreams: AI, Shame and Lots of Cameras," The New York Times, 8 July 2018.

42 ↩ En 2018, Australia dio a conocer un plan para conectar su red de cámaras de vídeovigilancia a las bases de datos de reconocimiento facial y biométricas existentes. La medida propuesta está pendiente en el Parlamento. Access Now: Human Rights in the Digital Era: An International Perspective on Australia

43 ↩ Recientemente, Amazon ha sido objeto de críticas por comercializar directamente un producto de reconocimiento facial llamado Rekognition a las fuerzas del orden para su uso junto con las cámaras corporales de la policía, lo que permitiría a la policía identificar a las personas en tiempo real. El producto se probó en los departamentos de policía de Orlando (Florida) y del condado de Washington (Oregón). Ver: Julia carrie Wong: 'Recipe for authoritarianism': Amazon under fire for selling face-recognition software to police — The Guardian, 22 May 2018.

44 ↩ Un ejemplo es una empresa israelí llamada Faception, que se autodenomina "empresa de tecnología de análisis de la personalidad facial" y afirma que puede clasificar a las personas en tipos de personalidad basándose únicamente en sus rostros. Los clasificadores que utiliza incluyen "delincuente de cuello blanco", "alto coeficiente intelectual", "pedófilo" y "terrorista". La empresa no ha revelado ninguna información sobre cómo su tecnología puede etiquetar correctamente a las personas basándose únicamente en sus rostros. Véase: Paul Lewis, "'I was shocked it was so easy': meet the professor who says facial recognition can tell if you're gay," The Guardian, 7 July 2018.

compartiendo enlaces a fuentes falsas e interactuando activamente con los usuarios como chatbots mediante el procesamiento del lenguaje natural.[45]

Además, el espectro de las "falsificaciones profundas", sistemas de IA capaces de crear grabaciones de vídeo y audio de personas reales que parecen realistas, hace que muchos crean que la tecnología se utilizará en el futuro para crear vídeos falsos de líderes mundiales con fines maliciosos. Aunque parece que las falsificaciones profundas aún no se han utilizado como parte de verdaderas campañas de propaganda o desinformación, y el audio y el vídeo falsificados aún no son lo suficientemente buenos como para parecer completamente humanos, la IA que está detrás de las falsificaciones profundas sigue avanzando, y existe un potencial para sembrar el caos, instigar conflictos y provocar aún más una crisis de la verdad que no debe descartarse.[46]

Perpetuar los prejuicios en el mercado laboral: Los procesos de contratación han estado durante mucho tiempo plagados de prejuicios y discriminación. Como respuesta, ha surgido toda una industria que utiliza la IA con el objetivo de eliminar los prejuicios humanos del proceso. Sin embargo, muchos productos corren el riesgo de perpetuar el mismo sesgo que pretenden mitigar. Al igual que en otras áreas, una de las principales causas es el uso predominante de datos históricos de empleados anteriores "exitosos" para entrenar los modelos de inteligencia artificial, reproduciendo así de forma natural el sesgo de las contrataciones anteriores.[47]

Impulsar la discriminación financiera contra los marginados: Los algoritmos se han utilizado durante mucho tiempo para crear puntuaciones de crédito e informar sobre la selección de préstamos. Sin embargo, con el auge de los macrodatos, los sistemas utilizan ahora el aprendizaje de máquina para incorporar y analizar datos no financieros con el fin de determinar la solvencia, desde el lugar de residencia de las personas hasta sus hábitos de navegación por Internet o sus decisiones de compra. Los resultados de estos sistemas se conocen como puntuaciones electrónicas y, a diferencia de las puntuaciones de crédito formales, no están reguladas. Como ha señalado la científica de datos Cathy O'Neil, estas puntuaciones son a menudo discriminatorias y crean circuitos de retroalimentación perniciosos.[48]

## IA y Derechos Humanos - ¿Por qué son Importantes los Derechos?

La IA ha "creado nuevas formas de opresión y, en muchos casos, afecta de forma desproporcionada a los más impotentes y vulnerables". El concepto de derechos humanos aborda las diferencias de poder y proporciona a los

---

[45] ↩ Dado que se estima que los bots representan al menos la mitad de todo el tráfico de Internet, no hay que subestimar su alcance. Véase: Michael Horowitz, Paul Scharre, Gregory C. Allen, Kara Frederick, Anthony Cho and Edoardo Saravalle, "Artificial Intelligence and International Security," Center for a New American Security, 10 July 2018.

[46] ↩ ibid.

[47] ↩ Monica Torres, "Companies Are Using AI to Screen Candidates Now with HireVue," Ladders, 25 August 2017.

[48] ↩ Por ejemplo, un aspirante a préstamos que vive en una zona conflictiva de la ciudad, donde hay más gente que no paga sus préstamos, puede recibir una puntuación baja y ser objeto de productos financieros que ofrecen menos crédito y tipos de interés más altos. Esto se debe a que estos sistemas agrupan a las personas en función de los hábitos observados de la mayoría. En este caso, a una persona responsable que intente iniciar un negocio se le podría negar el crédito o conceder un préstamo en condiciones desfavorables, perpetuando el sesgo y la desigualdad social existentes. O'Neil, 141-160. Una de las empresas que O'Neil destaca es ZestFinance, que utiliza el aprendizaje de máquina para ofrecer préstamos de día de pago a tipos más bajos que los típicos prestamistas de día de pago. La filosofía de la empresa es que "todos los datos son datos de crédito". Se ha descubierto que algunos de los datos son una aproximación a la raza, la clase social y el origen nacional. Esto incluye si los solicitantes utilizan la ortografía y las mayúsculas adecuadas en su solicitud, y cuánto tiempo tardan en leerla. Los errores de puntuación y ortografía se analizan para sugerir que el solicitante tiene menos educación y/o no es un hablante nativo de inglés, lo que está altamente correlacionado con el estatus socioeconómico, la raza y el origen nacional. Esto significa que los que se consideran que tienen un mal dominio del idioma -incluidos los que no son nativos- tendrán tasas de interés más altas. Esto puede dar lugar a un bucle de retroalimentación que afiance las prácticas de préstamo discriminatorias existentes: si los solicitantes tienen problemas para pagar estas tasas más altas, esto indica al sistema que eran en efecto de mayor riesgo, lo que dará lugar a puntuaciones más bajas para otros solicitantes similares en el futuro. O'Neil, 157-158.

individuos, y a las organizaciones que los representan, el lenguaje y los procedimientos para impugnar las acciones de los actores más poderosos, como los Estados y las empresas".[49]

Los derechos humanos son universales y vinculantes, y están codificados en un cuerpo de derecho internacional. El respeto de los derechos humanos se exige tanto a los gobiernos como a las empresas, aunque los gobiernos tienen obligaciones adicionales de proteger y cumplir los derechos humanos.[50] Existe todo un sistema de instituciones y organizaciones regionales, internacionales y nacionales que proporcionan marcos bien desarrollados para remediar y articular la aplicación de la legislación sobre derechos humanos a las circunstancias cambiantes, incluidos los avances tecnológicos. Y en los casos en los que se carece de legislación nacional, la legitimidad moral de los derechos humanos conlleva un importante poder normativo.[51] La violación de los derechos humanos conlleva costes políticos y de reputación a nivel mundial, por lo que nombrar y avergonzar a los violadores de los derechos humanos suele ser una herramienta eficaz. La legislación sobre derechos humanos puede abordar algunos de los daños sociales más atroces causados por la IA y evitar que se produzcan en el futuro.

***La ética y su papel como área complementaria:*** *Hasta ahora, el discurso ético ha dominado en gran medida el debate sobre la IA "buena" y "mala". Este enfoque es comprensible, y la ética desempeña un papel importante. La inteligencia artificial ha suscitado más debates sobre la interacción entre los seres humanos y las máquinas que quizás cualquier otro desarrollo tecnológico anterior. La consideración de conceptos éticos como la justicia, la equidad, la transparencia y la responsabilidad permite un valioso debate sobre las repercusiones sociales de la IA y el papel de ésta en nuestras vidas.[52] También existe una comunidad de investigación académica dedicada a abordar las cuestiones éticas.[53] La ética ha ayudado a quienes investigan y desarrollan la IA a definir sus propios límites. Los principales actores de la IA, como Google, Microsoft y DeepMind, han desarrollado principios éticos para guiar sus iniciativas de IA.[54]*

*Los derechos humanos son más universales y están mejor definidos que los principios éticos, y prevén la rendición de cuentas y la reparación. De este modo, los derechos humanos y la ética pueden reforzarse mutuamente. Por ejemplo, una empresa puede desarrollar principios éticos de la IA, como evitar el refuerzo de los prejuicios sociales negativos y asegurarse de que sus sistemas son responsables de la supervisión humana. Los derechos humanos a la privacidad y a la no discriminación, entre otros, pueden definir aún más esos principios éticos, y el régimen internacional de derechos humanos puede prever recursos en caso de que se violen esos principios. Además, si un uso de la IA se considera poco ético, es probable que también viole los derechos humanos, y los principios y procedimientos integrados en el régimen internacional de derechos humanos pueden aprovecharse para combatir ese uso poco ético de la IA. A continuación se exponen recomendaciones sobre cómo las partes interesadas pueden utilizar conjuntamente la ética y los derechos humanos en sus políticas internas.*

---

49 ↩ Christiaan van Veen: Artificial Intelligence: What's Human Rights Got To Do With It? — Points | Data Society, 14 May 2018

50 ↩ Según los Principios de las Naciones Unidas sobre las Empresas y los Derechos Humanos, los Estados deben proteger contra los abusos de los derechos humanos por parte de las empresas dentro de su jurisdicción, las empresas son responsables de respetar los derechos humanos dondequiera que operen, las víctimas deben tener acceso a recursos judiciales y no judiciales. Para más información, véase UN: Guiding Principles on Human Rights, 2011.

51 ↩ ibid.

52 ↩ Considerati: Marrying Ethics and Human Rights for AI Scrutiny

53 ↩ The Fairness, Accountability and Transparency in Machine Learning initiative

54 ↩ For each policy see: Microsoft, Google, DeepMind.

## CÓMO AFECTA LA IA A LOS DERECHOS HUMANOS

El papel de la IA en la facilitación de la discriminación está bien documentado, y es uno de los temas clave en el debate ético actual. Para reconocer estas cuestiones, Access Now se asoció con organizaciones de derechos humanos y empresas de IA para publicar "La Declaración de Toronto" en marzo de 2018.[55] Sin embargo, el derecho a la no discriminación no es el único derecho humano implicado por la IA. Dado que los derechos humanos son interdependientes y están interrelacionados, la IA afecta a casi todos los derechos humanos reconocidos internacionalmente.

A continuación examinamos muchos de los derechos humanos afectados por la IA.[56] Los derechos que se analizan son, en gran medida, los plasmados en los tres documentos que constituyen la base de la legislación internacional sobre derechos humanos, la llamada "Carta Internacional de Derechos Humanos".[57] Esto incluye la Declaración Universal de Derechos Humanos (DUDH), el Pacto Internacional de Derechos Civiles y Políticos (PIDCP) y el Pacto Internacional de Derechos Económicos, Sociales y Culturales (PIDESC).[58] A ellos, este informe añade el derecho a la protección de datos definido en la Carta de Derechos Fundamentales de la UE.[59] Para cada uno de los derechos humanos implicados, analizamos cómo los usos actuales de la IA violan o corren el riesgo de violar ese derecho, así como los riesgos que plantean los futuros desarrollos de la IA. Es importante señalar que los problemas de derechos humanos que se analizan a continuación no son necesariamente exclusivos de la IA. Muchos de ellos ya existen en el ámbito de los derechos digitales, pero la capacidad de la IA para identificar, clasificar y discriminar aumenta la posibilidad de que se produzcan abusos de los derechos humanos tanto en su escala como en su alcance.

Al igual que los daños a los derechos humanos en otros usos de la tecnología que aprovechan los datos, los daños relacionados con el uso de la IA suelen afectar de forma desproporcionada a las poblaciones marginadas.[60] Esto puede incluir a las mujeres y los niños, así como a determinados grupos étnicos, raciales o religiosos, los pobres, las personas con capacidades diferentes y los miembros de la comunidad LGBTQ. La arraigada marginación de estos grupos se refleja en los datos y se reproduce en los resultados que afianzan las pautas históricas.

---

55 ↩ "The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems.

56↩ Los derechos humanos no incluidos son: el derecho a no ser torturado, el derecho a no ser esclavizado, los derechos de los detenidos, el derecho a no ser encarcelado simplemente por no poder cumplir una obligación contractual, los derechos de los extranjeros y el derecho a la seguridad social. Esto no significa que la IA no pueda afectar en última instancia a estos derechos, simplemente que no hemos encontrado violaciones documentadas en la actualidad, ni tampoco violaciones potenciales que creamos que puedan producirse en un futuro próximo.

57 ↩ Hay que tener en cuenta que, aunque hay muchos sistemas regionales de derechos humanos que son más completos, hemos limitado nuestro análisis al sistema basado en la ONU en aras de la aplicabilidad universal. La excepción es el derecho a la protección de datos, que Access Now reconoce como un derecho y es especialmente relevante en el contexto de la IA. Merece la pena seguir analizando la IA en relación con los derechos enumerados en estos sistemas regionales. Por ejemplo, el Convenio Europeo de Derechos Humanos y la Carta de Derechos Fundamentales de la UE son mucho más amplios en lo que respecta a los derechos de los trabajadores, y el uso de la IA por parte de los empresarios para controlar y vigilar la actividad de los empleados puede violar los derechos humanos europeos.

58 ↩ Véase UN: International Human Rights Law for more information

59↩ CHARTER OF FUNDAMENTAL RIGHTS OF THE EUROPEAN UNION (2012/C 326/02)

60↩ Para un examen más detallado de cómo se desarrolla esto en la EUA, véase Automating Inequality, de Virginia Eubanks. Véase también https://harpers.org/archive/2018/01/the-digital-poorhouse/ ("los más marginados de nuestra sociedad se enfrentan a mayores niveles de recopilación de datos cuando acceden a prestaciones públicas, caminan por barrios muy vigilados, entran en el sistema de salud o cruzan las fronteras nacionales. Esos datos refuerzan su marginalidad cuando se utilizan para someterlos a un escrutinio adicional. Los grupos considerados como no merecedores de apoyo social e inclusión política son señalados para una política pública punitiva y una vigilancia más intensa, y el ciclo comienza de nuevo. Es un bucle de retroalimentación de la injusticia").

### Derechos a la vida, a la libertad y a la seguridad, a la igualdad ante los tribunales y a un juicio justo[61]

*"Todo individuo tiene derecho a la libertad y a la seguridad de su persona. Nadie podrá ser sometido a detención o prisión arbitrarias. Nadie podrá ser privado de su libertad, salvo por las causas y con arreglo al procedimiento que establezca la ley."* - Artículo 9 del PIDCP.

*"Todas las personas son iguales ante los tribunales y cortes de justicia. Toda persona tendrá derecho a ser oída públicamente y con las debidas garantías por un tribunal competente, independiente e imparcial, establecido por la ley, en la substanciación de cualquier acusación de carácter penal formulada contra ella o para la determinación de sus derechos u obligaciones de carácter civil [...] Toda persona acusada de un delito tendrá derecho a que se presuma su inocencia mientras no se pruebe su culpabilidad conforme a la ley."* - Artículo 14 del PIDCP.

*"Todo ser humano tiene el derecho inherente a la vida. Este derecho estará protegido por la ley. Nadie podrá ser privado de la vida arbitrariamente". En los países que no hayan abolido la pena de muerte, sólo podrá imponerse la pena de muerte por los delitos más graves, de conformidad con la legislación vigente en el momento de cometerse el delito y no en contra de las disposiciones del presente Pacto."* - Artículo 6 del PIDCP.

El creciente uso de la IA en el sistema de justicia penal corre el riesgo de interferir con el derecho a no sufrir injerencias en la libertad personal. Un ejemplo es el programa de puntuación del riesgo de reincidencia que se utiliza en todo el sistema de justicia penal de EUA para fundamentar las decisiones de detención en casi todas las etapas, desde la asignación de la fianza hasta la sentencia penal.[62] Este programa ha dado lugar a que un mayor número de acusados de raza negra sean etiquetados falsamente como de alto riesgo y se les impongan condiciones de fianza más elevadas, se les mantenga en prisión preventiva y se les condene a penas de cárcel más largas. Además, dado que los sistemas de puntuación de riesgos no están prescritos por la ley y utilizan datos que pueden ser arbitrarios, las decisiones de detención basadas en estos sistemas pueden ser ilegales o arbitrarias.

Los programas de evaluación de riesgos penales se presentan como una herramienta que simplemente ayuda a los jueces a tomar decisiones sobre las sentencias. Sin embargo, al calificar a un acusado como de alto o bajo riesgo de reincidencia, le atribuyen un nivel de culpabilidad futura, lo que puede interferir con la presunción de inocencia requerida en un juicio justo.[63] Los programas de predicción policial también corren el riesgo de imputar erróneamente la culpabilidad, incorporando los prejuicios policiales existentes mediante el uso de datos pasados. Los informes sugieren que los jueces saben muy poco sobre el funcionamiento de estos sistemas de puntuación de riesgos, pero muchos confían en los resultados porque el programa se considera imparcial.[64] Esto plantea la cuestión de si las decisiones judiciales tomadas sobre la base de estos programas pueden considerarse realmente justas.[65]

---

61 ↩Article 3, 6, 7, 8, and 10 of UDHR, Articles 9 and 14 of the ICCPR

62 ↩Angwin et. al, "Machine Bias.

63 ↩According to the General Comment 32 on Article 14 of the ICCPR

64↩ Angwin et. al, "Machine Bias."

65↩ Como ya se ha dicho, no todas las comunidades están vigiladas por igual y, debido a este sesgo, los programas basados en la inteligencia artificial acaban creando circuitos de retroalimentación negativos que pueden "predecir" el aumento de la actividad delictiva en determinadas zonas, lo que da lugar a comunidades continuamente vigiladas en exceso. Véase: Comité de Expertos en Intermediarios de Internet, "Algorithms and Human Rights: Estudio sobre las dimensiones de los derechos humanos de las técnicas de procesamiento automatizado de datos y las posibles implicaciones regulatorias" Council of Europe, March 2018, pg. 10-12, https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5. 66 Robert Brauneis and Ellen P. Goodman, "Algorithmic Transparency for the Smart City," The Yale Journal of Law and Technology, Vol. 20 (2018), 103-176.

Cuando utilizan estas herramientas, los gobiernos ceden esencialmente la toma de decisiones a proveedores privados. Los ingenieros de estos proveedores, que no son funcionarios elegidos, utilizan el análisis de datos y las opciones de diseño para codificar las decisiones políticas que a menudo no ven ni el organismo gubernamental ni el público. Cuando se deniega la libertad condicional a una persona o se le impone una determinada condena por razones que nunca conocerán y que no pueden ser articuladas por la autoridad gubernamental encargada de tomar esa decisión, los juicios pueden no ser justos y este derecho puede ser violado.[66]

*De cara al futuro: El uso generalizado de programas de reconocimiento facial en las fuerzas de seguridad aumenta el riesgo de que se produzcan detenciones ilegales debido a errores y a la exageración. La historia está plagada de ejemplos de personas que han detenido erróneamente a personas con un aspecto similar al de los delincuentes buscados.[67] Dados los índices de error de la actual tecnología de reconocimiento facial, estas imprecisiones podrían conducir a un aumento de las detenciones erróneas debido a la identificación incorrecta, exacerbada por los índices de precisión más bajos para los rostros no blancos.[68]*

La incapacidad de la IA para manejar los matices probablemente causará más problemas en el futuro. Las leyes no son absolutas; hay ciertos casos en los que está justificado infringir la ley. Por ejemplo, probablemente sea aceptable saltarse un semáforo en rojo para evitar una colisión por detrás con un coche que le sigue. Mientras que un agente de policía humano puede hacer esa distinción y decidir no multar al conductor, las cámaras de semáforo en rojo no tienen esa capacidad de juicio. En un futuro de ciudades inteligentes impulsadas por la inteligencia artificial y de "policías robóticos", existe el riesgo de que esta pérdida de matices conduzca a un aumento drástico de personas detenidas, multadas o multadas injustamente, con un recurso limitado. Con el tiempo, estas circunstancias podrían empujarnos a un mundo en el que la gente prefiera seguir estrictamente cualquier ley o norma a pesar de las circunstancias atenuantes, perdiendo la capacidad de tomar decisiones necesarias.

Con la disponibilidad de cada vez más datos sobre nuestras vidas, es previsible que la información, como las publicaciones y la actividad en las redes sociales, se incluya en los sistemas basados en la IA que informan de las decisiones policiales y judiciales. La Inteligencia Artificial podría utilizarse para identificar el lenguaje o los comportamientos que muestren una propensión a la violencia o un riesgo de cometer determinados tipos de delitos. Este uso implicaría además el derecho a la igualdad ante la ley y a un juicio justo.

### Derechos a la privacidad y a la protección de datos[69]

*"Nadie podrá ser objeto de injerencias arbitrarias o ilegales en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques ilegales a su honra y a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques"*. - Artículo 17 del PIDCP.

*"Toda persona tiene derecho al respeto de su vida privada y familiar, de su domicilio y de sus comunicaciones"*. - Artículo 7 de la Carta de los Derechos Fundamentales de la UE.

---

66 ↩Robert Brauneis and Ellen P. Goodman, "Algorithmic Transparency for the Smart City," The Yale Journal of Law and Technology, Vol. 20 (2018), 103-176.

67 ↩ "Face Value," IRL: Online Life is Real Life, Podcast audio, 4 February 2018.

68 ↩Lauren Goode, "Facial recognition software is biased towards white men, researcher finds," the Verge, 11 Feb 2018.

69 ↩Article 12 of UDHR, Article 17 of ICCPR, Article 8 of the EU Charter of Fundamental Rights

*"Toda persona tiene derecho a la protección de los datos personales que le conciernen. Estos datos deben ser tratados de forma leal, para fines concretos y sobre la base del consentimiento de la persona afectada o de otra base legítima establecida por la ley. Toda persona tiene derecho a acceder a los datos recogidos que le conciernen, así como a su rectificación".* - Artículo 8 de la Carta de Derechos Fundamentales de la UE.

La privacidad es un derecho fundamental que es esencial para la dignidad humana. El derecho a la privacidad también refuerza otros derechos, como el de la libertad de expresión y asociación.[70] Muchos gobiernos y regiones reconocen ahora un derecho fundamental a la protección de datos. La protección de datos consiste principalmente en proteger cualquier dato personal relacionado con usted.[71] Está estrechamente relacionada con el derecho a la privacidad, e incluso puede considerarse parte del derecho a la privacidad dentro del sistema de derechos humanos de la ONU.

Los sistemas de IA suelen entrenarse mediante el acceso y el análisis de grandes conjuntos de datos. También se recopilan datos para crear mecanismos de retroalimentación y permitir la calibración y el perfeccionamiento continuo. Esta recopilación de datos interfiere con el derecho a la privacidad y a la protección de datos. El análisis de datos mediante sistemas de IA puede revelar información privada sobre las personas, información que se califica como información protegida y que debe tratarse como sensible, incluso si se deriva de conjuntos de grandes datos alimentados con información disponible públicamente. Por ejemplo, los investigadores han desarrollado modelos de inteligencia artificial que pueden estimar con precisión la edad, el sexo, la ocupación y el estado civil de una persona sólo a partir de los datos de localización de su teléfono móvil. También son capaces de predecir la ubicación futura de una persona a partir de su historial y de los datos de localización de sus amigos.[72] Para proteger los derechos humanos, esta información debe ser tratada igual que cualquier otro dato personal.

*Se ha comprobado que la recopilación masiva de este tipo viola intrínsecamente los derechos humanos.*

Otro ejemplo de la delgada línea que separa los datos públicos de los privados es el creciente uso de los programas gubernamentales de vigilancia de las redes sociales, en los que las fuerzas del orden recogen una gran cantidad de información de las redes sociales y la transmiten a programas basados en IA para detectar supuestas amenazas. Mientras que las comprobaciones aisladas de las redes sociales públicas de un objetivo pueden parecer a algunos una estrategia policial inteligente, estos programas, en cambio, implican la ingesta masiva e injustificada de todo el ciclo de vida de las redes sociales de una cuenta, grupo de cuentas, o más. Se ha comprobado que la recopilación masiva de este tipo viola intrínsecamente los derechos humanos. Además, si los sistemas utilizados para procesar los datos no son lo suficientemente transparentes o responsables, de modo que no está claro en términos humanos cómo se toman las decisiones, los sistemas violan elementos clave del derecho a la protección de datos.

*De cara al futuro: Los riesgos debidos a la capacidad de la IA para rastrear y analizar nuestras vidas digitales se agravan debido a la enorme cantidad de datos que producimos hoy en día al utilizar Internet. Con el aumento del uso de los dispositivos del Internet de las Cosas (IdC) y los intentos de cambiar hacia las "ciudades inteligentes", las personas pronto crearán un rastro de datos para casi todos los aspectos de sus vidas. Aunque las piezas individuales de estos datos pueden parecer inocuas, cuando se agregan revelan detalles minúsculos sobre nuestras vidas. La IA se utilizará para procesar y analizar todos estos datos para todo, desde la publicidad*

70 ↩"Necessary and Proportionate Principles"

71 ↩Estelle Masse, "Data Protection: why it matters and how to protect it," Access Now, 25 January 25 2018.

72 ↩Steven M. Bellovin, et. al, "When enough is enough: Location tracking, mosaic theory, and machine learning," NYU Journal of Law and Liberty, 8(2) (2014) 555–628.

*microdirigida hasta la optimización del transporte público, pasando por la vigilancia gubernamental de los ciudadanos. En un mundo así, no sólo existen enormes riesgos para la privacidad, sino que la situación plantea la cuestión de si la protección de datos será siquiera posible.*

La vigilancia gubernamental se ha ampliado con el crecimiento de Internet y el desarrollo de nuevas tecnologías, y la IA está permitiendo herramientas de vigilancia más invasoras que nunca. Por ejemplo, aunque todavía no se conoce la existencia de ningún sistema gubernamental de reconocimiento facial totalmente centralizado, el trabajo de China para instalar más cámaras de CCTV en lugares públicos y centralizar sus sistemas de reconocimiento facial muestra que esto podría cambiar pronto. En EUA, la mitad de los adultos ya figuran en las bases de datos de reconocimiento facial de las fuerzas de seguridad.[73] Su uso amenaza con acabar con el anonimato, y el miedo a ser vigilado puede impedir que la gente ejerza otros derechos, como la libertad de asociación. El impacto negativo de la vigilancia impulsada por la IA se dejaría sentir con mayor intensidad en las poblaciones marginadas, que son objetivo desproporcionado de las fuerzas de seguridad.[74] Además, dado que la vigilancia de la población en general las 24 horas del día, los 7 días de la semana, no es necesaria ni proporcionada para el objetivo de la seguridad pública o la prevención de la delincuencia,[75] es casi seguro que violaría el derecho fundamental a la privacidad.

*En EUA, la mitad de los adultos ya figuran en las bases de datos de reconocimiento facial de las fuerzas de seguridad. Su uso amenaza con acabar con el anonimato, y el miedo a ser vigilado puede impedir que la gente ejerza otros derechos, como la libertad de asociación. El impacto negativo de la vigilancia impulsada por la IA se dejaría sentir con mayor intensidad en las poblaciones marginadas, que son objetivo desproporcionado de las fuerzas de seguridad.*

### Derecho a la libertad de circulación

*"Toda persona que se halle legalmente en el territorio de un Estado tiene derecho a circular libremente por él y a escoger libremente su residencia. Toda persona es libre de salir de cualquier país, incluido el suyo. Los derechos antes mencionados no podrán ser objeto de restricciones salvo cuando éstas se hallen previstas en la ley, sean necesarias para proteger la seguridad nacional, el orden público, la salud o la moral públicas o los derechos y libertades de los demás, y sean compatibles con los demás derechos reconocidos en el presente Pacto. Nadie podrá ser privado arbitrariamente del derecho a entrar en su propio país".* - Artículo 12 del PIDCP.

El potencial de la IA para restringir la libertad de movimiento está directamente ligado a su uso para la vigilancia. En los sistemas que combinan datos de imágenes por satélite, cámaras con reconocimiento facial e información de localización de teléfonos móviles, entre otras cosas, la IA puede proporcionar una imagen detallada de los movimientos de las personas, así como predecir su ubicación futura. Por tanto, podría ser utilizada fácilmente por los gobiernos para facilitar una restricción más precisa de la libertad de movimiento, tanto a nivel individual como de grupo.

*Mirando hacia el futuro: En la actualidad, la falta de cartografía formal en muchas comunidades pobres y desatendidas de todo el mundo ha provocado su exclusión de las aplicaciones de cartografía GPS. Dada la creciente tendencia de la IA a la vigilancia policial predictiva, es posible que el aumento de la cartografía de estas*

73 ↩Jordan G. Telcher, "What Do Facial Recognition Technologies Mean for Our Privacy?" The New York Times, 18 July 2018

74 ↩Evan Selinger, "Amazon Needs to Stop Providing Facial REcognition Tech for the Government," Medium, 21 June 2018,

75 ↩Véase "The Necessary and Proportionate Principles," and Privacy International, "Guide to International Law and Surveillance," August 2017.

*zonas y la combinación del uso de esa información con los datos de las aplicaciones de las fuerzas del orden, como las que califican los niveles de delincuencia y la seguridad de los barrios, puedan cerrar efectivamente el turismo o inhibir el movimiento alrededor o dentro de una zona. Incluso si esto se hace por razones legítimas de seguridad pública, se corre el riesgo de violar la libertad de movimiento.*

A medida que el IdC se extienda a los sistemas de infraestructura y transporte, desde las autopistas inteligentes hasta los sistemas de transporte público etiquetados biométricamente, la IA seguirá utilizándose para localizar a las personas en tiempo real, lo que permitirá a los gobiernos restringir aún más la libertad de movimiento. Además, si la IA se utiliza para automatizar las decisiones sobre quién puede viajar -por ejemplo, colocando a las personas en una lista de "No volar" o en otra lista de viajes prohibidos-, los errores podrían dar lugar a que las personas vieran restringida injustamente su libertad de movimiento.

### Derechos a la libertad de expresión, pensamiento, religión, reunión y asociación[76]

*"Toda persona tiene derecho a la libertad de pensamiento, de conciencia y de religión. Este derecho comprende la libertad de tener o de adoptar la religión o las creencias de su elección, así como la libertad de manifestar su religión o sus creencias, individual o colectivamente, tanto en público como en privado, mediante el culto, la celebración de los ritos, las prácticas y la enseñanza. Nadie será objeto de medidas coercitivas que puedan menoscabar su libertad de tener o de adoptar la religión o las creencias de su elección"*. - Artículo 18 del PIDCP y artículo 18 de la DUDH.

*"Toda persona tiene derecho a opinar sin ser molestada. Toda persona tiene derecho a la libertad de expresión; este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección."* - Artículo 19 del PIDCP.

*"Se reconoce el derecho de reunión pacífica. [...] Toda persona tiene derecho a asociarse libremente con otras, incluso el derecho a fundar sindicatos y afiliarse a ellos para la defensa de sus intereses. El ejercicio de este derecho sólo podrá estar sujeto a las restricciones previstas por la ley que sean necesarias en una sociedad democrática, en interés de la seguridad nacional, de la seguridad pública o del orden público, o para proteger la salud o la moral públicas o los derechos y libertades de los demás"*. - Artículos 21 y 22 del PIDCP.

Amenazas directas: Las empresas de Internet que alojan contenidos utilizan la IA para señalar las publicaciones que infringen sus condiciones de servicio. La presión formal e informal de los gobiernos sobre las empresas para que aborden el problema de los supuestos contenidos terroristas, la incitación al odio y las denominadas "noticias falsas", pero sin normas ni definiciones claras, ha llevado a un mayor uso de los sistemas automatizados.[77] Una ley aprobada recientemente en Alemania obliga a los sitios de redes sociales a eliminar una amplia gama de contenidos en un plazo de 24 horas después de haber sido señalados (o hasta siete días en casos menos claros).[78] Como la IA es imperfecta y las empresas se ven presionadas para retirar los contenidos dudosos con tanta rapidez, gran parte de los contenidos se eliminan por error.[79] YouTube eliminó más de 100.000 vídeos que documentaban atrocidades en Siria después de que

[76] ↩Article 19 of the UDHR and Article 19 of the ICCPR; Article 18 of the ICCPR and UDHR, Articles 21 and 22 of the ICCPR, Article 20 of the UDHR

[77] ↩Una encuesta de Freedom House survey halló que 30 de 65 of gobiernos intentaron controlar las discusiones en línea.

[78] ↩ "Germany starts enforcing hate speech law," The BBC, 1 January 2018.

[79] ↩Denis Nolasco and Peter Micek, "Access Now responds to Special Rappoteur Kaye on 'Content Regulation in the Digital Age, Access Now, 11 January 2018

fueran marcados. Estos vídeos suelen ser la única prueba de crímenes horribles y violaciones de los derechos humanos, y la política de YouTube contempla excepciones para los contenidos violentos cuando tienen un importante valor educativo o documental.[80] Empero, fueron retirados.

Los gobiernos autoritarios pueden utilizar una tecnología similar para aumentar la censura. El gobierno chino ya está sustituyendo algunos de sus censores humanos por IA. La popular plataforma de vídeo china iQiyi utiliza AM para identificar el contenido pornográfico y violento, así como el contenido considerado "políticamente sensible". Debido a que el AM no puede manejar los matices, el contenido marcado es revisado actualmente por humanos, aunque esto puede cambiar a medida que la tecnología se vuelve más sofisticada y la industria vea los recursos humanos requeridos para la revisión como un gasto innecesario.[81]

En los países en los que la libertad de religión está amenazada, la IA podría ayudar a los funcionarios del gobierno a vigilar y perseguir a los miembros de los grupos religiosos perseguidos. Esto no sólo podría obligar a estos grupos a mantenerse en secreto por miedo a ser identificados, sino que podría producir consecuencias físicas, desde la violencia hasta la detención o la muerte. La IA también podría utilizarse para identificar y retirar contenidos religiosos. Esto constituiría una violación directa de la libertad de religión si la gente no puede mostrar símbolos religiosos, rezar o enseñar sobre su religión en línea.

*La censura mediante IA puede utilizarse para restringir la libertad de asociación, eliminando grupos, páginas y contenidos que faciliten la organización de reuniones y la colaboración en persona.*

Por último, la censura mediante IA puede utilizarse para restringir la libertad de asociación, eliminando grupos, páginas y contenidos que faciliten la organización de reuniones y la colaboración en persona. Dado el importante papel que desempeñan las redes sociales en la organización de movimientos de protesta a nivel mundial, el uso de la IA podría tener el efecto generalizado de obstaculizar las reuniones en todo el mundo.[82]

*Cuando las personas se sienten vigiladas o carecen de anonimato, se ha demostrado que se autocensuran y modifican su comportamiento. La vigilancia con IA no hace sino agravar este efecto, que tendrá graves repercusiones en la libertad de expresión.*

Amenazas indirectas: Las violaciones del derecho a la privacidad tienen un efecto amedrentador sobre la libertad de expresión. Cuando las personas se sienten vigiladas o carecen de anonimato, se ha demostrado que se autocensuran y modifican su comportamiento. La vigilancia con IA no hace sino agravar este efecto, que tendrá graves repercusiones en la libertad de expresión.[83] Un poderoso ejemplo es el reconocimiento facial. Si se utiliza en los espacios públicos para identificar a los individuos en una protesta, puede tener un importante efecto amedrentador en las reuniones. La implantación de un sistema de este tipo en países que restringen la libertad de reunión impediría de hecho el disfrute de este derecho, ya que muchas personas confían en el nivel de seguridad que proporciona el anonimato para reunirse en público y expresar sus opiniones.

---

80 ↩ Kate O'Flaherty, "YouTube keeps deleting evidence of Syrian chemical weapon attacks," Wired, 26 June 26 2018.

81 ↩ Yuan Tang, "Artificial intelligence takes jobs from Chinese censors," Financial Times, May 21, 2018.

82 ↩ Alex Comninos, "Freedom of Peaceful Assembly and Freedom of Association and the Internet," APC.

83 ↩ Privacy International and Article 19, "Privacy and Freedom of Expression in the Agen of Artificial Intelligence," April 2018.

Otra amenaza indirecta es el impacto de los algoritmos de búsqueda y redes sociales impulsados por la IA. Por ejemplo, el algoritmo de Facebook determina el contenido de las noticias de un usuario e influye en la amplitud de los contenidos y en quién los comparte. El algoritmo de búsqueda de Google indexa el contenido y decide qué aparece en la parte superior de los resultados de búsqueda. Estos algoritmos han desempeñado un papel importante en el establecimiento y el refuerzo de las cámaras de eco, y en última instancia, corren el riesgo de afectar negativamente al pluralismo de los medios de comunicación y a la inhibición de la diversidad de opiniones.[84]

El papel de la IA en la clasificación de contenidos y en la creación y refuerzo de burbujas de filtros supone una amenaza indirecta para la libertad de pensamiento, ya que determina el tipo de información a la que la gente tiene acceso. Aunque las personas suelen tener la posibilidad de acceder a otras fuentes de información o de buscar opiniones diferentes, el tiempo y la atención limitados de los seres humanos hacen que la mayoría no lo haga. Y en los países que carecen de una prensa libre sólida y tienen un acceso limitado a Internet, las plataformas de medios sociales como Facebook son a menudo la única fuente de información no regulada.

*Mirando hacia el futuro: Una amenaza directa que se cierne sobre la libertad de expresión es el acoso en línea mediante bots. Aunque el acoso no es algo nuevo, cada vez lo perpetran más los bots en lugar de los humanos. Estas cuentas de bots se hacen pasar por usuarios reales y envían respuestas automáticas a cuentas identificadas o a cualquiera que comparta una determinada opinión.[85] Este tipo de acoso implacable en línea tiene un efecto escalofriante sobre la libertad de expresión, en particular para las poblaciones marginadas, que son objeto de un ataque desproporcionado.[86] Como los diseñadores de bots emplean cada vez más el procesamiento del lenguaje natural, los bots de acoso seguirán su ejemplo. Esto hará más difícil detectar, denunciar y eliminar las cuentas de bots.*

El poder predictivo de la IA ya se utiliza para predecir y ayudar a prevenir conflictos armados. Este mismo enfoque también podría ser utilizado de forma preventiva por los gobiernos para predecir y evitar manifestaciones o protestas públicas antes de que se produzcan.[87]

### Derecho a la igualdad y a la no discriminación[88]

*"Todas las personas son iguales ante la ley y tienen derecho, sin discriminación alguna, a igual protección de la ley. A este respecto, la ley prohibirá toda discriminación y garantizará a todas las personas protección igual y efectiva contra toda 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."* - Artículo 26 del PIDCP.

*"En los Estados en que existan minorías étnicas, religiosas o lingüísticas, no se negará a las personas que pertenezcan a dichas minorías el derecho que les corresponde, en común con los demás miembros de su grupo, a tener su propia vida cultural, a profesar y practicar su propia religión y a emplear su propio idioma."* - Artículo 27 del PIDCP.

---

84 ↩ Council of Europe, "Algorithms and Human Rights."

85 ↩ Michael Bernstein, "Identifying Harassment Bots on Twitter," Daemo, August 17, 2017, https://www.daemo.org/demo/botcheck

86 ↩ Megan White, "How do you solve a problem like troll armies?" Access Now, 21 April 21, and Constance Grady, "Online harassment threatens free speech. Now there's a field guide to help survive it," Vox, 2 May 2018.

87 ↩ Council of Europe, "Algorithms and Human Rights."

88 ↩ Articles 3, 26 and 27 of the ICCPR. Article 3 of the ICESCR

"Los Estados Partes en el presente Pacto se comprometen a garantizar a hombres y mujeres la igualdad en el goce de todos los [...] derechos enunciados en el presente Pacto." - Artículo 3 del PIDCP y del PIDESC

Los modelos de IA están diseñados para ordenar y filtrar, ya sea clasificando los resultados de las búsquedas o clasificando a las personas en categorías. Esta discriminación puede interferir con los derechos humanos cuando trata de forma diferente a distintos grupos de personas. A veces esta discriminación tiene objetivos sociales positivos, por ejemplo, cuando se utiliza en programas para promover la diversidad. En la justicia penal, esta discriminación suele ser el resultado de formas de sesgo. El uso de la IA en algunos sistemas puede perpetuar la injusticia histórica en todo tipo de casos, desde las sentencias de prisión hasta las solicitudes de préstamos.

*Los algoritmos de anuncios personalizados de Google se basan en la IA y aprenden del comportamiento de los usuarios. Cuanto más haga la gente clic, busque y utilice Internet de forma racista o sexista, más traducirá el algoritmo eso en anuncios. A esto se suman las preferencias discriminatorias de los anunciantes, y se convierte en parte de un círculo vicioso.*

Aunque la gente puede pensar que los anuncios en línea no tienen mucho impacto en sus vidas, la investigación sugiere que el espacio publicitario en línea puede dar lugar a la discriminación y perpetuar los prejuicios históricos. En 2013, la investigadora Latanya Sweeney descubrió que una búsqueda en Google de nombres que suenan estereotípicamente afroamericanos daba lugar a anuncios que sugerían un registro de arresto (como "Trevon Jones, ¿arrestado?") en la gran mayoría de los casos.[89] En 2015, investigadores de Carnegie Mellon descubrieron que Google mostraba muchos menos anuncios de empleos ejecutivos bien pagados a las mujeres. Los algoritmos de anuncios personalizados de Google se basan en la IA y aprenden del comportamiento de los usuarios. Cuanto más haga la gente clic, busque y utilice Internet de forma racista o sexista, más traducirá el algoritmo eso en anuncios. A esto se suman las preferencias discriminatorias de los anunciantes, y se convierte en parte de un círculo vicioso. "La forma en que la gente percibe las cosas afecta a los resultados de las búsquedas, que afectan a la forma en que la gente percibe las cosas".[90]

*Mirando al futuro: Dado que el programa de reconocimiento facial tiene mayores tasas de error para los rostros de piel más oscura, es probable que la identificación errónea afecte de forma desproporcionada a las personas de color. La gravedad del problema queda demostrada por la prueba realizada por la ACLU del programa de reconocimiento facial Rekognition de Amazon. La ACLU cotejó los rostros de los 535 miembros del Congreso de la EUA con 25.000 fotografías públicas de delincuentes utilizando la API de Rekognition con el nivel de confianza predeterminado del 80%. Ningún miembro del Congreso de la EUA está realmente en la base de datos de fotos policiales, y sin embargo hubo 28 coincidencias falsas. De estas coincidencias, el 38% eran personas de color, aunque sólo el 20% de los miembros del Congreso son personas de color.*[91]

Los programas de vigilancia basados en la IA también pueden utilizarse con el propósito expreso de discriminar, permitiendo a los gobiernos identificar, seleccionar y negar servicios a personas de diferentes grupos. En 2017, un controvertido estudio descubrió que un sistema de inteligencia artificial podía adivinar con exactitud si alguien era gay o heterosexual, supuestamente basándose únicamente en las fotos de sus rostros. Otros expertos refutaron enérgicamente

89 ↩ Latanya Sweeney, "Discrimination in Online Ad Delivery," Harvard University, 28 January 2013.

90 ↩ Julia Carpenter, "Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you." The Washington Post, 6 July 2015.

91 ↩ Esto demostró que, al igual que muchos sistemas de reconocimiento facial, Rekognition afectaba de forma desproporcionada a las personas de color. Véase: Russell Brandom, "Amazon's facial recognition matched 28 members of Congress to criminal mugshots," The Verge, 26 July 2018.

los resultados, señalando que hay numerosos indicios no faciales que el AM podría haber detectado en las fotos. Sin embargo, independientemente de la calidad del estudio, el modelo fue capaz de identificar con precisión la sexualidad del 81% de los hombres y del 74% de las mujeres. Los gobiernos podrían utilizar sistemas como éste para identificar y discriminar a las personas LGBTQ en lugares donde la homosexualidad y la no conformidad de género son ilegales o socialmente inaceptables. La cuestionable ciencia que hay detrás del estudio de los rostros, o las altas tasas de error de este tipo de sistemas, pueden no importar a quienes manejan la tecnología.[92]

### Derechos de participación política y autodeterminación[93]

*"Todo ciudadano tendrá el derecho y la oportunidad [...] de participar en la dirección de los asuntos públicos, directamente o por medio de representantes libremente elegidos; de votar y ser elegido en elecciones periódicas, auténticas, realizadas por sufragio universal e igual y por voto secreto que garantice la libre expresión de la voluntad de los electores; de tener acceso, en condiciones generales de igualdad, a las funciones públicas de su país."* - Artículo 25 del PIDCP.

*El papel de IA en la creación y difusión de desinformación desafía la noción de elecciones justas y crea una amenaza para el derecho a la participación política y a la autodeterminación.*

El papel de IA en la creación y difusión de desinformación desafía la noción de elecciones justas y crea una amenaza para el derecho a la participación política y a la autodeterminación. Las elecciones presidenciales de 2016 en EUA demostraron cómo una potencia extranjera puede aprovechar los bots y los algoritmos de las redes sociales para aumentar el alcance de la información falsa e influir potencialmente en los votantes. Aunque las plataformas están trabajando para evitar este tipo de actividad, un futuro de chatbots potenciados por la IA y de falsificaciones profundas probablemente hará que este tipo de contenido sea más convincente para los votantes y más difícil de detectar para las empresas. Esto puede enfriar la participación política, especialmente si los votantes pierden la confianza en la legitimidad de las elecciones.

*De cara al futuro: La vigilancia impulsada por la IA podría utilizarse para restringir e inhibir la participación política, incluso identificando y disuadiendo a ciertos grupos de personas de votar. El uso del reconocimiento facial en los colegios electorales o en las cabinas de votación podría comprometer el secreto del voto. Los gobiernos que deseen disuadir a los votantes de votar por la oposición ni siquiera tienen que vigilar directamente el acto de votar; la mera significación de la vigilancia podría ser suficiente para convencer a los votantes de que sus votos no son secretos, y podría influir en sus decisiones de voto en consecuencia.*

### Prohibición de la propaganda[94]

*"Queda prohibida por ley toda propaganda en favor de la guerra. Toda apología del odio nacional, racial o religioso que constituya una incitación a la discriminación, la hostilidad o la violencia estará prohibida por la ley."* - Artículo 20 del PIDCP.

---

92 ↩ Sam Levin, "New AI can guess whether you're gay or straight from a photograph," The Guardian, 7 September 2017, and Lewis, "Facial recognition can tell if you're gay.".

93 ↩ Article 21 of the UDHR, Article 25 of the ICCPR

94 ↩ Article 20 of the ICCPR

*Mirando hacia el futuro: Al igual que las personas pueden utilizar la tecnología impulsada por la IA para facilitar la difusión de desinformación o influir en el debate público, pueden utilizarla para crear y propagar contenidos diseñados para incitar a la guerra, la discriminación, la hostilidad o la violencia. Podemos ver el potencial de este escenario en las disputas entre Rusia y Ucrania por Crimea. Rusia ha puesto a prueba tácticas de desinformación ya conocidas para intentar socavar la fe del público en los medios de comunicación.[95] Los gobiernos de todo el mundo han desplegado "ejércitos de trolls" para avivar las llamas del conflicto con fines políticos.[96] En un futuro próximo, podrían utilizar chatbots para incitar a la violencia racial y étnica en regiones que ya están plagadas de tensiones, o desplegar falsificaciones profundas para simular que los líderes mundiales declaran la guerra o instigan un conflicto armado.*

### Derecho al trabajo, un nivel de vida digno[97]

*"Los Estados Partes en el presente Pacto reconocen el derecho al trabajo, que comprende el derecho de toda persona a tener la oportunidad de ganarse la vida mediante un trabajo libremente escogido o aceptado, y tomarán las medidas adecuadas para garantizar este derecho. Entre las medidas que habrá de adoptar todo Estado Parte en el presente Pacto para lograr la plena efectividad de este derecho figurarán programas, políticas y técnicas de orientación y formación técnica y profesional a fin de conseguir un desarrollo económico, social y cultural constante y la ocupación plena y productiva en condiciones que garanticen las libertades políticas y económicas fundamentales del individuo."* - Artículo 6 del PIDESC.

*"Los Estados Partes en el presente Pacto reconocen el derecho de toda persona a un nivel de vida adecuado para sí y su familia, incluso alimentación, vestido y vivienda adecuados, y a una mejora continua de las condiciones de existencia."* - Artículo 11 del PIDESC.

Aunque el derecho al trabajo no constituye el derecho absoluto e incondicional a obtener un empleo, sí exige a los Estados que trabajen para lograr el pleno empleo.[98] El papel de la IA en la automatización de los puestos de trabajo podría suponer una amenaza real para el derecho al trabajo, ya que puede impedir que algunas personas accedan al mercado laboral en primer lugar. La automatización ha provocado la pérdida de puestos de trabajo en determinados sectores, y se prevé que la IA acelere esta tendencia. Aunque existe un importante desacuerdo sobre el grado de automatización de los puestos de trabajo, no cabe duda de que la IA provocará algunos cambios en el mercado laboral, tanto por la creación como por la destrucción de puestos de trabajo.[99]

*No cabe duda de que la IA provocará algunos cambios en el mercado laboral, tanto por la creación como por la destrucción de puestos de trabajo.*

*De cara al futuro: Si la automatización desplaza el mercado de trabajo de forma significativa y un gran número de personas no puede encontrar trabajo, tendrán dificultades para mantenerse a sí mismas y a sus familias. Los investigadores están estudiando formas de garantizar que las personas puedan mantener un nivel de vida digno*

---

[95] ↩ Después de las masivas protestas prodemocráticas de 2014 en Ucrania que resultaron en la destitución del presidente prorruso Víktor Yanukóvich, Rusia comenzó una campaña de desinformación masiva para desacreditar al nuevo gobierno y alentar a los separatistas a iniciar la actual guerra civil. Véase: Gregory Warner, "What Americans Can Lean From Fake News in Ukraine," Rough Translation, audio podcast, August 21, 2017.

[96] ↩ White, "troll armies," Access Now.

[97] ↩ Articles 23 and 25 of the UDHR, Articles 6, 7, 11 of the ICESCR

[98] ↩ See http://hrlibrary.umn.edu/gencomm/escgencom18.html

[99] ↩ See https://www.technologyreview.com/s/610005/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/

*con la volatilidad del mercado laboral. Un enfoque es la renta básica universal, un ingreso fijo que proporcionan los gobiernos. Canadá, Finlandia y California están probando sistemas de renta básica, y hay más pruebas previstas en otros países.*[100]

La automatización del trabajo puede plantear una serie de retos que los gobiernos tendrán que abordar para garantizar un nivel de vida digno. En EUA, el gobierno utiliza sistemas automatizados de toma de decisiones en programas para abordar la pobreza, para todo, desde la elegibilidad para la atención médica financiada por el gobierno hasta la asistencia alimentaria.[101] Durante su visita a EUA en 2017, el Relator Especial de la ONU sobre la extrema pobreza y los derechos humanos descubrió que las autoridades municipales de todo el país utilizan sistemas automatizados para poner en contacto a la población sin hogar con los servicios disponibles. Estos sistemas utilizan algoritmos estadísticos deterministas tradicionales para asignar a un encuestado sin hogar una "puntuación de vulnerabilidad" y luego conectar a la persona con las oportunidades de alojamiento adecuadas.[102] La existencia de estos sistemas plantea importantes cuestiones sobre la automatización de estas decisiones cruciales, pero al menos producen resultados rastreables. Sin embargo, si se produce un cambio hacia el uso del AM, la falta inherente de transparencia y explicación del AM podría hacer que las decisiones automatizadas sobre la prestación de un servicio público sean algo que ni el organismo gubernamental encargado de tomar la decisión ni el público comprendan plenamente.

### Derecho a la salud

*"Los Estados Partes en el presente Pacto reconocen el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental. Las medidas que deberán adoptar los Estados Partes en el presente Pacto para lograr la plena efectividad de este derecho comprenderán las necesarias para (a) La reducción de la tasa de mortandad de fetos al nacer e infantil, así como el sano desarrollo del niño; (b) El mejoramiento en todos sus aspectos de la higiene del trabajo y del medio ambiente; (c) La prevención y el tratamiento de las enfermedades epidémicas, endémicas, profesionales y de otra índole, y la lucha contra ellas; (d) La creación de condiciones que aseguren a todos asistencia médica y servicios médicos en caso de enfermedad."* - Artículo 12 del PIDESC.

Algunas de las aplicaciones más prometedoras e impactantes de la IA se encuentran en la atención sanitaria, desde ayudar a los médicos a diagnosticar con mayor precisión las enfermedades, hasta proporcionar recomendaciones de tratamiento más individualizadas a los pacientes, pasando por hacer más accesible el asesoramiento médico especializado. Sin embargo, también hay formas en las que la IA podría poner en peligro el derecho a la salud. Una de ellas es la posibilidad de que los sistemas impulsados por la IA sean discriminatorios o estén programados de forma que los resultados (como la reducción de costes) prevalezcan sobre el bienestar del paciente.

*Un sistema de IA podría estar diseñado para recomendar diferentes tratamientos en función de la situación del seguro del paciente o de su capacidad de pago, lo que podría negar la atención vital a alguien por su situación socioeconómica, perjudicando a los grupos marginados que ya sufren un acceso insuficiente a la atención sanitaria de calidad.*

---

[100] ↩ Leonid Bershidsky, "Finland's Basic Income Test Wasn't Ambitious Enough," Bloomberg Opinion, 26 April 2018,, Chis Weller, "One of the biggest VCs in Silicon Walley is launching an experiment that will gie 3,000 people free money until 2022," Business Insider, September 21, 2017, and Jordan Pearson, "Basic Income Is Already Transforming Life and Work In a Postindustrial Canadian City," Motherboard, 23 April 2018.

[101] ↩ Eubanks, "The Digital Poorhouse."

[102] ↩ "Statement on Visit to the USA, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights," Office of the High Commissioner for Human Rights, United Nations, December 15, 2017.

Por ejemplo, un sistema de IA podría estar diseñado para recomendar diferentes tratamientos en función de la situación del seguro del paciente o de su capacidad de pago, lo que podría negar la atención vital a alguien por su situación socioeconómica, perjudicando a los grupos marginados que ya sufren un acceso insuficiente a la atención sanitaria de calidad. Otro problema potencial son los bucles de retroalimentación negativos que podrían resultar de la excesiva dependencia de la orientación de un sistema de IA. Por ejemplo, si los médicos tienden a retirar la atención a los pacientes con determinados diagnósticos, como un parto prematuro extremo o lesiones cerebrales graves, un sistema basado en AM puede aprender que esos diagnósticos son casi siempre mortales y recomendar al médico que no los trate, aunque en algunos casos el tratamiento pueda ser eficaz.[103]

*¿qué tipo de responsabilidad hay en una decisión médica de vida o muerte tomada por una máquina frente a un médico?*

Y, por supuesto, está el impacto de las inevitables tasas de error de cualquier sistema. Aunque, por ejemplo, el "Watson" de IBM sea más preciso que los médicos humanos a la hora de diagnosticar enfermedades, seguirá equivocándose de diagnóstico en alguna ocasión, o recomendando un tratamiento erróneo. En este caso, ¿qué tipo de responsabilidad hay en una decisión médica de vida o muerte tomada por una máquina frente a un médico? El mismo problema podría plantearse en los sistemas de IA que predicen brotes de enfermedades y recomiendan respuestas. ¿Qué ocurre cuando se despliegan recursos en una zona considerada de alto riesgo y se deja a otros sin asistencia? Los trabajadores sanitarios ya toman esta decisión, pero la IA lo haría de forma preventiva, y a veces podría equivocarse. Esto plantea cuestiones más amplias sobre hasta qué punto deben automatizarse ciertas cosas, cómo y cuándo requerir un "humano en el bucle", y cuánta responsabilidad deben tener los médicos humanos frente a los sistemas de IA para hacer las recomendaciones.

*De cara al futuro: Otra preocupación se refiere al uso de la IA para determinar quién tiene acceso a la atención sanitaria y lo que paga por el seguro médico. Existe el peligro de que los proveedores de seguros sanitarios utilicen la IA para elaborar perfiles basados en determinados comportamientos e historial. Un sistema de IA podría utilizar puntos de datos sobre usted para recomendar tarifas de seguro médico individualizadas. Podría ver que usted tiene antecedentes de enfermedad en la familia, que no es físicamente activo, que le gusta comer en restaurantes de comida rápida y que fuma, y recomendar que se le apliquen tarifas más altas en función de estos factores.*

### Derecho a la educación[104]

*"Los Estados Partes en el presente Pacto reconocen el derecho de toda persona a la educación. [...] Los Estados Partes en el presente Pacto reconocen que, para lograr la plena efectividad de este derecho (a) La enseñanza primaria debe ser obligatoria y asequible a todos gratuitamente; b) La enseñanza secundaria, en sus diferentes formas, incluso la enseñanza secundaria técnica y profesional, debe ser generalizada y hacerse accesible a todos, por cuantos medios sean apropiados, y en particular por la implantación progresiva de la enseñanza gratuita; c) La enseñanza superior debe hacerse igualmente accesible a todos, sobre la base de la capacidad de cada uno, por cuantos medios sean apropiados, y en particular por la implantación progresiva de la enseñanza gratuita; (d) Se fomentará o intensificará, en la medida de lo posible, la enseñanza fundamental para aquellas personas que no hayan recibido o completado la totalidad de su educación primaria; (e) Se proseguirá activamente el desarrollo de un sistema*

*La IA puede violar fundamentalmente el principio de igualdad de acceso.*

[103] ↩ Patricia Hannon, "Researchers say use of artificial intelligence in medicine raises ethical questions," Stanford Medicine News Center, 14 March 2018,.

[104] ↩ Article 25 of the UDHR. Article 13 and 14 of the ICESCR

*de escuelas en todos los niveles, se establecerá un sistema adecuado de becas y se mejorarán continuamente las condiciones materiales del personal docente."* - Artículo 13 del PIDESC.

La IA puede violar fundamentalmente el principio de igualdad de acceso. Las universidades de EUA utilizan sistemas algorítmicos deterministas para recomendar a los solicitantes que deben admitir. Estos sistemas suelen estar hechos a medida para satisfacer las preferencias de la universidad, y presentan una serie de problemas que pueden dar lugar a la discriminación, como el uso de datos históricos de estudiantes previamente admitidos para informar al modelo. Dado que muchas universidades de élite han sido históricamente frecuentadas por hombres blancos ricos, cualquier modelo que utilice estos datos corre el riesgo de perpetuar las tendencias del pasado.[105] Es probable que estos sistemas empleen AM en el futuro, lo que haría más difícil detectar el sesgo. Esto podría dar lugar a que las universidades discriminen bajo la apariencia de objetividad.

*De cara al futuro: Si la IA se utiliza para seguir y predecir el rendimiento de los estudiantes de forma que se limite la posibilidad de estudiar ciertas asignaturas o de acceder a determinadas oportunidades educativas, se pondrá en riesgo el derecho a la educación. Dado el aumento de las investigaciones sobre los factores de predicción del éxito en la primera infancia, es probable que un sistema de este tipo pueda utilizarse para restringir las oportunidades de los estudiantes a edades cada vez más tempranas, lo que daría lugar a una importante discriminación, ya que a los estudiantes procedentes de entornos desfavorecidos se les acabarían negando las oportunidades porque las personas de ese origen tienden a tener resultados más negativos. Un sistema de este tipo ignoraría a los estudiantes que se sobreponen a la adversidad para alcanzar el éxito académico y profesional, y afianzaría las desigualdades educativas existentes.*

### Derecho a participar en la vida cultural y a gozar de los beneficios del progreso científico[106]

*"Los Estados Partes en el presente Pacto reconocen el derecho de toda persona (a) Participar en la vida cultural; (b) Gozar de los beneficios del progreso científico y de sus aplicaciones; (c) Beneficiarse de la protección de los intereses morales y materiales que le correspondan por razón de las producciones científicas, literarias o artísticas de que sea autora."* - Artículo 15 del PIDESC.

*Existe el riesgo de que la IA se utilice para "criminalizar" determinadas culturas.*

El uso de tecnologías de IA que permiten a los gobiernos identificar y reprimir a los grupos culturales podría impedir que las personas participen en la vida cultural, ya sea directa o indirectamente (por ejemplo, a través de la vigilancia que inspira el miedo a ser identificado o a sufrir represalias por la identidad cultural, lo que lleva a las personas a evitar por completo las expresiones culturales). Existe el riesgo de que la IA se utilice para "criminalizar" determinadas culturas. Cuando los miembros de una determinada cultura son detenidos de forma desproporcionada o son objeto de otro tipo de acciones por parte de las fuerzas de seguridad, los comportamientos y costumbres asociados a estas culturas podrían vincularse con actividades delictivas. Por ejemplo, un sistema de inteligencia artificial que analice imágenes de vídeo o fotográficas podría aprender a asociar determinados tipos de vestimenta, formas de hablar o gestos con actividades delictivas, y podría utilizarse para justificar la persecución de estos grupos con el pretexto de prevenir la delincuencia.

[105] ↩ O'Neil, Weapons of Math Destruction, 50-67.

[106] ↩ Article 27 of the UDHR, Article 15 of the ICESCR

A muchos de los países en vías de desarrollo les preocupa quedarse atrás en la carrera mundial de la IA y el correspondiente cambio económico transformador. Pero los habitantes de los países en desarrollo se convertirán en consumidores pasivos de sistemas de IA desarrollados en China u Occidente para personas, culturas y situaciones diferentes. La IA desarrollada en el extranjero corre el riesgo de profundizar la desigualdad y la división social existentes en lugares donde el acceso a Internet y a la tecnología está restringido en gran medida a los ricos y a los urbanos. Este riesgo de profundizar la desigualdad se ve agravado por el riesgo de que la automatización del trabajo provoque la pérdida de puestos de trabajo al desplazar el papel de la industria manufacturera en el desarrollo económico.

### Derecho al matrimonio, derechos de los niños y derechos de la familia[107]

*"La familia es el núcleo natural y fundamental de la sociedad y tiene derecho a la protección de la sociedad y del Estado. Se reconoce el derecho del hombre y de la mujer en edad de contraer matrimonio a casarse y a fundar una familia. No se celebrará ningún matrimonio sin el libre y pleno consentimiento de los futuros cónyuges"*. - Artículo 23 del PIDCP.

*"Todo niño tiene derecho, sin discriminación alguna por motivos de raza, color, sexo, idioma, religión, origen nacional o social, posición económica o nacimiento, a las medidas de protección que su condición de menor requiere por parte de su familia, de la sociedad y del Estado."* - Artículo 24 del PIDCP.

*Mirando hacia el futuro: Si la tecnología de IA se utiliza para la detección de la salud y la reproducción, y se descubre que algunas personas tienen pocas probabilidades de tener hijos, la detección podría impedir que se casen, o que se casen con una determinada persona si se considera que la pareja tiene pocas probabilidades de concebir. Del mismo modo, las pruebas genéticas y de ADN impulsadas por la IA podrían utilizarse para producir niños con las cualidades deseadas.*

## ROBÓTICA Y LA IA

El uso de la IA en la robótica representa un pequeño porcentaje del uso de la IA en la actualidad. Sin embargo, la robótica es un campo en crecimiento y los robots desempeñarán un papel cada vez más importante en nuestras vidas. En muchos casos, un robot simplemente proporciona el cuerpo físico para los tipos de sistemas de IA explorados en este informe. Sin embargo, este cuerpo físico y el contexto en el que se utilizan los robots dotados de IA pueden plantear nuevos retos.[108]

### Derecho a la vida

En muchos países se están desarrollando sistemas de armas totalmente autónomos. El creciente uso de drones y armamento similar significa que es probable que las armas autónomas sean accesibles a los actores no estatales que no están sujetos a las leyes tradicionales de los conflictos armados. Es probable que las armas autónomas del futuro próximo adolezcan de la incapacidad de la IA para hacer frente a los matices o a los imprevistos. En una situación de conflicto, esto podría provocar la muerte o las lesiones de civiles inocentes que un operador humano podría haber evitado.[109]

---

107 ↩ Article 16 of the UDHR, Articles 23 and 24 of the ICCPR, Article 10 of the ICESC

108 ↩ Un punto de partida útil para pensar en los límites deseados de los robots en la sociedad son las tres leyes de la robótica del autor de ciencia ficción Isaac Asimov.# Las dos primeras leyes tienen especial relevancia para los derechos humanos: 1) Un robot no puede herir a un ser humano o, por inacción, permitir que un ser humano sufra daños; 2) Un robot debe obedecer las órdenes que le den los seres humanos, excepto cuando dichas órdenes entren en conflicto con la Primera Ley. A continuación exploramos algunas amenazas potenciales de los robots impulsados por IA para los derechos humanos.

109 ↩ OMEST, "Report of COMEST on Robotics Ethics."

Otra amenaza para el derecho a la vida podría surgir del uso de robots con IA en la atención sanitaria. Actualmente se utilizan robots para ayudar en la cirugía, y es fácil imaginar la existencia de robots quirúrgicos totalmente autónomos en un futuro próximo, al igual que los robots que se utilizan para la terapia de rehabilitación y los entornos de atención general. Es inevitable que los robots se equivoquen. ¿Qué ocurre cuando lo hacen? ¿Y quién es el responsable?[110] Además, si los malos actores interfieren con los robots sanitarios y éstos causan daños físicos, las vías para remediar o reparar el daño están lejos de estar establecidas.

### Derecho a la privacidad

Los drones de vigilancia u otros robots han sido utilizados durante mucho tiempo por los militares, y ahora son utilizados cada vez más por las fuerzas del orden o por agentes no estatales. Cuando están equipados con tecnología basada en la IA, como la tecnología de reconocimiento facial, y se hacen semiautónomos o totalmente autónomos -por ejemplo, utilizados para seguir a un determinado grupo o a una persona de forma independiente-, estos drones podrían profundizar en el impacto de la vigilancia generalizada e invasiva que viola los principios "necesarios y proporcionados" que rigen la vigilancia estatal.

### Derecho al trabajo

Los robots impulsados por la IA pueden permitir la automatización de los puestos de trabajo y, por lo tanto, pueden amenazar el derecho al trabajo de las formas que hemos analizado anteriormente.

### Derecho a la educación

Aunque todavía está en fase inicial, el uso de la robótica en la educación es un área de investigación activa. Esto incluye robots utilizados para tareas como la enseñanza de segundas lenguas en las escuelas primarias, y para la narración de cuentos.[111] Al igual que ocurre con la IA en general, los riesgos que plantean los robots impulsados por la IA tienen que ver con resultados que violan la igualdad de acceso. Por ejemplo, en las zonas en las que los robots sustituyen a los profesores en las escuelas, los alumnos recibirían un tipo de educación diferente al de los que tienen profesores humanos, y eso puede constituir una violación de la igualdad de acceso.

## Recomendaciones: Cómo abordar los daños a los derechos humanos relacionados con la IA

Actuar con rapidez ahora para hacer frente a los riesgos para los derechos humanos puede ayudar a prevenir los impactos perjudiciales previsibles de la IA, al tiempo que proporciona un espacio y un marco para abordar los problemas que no podemos predecir. Dado que la IA es un campo tan amplio y diverso, cualquier enfoque tendrá que ser hasta cierto punto específico para cada sector. Sin embargo, hay cuatro enfoques políticos generales que podrían abordar muchos de los riesgos para los derechos humanos que plantea la IA.

1. Una legislación exhaustiva de protección de datos puede anticipar y mitigar muchos de los riesgos para los derechos humanos que plantea la IA. Sin embargo, al ser específica para los datos, también son necesarias medidas adicionales.

110 ↩ ibid.

111 ↩ ibid.

2. El uso de la IA por parte de los gobiernos debe regirse por un estándar elevado, que incluya normas de contratación pública, evaluaciones de impacto sobre los derechos humanos, plena transparencia y procesos de explicación y rendición de cuentas.
3. Dado el deber del sector privado de respetar y defender los derechos humanos, las empresas deberían ir más allá del establecimiento de políticas éticas internas y desarrollar procesos de transparencia, explicación y rendición de cuentas.
4. Debería investigarse mucho más sobre los posibles daños a los derechos humanos de los sistemas de IA y debería invertirse en la creación de estructuras para responder a estos riesgos.

## EL PAPEL DE LAS LEYES INTEGRALES DE PROTECCIÓN DE DATOS

Las leyes integrales de protección de datos, que deberían aplicarse tanto al gobierno como al sector privado, pueden contribuir en gran medida a abordar muchos de los riesgos para los derechos humanos que plantea la IA. Dado que los datos son el motor de la IA, cualquier ley que ordene la protección de los datos personales implicará necesariamente a los sistemas de IA. Dada la presión global hacia la legislación de protección de datos, esto es alentador y práctico.

Consideremos el impacto del Reglamento General de Protección de Datos (RGPD) de la Unión Europea. El RGPD es un marco positivo que prevé el control de la información personal de una persona y faculta a las personas a tomar decisiones informadas sobre el uso de sus datos. El RGPD limita el tratamiento de datos a los fines permitidos, con una mayor protección de los datos sensibles. También exige el consentimiento previo,[112] lo que limita el uso de datos personales para el entrenamiento de sistemas de IA.

Los derechos previstos en el RGPD, y otras leyes similares, ofrecen un marco para evitar los usos no responsables de la IA que afectan a los derechos individuales, al tiempo que garantizan un nivel de control de los datos personales y la responsabilidad por el uso de los sistemas de IA y AM.

Algunos han sugerido que las leyes de protección de datos son incompatibles con la IA y que deberíamos hacer amplias excepciones para su desarrollo y uso. Esto es un error. Si bien es cierto que las leyes de protección de datos pueden impedir el despliegue de ciertos sistemas de IA, las empresas nunca han podido "innovar" sin tener en cuenta el daño potencial. Si los sistemas de IA se utilizan para tomar decisiones sobre una base o fundamento que ni siquiera sus desarrolladores pueden explicar completamente, los individuos en riesgo -o los "conejillos de indias" de la IA- serán los primeros en sufrir las consecuencias negativas. Los derechos de protección de datos no sólo proporcionan estructuras de responsabilidad para mitigar el daño, sino que también protegen a las personas contra la cooptación encubierta de sus datos personales, su mercantilización y cualquier otra explotación que perjudique a otros o a la sociedad en general.

*Los derechos de protección de datos no sólo proporcionan estructuras de responsabilidad para mitigar el daño, sino que también protegen a las personas contra la cooptación encubierta de sus datos personales, su mercantilización y cualquier otra explotación que perjudique a otros o a la sociedad en general.*

[112] ↩ https://gdpr-info.eu/art-9-gdpr/

### Innovación e IA

Algunas industrias han empezado a tomar nota de estas cuestiones y están desarrollando marcos voluntarios para el uso de la IA.[113] Sin embargo, la historia demuestra que la autorregulación de la industria puede ser lamentablemente inadecuada para proteger a las personas, especialmente a las de las comunidades marginadas que con frecuencia son objeto de campañas de manipulación. A medida que la IA aumenta su sofisticación, la sociedad no puede permitirse sacrificar los derechos individuales en aras de la innovación. En cambio, en las jurisdicciones que no cuentan con leyes de protección de datos, los funcionarios públicos deberían adoptar medidas neutrales desde el punto de vista tecnológico para abordar el impacto de la revolución de la IA en los derechos humanos. Del mismo modo, en las zonas con leyes ya en vigor, los supervisores y los organismos de control deben garantizar que la ley se cumpla y siga siendo pertinente a medida que avanza la tecnología de la IA.

*La historia demuestra que la autorregulación de la industria puede ser lamentablemente inadecuada para proteger a las personas, especialmente a las de las comunidades marginadas que con frecuencia son objeto de campañas de manipulación.*

### Derechos de protección de datos y IA[114]

El derecho a la información y el derecho de acceso funcionan conjuntamente para permitir que las personas obtengan información sobre los datos que recopila una entidad, cómo los recopila, cómo los utilizará y si los datos se utilizarán para la toma de decisiones automatizada. Estos derechos aumentan la conciencia pública sobre la existencia de los sistemas de IA y el papel que desempeñan.[115] Además, estos derechos permiten que la gente descubra y comprenda los posibles daños a los derechos humanos y empujan a las entidades a ser más transparentes sobre cómo utilizan la IA.

El Derecho de Rectificación permite a las personas enmendar y modificar su información en poder de un tercero si es incorrecta, incompleta o inexacta. Este derecho puede ayudar a mitigar el impacto de los índices de error en los sistemas de IA.

El Derecho a Restringir el Procesamiento da a las personas la posibilidad de solicitar que una entidad deje de utilizar o limite el uso de la información personal, mientras que el derecho a la supresión ofrece una vía para eliminar los datos personales de una persona en poder de una entidad tercera cuando ya no es necesario, la información se ha utilizado de forma indebida o la relación entre el usuario y la entidad ha terminado. Estos derechos podrían utilizarse para detener temporalmente el uso de un sistema de IA cuestionado, o para presionar a una entidad para que utilice un sistema de IA de forma más responsable.

*El derecho de objeción ofrece a las personas la posibilidad de impugnar la mayor parte del tratamiento de sus datos personales por parte de una entidad cuando los datos se utilizan para el marketing directo, la toma de decisiones automatizada la investigación y las estadísticas, o para el "interés legítimo" de una entidad.*

---

113 ↩ Véase e.g. Rian Wanstreet: America's Farmers Are Becoming Prisoners to Agriculture's Technological Revolution — Vice, 8 March 2018. ("La Oficina Agrícola estadounidense ayudó a elaborar los "Principios de privacidad y seguridad de los datos agrícolas", que abordan cuestiones relativas a la propiedad, la portabilidad, el uso y el intercambio de datos. Empresas como Deere y Monsanto fueron las primeras en firmarlos, pero siguen existiendo dudas sobre el grado de protección de estos principios en la práctica").

114 ↩ https://www.accessnow.org/cms/assets/uploads/2018/07/GDPR-User-Guide_digital.pdf

115 ↩ En la EUA, muchos de los detalles del uso gubernamental del sistema de toma de decisiones algorítmicas se ocultan tras acuerdos de no divulgación y memorandos de entendimiento con los proveedores. Véase Tom Simonite: When Government Rules by Software, Citizens Are Left in the Dark — Wired, 17 August 2017.

El Derecho a Una Explicación permite a una persona obtener una explicación sobre cómo se toma una decisión automatizada que le concierne. Este derecho garantiza que las entidades entiendan cómo funcionan realmente los sistemas que utilizan, y empuja a los desarrolladores de IA a seguir trabajando para que la IA sea comprensible.

El Derecho de Objeción ofrece a las personas la posibilidad de impugnar la mayor parte del tratamiento de sus datos personales por parte de una entidad cuando los datos se utilizan para el marketing directo, la toma de decisiones automatizada (en la que no hay intervención humana), la investigación y las estadísticas, o para el "interés legítimo" de una entidad. Este derecho permite impugnar directamente las decisiones tomadas con sistemas de IA. Es especialmente importante para el uso gubernamental de la IA en formas que pueden ser discriminatorias. También garantiza que haya un humano en el bucle de los sistemas importantes de toma de decisiones automatizadas, lo que añade una capa de responsabilidad.

## RECOMENDACIONES ESPECÍFICAS DE LA IA PARA EL GOBIERNO Y EL SECTOR PRIVADO

Dado que la IA es un campo diverso, el potencial de interferencia con los derechos humanos depende tanto del tipo de datos que utiliza un sistema como del contexto en el que se aplica. Por ejemplo, el uso de la IA por parte del gobierno de una ciudad para optimizar el uso del agua presenta menos y diferentes riesgos para los derechos humanos que el uso por parte de un departamento de policía de una herramienta de evaluación del riesgo criminal. Teniendo esto en cuenta, recomendamos enfoques diferentes para el gobierno y el sector privado.

### Recomendaciones para el uso de la IA por parte de las administraciones públicas

Los sistemas de IA para la administración pública suelen implicar juicios de valor que están necesariamente vinculados al proceso político en los sistemas de gobierno libres y democráticos. Por esta razón, y porque los gobiernos pueden privar directamente a las personas de su libertad, este informe recomienda normas más estrictas para el sector público en lo que respecta al uso de la IA. Los Estados tienen el deber primordial de promover, proteger, respetar y cumplir los derechos humanos según el derecho internacional, y no deben participar ni apoyar prácticas que violen los derechos, ya sea en el diseño o en la aplicación de los sistemas de IA. Están obligados a proteger a las personas contra los abusos de los derechos humanos, así como a adoptar medidas positivas para facilitar el disfrute de los derechos.[116] Las recomendaciones que se exponen a continuación articulan un marco para la toma de decisiones de los gobiernos en general, no sólo en materia de IA. Se aplican a cualquier tipo de sistema algorítmico de toma de decisiones, independientemente de que utilice IA.

**1. Seguir las normas de contratación pública.** Cuando un organismo público desee adquirir un sistema de IA o sus componentes, la contratación debe hacerse de forma abierta y transparente, de acuerdo con las normas de contratación pública. Esto incluye la publicación del propósito del sistema, los objetivos, los parámetros y otra información para facilitar la comprensión del público. Las adquisiciones deben incluir un periodo de comentarios públicos, y los Estados deben ponerse en contacto con los grupos potencialmente afectados cuando sea pertinente para garantizar la oportunidad de hacer aportaciones.

**2. Ordenar evaluaciones de impacto sobre los derechos humanos.** Los Estados deben investigar a fondo los sistemas de IA para identificar los riesgos para los derechos humanos antes de su desarrollo o adquisición, así como de forma regular y continua a lo largo del ciclo de vida del sistema. Una evaluación del impacto sobre los derechos humanos puede ser una parte necesaria de un proceso más amplio de evaluación del impacto algorítmico que examine amenazas

116 ↩ Véase un resumen de los derechos humanos de la ONU state' human rights obligations under international law.

más amplias, incluidas las amenazas que plantea el uso de la IA para llevar a cabo la vigilancia u otras actividades que interfieren con los derechos humanos.[117] Deben existir leyes adecuadas que regulen los usos de la IA para estos fines.[118] Todo proceso de evaluación debe incluir:

- Pruebas y auditorías realizadas por expertos independientes
- La identificación de medidas para mitigar los riesgos identificados y evitar que se produzcan violaciones de derechos, y la medición del cumplimiento y la eficacia
- Un mecanismo de seguridad para poner fin a la adquisición, el despliegue o cualquier uso continuado si en algún momento una violación de los derechos humanos identificada es demasiado elevada o no puede mitigarse
- Identificación de cualquier nueva salvaguarda legal necesaria para proteger los derechos humanos en aplicaciones específicas de las herramientas de IA
- Determinaciones especiales de parcialidad, particularmente en el sector de la justicia penal debido a los riesgos para un juicio justo, el derecho a la libertad y la no discriminación
- Si se recurre a un tercero para desarrollar y/o aplicar el sistema, un requisito para que el tercero participe en el proceso de evaluación de los derechos humanos

**3. Garantizar la transparencia y la explicación.** Es necesaria la máxima transparencia posible para cualquier sistema de IA, incluida la transparencia respecto a su propósito, cómo se utiliza y cómo funciona, que debe continuar durante todo el ciclo de vida del sistema. Los acuerdos de no divulgación y otros contratos con terceros con el pretexto de proteger la propiedad intelectual son una violación de este principio porque impiden la supervisión pública y la rendición de cuentas. En concreto, la transparencia y la explicación adecuadas deben incluir:

- Información periódica sobre dónde y cómo los gobiernos utilizan y gestionan los sistemas de IA
- El uso de estándares de datos abiertos tanto en los datos de formación como en el código en la mayor medida posible, respetando al mismo tiempo las normas de privacidad[119]
- Permitir auditorías independientes de los sistemas y los datos
- Informes claros y accesibles sobre el funcionamiento de cualquier sistema de IA. Esto significa proporcionar información relevante sobre cómo se obtienen los resultados y qué medidas se adoptan para minimizar los efectos perjudiciales para los derechos.
- Notificación específica cuando un sistema de IA gubernamental toma una decisión que afecta a los derechos de una persona
- Evitar los "sistemas de caja negra", es decir, evitar cualquier sistema de IA cuando una persona no puede entender de forma significativa su funcionamiento

**4. Establecer la responsabilidad y los procedimientos de reparación.** El uso de un sistema de IA para realizar una tarea que antes realizaba un ser humano no elimina los requisitos estándar de responsabilidad y rendición de cuentas en los procesos de toma de decisiones del gobierno; siempre debe haber un ser humano en el bucle, y para las áreas de alto riesgo, incluida la justicia penal, es necesaria una supervisión humana significativa. Los gobiernos deben establecer políticas relativas a la automatización de los procesos, teniendo en cuenta las repercusiones sobre los derechos

---

117 ↩ El IA Now Institute ha esbozado un marco práctico para las evaluaciones de impacto de los algoritmos por parte de los organismos públicos, https://ainowinstitute.org/aiareport2018.pdf. El artículo 35 del Reglamento General de Protección de Datos (RGPD) de la UE establece la obligación de realizar una evaluación de impacto de la protección de datos (EIPD); además, el artículo 25 del RGPD exige que los principios de protección de datos se apliquen por diseño y por defecto desde la fase de concepción de un producto, servicio o prestación y a lo largo de su ciclo de vida.

118 ↩ Véase, e.g., International Principles on the Application of Human Rights to Communications Surveillance, last accessed June 15 2018, available at https://necessaryandproportionate.org/.

119 ↩ Véase https://www.opengovpartnership.org/sites/default/files/open-gov-guide_summary_all-topics.pdf for more information on open data standards for government data

humanos. Además, las personas deben tener derecho a impugnar el uso de un sistema de IA o a recurrir una decisión informada o tomada en su totalidad por un sistema de IA. Más concretamente, la rendición de cuentas y el recurso requieren:

- Una formación adecuada para los operadores de un sistema de IA. Los empleados de la administración que utilizan y gestionan un sistema de IA deben entender cómo funciona, los límites de su uso y el potencial de daño. Una formación adecuada garantiza que los humanos permanezcan en el bucle de manera importante y aumenta la probabilidad de detectar resultados perjudiciales.
- Establecer la responsabilidad de los resultados de un sistema de IA. Aunque los Estados suelen recurrir a terceros para diseñar y aplicar sistemas de IA, la responsabilidad última de las interferencias en los derechos humanos debe recaer en los Estados. Para protegerse contra los abusos, las entidades gubernamentales deben adquirir los conocimientos técnicos necesarios para examinar a fondo un determinado sistema.
- Establecer mecanismos para apelar cualquier uso o determinación específica de un sistema de IA. Incluso los sistemas adquiridos y aplicados de forma transparente y con la participación de las partes interesadas pueden seguir corriendo el riesgo de que se produzcan importantes violaciones de los derechos humanos.[120] Para ello, debe existir un proceso que permita al público impugnar el uso de un sistema de IA en su totalidad.

### Recomendaciones Para el Uso de la IA por parte del Sector Privado y las Entidades No Estatales

Los actores del sector privado también tienen la responsabilidad de respetar los derechos humanos, independientemente de las obligaciones del Estado. Para cumplir con su deber, los actores del sector privado deben tomar medidas constantes para garantizar que no causan ni contribuyen a los abusos de los derechos humanos.[121] El establecimiento de políticas éticas en materia de IA por parte de muchos de los grandes actores del sector privado es loable, pero debería integrarse una evaluación del impacto sobre los derechos humanos en los procesos más amplios de revisión ética. Además, los actores del sector privado deberían aplicar medidas de transparencia y explicación, así como establecer procedimientos para garantizar la rendición de cuentas y el acceso a la reparación. En conjunto, esta diligencia debida en materia de derechos humanos, informada por expertos interesados, ayuda a las empresas a prevenir y mitigar los abusos. Sin embargo, las empresas también deben cumplir su obligación de reparar los daños directa o indirectamente resultantes de sus operaciones, a través de procesos respetuosos con los derechos desarrollados en consulta con las comunidades afectadas. Reconocemos que no todos los usos de la IA presentan el mismo riesgo de daños a los derechos humanos, y que las acciones necesarias para prevenir y responder a las violaciones de los derechos humanos dependerán del contexto. En concreto, los actores del sector privado deberían:

**1. Llevar a cabo la debida diligencia en materia de derechos humanos según los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos, y que consiste en los siguientes tres pasos fundamentales:**[122]

a. Identificar los resultados potencialmente adversos para los derechos humanos. Los actores del sector privado deben evaluar los riesgos de que un sistema de IA pueda causar o contribuir a violaciones de los derechos humanos. Para ello, los agentes deben:
   - Identificar los daños directos e indirectos, así como los daños emocionales, sociales, medioambientales u otros daños no financieros.
   - Consultar con las partes interesadas de forma incluyente, especialmente con los grupos afectados, las organizaciones de derechos humanos y los expertos independientes en derechos humanos y en IA. Si el sistema

120 ↩ El estado de Pensilvania se ha esforzado por aplicar la transparencia algorítmica en el uso de los sistemas automatizados de toma de decisiones. Sin embargo, el proceso de comentarios públicos y las normas de datos abiertos no han impedido que se utilicen sistemas problemáticos.

121 ↩ Véase UN Guiding Principles on Business and Human Rights

122 ↩ Adaptado de Toronto Declaration

está destinado a ser utilizado por una entidad gubernamental, tanto los actores públicos como los privados deben realizar una evaluación.

b. Adoptar medidas eficaces para prevenir y mitigar los daños, así como hacer un seguimiento de las respuestas. Una vez identificados los riesgos para los derechos humanos, los agentes del sector privado deben mitigarlos y hacer un seguimiento a lo largo del tiempo. Esto requiere que los actores del sector privado:
   - Corrijan el sistema, incluyendo dónde se encuentran los riesgos con los datos de formación, el diseño del modelo o el impacto del sistema.
   - Garanticen la diversidad y la inclusión de los conocimientos pertinentes para evitar el sesgo por diseño y los daños involuntarios.
   - Sometan los sistemas de IA con un riesgo significativo de abusos de los derechos humanos a auditorías de terceros independientes.
   - Detengan el despliegue de cualquier sistema de IA en un contexto en el que el riesgo de violación de los derechos humanos sea demasiado alto o imposible de mitigar.
   - Realicen un seguimiento de las medidas adoptadas para mitigar los daños a los derechos humanos y evalúen su eficacia. Esto incluye controles de calidad y auditorías periódicas a lo largo del ciclo de vida del sistema. Esto es especialmente importante dado el papel de los bucles de retroalimentación negativos que pueden exacerbar los resultados perjudiciales.

c. Ser transparente sobre los esfuerzos para identificar, prevenir y mitigar los daños en los sistemas de IA. La transparencia hacia todas las personas y grupos afectados, así como hacia otras partes interesadas, es una parte fundamental de la diligencia debida en materia de derechos humanos, y conlleva la comunicación.[123] En la práctica, esto significa que los actores del sector privado deben:
   - Divulgar públicamente información sobre los riesgos identificados en materia de derechos humanos, incluyendo tanto la forma en que está diseñado el sistema como el contexto en el que se utiliza.
   - Publicar detalles técnicos sobre el sistema de IA, incluyendo muestras de datos de entrenamiento y detalles sobre las fuentes de los datos.

**2. Proporcionar transparencia y explicación en la medida de lo posible.** Los actores del sector privado deben ser muy transparentes y proporcionar información significativa sobre el funcionamiento de los sistemas de IA. La transparencia es especialmente importante cuando los sistemas de IA pueden tener un impacto público o personal significativo, por ejemplo en medicina o en la recomendación y moderación de contenidos. En concreto, esto incluye:

- Adhesión a estándares de código abierto y datos abiertos.
- La publicación de explicaciones significativas y accesibles sobre el funcionamiento de un sistema de IA para que la gente pueda estar informada de manera relevante sobre cómo puede afectarles.

**3. Establecer mecanismos adecuados de responsabilidad y reparación.** Los actores del sector privado deben establecer mecanismos internos de rendición de cuentas sobre el funcionamiento de los sistemas de IA. Aunque los Estados tienen el deber primordial de proporcionar acceso a un recurso formal en caso de violaciones de los derechos humanos, las empresas deben tomar medidas adicionales para garantizar el acceso a un recurso y reparación no judicial significativo y efectivo.[124] Como mínimo, esto incluye:

---

123 ↩ UN Guiding Principles on Business and Human Rights, Principle 21

124 ↩ Para más información sobre los aspectos procesales y sustantivos del recurso en el sector de las tecnologías de la información y las comunicaciones, véase Access Now "Telco Remedy Plan"

- Responsabilidad interna en el desarrollo e implementación de un sistema de IA.
- Compromiso por parte de terceros que desarrollen sistemas de IA para terceros de delimitar claramente la responsabilidad y la obligación de rendir cuentas entre el proveedor y el cliente, incluida la obligación del proveedor de garantizar una formación adecuada de los riesgos del sistema, así como de mitigar el riesgo de la proliferación de funciones y el uso indebido de un sistema de IA.
- Creación de procesos claros y transparentes mediante los cuales una persona pueda presentar directamente sus quejas y solicitar reparación por daños a los derechos humanos de manera oportuna. Estos procesos podrían administrarse internamente, en colaboración con otras partes interesadas, o a través de un organismo externo mutuamente aceptable.[125] Los resultados deben retroalimentar el desarrollo de productos y políticas para prevenir y mitigar mejor los daños.

## LA NECESIDAD DE INVESTIGAR MÁS LOS FUTUROS USOS DE LA IA

Aunque la protección de datos, la transparencia y los mecanismos de rendición de cuentas contribuyen en gran medida a mitigar los abusos de los derechos humanos en el uso de la IA, no resuelven todos los problemas previsibles. Por ejemplo, como hemos identificado, los sistemas de IA pueden afectar sustancialmente a las oportunidades económicas o facilitar la guerra o los conflictos a nivel mundial en el futuro. Por estas razones, recomendamos que los Estados y las entidades del sector privado, incluidas las organizaciones de la sociedad civil y los individuos del mundo académico, trabajen juntos para investigar los futuros usos de la IA y sigan explorando las posibles repercusiones sobre los derechos humanos aquí identificadas. Debe hacerse hincapié en la identificación y creación de mecanismos de respuesta a las posibles amenazas para garantizar que las implicaciones negativas se mitiguen en la mayor medida posible. Estos foros deben ser multi partes interesadas y pluralistas para garantizar que se identifican todas las amenazas potenciales y que las soluciones no den preferencia a ningún grupo específico sobre otro o disminuyan aún más las voces marginadas.

## REFUTACIÓN: LA TRANSPARENCIA Y LA EXPLICACIÓN NO ACABARÁN CON LA INNOVACIÓN EN IA

Hay dos argumentos que se escuchan con frecuencia en contra de los requisitos de transparencia y explicación de los sistemas de IA.[126] Uno de ellos es que la IA es demasiado compleja para exigir transparencia y que hacerlo podría perjudicar la innovación. El segundo argumento es que la explicación es imposible y, si se impone, obligaría a los desarrolladores de IA a sacrificar la complejidad de sus sistemas y (de nuevo) obstaculizaría la innovación. Estos argumentos son exagerados e incoherentes con la evolución de la IA en la actualidad. A continuación se aborda cada uno de estos argumentos.

### Publicar el código y los datos permite a terceros expertos identificar posibles problemas.

Quienes se oponen a la transparencia de los sistemas de IA cuestionan la utilidad de publicar el código y los datos de entrenamiento de sistemas tan complejos que pueden basarse en millones de puntos de datos y tener modelos que cambian con el tiempo.

Aunque la auditoría de los sistemas de IA puede plantear una nueva serie de retos técnicos, esto no hace que la transparencia carezca de sentido. Los desarrolladores de IA examinan los datos de entrenamiento y los resultados de la IA para identificar las fuentes de sesgo y comprobar la equidad de los resultados. La transparencia garantizaría el acceso a los datos de entrenamiento y a los resultados de la IA necesarios para que expertos independientes identifiquen las fuentes de sesgo y comprueben la equidad de los resultados, incluyendo la identificación de cualquier problema que los

125 ↩ Véase Principles 29 and 31 of the UN Guiding Principles on Business and Human Rights for more information.

126 ↩ Véase Joshua New, "How (and how not) fo fix AI," Tech Crunch, 26 July 2018.

desarrolladores hayan pasado por alto o enmascarado. Este proceso aumenta necesariamente la responsabilidad y fomenta la confianza de los usuarios.[127]

**La plena transparencia es vital en los casos de alto riesgo y no tiene por qué perjudicar a las empresas.**
Los que se oponen a la transparencia también argumentan que los requisitos de transparencia reducen los incentivos para invertir en nuevos sistemas de IA porque permitiría la réplica.

Aunque la historia ha demostrado que los proyectos de código abierto no sólo suelen tener éxito, sino que facilitan la innovación, puede haber una opción en circunstancias limitadas en las que los actores del sector privado determinen que esa vía es insostenible. En tales casos, los agentes del sector privado podrían facilitar el acceso al código pertinente a terceros identificados y de confianza para realizar auditorías y pruebas. Recientemente, Facebook ha dado a determinados investigadores acceso a los datos para estudiar la interferencia electoral en la plataforma.[128] Dado el creciente escrutinio público del papel de los algoritmos en nuestras vidas, es concebible que otras empresas sigan su ejemplo. Sin embargo, la información sobre los conjuntos de datos utilizados y los resultados debería seguir publicándose, así como cualquier otra información que pueda facilitar la comprensión y la medición del sesgo.

Sin embargo, siempre se debe exigir a los Estados que proporcionen total transparencia en el uso gubernamental de los sistemas de IA. Esto es especialmente importante en ámbitos como la aplicación de la ley y el sistema judicial. Aunque las empresas que suministran el programa pueden tener razones para no publicar el código y los datos de entrenamiento en estos casos, los derechos humanos fundamentales no pueden sacrificarse en aras de los intereses empresariales.

**La explicación relevante es cada vez más posible, y el mundo de la IA está en gran medida detrás de ella.**
Quienes se oponen a la regulación argumentan que exigir la explicación significaría que los sistemas tendrían que ser sustancialmente menos complejos y, por tanto, menos precisos, lo que en última instancia ahogaría la innovación. Este argumento es erróneo en varios sentidos.

En primer lugar, se pueden alcanzar valiosos niveles de explicación. Aunque Facebook no entienda del todo cómo funciona su algoritmo de publicidad dirigida, sabe lo suficiente como para decir a los usuarios qué acciones les llevaron a recibir un determinado anuncio. Este tipo de información es importante, y es fácil de proporcionar. La investigación sugiere que incluso es posible que los sistemas midan cómo una entrada determinada afectó a la salida. En un sistema utilizado por las universidades para clasificar a los solicitantes, se podría decir, por ejemplo, que el 20% de la clasificación se debe al GPA, el 25% a las pruebas estandarizadas y el 10% a la clasificación de la escuela y otros factores. Este nivel de explicación contribuiría en gran medida a resolver el problema de la "caja negra" y a identificar posibles fuentes de sesgo.

Además, la explicación es técnicamente valiosa. Los desarrolladores deben ser capaces de determinar si un sistema está resolviendo el problema correcto. Hay muchos ejemplos de sistemas de IA que "hacen trampa" para llegar al resultado deseado. Por ejemplo, los investigadores de la Universidad de Washington crearon un algoritmo deliberadamente malo que debía clasificar imágenes de perros husky y lobos. El sistema etiquetó correctamente las imágenes, pero en lugar de

---

[127] ↩ Véase e.g. Rian Wanstreet: America's Farmers Are Becoming Prisoners to Agriculture's Technological Revolution — Vice, 8 March 2018. ("La Oficina Agrícola estadounidense ayudó a elaborar los "Principios de privacidad y seguridad de los datos agrícolas", que abordan cuestiones relativas a la propiedad, la portabilidad, el uso y el intercambio de datos. Empresas como Deere y Monsanto fueron las primeras en firmarlos, pero siguen existiendo dudas sobre el grado de protección de estos principios en la práctica").

[128] ↩ https://www.accessnow.org/cms/assets/uploads/2018/07/GDPR-User-Guide_digital.pdf

aprender la diferencia entre la apariencia de los huskies y la de los lobos, el sistema detectó la presencia de nieve porque la mayoría de las imágenes de lobos tenían nieve de fondo.[129] Si los sistemas de IA en campos de gran importancia acaban resolviendo el problema equivocado, el resultado podría ser mortal.

Dado que la explicación es necesaria para la adopción de la IA en determinados campos, en cierto modo la búsqueda de la explicación está estimulando la innovación en IA. Por razones tanto éticas como técnicas, tanto los académicos como las principales empresas de IA están dedicando un esfuerzo significativo hacia la explicación, y están haciendo serios progresos. En agosto de 2018, DeepMind de Google publicó un estudio sobre un sistema de IA que desarrolló para identificar enfermedades oculares en escaneos oculares 3D. Cuando el sistema hace un diagnóstico, señala las partes del escáner que utilizó para que los médicos puedan ver cómo llegó a ese diagnóstico, así como la confianza que tiene en el mismo.[130] Avances como éste demuestran que la capacidad de explicación puede impulsar la innovación en la IA.

## Conclusión

Los sistemas de inteligencia artificial están cambiando la forma de hacer las cosas en las empresas y los gobiernos de todo el mundo, y traen consigo un potencial de interferencia significativa con los derechos humanos. Las leyes de protección de datos y las salvaguardias para la rendición de cuentas y la transparencia, como las que hemos descrito en este documento, pueden mitigar algunos de los peores usos conocidos hoy en día, pero es necesario seguir trabajando para salvaguardar los derechos humanos a medida que la tecnología de IA se vuelve más sofisticada y se expande a otras áreas. Esperamos que este informe ayude a inspirar diálogos más profundos en esta área crucial para aquellos que se preocupan por el futuro de los derechos humanos, y esperamos participar en esas discusiones.

## Vínculos relacionados:

- La Alianza Global Jus Semper
- Access Now
- John Bellamy Foster, Brett Clark y Hannah Holleman: Capitalismo y Robo
- John Bellamy Foster, R. Jamil Jonna y Brett Clark: El Contagio del Capital
- Samir Amin: La Nueva Estructura Imperialista
- Álvaro de Regil Castilla: Mercadocracia y el Secuestro de la Gente y el Planeta
- Álvaro de Regil Castilla: Transitando a Geocracia — Paradigma de la Gente y el Planeta y No el Mercado — Primeros Pasos

[129] ↩ En EUA, muchos de los detalles del uso gubernamental del sistema de toma de decisiones algorítmicas se ocultan tras acuerdos de no divulgación y memorandos de entendimiento con los proveedores.

[130] ↩ Vése https://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx para un resumen de las obligaciones de los Estados en materia de derechos humanos según el derecho internacional.

- **Acerca de Jus Semper:** La Alianza Global Jus Semper aspira a contribuir a alcanzar un etos sostenible de justicia social en el mundo, donde todas las comunidades vivan en ámbitos verdaderamente democráticos que brinden el pleno disfrute de los derechos humanos y de normas de vida sostenibles conforme a la dignidad humana. Para ello, coadyuva a la liberalización de las instituciones democráticas de la sociedad que han sido secuestradas por los dueños del mercado. Con ese propósito, se dedica a la investigación y análisis para provocar la toma de conciencia y el pensamiento crítico que generen las ideas para la visión transformadora que dé forma al paradigma verdaderamente democrático y sostenible de la Gente y el Planeta y NO del mercado.

- **Acerca de los autores: Access Now** es una organización internacional que defiende y amplía los derechos digitales de los usuarios en peligro en todo el mundo. Combinando el apoyo técnico directo, el compromiso político integral, la promoción global, la concesión de subvenciones de base y las convocatorias como la RightsCon, Access Now lucha por los derechos humanos en la era digital.

- **Acerca de este trabajo:** Este informe es un producto de Access Now, con Lindsey Andersen como autora principal. Este documento fue publicado originalmente en inglés por Access Now en noviembre de 2018. Este ensayo ha sido publicado bajo Creative Commons, CC-BY-NC-ND 4.0. Se puede reproducir el material para uso no comercial, acreditando al autor y proporcionando un enlace al editor original.

- **Cite este trabajo como:** Access Now: Derechos Humanos en la Era de la Inteligencia Artificial – La Alianza Global Jus Semper, septiembre de 2021.

- **Etiquetas:** Capitalismo, Democracia, Derechos Humanos, Inteligencia Artificial (IA), Aprendizaje de Maquina (AM), Responsabilidad Social Corporativa (RSC), Trabajo.

- La responsabilidad por las opiniones expresadas en los trabajos firmados descansa exclusivamente en su(s) autor(es), y su publicación no representa un respaldo por parte de La Alianza Global Jus Semper a dichas opiniones.