



Organización  
de las Naciones Unidas  
para la Educación,  
la Ciencia y la Cultura

Ediciones  
UNESCO

# El aporte de la inteligencia artificial y las TIC avanzadas a las sociedades del conocimiento

Una perspectiva de Derechos,  
Apertura, Acceso y Múltiples  
actores

Serie de la UNESCO sobre la  
libertad en Internet

Publicado en 2021 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), 7, Place de Fontenoy, 75352 Paris 07 SP, Francia y Oficina fuera de la sede de la UNESCO.

© UNESCO 2021

ISBN 978-92-3-300155-8



Esta publicación está disponible en Acceso Abierto bajo la licencia Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo>). Al utilizar el contenido de la presente publicación, los usuarios aceptan las condiciones de utilización del Repositorio de Acceso Abierto de la UNESCO. (<http://www.unesco.org/open-access/terms-use-ccbysa-sp>).

Titulo original: *Steering AI and Advanced ICTs in Knowledge Societies: a Rights, Openness, Access and Multi-Stakeholder Perspective*. Publicado en 2019 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura y Oficina fuera de la sede de la UNESCO.

Los términos empleados en esta publicación y la presentación de los datos que en ella aparecen no implican toma alguna de posición de parte de la UNESCO en cuanto al estatuto jurídico de los países, territorios, ciudades o regiones ni respecto de sus autoridades, fronteras o límites.

Las ideas y opiniones expresadas en esta obra son las de los autores y no reflejan necesariamente el punto de vista de la UNESCO ni comprometen a la Organización.

Los autores de esta publicación son:

Xianhong Hu, Bhanu Neupane, Lucía Flores Echaiz, Prateek Sibal y Macarena Rivera Lam.

El informe completo y el folleto con el resumen se encuentran disponibles en: <https://en.unesco.org/internet-universality-indicators/ai-roam>

Diseño gráfico, diseño de tapa y composición: Josselyn Guillarmou

Ilustración de capítulos: © Josselyn Guillarmou

Ilustración de tapa: © Shutterstock/greiss design

Corrección: Yahia Dabbous

Impreso por UNESCO

Impreso en Francia

# EL APORTE DE LA INTELIGENCIA ARTIFICIAL Y LAS TIC AVANZADAS A LAS SOCIEDADES DEL CONOCIMIENTO

*Una perspectiva de Derechos, Apertura, Acceso  
y Múltiples actores*

# ÍNDICE

LIST OF FIGURES, BOXES AND TABLES	5
PREFACIO	7
AGRADECIMIENTOS	9
GLOSARIO	10
RESUMEN EJECUTIVO	17

## **INTRODUCCIÓN** **25**

1. CONCEPTUALIZACIÓN DE LA IA EN ESTA PUBLICACIÓN	26
2. ¿POR QUÉ LA UNESCO ESTÁ INTERESADA EN LA IA?	27
3. LA REFLEXIÓN EN CURSO DE LA UNESCO SOBRE LA IA	29
4. LA UNESCO ABORDA LA IA Y LAS TIC AVANZADAS A TRAVÉS DE LA PERSPECTIVA DAAM	30

## **CAPÍTULO 1: DERECHOS HUMANOS E INTELIGENCIA ARTIFICIAL** **35**

1. DERECHO A LA LIBERTAD DE EXPRESIÓN	39
1.1. Información personalizada, libertad de opinión y el derecho a buscar información	40
1.2. Moderación de contenidos en línea por parte de la IA y el derecho a impartir información	42
1.3. Plataformas de Internet bajo presión para regular los contenidos en línea	44
1.4. Detección y eliminación de contenidos por parte de la IA: casos y desafíos	45
1.5. Falta de transparencia y debido proceso en la moderación de contenidos	47
2. DERECHO A LA PRIVACIDAD Y LA PROTECCIÓN DE DATOS PERSONALES	49
2.1. Recolección de datos e Internet de las cosas	52
2.2. Vigilancia, monitoreo y reconocimiento facial con tecnología de IA	53
2.3. Seguimiento en línea y desanonimización de personas	56
2.4. Análisis predictivo y elaboración de perfiles con IA	57
2.5. Protección de la privacidad en el contexto de la IA: paradigmas actualizados y soluciones de Tecnologías de Mejora de la Privacidad	59

3.	DESARROLLO DEL PERIODISMO Y LOS MEDIOS DE COMUNICACIÓN COMO EJERCICIOS DE LA LIBERTAD DE EXPRESIÓN	61
3.1.	La era del periodismo computacional, el periodismo de datos, el periodismo automatizado y el periodismo robótico	61
3.2.	La inteligencia artificial amenaza el pluralismo de los medios: priorización de contenidos, personalización de información y microfocalización	63
3.3.	Aumento de la desinformación automatizada e iniciativas para contrarrestarla	65
3.4.	La protección de los periodistas y las fuentes periodísticas en la era de la IA	67
4.	DERECHO A LA IGUALDAD	69
4.1.	¿Qué es la discriminación?	70
4.2.	¿Cómo se diseña la discriminación en los algoritmos?	72
5.	CONCLUSIÓN Y OPCIONES DE POLÍTICA	75

## **CAPÍTULO 2: APERTURA E INTELIGENCIA ARTIFICIAL** **81**

1.	APERTURA EN LA INVESTIGACIÓN EN IA	84
2.	DATOS ABIERTOS E IA	88
3.	APERTURA EN LA IA: CAJA NEGRA Y PROBLEMAS DE TRANSPARENCIA	90
4.	PAPEL DE LOS MERCADOS EN LA IA ABIERTA	93
5.	RIESGOS DE LA APERTURA Y RESPUESTAS	94
6.	CONCLUSIÓN Y OPCIONES DE POLÍTICA	97

## **CAPÍTULO 3: ACCESO E INTELIGENCIA ARTIFICIAL** **101**

1.	ACCESO A LA INVESTIGACIÓN	107
2.	ACCESO A CONOCIMIENTO, EDUCACIÓN Y RECURSOS HUMANOS	110
3.	ACCESO A <i>SOFTWARE</i> Y DATOS PARA ENTRENAMIENTO DE ALGORITMOS	114
4.	ACCESO A CONECTIVIDAD Y <i>HARDWARE</i>	117
5.	CONCLUSIÓN Y OPCIONES DE POLÍTICA	120

## **CAPÍTULO 4: ENFOQUE DE MÚLTIPLES ACTORES PARA LA GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL** **125**

1.	CONTEXTUALIZACIÓN DE LA IA EN LA GOBERNANZA DE INTERNET POR MÚLTIPLES ACTORES	129
----	---	-----

2.	TOMA DE DECISIONES COMPLEJAS Y EQUILIBRIO DE INTERESES EN EL DESARROLLO DE LA IA – LA NECESIDAD DE QUE PARTICIPEN MÚLTIPLES ACTORES	131
3.	GOBERNANZA DE LA IA MEDIANTE UN ENFOQUE DE MÚLTIPLES ACTORES: PRÁCTICAS, VALORES E INDICADORES	134
4.	CONCLUSIÓN Y OPCIONES DE POLÍTICA	137

## **CAPÍTULO 5: IGUALDAD DE GÉNERO E INTELIGENCIA ARTIFICIAL** **141**

1.	PERSPECTIVAS DE ESTUDIOS DE GÉNERO SOBRE TECNOLOGÍA	145
2.	IMPLICACIONES DE GÉNERO DE LAS TECNOLOGÍAS DE IA	147
2.1.	Predominio masculino en conocimientos y fuerza de trabajo de IA	147
2.2.	Culturas de patriarcado y sexismo	149
2.3.	Consecuencias económicas y sistemas de IA sesgados	150
3.	DISCRIMINACIÓN ALGORÍTMICA	151
3.1.	Exclusión, sesgo y discriminación	151
3.2.	Revelación de la identidad sexual de personas LGBTI	154
4.	ASISTENTES DE VOZ "FEMENINOS"	156
4.1.	Refuerzo de estereotipos de género de larga data	156
4.2.	Respuestas permisivas al acoso sexual	158
5.	LA INDUSTRIA DE LOS ROBOTS SEXUALES	159
5.1.	Objetivación de las mujeres	159
5.2.	Pornografía con falsificación profunda (deepfake)	161
5.3.	Avanzar poniendo en foco en el género y la IA	162
6.	CONCLUSIÓN Y OPCIONES DE POLÍTICA	163

## **CAPÍTULO 6: INTELIGENCIA ARTIFICIAL EN ÁFRICA** **167**

1.	LOS DESAFÍOS DE CIENCIA, TECNOLOGÍA E INNOVACIÓN EN ÁFRICA	171
2.	INICIATIVAS DE LOS GOBIERNOS DE ÁFRICA EN RELACIÓN CON LA IA	173
3.	INICIATIVAS DEL SECTOR PRIVADO, LA COMUNIDAD TÉCNICA Y LA SOCIEDAD CIVIL PARA LA IA EN ÁFRICA	175
3.1.	Sector privado	175
3.2.	Universidades e institutos educativos	178
3.3.	La sociedad civil y la comunidad técnica	179
4.	CONCLUSIÓN Y OPCIONES DE POLÍTICA	182
	BIBLIOGRAFÍA	190

# FIGURAS, BOXES Y TABLAS

## FIGURAS

Figura 1: Cantidad de artículos de IA en arXiv por subcategoría	85
Figura 2: Ejemplo de una capa de entrada conflictiva agregada a la imagen de un panda que lleva a su clasificación errónea como un gibón	95
Figura 3: Evolución en imágenes generadas por IA	96
Figura 4: Tasa de penetración de Internet para hombres y mujeres, 2017	104
Figura 5: Brecha de género en los usuarios de Internet (%), 2013 y 2017	105
Figura 6: Artículos de IA publicados anualmente en Scopus por región	108
Figura 7: Impacto de la cita de los autores de IA, por región (Shoham y otros, 2018)	109
Figura 8: Cantidad de artículos aceptados y presentados en la Conferencia de la AAIA de 2018	110
Figura 9: Diferencia en capacidades de IA entre diferentes grupos de países	111
Figura 10: Diferencia en las capacidades de IA entre diferentes regiones	112
Figura 11: Crecimiento de las ofertas laborales según las habilidades en IA necesarias	113
Figura 12: Marco de Datos Comunes	116
Figura 13: Exportadores de dispositivos semiconductores por continente. El comercio de semiconductores se muestra como una representación para el hardware de computación	119
Figura 14: Importadores de dispositivos semiconductores por continente. El comercio de semiconductores se muestra como una representación para el hardware de computación	119
Figura 15: Índice de paridad de género entre adultos que realizaron una actividad relacionada con la computadora en los 3 meses anteriores	148
Figura 16: Brecha de género en conocimientos de programación informática	149
Figura 17: El porcentaje de mujeres con piel más oscura, mujeres con piel más clara, hombres con piel más oscura y hombres con piel más clara en los conjuntos de datos	152
Figura 18: Primer plano de Harmony	159
Figura 19: Imagen <i>deepfake</i> de la actriz Natalie Portman	161

## RECUADROS

Recuadro 1: Mandato de la UNESCO en la Cumbre Mundial sobre la Sociedad de la Información (CMSI)	28
Recuadro 2: Posición de la UNESCO respecto de los derechos humanos para la universalidad de Internet	37
Recuadro 3: La UNESCO y la libertad de expresión	38
Recuadro 4: Facebook prohíbe la fotografía de la "Niña del napalm", ganadora del Premio Pulitzer	46
Recuadro 5: Los Principios de Santa Clara	48
Recuadro 6: Opciones de la UNESCO en el ámbito de la privacidad	50
Recuadro 7: Los asistentes virtuales escuchan a escondidas: Alexa de Amazon	54
Recuadro 8: Software de reconocimiento facial en centros comerciales	55
Recuadro 9: Seguimiento en línea	56
Recuadro 10: Efecto "inhibidor" en otros derechos humanos	58
Recuadro 11: La inteligencia artificial ayuda a medir la calidad del periodismo	62
Recuadro 12: El asunto Cambridge Analytica	65
Recuadro 13: Nuevas formas de ataque contra los periodistas	68
Recuadro 14: Sesgos controlados por datos que implican una discriminación basada en la raza	74
Recuadro 15: Posición de la UNESCO respecto de la apertura para la universalidad de Internet	84
Recuadro 16: Tecnologías y plataformas clave	88
Recuadro 17: Posición de la UNESCO respecto del acceso a la universalidad de Internet	103
Recuadro 18: Mandato de la Cumbre Mundial sobre la Sociedad de la Información (CMSI)	106
Recuadro 19: Posición de la UNESCO en un enfoque de múltiples actores para la universalidad de Internet	128
Recuadro 20: Informe "La era de la interdependencia digital"	130
Recuadro 21: Prioridad de la UNESCO respecto de la igualdad de género y las TIC	144
Recuadro 22: Representación de la diáspora africana en la comunidad occidental de IA	179
Recuadro 23: IA para los idiomas africanos: reforzar el multilingüismo	181

## TABLAS

Tabla 1: Respuestas de los asistentes de voz al acoso sexual	158
Tabla 2: Iniciativas en África que usan IA en salud, agricultura, tecnología financiera y transporte	178

# PREFACIO



La Inteligencia Artificial —abreviada IA— puede ayudar a allanar el camino para nuevas oportunidades en materia de desarrollo sostenible, incluso cuando se trata de los ámbitos de competencia de la UNESCO. Los avances en este campo ya tienen un impacto directo en nuestro trabajo en las áreas de educación, ciencias naturales y humanas, cultura, y comunicación e información.

La UNESCO ha jugado durante mucho tiempo un papel destacado en la cooperación y el establecimiento de normas internacionales, y los Estados Miembros ahora han reconocido la necesidad de que desarrollemos principios éticos para la IA. De este modo, podemos ayudar a garantizar que el desarrollo tecnológico se ajuste a una visión centrada en el ser humano, con respeto por los derechos humanos y adelantos en lo que respecta al desarrollo sostenible. Este es el motivo por el cual la UNESCO participa en el análisis y la reflexión sobre la IA, incluso desde la base de su mandato de promover la libertad de expresión y construir sociedades del conocimiento integradoras.

Sin dejar de reconocer que no hay respuestas simples a la hora de analizar qué le depara el futuro a la humanidad, este informe de investigación es una contribución al debate más amplio sobre la ética y la gobernanza de la IA. Es un intento de evitar tanto el utopismo tecnológico como el pensamiento distópico. En lugar del determinismo tecnológico y su idea de inevitabilidad, la UNESCO presta atención al rol de la agencia humana y los valores centrados en el ser humano en el desarrollo de la IA y otras tecnologías de la información y las comunicaciones (TIC) avanzadas.

Nuestro punto de partida es el reconocimiento de la IA como una oportunidad para alcanzar los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas y construir sociedades del conocimiento. Para la UNESCO, estas sociedades se basan en la libertad de expresión, el acceso a la información, una educación de calidad y el respeto a la diversidad cultural y lingüística. Representan una visión a la que podemos aspirar y un faro mientras recorremos el camino de dar forma al rol de la IA para la humanidad.

Esta misión se aplica al trabajo de la UNESCO no solo en torno a la IA, sino también respecto de avances como la Internet de las cosas, la cadena de bloques, la biométrica y la toma de decisiones algorítmica. Mientras examinamos oportunidades, también buscamos identificar y mitigar riesgos como los que plantean la vigilancia arbitraria y masiva, la elaboración de perfiles y las violaciones de la privacidad y la igualdad. Esto es importante a la hora de evaluar los posibles impactos de la digitali-

zación continua en la educación, las ciencias, la cultura y la comunicación y la información, así como en el empleo, la igualdad y el empoderamiento.

La investigación es fundamental si queremos entender cómo se están usando la IA y otras tecnologías avanzadas para incidir en gran parte de nuestra vida cotidiana. Es la base sobre la que podemos construir el conocimiento que necesitamos para dar forma a la evolución tecnológica y aprovechar el potencial de la IA de manera positiva.

La evaluación de la IA en el presente estudio se realiza mediante el marco DAAM de la universalidad de Internet de la UNESCO, acordado por los Estados Miembros en 2015. Por lo tanto, aborda de qué forma la IA y las TIC avanzadas incidirán en los **Derechos Humanos, la Apertura** y el **Acceso**, y cómo un enfoque de **Múltiples actores** sustenta la labor para enfrentar tanto los desafíos como las oportunidades que presenta la IA.

Los principios DAAM pueden ayudar a establecer los valores necesarios para orientar el desarrollo ético y basado en derechos, y la implementación de la IA en formas que mitigan riesgos y alcanzan los ODS. En este contexto, este estudio también ofrece un conjunto de opciones para la acción que pueden servir de inspiración para el desarrollo de nuevos marcos de política éticos y otras acciones, ya sea por parte de los Estados en sus distintos ámbitos de trabajo, los diversos actores del sector privado, los miembros del sector académico y la comunidad técnica, o la sociedad civil. Esperamos que las ideas plasmadas en estas páginas ayuden a encaminar la IA para que haga una importante contribución al desarrollo de sociedades del conocimiento integradoras que no dejen a nadie atrás.



Moez Chakchouk  
Subdirector General de Comunicación e Información  
UNESCO

# AGRADECIMIENTOS

Agradecemos sinceramente a quienes realizaron la revisión por pares por su tiempo y sus esclarecedoras sugerencias, que sirvieron para mejorar la publicación:

- Olubayo Adekanmbi, Director de Transformación, MTN Nigeria
- Alex Comminos, Investigador, Research ICT Africa
- Hugo Cyr, Decano, Facultad de Derecho y Ciencias Políticas, Université du Québec à Montréal
- Eileen Donahoe, Directora Ejecutiva, Global Digital Policy Incubator, Stanford University
- Jaco du Toit, Especialista del Programa, UNESCO
- Helani Galpaya, Directora Ejecutiva, LIRNEasia
- Damiano Giampaoli, Especialista del Programa, UNESCO
- Marcus Goddard, Socio, Inteligencia, NETEXPLO Observatory
- Helen Hester, Profesora Titular de Medios y Comunicación, University of West London
- Joe Hironaka, Especialista del Programa, UNESCO
- Joe F. Khalil, Profesor Titular, Northwestern University en Qatar
- Nnenna Nwakanma, Chief Web Advocate, World Wide Web Foundation
- Michael J. Oghia, Gerente de Promoción y Compromiso, Global Forum for Media Development (GFMD)
- Julia Pohle, Investigadora Sénior, WZB Berlin Social Science Center
- Rachel Pollack, Especialista del Programa Asociada, UNESCO
- Jan Rydzak, Director Asociado de Programa, Global Digital Policy Incubator, Stanford University
- Moses M. Thiga, Profesor, Informática y tecnologías de la información, Kabarak University
- Tim Unwin, Catedrático, Cátedra UNESCO de TIC para el Desarrollo, Royal Holloway University of London

Queremos agradecer a Alexandra Agay, Antoine Martin, Berenice Chaumont, Maud Barret, Valentin Leblanc, Melissa Tay Ru Jein y Joo Eun Chae por su asistencia en la investigación.

Queremos agradecer a nuestras colegas Kelly Christine Wong y Yahia Dabbous por su ayuda en la compilación de la publicación.

También queremos agradecer el apoyo de otros colegas del Sector de Comunicación e Información, el Sector de Ciencias Naturales, el Sector de Cultura, el Sector de Educación, el Sector de Ciencias Sociales y Humanas, la División para la Igualdad de Género y el Sector de Prioridad África y de Relaciones Exteriores, que forman parte del Equipo de Trabajo sobre IA de la UNESCO.

Rendimos homenaje al Sr. Indrajit Banerjee por su visión y apoyo para el comienzo de esta investigación en forma conjunta.

# GLOSARIO

Este glosario incluye los significados amplios de algunos términos clave como se usan en este estudio. La atribución de ideas a los autores no constituye una aprobación de sus definiciones.

## **Algoritmo**

Un conjunto de instrucciones paso a paso para resolver un problema (Negnevitsky, 2011)

## **Toma de decisiones algorítmica**

Una forma de tomar decisiones basada en resultados de algoritmos (Andersen, 2018).

## **Anonimización**

El proceso de eliminar irreversiblemente identificadores personales, directos e indirectos, que pueden llevar a la identificación de un individuo (Artículo 29 Data Protection Working Party, 2014)

## **Inteligencia artificial (IA)**

Si bien no hay una única definición de "inteligencia artificial" (IA), esta publicación tiende a definir la IA como un conjunto de TIC avanzadas que permiten el funcionamiento de "máquinas capaces de imitar ciertas funcionalidades de la inteligencia humana, incluidas características como la percepción, el aprendizaje, el razonamiento, la resolución de problemas, la interacción mediante el lenguaje y hasta la producción de trabajo creativo" (COMEST, 2019).

## **Inteligencia artificial estrecha (IAE)**

La capacidad de las máquinas para asemejarse a las capacidades humanas en dominios estrechos, con diferentes grados de sofisticación técnica y autonomía (Artículo 19 y Privacy International, 2018).

La inteligencia artificial débil o inteligencia artificial estrecha (IAE) es la forma de IA que la humanidad ha logrado hasta el momento: máquinas capaces de realizar ciertas tareas precisas de manera autónoma, pero sin conciencia, en un marco definido por los humanos y siguiendo decisiones tomadas solo por los humanos (UNESCO, 2018d).

## **Inteligencia artificial general (IAG)**

El objetivo global, y aún no alcanzado, de un sistema que muestra la inteligencia a través de múltiples dominios, con la capacidad de aprender nuevas habilidades, y que imitan o incluso superan la inteligencia humana (Artículo 19 y Privacy International, 2018).

La IA fuerte o IAG se refiere entonces a una máquina que tiene conciencia y es capaz de proporcionar una respuesta similar a la de los humanos (UNESCO, 2018d).

## Red neuronal artificial (RNA)

Un paradigma de procesamiento de información inspirado en la estructura y las funciones del cerebro humano. Una RNA consiste en un número de procesadores simples y altamente interconectados, llamados neuronas, que son análogos a las neuronas biológicas del cerebro. Las neuronas están conectadas por enlaces ponderados que pasan señales de una neurona a otra. Mientras que en una red neuronal biológica el aprendizaje implica ajustes en las sinapsis, las RNA aprenden a través de repetidos ajustes de los pesos. Estos pesos almacenan las reglas necesarias para resolver problemas específicos (Negnevitsky, 2011).

## Toma de decisiones automatizada

Un proceso de toma de decisiones por medios automatizados. Suele implicar el uso del razonamiento automatizado para ayudar o sustituir un proceso de toma de decisiones que de otro modo sería realizado por humanos. No implica necesariamente el uso de IA, pero por lo general implicará la recopilación y el procesamiento de datos (CoE CHR/Rec(2019)1, 2019).

## Sesgo

Una inclinación o prejuicio a favor o en contra de una persona o grupo, especialmente de una manera que se considera injusta (definición social); la diferencia entre el valor estimado —o pronosticado— y el valor real; en otras palabras, la diferencia entre lo que un sistema predice y lo que realmente sucede (definición estadística) (Andersen, 2018).

## Macrodatos

Conjuntos de datos que son demasiado grandes o complejos para que un *software* de procesamiento de datos tradicional los analice (Andersen, 2018). La mayoría de los sistemas de IA dependen de la recopilación, el procesamiento y el intercambio de macrodatos para poder cumplir sus funciones.

## Caja negra

Un modelo que es opaco para el usuario. Aunque el modelo puede producir resultados correctos, se desconoce cómo se producen estos resultados. Una red neuronal es un ejemplo de caja negra. Para comprender las relaciones entre las entradas y salidas de una caja negra se puede utilizar el análisis de sensibilidad (Negnevitsky, 2011).

## Bots

Aplicaciones de *software* que ejecutan tareas automatizadas, cada vez más impulsadas por el aprendizaje automático (Andersen, 2018).

## Nube

Una metáfora que describe el acceso a la red a un conjunto escalable y elástico de recursos físicos o virtuales compartibles con aprovisionamiento de autoservicio y administración a demanda (UTI, 2014).

## Dilema de Collingridge

El problema de que cuando el cambio es fácil, no se puede prever su necesidad; cuando la necesidad de cambio es evidente, el cambio se ha convertido en algo caro, difícil y que requiere mucho tiempo (Collingridge, 1980).

## Datos

Hechos, medidas u observaciones. Además, una representación simbólica de hechos, medidas u observaciones (Negnevitsky, 2011).

## Base de datos

Una colección de datos estructurados (Negnevitsky, 2011).

## Minería de datos

Extracción de información y conocimiento de los datos. Además, la exploración y el análisis de grandes cantidades de datos para descubrir reglas y patrones significativos. El objetivo final de la minería de datos es descubrir información y conocimiento (Negnevitsky, 2011).

## Aprendizaje profundo

Esta técnica permite que una máquina reconozca independientemente variaciones complejas. Un ejemplo es el rastreo y la clasificación automatizados de millones de imágenes tomadas de Internet que no han sido etiquetadas exhaustivamente por los humanos. El aprendizaje profundo, que es el resultado de una combinación de algoritmos de aprendizaje y redes neuronales formales, y el uso de cantidades masivas de datos, impulsa la IA (UNESCO, 2018d).

## Tecnologías de la información y las comunicaciones (TIC)

Conjunto variado de herramientas y recursos tecnológicos usados para transmitir, almacenar, crear, compartir o intercambiar información. Esta herramientas y recursos tecnológicos incluyen *software*, computadoras, Internet (sitios web, blogs y correos electrónicos), tecnologías de radiodifusión en directo (radio, televisión y difusión por Internet), tecnologías de radiodifusión grabada (*podcasting*, reproductores de audio y video, y dispositivos de almacenamiento) y telefonía (fija o móvil, satelital, visio/videoconferencia, etc.) (Instituto de Estadística de la UNESCO, 2019).

## Programa Información para Todos (PIPT)

El Programa Información para Todos es un programa intergubernamental exclusivo de la UNESCO que fue establecido en 2001. A través de PIPT, los gobiernos miembros y asociados se comprometen a aprovechar las nuevas oportunidades de la era de la

información para crear sociedades equitativas mediante un mejor acceso a la información.

## **Inteligencia**

La capacidad de aprender y entender, de definir problemas y de tomar decisiones para resolverlos. Se estima que una máquina es inteligente si puede lograr un rendimiento de nivel humano en alguna tarea cognitiva (Negnevitsky, 2011).

## **Intermediarios de Internet**

El término intermediarios de Internet habitualmente se refiere a una variedad de proveedores de servicios amplia, diversa y de rápida evolución que facilita las interacciones en Internet entre personas físicas y jurídicas ofreciendo y realizando una diversidad de funciones y servicios. Algunos conectan a los usuarios a Internet, permiten el procesamiento de información y datos, o alojan servicios basados en la Web, incluso para el contenido generado por el usuario.

Otros agregan información y permiten realizar búsquedas; dan acceso, alojan e indexan contenidos y servicios diseñados o gestionados por terceros. Algunos incluso facilitan la venta de bienes y servicios, incluidos servicios audiovisuales, y permiten otras transacciones comerciales, incluso pagos (CoE, 2018).

## **Internet de las cosas**

Una infraestructura global que permite servicios avanzados mediante la interconexión de cosas (físicas y digitales) sobre la base de tecnologías de la información y las comunicaciones interoperables existentes y en evolución (UTI, 2012).

## **Sociedades del conocimiento**

Las sociedades del conocimiento abarcan la capacidad de identificar, producir, procesar, transformar, difundir y utilizar la información para construir y aplicar conocimientos para el desarrollo humano. Requieren una visión social empoderadora que abarque la pluralidad, la inclusión, la solidaridad y la participación. Los cuatro principios que son esenciales para el desarrollo de una sociedad del conocimiento equitativa son: i) diversidad cultural, ii) educación para todos, iii) acceso universal a la información y iv) libertad de expresión (UNESCO, 2005).

## **Aprendizaje automático**

Un mecanismo de adaptación que permite a las computadoras aprender de la experiencia, aprender con el ejemplo y aprender por analogía. Las capacidades de aprendizaje mejoran el rendimiento de un sistema inteligente a lo largo del tiempo. El aprendizaje automático es la base de los sistemas que pueden adaptar su respuesta continuamente (Negnevitsky 2011).

## **Metadatos**

Datos usados para definir, contextualizar o caracterizar datos (CoE, 2018).

## Datos abiertos

Bases de datos que están a disposición del público para que cualquiera pueda acceder, utilizarlas y compartirlas<sup>1</sup>.

## Datos personales

Información relacionada con una persona física identificada o identificable, directa o indirectamente, por referencia a uno o más elementos específicos de esa persona (CoE, 2018).

## Procesamiento de datos personales

Toda operación o conjunto de operaciones realizadas mediante procesos automatizados y aplicadas a datos o conjuntos de datos personales, como la recopilación, el registro, la organización, la estructuración, el almacenamiento, la adaptación o modificación, la recuperación, la consulta, la utilización, la comunicación por transmisión, la difusión o cualquier otra forma de puesta a disposición, vinculación o interconexión, limitación, supresión o destrucción (CoE, 2018).

## Creación de perfiles

El procesamiento de datos personales con el fin de evaluar determinados aspectos de la vida de una persona física, en particular para analizar o predecir aspectos relativos al desempeño de esa persona física en el trabajo, la situación económica, la salud, las preferencias personales, los intereses, la confiabilidad, el comportamiento, la ubicación o los movimientos (CoE, 2018).

## Seudonimización

El procesamiento de los datos personales de manera que estos ya no puedan atribuirse a un sujeto de datos específico sin el uso de información adicional. Esa información adicional se guarda por separado y está sujeta a medidas técnicas y de organización para garantizar que los datos personales no se atribuyan a una persona física identificada o identificable (GDPR art. IV, cl. 4, 2016).

## Principios DAAM

Los principios DAAM (Derechos, Apertura, Accesibilidad y Múltiples actores) de la UNESCO son una lente a través de la cual se puede evaluar el carácter inclusivo del desarrollo y el despliegue de las tecnologías de IA, destacando la relevancia de los derechos humanos, así como la importancia de la apertura, la accesibilidad y la participación de múltiples actores<sup>2</sup>.

---

1 Open Data Institute: <https://theodi.org/article/what-is-open-data-and-why-should-we-care/>.

2 Sobre la base del marco de universalidad de Internet de la UNESCO, como fuera adoptado por la resolución 38 C/53 de la Conferencia General en el documento final "Conectando los puntos: opciones para la acción futura" en 2015. Disponible en: [https://unesdoc.unesco.org/ark:/48223/pf0000234090\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000234090_spa).

## Indicadores DAAM-X

Los indicadores DAAM-X sobre la universalidad de Internet constituyen un instrumento de investigación que contiene 303 indicadores (109 identificados como básicos) relativos a las categorías DAAM (Derechos, Apertura, Accesibilidad, Múltiples actores), así como indicadores contextuales y transversales (X) para abordar la igualdad de género y las necesidades de los niños y los jóvenes, las dimensiones económicas, la confianza y la seguridad, así como los aspectos jurídicos y éticos de Internet. Las evaluaciones nacionales sobre el desarrollo de Internet pueden realizarse usando estos parámetros internacionales, cuya aplicación voluntaria se aprobó en la 31ª Sesión del Consejo del Programa Internacional para el Desarrollo de la Comunicación (PIDC) de la UNESCO, celebrada en noviembre de 2018<sup>3</sup>.

## Aprendizaje supervisado

Un tipo de aprendizaje automático que requiere un profesor externo, que presenta una secuencia de ejemplos de capacitación a las Redes Neuronales Artificiales. Cada ejemplo contiene el patrón de entrada y el patrón de salida que se desea que genere la red. La red determina su salida real y la compara con la salida deseada del ejemplo de capacitación. Si la salida de la red difiere de la salida deseada especificada en el ejemplo de capacitación, se modifican los pesos de la red (Negnevitsky, 2011).

## Determinismo tecnológico

Según la Enciclopedia Internacional de Ciencias Sociales y del Comportamiento, "determinismo tecnológico" es un término que se usa para describir un conjunto de afirmaciones sobre la relación entre lo que habitualmente se denomina "tecnología" y "sociedad". Se usa con dos significados: i) una lógica técnica interna determina el diseño de los artefactos y sistemas tecnológicos; y ii) el desarrollo de artefactos y sistemas tecnológicos determina grandes cambios sociales. Ambos significados con frecuencia se unen en la afirmación de que una tecnología autónoma (tanto en su desarrollo como en su uso) da forma a las relaciones sociales (Kline, 2015).

## Prueba de Turing

Una prueba diseñada por el científico informático Alan Turing para determinar si una máquina puede pasar una prueba de comportamiento inteligente. Turing definió el comportamiento inteligente de un ordenador como la capacidad de lograr un rendimiento similar al de un humano en tareas cognitivas. Durante la prueba, un humano interroga a dos interlocutores —una máquina y un humano—, a través de un medio neutro como una terminal remota. La computadora pasa la prueba si el interrogador no puede distinguir la máquina del humano (Negnevitsky, 2011).

---

3 Decisiones adoptadas en la 31a Sesión del Consejo del Programa Internacional para el Desarrollo de la Comunicación (PIDC), 21 y 22 de noviembre de 2018. Disponible [en inglés] en: [https://en.unesco.org/system/files/private\\_documents/266235eng.pdf](https://en.unesco.org/system/files/private_documents/266235eng.pdf).

## **Programa UNITWIN y de Cátedras UNESCO**

El Programa UNITWIN y de Cátedras UNESCO, que fue lanzado en 1992, cuenta con la participación de más de 700 instituciones en 116 países, promueve la cooperación y el establecimiento de redes internacionales entre universidades para mejorar las capacidades institucionales mediante el intercambio de conocimientos y la labor de colaboración. El programa apoya la creación de Cátedras UNESCO y Redes UNITWIN en los principales ámbitos prioritarios correspondientes a las esferas de competencia de la UNESCO: educación, ciencias naturales y sociales, cultura y comunicación.

## **Cumbre Mundial sobre la Sociedad de la Información (CMSI)**

La Resolución 56/183 de la Asamblea General de las Naciones Unidas (21 de diciembre de 2001) aprobó la celebración de la Cumbre Mundial sobre la Sociedad de la Información (CMSI) en dos etapas. La CMSI, celebrada en dos fases (en Ginebra en diciembre de 2003 y en Túnez en noviembre de 2005), fue la primera oportunidad para que la comunidad internacional evaluara el potencial de las nuevas tecnologías de la información y las comunicaciones (TIC) para alcanzar los objetivos de desarrollo acordados internacionalmente, y considerara los nuevos retos que estos presentaban. En los cuatro documentos finales de la CMSI —la Declaración de Principios de Ginebra, el Plan de Acción de Ginebra, el Compromiso de Túnez y la Agenda de Túnez para la Sociedad de la Información— se expone la visión de una sociedad de la información centrada en las personas, integradora y orientada al desarrollo, que mejoraría las oportunidades y la calidad de vida de las personas en todo el mundo y facilitaría el desarrollo sostenible (CSTD/UNCTAD, 2015).

# RESUMEN EJECUTIVO

El mandato de la UNESCO de construir sociedades del conocimiento integradoras se centra en sus esfuerzos por promover la libertad de expresión y el acceso a la información, junto con una educación de calidad y el respeto de la diversidad cultural y lingüística. La transformación digital que está teniendo lugar en la sociedad afecta a todas las esferas de la actividad humana, y es oportuno reflexionar sobre los principales desafíos y oportunidades que crean las tecnologías digitales, como la inteligencia artificial (IA).

El título de esta publicación es un llamado a la "Dirigir la IA y las TIC avanzadas para las sociedades del conocimiento" desde la perspectiva de los derechos humanos, la apertura, el acceso y la gobernanza de múltiples actores (los principios DAAM). Dicha dirección también debe apoyar la igualdad entre los géneros y a África, las dos prioridades globales de la UNESCO. El cambio y el avance tecnológicos son importantes para el desarrollo sostenible, pero con la creencia en el determinismo tecnológico se corre el riesgo de descuidar los impulsores sociales, económicos y de otro tipo. En cambio, el desafío consiste en aprovechar la agencia humana para configurar la trayectoria de IA y las tecnologías de la información y las comunicaciones (TIC) conexas.

Si bien no hay una única definición de "inteligencia artificial" (IA), esta publicación se centra en lo que la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST) de la UNESCO describe como "máquinas capaces de imitar ciertas funcionalidades de la inteligencia humana, incluidas características como la percepción, el aprendizaje, el razonamiento, la resolución de problemas, la interacción mediante el lenguaje y hasta la producción de trabajo creativo" (COMEST, 2019).

La IA y sus elementos constitutivos de datos, algoritmos, *hardware*, conectividad y almacenamiento aumentan exponencialmente la potencia de las TIC. Esta es una gran oportunidad para el desarrollo sostenible, con riesgos simultáneos que también deben abordarse. Para dirigir la IA de conformidad con estas características es necesario reconocer la distribución desigual, pero dinámica, de la potencia de la IA en múltiples centros dispersos en las distintas áreas del gobierno, el sector privado, la comunidad técnica, la sociedad civil y otras partes interesadas en todo el mundo. Es por esta razón que el compromiso de los múltiples actores en torno a la IA es vital. Esta perspectiva está en consonancia con el enfoque de gobernanza de las TIC según los principios y procesos de la Cumbre Mundial sobre la Sociedad de la Información (CMSI) que dirigen las Naciones Unidas.

Utilizando el marco y los indicadores de la universalidad de Internet<sup>1</sup>, en esta publicación se examinan las múltiples repercusiones para la IA y la forma en que los principios DAAM pueden orientar el desarrollo y la utilización de IA a partir de las siguientes dimensiones:

- ▶ Derechos humanos como libertad de expresión, privacidad e igualdad;
- ▶ Apertura respecto del conocimiento, los datos abiertos y los mercados abiertos y pluralistas;
- ▶ Acceso inclusivo respecto de la investigación, los recursos humanos, los datos y el *hardware*:
  - ▶ Gobernanza de múltiples actores;
  - ▶ Temas transversales: igualdad de género y África

En resumen, los resultados del estudio se estructuran de acuerdo con estas seis dimensiones. Cabe señalar que muchas de las ilustraciones incluidas en el texto se refieren a experiencias en determinadas zonas del mundo que están comparativamente avanzadas en lo que respecta al desarrollo y la aplicación de la IA. Esto es consecuencia de la disponibilidad de datos y materiales de investigación secundarios para el estudio, y sirve para alertar a otras zonas geográficas sobre los tipos de problemas que pueden llegar a experimentar. Por lo tanto, el estudio no es un examen global de la IA.



## CONSECUENCIAS EN LOS DERECHOS HUMANOS Y ENFOQUE DE DERECHOS HUMANOS POR DISEÑO PARA EL DESARROLLO DE LA IA

Como parte de la familia de las Naciones Unidas, la UNESCO defiende los derechos humanos. Desde los algoritmos que se diseñan y utilizan para dar forma a la manera en que se muestran nuestras noticias en las redes sociales, hasta los que crean los perfiles de los usuarios y curan la información que reciben, incidiendo así en las opciones de voto en las elecciones, la IA ya influye en todos los derechos humanos, incluidos los relativos a la comunicación y la información, tanto de forma positiva como negativa:

---

1 Véase más información sobre los indicadores de la universalidad de Internet en <https://es.unesco.org/internetuniversality>

## Derecho a la libertad de expresión

- La personalización del contenido en línea por parte de la IA ofrece contenidos y conexiones relevantes. No obstante, puede manipular la forma en que la gente usa su derecho a buscar información y su derecho a formarse una opinión. Esto podría debilitar el pluralismo de ideas y el grado de exposición a la información verificada.
- La IA se está utilizando para restar importancia o eliminar el contenido en línea que es discriminatorio o que incita al odio y la violencia. Con la moderación automatizada de contenido, sin embargo, también se corre el riesgo de bloquear la legítima libertad de expresión, y los niveles de transparencia así como los canales de apelación de reparación existentes son inadecuados.

## Derecho a la privacidad

- Hay violación de la privacidad cuando la IA implica la recopilación de datos opacos, la desanonimización, el intercambio no autorizado de datos por parte de terceros y el seguimiento y la elaboración de perfiles de personas. Sin embargo, la IA también podría ayudar a vigilar las violaciones y los abusos respecto de la privacidad personal.
- La protección de datos basada en el consentimiento y la transparencia es vital en la IA, pero la disponibilidad de esta protección es desigual en todo el mundo y no se ocupa de toda la gama de problemas de privacidad.

## Desarrollo del periodismo y los medios de comunicación (como ejercicios del derecho de expresión)

- La IA se puede usar para fortalecer el periodismo en sus operaciones de recolección, verificación, análisis y distribución de información.
- Sin embargo, el uso de la IA también tiene el efecto secundario de debilitar las instituciones del periodismo y reducir su diversidad al facilitar la migración de la publicidad a intermediarios de Internet ricos en datos. Hay elementos de la IA que también intervienen en muchos ataques digitales contra periodistas, sus dispositivos y sus sitios web.
- La IA puede utilizarse para difundir contenidos falsos fabricados deliberadamente con intención de dañar y, de hecho, se usa para opacar el contenido periodístico amplificando esa desinformación. Sin embargo, la IA también podría utilizarse para ayudar a identificar la coordinación encubierta de campañas en línea.

## Derecho a la igualdad y derecho a participar en la vida pública

- Sigue siendo un reto eliminar el sesgo en los sistemas automatizados de adopción de decisiones, lo que plantea riesgos para el disfrute de los derechos humanos en condiciones de igualdad por parte de las mujeres y los niños, así como de

las minorías, los grupos indígenas, las personas con discapacidad, los grupos que sufren discriminación por su identidad y expresión de género, y las personas económicamente desfavorecidas. Sin embargo, los sistemas de valor explícito y bien entrenados, basados en datos no sesgados, pueden disminuir el riesgo de sesgos humanos en ciertas decisiones.

- La participación en la vida pública ha sido posible gracias a la tecnología de la información y las comunicaciones (TIC), pero al mismo tiempo se han utilizado sistemas de IA para elaborar perfiles y manipular el acceso de los votantes a la información e influir en sus elecciones electorales, mediante la manipulación del comportamiento y la difusión de desinformación por medios tales como la microdeterminación de los usuarios.

## 2

## APERTURA E INTELIGENCIA ARTIFICIAL

La UNESCO aboga por el acceso abierto a la investigación científica, los datos abiertos, los recursos educativos abiertos y la ciencia abierta para garantizar la igualdad de acceso y de oportunidades. Esto tiene por objeto fortalecer el acceso universal a la información, reducir las desigualdades en materia de información y promover la transparencia. La apertura en la IA plantea desafíos y oportunidades.

### Explicabilidad y transparencia para el problema de la “caja negra” de la IA

- El problema de la “caja negra” de los sistemas de IA, entendido como la opacidad en la forma en que los sistemas de IA adoptan decisiones, genera preocupación respecto de la transparencia y la responsabilidad en la adopción de decisiones automatizada. Para abordar el problema de la “caja negra” se han propuesto varias soluciones, tanto técnicas como operativas, en relación con la transparencia en el uso de la adopción de decisiones automatizada y la generación de explicaciones respecto de por qué se han adoptado determinadas decisiones. Pero estas pueden chocar con algunas cuestiones vinculadas a la propiedad intelectual.
- Las normas de divulgación y transparencia son útiles para esclarecer los algoritmos del propósito previsto, pero son insuficientes para resolver el problema de la opacidad de la IA. No obstante, la IA también puede aprovecharse para explicar, al menos en parte, sus propios trabajos, y sus resultados se pueden auditar.

## Datos abiertos

- En oposición a los conjuntos de datos de propiedad exclusiva, los repositorios de datos abiertos juegan un importante papel en la reducción de las barreras de entrada para el desarrollo inclusivo de la IA.
- Los datos abiertos y disponibles, no obstante, generan preocupación con respecto a la privacidad debido a la posible desanonimización de las personas a través de la triangulación basada en distintos conjuntos de datos públicos.

## Mercados abiertos

- Los mercados abiertos y pluralistas son una forma de promover la innovación en el desarrollo de la IA y para la asignación eficiente de recursos.
- Al mismo tiempo, a fin de obtener una mayor cuota de mercado, las empresas pueden optar por prácticas que no se ajusten a los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas y, por consiguiente, apartarse de las prácticas éticas necesarias para el uso seguro y beneficioso de la IA.



## ACCESO E INTELIGENCIA ARTIFICIAL

La capacidad de todas las personas de acceder y contribuir a la información, las ideas y el conocimiento es esencial para las sociedades del conocimiento integradoras. El acceso a la información y el conocimiento puede promoverse aumentando la conciencia de las posibilidades que ofrece la IA entre todas las partes interesadas. Estas posibilidades incluyen el desarrollo de *software* gratuito y de código abierto para mejorar las aptitudes, la cooperación y la competencia, el acceso por parte de los usuarios, la diversidad de opciones y permitir que todos los usuarios puedan elaborar las soluciones que mejor se ajusten a sus necesidades.

### Acceso a la investigación

- Se registra un notorio incremento en la cantidad de publicaciones de investigación sobre IA y tecnologías asociadas. No obstante, también hay una importante diferencia en cuanto a la calidad, y el resultado de las investigaciones varía de un país a otro. Las desigualdades en el acceso a la investigación sobre la IA están aumentando entre los países y las instituciones de investigación.
- Cuando existen, las políticas nacionales y el apoyo internacional a la investigación relacionada con la IA ayudan a fortalecer el producto de la investigación en los países en desarrollo y proporcionan una base a partir de la cual la innovación local puede crecer.

## Acceso a conocimiento, educación y recursos humanos

- El acceso a educación y capacitación para el desarrollo y la implementación de la IA sigue siendo limitado en muchos países. Es necesario reforzar las capacidades y la infraestructura de las instituciones que ofrecen educación y capacitación en materia de IA.
- Los principales centros de investigación y desarrollo atraen talentos de todo el mundo, lo que a menudo causa una fuga de cerebros en algunos países y también una fuga de cerebros desde el sector académico hacia el sector privado. Los esfuerzos orientados a aumentar la disponibilidad de recursos humanos incluyen iniciativas locales para mejorar las aptitudes de los empleados, para ofrecer soluciones de colaboración masiva, aprovechando así un conjunto más amplio de conocimientos para resolver problemas, y para ofrecer plataformas de servicios de IA sin necesidad de realizar costosas inversiones en infraestructura y recursos humanos.
- La gente necesita comprender su propio compromiso con la IA para que las tecnologías sean accesibles a todos. La alfabetización mediática e informacional en relación con la IA y otras tecnologías digitales está lejos de ser universal, pero será necesaria para empoderar e informar a las personas.

## Acceso a datos

- Las empresas de tecnología y los agentes estatales tienen acceso a grandes cantidades de datos de los usuarios y los utilizan para entrenar los algoritmos, pero este acceso desigual a los datos crea barreras de entrada para los nuevos participantes (incluidas las empresas emergentes). Las instituciones académicas y los centros de investigación enfrentan dificultades para acceder a datos de alta calidad que se encuentran disponibles para empresas del sector privado.
- Basado en repositorios de datos abiertos, Data Commons puede permitir el entrenamiento de algoritmos que puedan fortalecer el acceso a los datos para el desarrollo inclusivo de la IA.

## Acceso a conectividad y hardware

- El desarrollo de la IA depende de la disponibilidad de banda ancha, almacenamiento en la nube y *hardware* computacional especial que pueda ejecutar algoritmos en procesadores diseñados para realizar grandes cantidades de cálculos.
- Las soluciones emergentes basadas en la nube, combinadas con conexiones de banda ancha asequibles y universales, reducen la necesidad de grandes gastos generales o inversiones de costo fijo para los usuarios y los desarrolladores de IA más pequeños.

# 4

## ENFOQUE DE MÚLTIPLES ACTORES PARA LA GOBERNANZA DE LA IA

Todas las partes interesadas —desde los gobiernos, el sector privado, la comunidad técnica, las organizaciones intergubernamentales, la sociedad civil y el sector académico hasta los usuarios individuales— se ven cada vez más afectadas por la IA y tienen un interés común en definir la forma en que se gobierna la IA.

Los procesos eficaces de múltiples actores son:

- Inclusivos
- Diversos
- Colaborativos
- Transparentes
- Iguales
- Bien informados
- Flexibles y relevantes
- Seguros y privados
- Responsables y legítimos
- Receptivos
- Oportunos

Los foros para los intercambios de ideas de los múltiples actores en la IA incluyen: debates participativos a nivel legislativo y de políticas regulatorias; marcos de cooperación en IA nacionales, regionales e internacionales, y consultas a empresas de tecnología para elaborar condiciones de servicio y procedimientos operativos.

# 5

## INTELIGENCIA ARTIFICIAL E IGUALDAD DE GÉNERO

La igualdad de género es importante para garantizar que todas las personas, sin discriminación por motivos de sexo, género u orientación sexual, disfruten del derecho a acceder a la sociedad, así como a participar y contribuir a ella. Apoyándose en la vasta literatura sobre la relación entre el género y la tecnología, en este capítulo se reconoce que las tecnologías impulsadas por la IA pueden hacer tanto retroceder como avanzar la lucha por la igualdad de género. La IA plantea desafíos tales como la falta de representatividad de quienes trabajan en este campo, la discriminación algorítmica, la subordinación de los robots sexuales y los asistentes de voz "femeninos", y la pornografía "ultrafalsa", que puede perpetuar los estereotipos de género negativos y poner en desventaja a las mujeres y a las personas LGBTI. Simultáneamente, los miembros de la comunidad de IA han comenzado a pensar soluciones para estos desafíos, algunas de los cuales hacen uso de la IA y tecnologías asociadas.



## INTELIGENCIA ARTIFICIAL Y ÁFRICA

La importancia de la ciencia, la tecnología y la innovación es ampliamente reconocida por los países africanos y forma parte esencial de la visión 2063 de la Unión Africana. Sin embargo, hay grandes problemas de capacidad, infraestructura y gobernanza para la creación de un entorno sólido y propicio para el desarrollo de la IA. Cada vez más gobiernos africanos son conscientes de estos problemas y están poniendo en marcha distintas iniciativas, algunas a través de políticas específicas de IA para empoderar al sector privado, los investigadores y la sociedad civil a fin de que puedan aprovechar la IA para el desarrollo. La velocidad y la escala de las iniciativas hasta la fecha son limitadas. No obstante, muchos agentes del sector privado, la comunidad técnica y la sociedad civil están tratando activamente de hacer frente a los problemas inmediatos de acceso a los conocimientos, las aptitudes, la tutoría y las oportunidades comerciales.

# INTRODUCCIÓN

# 1. CONCEPTUALIZACIÓN DE LA IA EN ESTA PUBLICACIÓN

Si bien no hay una única definición de "inteligencia artificial" (IA), esta publicación se centra en la combinación de tecnologías que permiten lo que la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST) de la UNESCO llama "máquinas capaces de imitar ciertas funcionalidades de la inteligencia humana, incluidas características como la percepción, el aprendizaje, el razonamiento, la resolución de problemas, la interacción mediante el lenguaje y hasta la producción de trabajo creativo" (COMEST, 2019). Esta idea se acerca más al alcance estrecho de la denominada "inteligencia artificial estrecha" (IAE), que significa la capacidad de las máquinas para asemejarse a las capacidades humanas en dominios estrechos, con diferentes grados de sofisticación técnica y autonomía (Artículo 19 y Privacy International, 2018). A los efectos de la simplificación, en ocasiones en esta publicación también se usan los términos IA y toma de decisiones algorítmica/automatizada de manera intercambiable, siempre teniendo en cuenta, no obstante, que la IA propiamente dicha debe entenderse como un complejo más amplio de tecnologías, relaciones y prácticas que incluyen el aprendizaje profundo.

A menudo somos ajenos a las huellas de nuestros datos y a la existencia de algoritmos a nuestro alrededor, en mayor o menor medida, en todo el mundo. Mientras tanto, la IA o sus elementos se despliegan cada vez más, como lo demuestran los siguientes ejemplos:

- Los algoritmos de los motores de búsqueda nos ayudan a acceder a la información que queremos consultando rápidamente datos en la World Wide Web, y los resultados de búsqueda tienden a ser cada vez más personalizados en virtud de la ubicación, el sexo, el idioma, el historial de búsqueda y otros rastros de datos de un usuario.
- Los algoritmos de coincidencia de trabajo analizan las competencias de las personas para mostrar a los empleadores candidatos adecuados para un empleo;
- Las plataformas de video a demanda brindan sugerencias de películas especialmente diseñadas basándose en nuestros patrones de consumo de películas y en los de otros millones de usuarios, y ofrecen a los anunciantes la capacidad de predecir y alentar nuestras actitudes y acciones;
- Los algoritmos ayudan a los jueces a determinar la posibilidad de reincidencia y sugieren el tiempo para las sentencias de cárcel;
- Los algoritmos de riesgo crediticio deciden a quién se debe ofrecer un préstamo y en qué condiciones;
- Las autoridades de inmigración usan los perfiles digitales para aprobar o rechazar solicitudes visa.

En algunos ámbitos se recurre a la frase "Cuarta Revolución Industrial" para describir la importancia de la IA y otras tecnologías avanzadas. No obstante, el término es cuestionado, y no forma parte del vocabulario de este estudio, que va mucho más allá de las cuestiones vinculadas a la relación entre la tecnología y los factores económicos.

## 2. ¿POR QUÉ LA UNESCO ESTÁ INTERESADA EN LA IA?

Dada la amplia aplicación de la IA en los ámbitos de trabajo de la UNESCO (entre ellos la educación, las ciencias, la cultura, el acceso a la información, la libertad de expresión y la ética), la UNESCO tiene un importante papel que desempeñar en estos tiempos de cambio. La Directora General de la UNESCO, Audrey Azoulay, ha destacado que "la humanidad está en el umbral de una nueva era" y que la "transformación ya ha comenzado" (Azoulay, 2018).

La UNESCO también juega un papel fundamental en el fomento de mecanismos de múltiples actores para proteger los derechos humanos, asegurar la apertura al conocimiento y la investigación, y reducir las desigualdades en sus 193 Estados Miembros y entre ellos para promover la inclusión. Las divisiones tecnológicas, incluso en la IA, pueden tener un efecto multiplicador en las desigualdades sociales y complicar la aspiración del desarrollo sostenible de no dejar a nadie atrás.

El Sector de Comunicación e Información de la UNESCO promueve la libertad de expresión, de conformidad con el mandato constitucional de la Organización de promover "la libre circulación de las ideas por medio de la palabra y de la imagen". Para llevar a cabo esta labor, el Sector de Comunicación e Información de la UNESCO adopta las siguientes dos líneas de acción principales:

- Fomentar la libertad de expresión en línea y fuera de línea, promover la seguridad de los periodistas, fomentar la diversidad y la participación en los medios de comunicación, y apoyar a los medios independientes.
- Construir sociedades del conocimiento mediante las TIC, posibilitando el acceso universal a la información y el conocimiento y su preservación.

Para la UNESCO, las sociedades del conocimiento se afirman en los pilares de libertad de expresión, acceso a la información, educación y diversidad cultural y lingüística. Promueven el conocimiento mediante el aprovechamiento de las TIC con el objetivo de mejorar el acceso a la educación, así como el conocimiento científico y la innovación, y de potenciar a las comunidades locales y marginadas. La UNESCO trabaja dentro de la amplia familia de las Naciones Unidas para promover sociedades del conocimiento capaces de afrontar el reto de cumplir con la Agenda 2030 para el Desarrollo Sostenible, que es de carácter universal.

## Recuadro 1 : Mandato de la UNESCO en la Cumbre Mundial sobre la Sociedad de la Información (CMSI)

La aproximación de la UNESCO a la IA se basa en ideas de la Cumbre Mundial sobre la Sociedad de la Información (CMSI) y su seguimiento, incluida la promoción por parte de la Organización de un enfoque de las tecnologías de la información y las comunicaciones (TIC) basado en los derechos humanos y centrado en el ser humano durante los eventos de la CMSI en Ginebra (2003) y Tunis (2005). También es relevante el proceso de seguimiento de la CMSI donde la UNESCO es el organismo de las Naciones Unidas responsable de dirigir/facilitar la implementación de las Líneas de Acción sobre el Acceso a información y conocimiento (C3); aprendizaje electrónico (C7); ciberciencia (C7); diversidad cultural e identidad, diversidad lingüística y contenido local (C8); medios de comunicación (Cg), y dimensiones éticas de la sociedad de la información (C10).

Como se establece en la Declaración de Principios de Ginebra de la CMSI (2003): "El uso de las TIC y la creación de contenidos debería respetar los derechos humanos y las libertades fundamentales de otros, lo que incluye la privacidad personal y el derecho a la libertad de opinión, conciencia y religión, de conformidad con los instrumentos internacionales relevantes".

En la Revisión de la CMSI en 2015 se reafirmó el deseo compartido de "crear una Sociedad de la Información centrada en la persona, integradora, abierta y orientada al desarrollo, en que todos puedan crear, consultar, utilizar y compartir la información y el conocimiento", permitiendo así que las personas desplieguen todo su potencial. En el documento final de la CMSI+10 también se pone énfasis en la importancia de los principios de derechos humanos y la cooperación entre múltiples actores para construir sociedades del conocimiento integradoras.

Estos debates son pertinentes en el contexto de la IA y requieren un firme compromiso de todas las partes interesadas para garantizar que el desarrollo de la IA se lleve a cabo de manera que se respeten los derechos humanos, la apertura, el acceso a la información y la participación de las múltiples actores.

### 3. LA REFLEXIÓN EN CURSO DE LA UNESCO SOBRE LA IA

A fin de dar sustancia al diálogo global sobre cómo aprovechar el potencial de la IA para alcanzar los ODS, la UNESCO organizó una serie de reuniones en 2018 y 2019. Estas incluyen el panel de discusión sobre "Respuesta a las oportunidades y los cambios de la era digital", durante el Foro de Asociados de la UNESCO; una mesa redonda sobre "Inteligencia artificial: reflexión sobre su complejidad e impacto en nuestra sociedad"; el debate sobre "IA para los derechos humanos y los ODS: promoción de enfoques de múltiples actores, integradores y abiertos", realizado como parte del Foro para la Gobernanza de Internet (FGI); un debate abierto sobre "Aprovechar la inteligencia artificial para promover las sociedades del conocimiento y la buena gobernanza" realizado en la Fundación Mozilla; y un panel sobre Reflexión filosófica sobre la IA durante el Día Mundial de la Filosofía.

La reflexión en curso de la UNESCO sobre el debate mundial en torno a la ética de la IA, centrado en las normas y los estándares, también quedó demostrada en el Foro sobre Inteligencia Artificial en África, que tuvo lugar en diciembre de 2018 en Benguerir (Marruecos). Al término del foro, los participantes aprobaron por unanimidad la Declaración de Benguerir, donde se pusieron de acuerdo respecto de la necesidad de promover una estrategia de IA para África, así como una IA centrada en el ser humano. La estrategia global de la UNESCO respecto de la IA se presentó a los Estados Miembros en una Reunión Informativa celebrada en enero de 2019, que fue seguida de un debate de expertos internacionales sobre "Futuros Tecnológicos: ¿Esperanza o Temor?". Esta reflexión culminó en la conferencia mundial "Principios de la inteligencia artificial: ¿hacia un enfoque humanista?" celebrada en marzo de 2019 en la Sede de la UNESCO en París.

La UNESCO también organizó una sesión sobre el Aprovechamiento de la inteligencia artificial para fortalecer el desarrollo del periodismo y los medios de comunicación, de conformidad con los principios DAAM de la universalidad de Internet de la UNESCO (véase debajo), en el marco de su proceso de facilitación de la Línea de Acción Cg (medios de comunicación) de la CMSI en la edición 2019 del Foro de la CMSI<sup>1</sup>. En el primer semestre de 2019 se presentó al Consejo Ejecutivo de la UNESCO un estudio realizado por el programa COMEST de la UNESCO, al que contribuyó el Sector de Comunicación e Información.

---

1 Enlace al evento [en inglés]: <https://www.itu.int/net4/wsis/forum/2019/Agenda/ViewSession/304#>.

## 4. LA UNESCO ABORDA LA IA Y LAS TIC AVANZADAS A TRAVÉS DE LA PERSPECTIVA DAAM

Si bien se reconoce que la IA es un conjunto de elementos que no son sinónimos de Internet, en esta publicación, la aplicación y el desarrollo inextricables de la IA se sitúan dentro del ecosistema de esta red de redes y la forma en que ha evolucionado su contexto social, político y económico.

En el primer período ordinario de sesiones de 2019 de la Junta de los Jefes Ejecutivos (JJE) del Sistema de las Naciones Unidas para la Coordinación se destacó el enfoque estratégico y la hoja de ruta de todo el sistema de las Naciones Unidas para apoyar el desarrollo de capacidades en materia de inteligencia artificial:

“La inteligencia artificial debe abordarse de manera ambiciosa y holística, promoviendo su uso como herramienta para implementar los Objetivos, y abordando también los derechos humanos y éticos emergentes, el trabajo decente y los desafíos técnicos y socioeconómicos”.

(CEB/2019/1/Add.3.2019)

Este es el motivo por el cual, como se estableció en el informe del Panel de Alto Nivel del Secretario General sobre la Cooperación Digital (2019), la UNESCO ha utilizado su marco DAAM (derechos, apertura, acceso y múltiples actores) para discutir las implicancias de la IA para los derechos, incluida la libertad de expresión, la privacidad, la igualdad y la participación en la vida pública.



El marco DAAM constituye la base del concepto de universalidad de Internet de la UNESCO, que fue aprobado por la 38° Conferencia General de la Organización celebrada en 2015<sup>2</sup>. Estos principios son derechos humanos, apertura, accesibilidad y participación de múltiples actores, y surgen de las piedras angulares del mandato de los Estados Miembros para promover el estudio de sociedades del conocimiento integradoras por parte de la UNESCO, que abarcaba los temas de privacidad, libertad de expresión, acceso y ética de una Internet global (2015a). El documento final "Conectando los puntos", también aprobado en la 38° Conferencia General de la UNESCO, la Organización se comprometió a promover una reflexión ética basada en los derechos humanos, la investigación y el diálogo público sobre las implicancias de las tecnologías nuevas y emergentes, y su posible impacto en la sociedad.

Sobre la base de estos acontecimientos, la UNESCO emprendió un proceso de tres años de elaboración del marco de la universalidad de Internet mediante un proceso de carácter global, abierto, inclusivo y de múltiples actores.

El resultado es el marco de indicadores DAAM-X de la universalidad de Internet, que fue recibido con beneplácito por los Estados Miembros en el Programa Internacional para el Desarrollo de la Comunicación (PIDC) de la UNESCO en noviembre de 2018 (UNESCO, 2019d; Consejo del PIDC CI/2018/COUNCIL.31/H/1). Los 303 indicadores se encuentran organizados en cinco categorías, cuatro de las cuales reflejan los cuatro principios DAAM. La quinta categoría abarca cuestiones intersectoriales como la igualdad de género y las necesidades de los niños y los jóvenes, las dimensiones económicas, la confianza y la seguridad, y los aspectos jurídicos y éticos de Internet.

El enfoque basado en DAAM sirve entonces de marco bien fundamentado y holístico para que la UNESCO y otras partes interesadas contribuyan al diseño, la aplicación y la gobernanza de la IA. El marco puede enriquecer las actividades para el establecimiento de principios normativos y éticos para la IA, y para la producción de directrices políticas y conjuntos de instrumentos innovadores.

Esto es relevante para el mandato de la 206° reunión del Consejo Ejecutivo de la UNESCO<sup>3</sup> (3 al 17 de abril de 2019, París) y el 40° Período de Sesiones de la Conferencia General (12 al 27 de noviembre de 2019, París), que reconoce que una recomendación podría ser un instrumento esencial para fortalecer la elaboración y aplicación de la legislación, las políticas y las estrategias nacionales e internacionales en la materia, así como para mejorar la cooperación internacional en el desarrollo y el uso ético de la IA en apoyo de los ODS. En ocasión de la 40° Conferencia General también se decidió que es oportuno y pertinente que la UNESCO prepare un instrumento normativo internacional sobre la ética de la inteligencia artificial en forma de recomendación y que el texto preliminar de una recomendación sobre la ética de la inte-

---

2 Enlace a la Conferencia General de la UNESCO [en inglés]: [https://en.unesco.org/system/files/private\\_documents/2340goe-1\\_0.pdf](https://en.unesco.org/system/files/private_documents/2340goe-1_0.pdf).

3 Enlace al proyecto de resolución de la 40° Conferencia General, Preliminary study on a possible standard-setting instrument on the ethics of artificial intelligence, aprobado el 21 de noviembre de 2019 [en inglés]: <https://unesdoc.unesco.org/ark:/48223/pf0000369455/PDF/369455esp.pdf.multi>.

ligencia artificial se debía someter a la consideración del 41° Período de Sesiones de la Conferencia General (UNESCO, 2019b) (206 EX/42). Esto se ajusta al papel de la UNESCO de brindar asesoramiento técnico a los Estados Miembros y otros actores, actuando como un centro de intercambio de información para la innovación y creando capacidad. De este modo, la ética de la IA, como se informara en los principios DAAM, puede ayudar a contribuir al beneficio de la humanidad, el desarrollo sostenible y la paz.

Desde el punto de vista de todo lo expuesto, este estudio se presenta de la siguiente manera:

El capítulo sobre **Derechos** se evalúa cómo se utiliza la IA para dar forma a la personalización de contenidos y cómo esto afecta la libertad de expresión y la libertad de opinión. También se examina la moderación de contenidos mejorados por la IA y la amenaza que representan para la libertad de expresión. La sección sobre privacidad analiza de qué manera la IA cuestiona los enfoques actuales de la privacidad. Amplía el debate de las leyes de privacidad que rigen la interacción entre los ciudadanos y las entidades jurídicas a las normas de privacidad que rigen las interacciones entre las personas por medio de algoritmos. Desde el periodismo de datos hasta la desinformación automatizada, el capítulo también explora los temas que se plantean en la intersección de la IA y el desarrollo del periodismo y los medios de comunicación. En la última sección, sobre el derecho a la igualdad, se descifra la palabra "discriminación", que se suele utilizar en el contexto de la evaluación de la adopción de decisiones por medios algorítmicos, diferenciando entre formas de discriminación directas, indirectas, institucionales, intencionales y no intencionales.

El capítulo sobre **Apertura** introduce el contexto más amplio de apertura en la IA y presenta un análisis de sus impulsores. Entre otras cosas, se explican las limitaciones a la hora de abordar el problema de la caja negra y la transparencia en procesos de toma de decisiones algorítmica. Como nota de cautela sobre la apertura y sus posibles riesgos, el capítulo invita a una reflexión más amplia sobre el uso incorrecto de la IA para causar daños potenciales. En el contexto del movimiento de datos abiertos, se presentan diferentes modos de recopilación de datos y los desafíos asociados a la disponibilidad de datos para el aprendizaje automático. El capítulo examina el papel de los mercados abiertos y pluralistas en la difusión de la innovación en la IA, al tiempo que marca el posible descuido de los derechos humanos en el contexto de la competencia comercial.

En el capítulo sobre **Acceso** se identifican elementos clave como el acceso a los algoritmos y a los datos de investigación, y los recursos humanos y computacionales disponibles para el desarrollo y la aplicación de la IA. El debate sobre cada elemento traza un mapa de la situación actual y señala las cuestiones que merecen atención si queremos mitigar las nuevas brechas digitales.

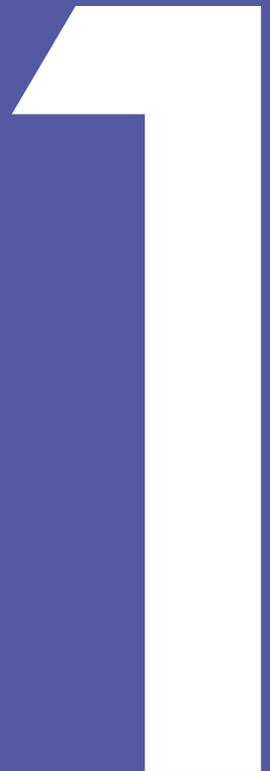
En el capítulo sobre la gobernanza de la IA con **múltiples actores** se presenta la evolución histórica de la labor y los instrumentos de la UNESCO en apoyo al diálogo entre múltiples actores, desde los gobiernos, las empresas, la comunidad técnica, las organizaciones intergubernamentales, la sociedad civil y el mundo académico.

Estos grupos se ven cada vez más afectados por la IA. En este capítulo también se comparten los valores, prácticas e indicadores conexos necesarios para fomentar y poner en práctica un enfoque de múltiples actores en la gobernanza de la IA.

El capítulo sobre **Igualdad de género** e IA se apoya en la vasta literatura sobre la relación entre el género y la tecnología. Reconoce que las tecnologías impulsadas por la IA pueden hacer tanto retroceder como avanzar la lucha por la igualdad de género.

En el capítulo sobre África se destaca que se reconoce la gran importancia que la ciencia, la tecnología y la innovación (CTI) tienen en los países africanos en tanto vía para crecer y desarrollarse. Sin embargo, hay grandes problemas de capacidad, infraestructura y gobernanza para la creación de un entorno sólido y propicio para el desarrollo de la IA. En el marco de la Estrategia de CTI de la Unión Africana, el capítulo describe algunas de las iniciativas adoptadas por los gobiernos, el sector privado, la comunidad técnica y la sociedad civil para fomentar el desarrollo de la IA en África.

DERECHOS  
HUMANOS E  
INTELIGENCIA  
ARTIFICIAL





# CAPÍTULO 1: DERECHOS HUMANOS E INTELIGENCIA ARTIFICIAL

La inteligencia artificial (IA) y los procesos de toma de decisiones algorítmica/automatizada relacionados están cada vez más incrustados en el tejido de las sociedades conectadas, adelantándose a una comprensión clara y subdesarrollada de las consecuencias para los derechos humanos. El diseño, la creación y el uso de la IA y las tecnologías conexas ofrecen oportunidades para mejorar el acceso a los derechos consagrados en la Declaración Universal de Derechos Humanos, para construir sociedades del conocimiento inclusivas y para alcanzar los Objetivos de Desarrollo Sostenible (ODS). Sin embargo, como ocurre con otros avances científicos y tecnológicos, el uso actual y la evolución futura de la IA también podría tener consecuencias negativas para los derechos y las libertades fundamentales, que deberían contrarrestarse o mitigarse (Access Now, 2018).

Este capítulo no pretende ser un estudio exhaustivo de todos los posibles impactos de la IA en los derechos humanos, sino que aborda algunas de las principales preocupaciones sobre la forma en que el uso de las tecnologías de IA puede obstaculizar los derechos humanos en el ámbito del mandato de la UNESCO. A medida que la tecnología avanza, estamos obligados a descubrir nuevos beneficios y riesgos para la protección y el disfrute de los derechos humanos.

## **Recuadro 2: Posición de la UNESCO respecto de los derechos humanos para la universalidad de Internet**

Al identificar la conexión de Internet con las normas basadas en los derechos humanos como elementos constitutivos de la libertad, la "universalidad de Internet" contribuye a poner de relieve la continua armonía entre el crecimiento y el uso de Internet y los derechos humanos. En este sentido, una Internet libre significa una Internet que respeta y permite la libertad de ejercer los derechos humanos. A este respecto, la universalidad

dad de Internet nos alienta a considerar la gama de interdependencias e interrelaciones entre Internet y los diferentes derechos humanos, como los derechos a la libertad de expresión, la privacidad, la diversidad de expresiones culturales, la participación pública y la asociación, la igualdad de género, la seguridad y la educación. La IA debe considerarse en el contexto holístico de Internet y los derechos humanos.

La IA y los derechos humanos están relacionados con los principios e indicadores de la universalidad de Internet de la UNESCO, que a su vez se mencionan en una Resolución del Consejo de Derechos Humanos de las Naciones Unidas de 2018 (A/HRC/38/L.10/Rev.1) sobre la promoción, la protección y el disfrute de los derechos humanos en Internet, en la que se destaca, en particular, la libertad de expresión y la privacidad en línea.

En este informe también se explora el potencial de la IA para respaldar al periodismo independiente y promover unos medios de comunicación libres, pluralistas e independientes. Esto se ajusta a la definición de libertad de expresión garantizada por el artículo 19 de la Declaración Universal de Derechos Humanos, que establece que todo individuo tiene derecho a la libertad de opinión y de expresión, lo que incluye el acceso a la información y a las ideas a través de cualquier medio de comunicación. La IA y las tecnologías relacionadas ya están dando forma a ciertos aspectos de las noticias, lo que repercute en las ideas del valor de los periodistas, la práctica del periodismo y la producción de otros tipos de contenidos.

No solo es importante que todas las partes interesadas reflexionen sobre los retos, sino que también formulen respuestas que protejan la libertad de expresión, la privacidad, el periodismo y los medios de comunicación, así como los derechos a la igualdad y la participación política. En el presente capítulo se examinará un conjunto de opciones para maximizar los beneficios y reducir al mínimo los riesgos que plantea IA para los derechos humanos.

### **Recuadro 3: La UNESCO y la libertad de expresión**

La UNESCO es el organismo especializado de las Naciones Unidas que tiene el mandato de defender la libertad de expresión, como lo establece su Constitución, de promover "la libre circulación de las ideas por medio de la palabra y de la imagen". La UNESCO reconoce que el derecho a la privacidad es la base de otros derechos y libertades, incluida la libertad de expresión, asociación y creencias. Asimismo, la Organización reconoce a la libertad de expresión como el pilar clave para construir sociedades del conocimiento.

# 1. DERECHO A LA LIBERTAD DE EXPRESIÓN

La libertad de expresión, tanto en línea como fuera de línea, desempeña un papel fundamental en las sociedades del conocimiento. La capacidad de expresar su propia opinión en la esfera pública es un componente esencial para que las personas puedan participar en los debates públicos. Tener acceso a un medio de expresión es "una condición necesaria para la participación en el proceso político del país" (Scanlon, 1972). De igual modo, la libertad de expresión es importante como forma de expresión personal del orador, lo que forma parte de la autorrealización individual (Gilmore, 2011). De este modo, además de permitir la participación social y política, la libertad de expresión también es un medio crucial de realización personal (Cannataci y otros, 2016).

La libertad de expresión y la libertad de opinión están estrechamente vinculadas. La Declaración Universal de Derechos Humanos establece que todo individuo tiene "derecho a la libertad de opinión y de expresión", y, de manera similar, el Artículo 19 del Pacto Internacional de Derechos Civiles y Políticos hace referencia al derecho a tener opiniones y el derecho a la libertad de expresión como derechos distintos pero adyacentes. El Comité de Derechos Humanos de las Naciones Unidas, en su Observación General núm. 34 (2011), aclara el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos al afirmar que la libertad de opinión y la libertad de expresión son condiciones indispensables para el pleno desarrollo de la persona y son los cimientos de toda sociedad libre. El derecho a la libertad de expresión en la Declaración Universal de Derechos Humanos y en el Pacto Internacional de Derechos Civiles y Políticos se afirma en direcciones complementarias: a buscar y recibir información e ideas, y a impartir información e ideas (de Zayas y Martin, 2012). En general, las personas tienen un derecho reconocido a intercambiar ideas, a informarse y a formarse y desarrollar opiniones personales.

Las tecnologías de la información y las comunicaciones (TIC) generalmente tienen el potencial de "permitir que un público mundial busque, reciba y difunda información e ideas y otros contenidos en particular para adquirir conocimientos, entablar debates y participar en procesos democráticos" (CoE CM/Rec(2012)3, 2012). De hecho, Internet facilita una estructura de comunicación en la que cada usuario tiene la posibilidad de buscar información y hacer valer su voz. Esto diferencia a Internet de la "estructura unidireccional de uno a muchos de los medios de comunicación tradicionales" (Hansen, 2018). La IA y las TIC avanzadas pueden ayudar a promover la libertad de expresión y también plantean problemas en diferentes dimensiones de este derecho fundamental.

Por ejemplo, los motores de búsqueda de Internet, respaldados por algoritmos de IA, son controladores cruciales para las personas que quieren buscar, recibir y difundir información (MSI-NET, 2016). Con sus algoritmos de clasificación mejorados por la IA, los motores de búsqueda son útiles para proporcionar enlaces a informa-

ción que de otro modo sería desconocida o inaccesible. Sin embargo, la vigilancia de los motores de búsqueda así como de las plataformas de redes sociales nunca puede ser completamente neutral. En muchos casos, los algoritmos dan visibilidad a la desinformación, la incitación al odio, etc., lo que a su vez puede afectar al derecho de las personas a formar sus opiniones de forma independiente (Solon y Levin, *How Google's search algorithm spreads false information with a rightwing bias*, 2016). En lo que respecta a muchas cuestiones relacionadas con la IA y los derechos humanos, las partes interesadas deben encontrar el equilibrio adecuado para beneficiarse de la IA y, al mismo tiempo, proteger las diferentes dimensiones de la libertad de expresión.

## 1.1. Información personalizada, libertad de opinión y el derecho a buscar información

Si bien las violaciones del derecho a impartir información son objeto de un amplio debate (es decir, la censura – véase el párrafo 1.2 infra), no se debe restar importancia a la dimensión de la búsqueda y recepción de información e ideas. El derecho a recibir información es un componente esencial para el ejercicio del derecho a la libertad de opinión, como se consagra en el Artículo 19 de la Declaración Universal de Derechos Humanos. David Kaye, Relator Especial de las Naciones Unidas, señaló que un elemento esencial del derecho a sostener una opinión es el derecho a formarse una opinión, y esto "requiere que la persona esté libre de coacción indebida en el desarrollo de sus creencias, ideologías, reacciones y posiciones" (Asamblea General de las Naciones Unidas, A/73/348, 2018). Si bien las normas internacionales sobre el derecho a buscar información permiten ciertas restricciones, la libertad de opinión no puede restringirse.

La IA se utiliza para influir en la forma en que las personas acceden a la información en línea. Los algoritmos de aprendizaje automático de los motores de búsqueda están diseñados para personalizar el contenido que se muestra al usuario. La forma en que se organizan las fuentes de las redes sociales también está dictada por el uso de estas predicciones algorítmicas (Flaxman, Goel y Rao, 2016). Estos algoritmos desplegados deciden lo que la gente ve y en qué orden. Combinando el historial de navegación de los usuarios, la geolocalización, "datos demográficos del usuario [y] análisis semánticos y de sentimientos y muchos otros factores", estos modelos algorítmicos se ponen en funcionamiento para personalizar la información a la que se da prioridad (Asamblea General de las Naciones Unidas A/73/348, 2018).

Esta experiencia personalizada tiene algunas ventajas, ya que puede aportar información pertinente, adaptada a las necesidades de cada persona (por ejemplo, un anuncio de un artículo que el usuario estaba pensando comprar). El contenido también puede estar disponible en la lengua principal de la persona.

Sin embargo, esta conformación automatizada del tipo de información a la que las personas tienen acceso mediante el uso de la IA o sus elementos nunca es puramente técnica, y su impacto puede ser distorsionar nuestra capacidad de conocer una serie de informaciones y opiniones, o qué valores editoriales o empresariales sustentan lo que se prioriza para ellos (Andersen, 2018). Como ya se ha dicho, la libertad de

expresión y de opinión es lo que permite la participación social y política, y actúa como medio de realización personal. Pero ambos aspectos pueden verse socavados por la personalización de los contenidos. A su vez, este uso de la tecnología puede conducir a la fragmentación de la esfera pública, así como al posible debilitamiento de la agencia individual y el autodesarrollo consciente.

Se ha dicho que esos posibles efectos negativos son el resultado de las burbujas de filtro (Pariser, 2011) y las cámaras de eco, dos conceptos relacionados con la personalización del contenido. Las burbujas de filtro se refieren a la forma en que las predicciones algorítmicas de las preferencias del usuario limitan el alcance de la información disponible para una persona (Flaxman, Goel y Rao, 2016). Los algoritmos se establecen para predecir el tipo de contenido que a un usuario le resultará interesante y mostrar ese contenido. Al hacerlo, excluyen otros contenidos sociales y políticos, socavando así la capacidad de una persona para encontrar ciertos tipos de información y opiniones. De este modo, dos usuarios distintos pueden hacer una solicitud idéntica en el mismo motor de búsqueda y recibir resultados diferentes, lo que les aporta información distinta. Esa personalización también se produce en las fuentes de noticias de los medios de comunicación social, determinando el orden y la visibilidad de las publicaciones. En este sentido, las características de los datos de un usuario y su actividad pasada han empezado a dictar los límites de su visión del mundo (Bezemek, 2018).

Las burbujas de filtro están estrechamente ligadas a la noción de cámaras de eco. Al definir el alcance de la información disponible en parte sobre la base del uso que el usuario hizo de Internet con anterioridad, las técnicas algorítmicas impulsadas por IA pueden reforzar "sus opiniones políticas anteriores debido a la exposición selectiva al contenido político" (Colleoni, Rozza y Arvidsson, 2014). Por consiguiente, la experiencia de Internet puede convertirse en una cámara de eco en la que se reafirma la orientación política. La información encontrada parece legitimar los puntos de vista y las opiniones que ya existen, presentándolos como hechos, creando así un entorno en el que los usuarios solo encuentran información que confirma lo que ya piensan (Sibal, 2016). Esta personalización del contenido también puede utilizarse para "empujar" a las personas hacia los extremos, e incluso contribuir al proceso de radicalización hacia el extremismo violento (Tschan y Bekkoenova, 2018).

Bill Gates, fundador de Microsoft Corporation y filántropo, ha expresado su preocupación por el hecho de que las burbujas de filtro se hayan convertido en un problema mayor de lo que muchos, incluido él mismo, habrían esperado, ya que "te permite estar con gente de ideas afines, por lo que no te mezclas, ni compartes, ni entiendes otros puntos de vista" (Delaney 2017). En este sentido, las burbujas de filtro y las cámaras de eco acentúan la fragmentación de la esfera pública (MSI-NET, 2018)<sup>1</sup>. El riesgo se ha resumido en términos de evaluación del desarrollo de algoritmos para

---

1 Documento MSI-NET – *Algorithms and human rights* – Study on the human rights dimensions of automated data processing techniques and possible regulatory implications (2018), disponible [en inglés] en: <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>.

"[...] cogobernar o codeterminar lo que se puede encontrar en Internet [...] se ve y se encuentra (aplicaciones de búsqueda, filtrado y agregación), se produce (aplicaciones de producción de contenido como periodismo algorítmico), se considera relevante (aplicaciones de búsqueda y puntuación; clasificación), se prevé (aplicaciones de pronóstico/predicción), y se elige o consume (aplicaciones de recomendación, puntuación y asignación; para opciones tanto económicas como sociales – que van desde bienes comerciales hasta amigos y socios)", [...] incidir en el comportamiento de productores y usuarios individuales [...]"

(Just y Latzer, 2017)

Aunque la toma de decisiones algorítmica sobre el contenido al que se accede no se traduce automáticamente en la determinación de las opiniones de las personas, es importante reconocer el papel que pueden desempeñar en la configuración de la agenda pública y otros elementos de la ecosfera de la información (Davies, 2018). La personalización de la IA puede amenazar la libertad de pensamiento en la medida en que determina el tipo de información a la que las personas tienen acceso (Andersen, 2018). El Consejo de Europa ha advertido, de hecho, que el uso de procesos algorítmicos y aprendizaje automático puede influir en las emociones y los pensamientos de las personas, a veces de forma subliminal (CoE Decl(13/02/2019)1, 2019).

No siempre se ofrecen a los usuarios alternativas a la personalización de la IA y el establecimiento de objetivos algorítmicos, como cuando pueden elegir conscientemente ver la información priorizada por fecha, fuente creíble u otras prioridades que puedan tener, a diferencia de otras lógicas integradas a nivel de los cálculos. La agenda no la establecen los medios de comunicación en la esfera pública con un enfoque editorial humano, sino que la establecen tecnologías con lógicas ocultas que a menudo se diseñan con el objetivo comercial de reunir datos personales y vender el acceso a los usuarios y sus datos. Lo que esto significa, en última instancia, es que los algoritmos de personalización de contenidos inciden en las oportunidades que se nos presentan, limitando también el alcance de las posibilidades que nos definen (Rouvroy, 2014).

## 1.2. Moderación de contenidos en línea por parte de la IA y el derecho a impartir información

En muchos países, las plataformas de Internet representan un lugar central en el que se celebran debates públicos (Latonero, 2018). Una proporción importante del discurso y la expresión tiene lugar en línea (Balkin, 2017). Por lo tanto, no es de extrañar que Internet sea un espacio en el que, entre otras expresiones, también hay desinformación, discurso de odio y propaganda de la violencia y la guerra. En los últimos años, las plataformas de Internet han confiado en la IA para moderar el contenido publicado en línea. Mediante prácticas como "la detección de correo basura, la tecnología de comparación de *hash*, los filtros de palabras clave, el procesamiento de lenguas naturales y otros algoritmos de detección" (Asamblea General de las Naciones Unidas A/73/348, 2018), las empresas de redes sociales y otras

compañías de Internet pueden eliminar o “bajar de rango” el contenido percibido como “indeseable”.

El artículo 19 del Pacto Internacional de Derechos Civiles y Políticos garantiza el derecho a difundir ideas de todo tipo a través de cualquier medio. Sin embargo, también establece que se pueden aplicar ciertas restricciones cuando estas se ajusten a la ley y sean necesarias para respetar los derechos y la reputación de los demás, además de la protección de la seguridad nacional, el orden público o la salud o la moral públicas. El artículo 20 del Pacto Internacional de Derechos Civiles y Políticos describe los casos en los que se espera que la libertad de expresión sea restringida de conformidad con la ley. Estas condiciones incluyen la propaganda en favor de la guerra y la apología del odio nacional, racial o religioso que constituya una incitación a la discriminación, la hostilidad o la violencia (Resolución 2200A(XXI) de la Asamblea General de las Naciones Unidas, 1966).

A la luz de estas consideraciones, en las secciones siguientes se examinan las ventajas de utilizar la IA para la moderación de contenidos en línea, los procesos que siguen los intermediarios de Internet para la moderación de contenidos y los retos asociados.

Los algoritmos de IA que filtran y señalan formas de expresión violentas, discriminatorias o que inducen al odio responden a un objetivo legítimo, ya que es obligación impedir la difusión de ese tipo de mensajes. Se ha planteado que con menos propaganda terrorista y menos expresiones de odio en línea podría haber menos personas que recurran al extremismo violento (Citron, 2017), aunque también es probable que ese contenido en línea sea el único, o incluso el principal, impulsor de esa conducta (Alava, Frau-Meigs y Ghayda, 2017).

La IA puede ayudar a responder a la escala del contenido, complementando los esfuerzos humanos. Otro posible beneficio de utilizar IA para la moderación de contenidos en lugar de moderadores humanos radica en el considerable “costo psicológico que supone el trabajo en la primera línea de revisión y moderación de contenidos”, ya que los revisores humanos están “expuestos, día tras día, a lo peor de la humanidad”, desde contenido sobre abuso infantil hasta horribles actos de violencia (Raso, Hilligoss, Krishnamurthy, Bavitz y Kim, 2018). La IA y las tecnologías relacionadas ayudan a reducir la carga de ver este tipo de contenido.

Estas razones hacen que el uso de IA se justifique en la moderación de contenidos en línea. No obstante, el uso de esas medidas técnicas limitadas debe respetar las limitaciones internacionales en relación con las posibles restricciones a la libertad de expresión. Muchos observadores temen que la IA actualmente no tenga el nivel de desarrollo necesario para diferenciar entre la cobertura de noticias, la promoción y la sátira por un lado, y la incitación real al daño por el otro. El papel que desempeñan las empresas de Internet al optar por utilizar la IA en la moderación del contenido plantea otros desafíos (como se explica en la sección 1.4 más adelante), y deberían implementarse mecanismos en caso de que la IA se extralimitara en la difusión que violara la expresión legítima.

### 1.3. Plataformas de Internet bajo presión para regular los contenidos en línea

Desde las empresas de redes sociales hasta los motores de búsqueda, pasando por las empresas con fines de lucro y las organizaciones sin fines de lucro como Wikipedia, una amplia gama de plataformas de Internet se enfrentan a la presión de los gobiernos nacionales y las autoridades regionales para que regulen el contenido publicado en sus plataformas (Citron, 2017). Algunos gobiernos les exigen que vigilen y eliminen el contenido sin esperar a que las autoridades nacionales lo soliciten por ley. El contenido terrorista y extremista, los discursos de incitación al odio y las "noticias falsas" se califican de inaceptables. Sin embargo, las definiciones suelen ser débiles. Según el Relator Especial David Kaye, las leyes restrictivas con una redacción de carácter muy general sobre el "extremismo", la blasfemia, la difamación, el discurso "ofensivo", las "noticias falsas" y la "propaganda" suelen servir como pretexto para exigir que las empresas supriman la expresión legítima (Asamblea General de las Naciones Unidas A/HRC/38/35, 2018).

La presión sobre los intermediarios de Internet "va desde la regulación directa, pasando por las amenazas, las sugerencias de que las cosas irán mejor para los operadores de infraestructura si cooperan, hasta las negociaciones sobre los términos de la cooperación" (Balkin, 2017). El Consejo Europeo espera explícitamente que las plataformas de Internet "desarrollen nuevas tecnologías e instrumentos para mejorar la detección y eliminación automáticas de contenidos que inciten a la comisión de actos terroristas" (Consejo Europeo CO EUR 8 CONCL 3, 2017). Algunos líderes políticos insisten en la prevención automática de contenidos que se cargan en primer lugar (Hope y McCann, 2017). De hecho, este tipo de demanda a las plataformas de Internet ha sido un factor que ha impulsado el uso de sistemas automatizados para la moderación de contenidos (Andersen, 2018).

La autorregulación de las empresas privadas junto con el uso de la IA no cumple necesariamente con las normas internacionales de libertad de expresión (Asamblea General de las Naciones Unidas A/HRC/38/35, 2018). En cambio, se basa en gran medida en sus Condiciones de Servicio y Directrices de la Comunidad.

Existe la preocupación de "que los gobiernos pueden violar los derechos de libertad de expresión al intimidar a las plataformas para que eliminen el discurso ofensivo pero legal [al tiempo que evitan] el proceso democrático y la responsabilidad" (Keller, 2018). De hecho, a diferencia de lo que ocurre en los casos de censura gubernamental directa, la censura privatizada significa que es difícil apelar las solicitudes de eliminación de contenidos en un tribunal de justicia. Las entidades privadas por lo general están legalmente facultadas para tomar sus propias decisiones, y, si bien deben respetar los derechos humanos en virtud de los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas (los "Principios Ruggie"), no tienen el mismo alcance que los Estados en materia de obligaciones para operar de acuerdo con normas universales (Consejo de Derechos Humanos de las Naciones Unidas A/HRC/17/31, 2011). Estas dinámicas se complican aún más cuando la IA interviene en cuestiones vinculadas a la moderación de contenidos.

## 1.4. Detección y eliminación de contenidos por parte de la IA: casos y desafíos

En el primer semestre de 2017, la IA identificó el 95 % de las 300.000 cuentas eliminadas de la plataforma de Twitter por violaciones relacionadas con la promoción del terrorismo (Dialani, 2019), utilizando designaciones de terrorismo a nivel nacional e internacional para evaluar lo que se constituye como terrorismo (Centro de Ayuda de Twitter, 2019). El director general de Facebook, Mark Zuckerberg, explicó en una publicación de diciembre de 2018 que el 99 % del contenido relacionado con el terrorismo en la plataforma fue retirado por sistemas de IA, antes de que los usuarios presentaran quejas o los organismos encargados de hacer cumplir la ley presentaran algún informe (Zuckerberg, 2018). Sin embargo, la Relatora Especial de las Naciones Unidas sobre la protección de los derechos humanos en la lucha contra el terrorismo, Ni Aolain, ve la definición de "terrorismo" de Facebook con preocupación por ser demasiado amplia (ACNUDH, 2018b). Al mismo tiempo, algunos videos de periodistas y grupos de defensa que estaban documentando presuntos crímenes de guerra también fueron eliminados cuando YouTube puso en marcha un sistema de IA para identificar y eliminar lo que considera "propaganda terrorista" (Citron y Jurecic, 2018). Si bien en virtud del Llamamiento de Christchurch existe un acuerdo generalizado para el uso de la tecnología con el fin de evitar que los terroristas transmitan imágenes en vivo y eliminar o bloquear cualquier carga de dicho contenido, la mayor parte del contenido impugnado se encuentra en una zona más gris.

En problema puntual es la probabilidad de que la IA se programe —o se use— para llevar adelante un bloqueo excesivo de contenido legítimo con el fin de proteger a las empresas. Además, en la actualidad, las herramientas de IA no son lo suficientemente precisas, y tienen tasas de error más altas que las de los humanos (Raso, Hilligoss, Krishnamurthy, Bavitz y Kim, 2018). Por lo tanto, se está eliminando contenido considerado lícito y que cumple con las normas comunitarias particulares de las plataformas en línea.

La naturaleza del desarrollo y el despliegue de las herramientas de IA también hacen que el riesgo de la libertad de expresión sea aún mayor. Esto se debe a que los particulares sesgos humanos significan que aún estamos lejos de desarrollar conjuntos de datos holísticos y dinámicamente actualizados que reflejen la complejidad del tono, el contexto y el sentimiento de las distintas culturas y subculturas (Artículo 19, 2018).

La IA no puede detectar satisfactoriamente consideraciones contextuales esenciales para determinar la naturaleza del contenido. Estas consideraciones abarcan la ironía, la sátira, los aspectos relacionados con la cultura o la intención del usuario, el tono, el público, el propósito artístico, etc. (Asamblea General de las Naciones Unidas A/73/348, 2018). Actualmente, el procesamiento del lenguaje natural no puede diferenciar entre una refutación sarcástica de un discurso de odio y un comentario de odio real (Citron y Jurecic, 2018). Mientras que el Plan de Acción de Rabat de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH) proporciona un matiz para evaluar cuándo la expresión de odio se vuelve

peligrosa, este tipo de evaluación es difícil de programar en IA (Asamblea General de las Naciones Unidas A/HRC/22/17/Add.4, 2013). Los mismos obstáculos enfrenta la automatización de la evaluación de la proporcionalidad y el propósito legítimo que determinan si ciertos contenidos merecen ser objeto de restricción y de qué manera esto debe llevarse a cabo.

Esta cuestión también afecta al contenido que no incita a un posible daño. Un ejemplo específico vinculado a la desnudez se presentó en 2017 cuando Facebook retiró una foto de una estatua de 500 años de antigüedad. Si bien se desconoce si esta remoción fue resultado de la IA, de la decisión de un empleado o de ambos, parece muy probable que se tratara de una eliminación automática basada en la regulación de Facebook sobre la desnudez (Shah, 2017). El hecho de que se desconozca el responsable es indicativo de la necesidad de que estas empresas brinden más explicaciones. (Véase la sección 1.5 a continuación).

### **Recuadro 4: Facebook prohíbe la fotografía de la “Niña del napalm”, ganadora del Premio Pulitzer**

En septiembre de 2016 Facebook decidió eliminar la publicación de la icónica fotografía de Phan Thị Kim Phúc, la niña de 9 años que corre desnuda después de un ataque con napalm durante la guerra de Vietnam. Facebook declaró que la foto violaba los estándares de la comunidad debido a la exhibición de genitales desnudos que podrían calificarse como imágenes de abuso infantil. Después de recibir críticas, Facebook revocó su decisión y reconoció la importancia mundial de la fotografía. Si bien no se trataba de un caso de remoción automatizada completamente autónoma, “la foto fue etiquetada para su remoción por uno de los algoritmos de Facebook, que luego fue seguido por un editor humano” como suele ocurrir (Scott e Isaac, 2016). En principio, la moderación supervisada por la tecnología puede tener un papel en la señalización del contenido, con un humano capacitado para decidir qué medidas tomar. En este caso, el error fue humano.

La cuestión se vuelve aún más compleja cuando los instrumentos de IA afectan de manera desproporcionada la libertad de expresión de los grupos vulnerables y minoritarios, así como a las personas con discapacidad, un ámbito que requiere un estudio más profundo (Asamblea General de las Naciones Unidas, A/73/348, 2018). Se ha demostrado que los algoritmos de procesamiento del lenguaje natural tienen tasas de error más altas hacia los hablantes y grupos marginales (Duarte, Llanso y Loup, 2017). Esta discriminación algorítmica contra ciertas perspectivas “favorece a los poderosos sobre los marginados” (Raso, Hilligoss, Krishnamurthy, Bavitz y Kim, 2018).

De hecho, se ha demostrado que las herramientas automatizadas “funcionan con menos precisión al analizar el lenguaje de las mujeres y de los afroamericanos” (Citron

y Jurecic, 2018). Además de restringir los derechos de estas personas a expresarse, las herramientas de IA pueden, de esta forma, limitar e incluso eliminar contenidos que son relevantes y necesarios para el debate público.

Además, el procesamiento del lenguaje natural todavía no funciona tan bien en otros idiomas como en inglés. Por lo tanto, las herramientas automatizadas son menos precisas a la hora de evaluar a personas que no son angloparlantes, lo que puede restringir desproporcionadamente su discurso. Las herramientas de traducción de idiomas, en particular, pueden tener este problema. Un ejemplo de esto fue la detención de un hombre palestino por fuerzas israelíes. El hombre escribió "buenos días" en árabe en una fotografía de sí mismo que publicó en Facebook, pero la herramienta de traducción de la compañía, que funciona con IA, lo tradujo como "atáquenlos" en hebreo o "hiéranlos" en inglés (Duarte, Llanso y Loup, 2017; Hern, 2017). Facebook reconoció que su sistema de traducción cometió un error y se disculpó con el hombre y su familia por los problemas causados (Fussell, 2017).

Como concluye un informe del Centro para la Democracia y la Tecnología, las herramientas actuales para automatizar el análisis del contenido de las redes sociales tienen una capacidad limitada para analizar el significado matizado de la comunicación humana o para detectar la intención o la motivación del orador. Los encargados de formular políticas deben entender estas limitaciones antes de aprobar o adoptar herramientas de análisis de contenido automatizado. Sin las salvaguardas apropiadas, estas herramientas pueden facilitar una censura demasiado amplia y una aplicación sesgada de las leyes y de las condiciones de servicio de las plataformas (Centro para la Democracia y la Tecnología, 2017).

## 1.5. Falta de transparencia y debido proceso en la moderación de contenidos

En gran medida, no está claro cuál es el papel de los algoritmos responsables de la personalización de la información, ni de los que participan en la moderación del contenido en línea. Las empresas de Internet se enfrentan al reto de dar transparencia a estos procesos para poder rendir cuentas

En relación con el tema de la detección y eliminación de contenidos, el Observatorio de Políticas de Internet llevó a cabo un proyecto de investigación que culminó con los Principios de Santa Clara sobre la transparencia y la responsabilidad en la moderación de contenidos. Uno de los tres principios se refiere a la notificación que las empresas deben dar a cada usuario cuando su contenido ha sido retirado o su cuenta suspendida. Para explicar plenamente la lógica detrás del retiro o la suspensión, las empresas deben proporcionar a la parte afectada información detallada sobre qué directrices específicas se violaron y cómo se detectó y eliminó el contenido (Principios de Santa Clara, 2018). Por ejemplo, ¿el contenido eliminado fue detectado por un algoritmo y posteriormente eliminado por un empleado? ¿Fue marcado por una autoridad gubernamental y después eliminado? Alternativamente, ¿fue identificado y eliminado de forma autónoma mediante la aplicación de un algoritmo?

Otro principio de Santa Clara se refiere a que las empresas deben publicar informes periódicos que contengan el número de publicaciones eliminadas o cuentas suspendidas debido a violaciones de las directrices de contenido (Principios de Santa Clara, 2018). El principio no exige que las empresas proporcionen el código fuente ni información técnica sobre sus algoritmos; de hecho, es poco probable que se exija a las empresas que pongan su código fuente real a disposición del público, dadas sus reivindicaciones de propiedad intelectual y sus intereses secretos, tanto por intereses económicos como para impedir que los actores "jueguen" con los sistemas. Por lo tanto, es necesario explorar otras formas de mejorar la transparencia algorítmica, incluso mediante la transparencia cualificada, la elección del consumidor y la educación (Pollack, 2016). Los límites técnicos y de otro tipo a la transparencia se discuten en el capítulo de Apertura de esta publicación.

La restricción de la libertad de expresión es un tema serio y el problema es que la censura de algoritmos en línea en la actualidad no respeta las normas del debido proceso. Preocupan las repercusiones en la presunción de inocencia, la evasión de la censura previa a la publicación, el derecho a ser informado rápidamente de la causa y la naturaleza de una acusación, el derecho a un juicio justo y el derecho a defenderse (MSI-NET, 2018). De hecho, los usuarios hoy no tienen la posibilidad de defenderse y no se les da una oportunidad significativa de impugnar las decisiones de bloquear o "rebajar el rango" de su contenido (Balkin, 2017; Citron y Jurecic, 2018).

## Recuadro 5: Los Principios de Santa Clara

En mayo de 2018, un grupo de organizaciones, defensores y expertos académicos propuso los Principios de Santa Clara como pasos iniciales que deben seguir las empresas y plataformas que se dedican a la moderación de contenidos, con el fin de garantizar la aplicación justa de las directrices sobre contenidos (The Royal Society, 2018). Los tres principios son:

1. **Números:** Las empresas deben publicar el número de publicaciones eliminadas y de cuentas suspendidas de forma permanente debido a violaciones de sus directrices de contenido.
2. **Notificación:** Las empresas deben notificar a cada usuario cuyo contenido sea retirado o cuya cuenta sea suspendida la razón de la retirada o suspensión.
3. **Apelación:** Las empresas deben proporcionar una oportunidad significativa para apelar oportunamente ante cualquier contenido eliminado o suspensión de cuenta.

Facebook anunció recientemente la creación de un consejo independiente para revisar las decisiones de moderación de contenidos (Hensel, 2018). Sin embargo, aún quedan preguntas sobre la independencia de los miembros y el proceso mediante el cual elegirán los casos a ser escuchados (Newton, 2019). La idea de los consejos de redes sociales para toda la industria también ha sido planteada por el Relator Especial de las Naciones Unidas sobre la protección y promoción del derecho a la libertad de opinión y de expresión, y elaborada por organizaciones no gubernamentales (ONG) como Artículo 19 (Asamblea General de las Naciones Unidas A/HRC/38/35, 2018; Artículo 19, 2019). Estos podrían contribuir a mejorar la rendición de cuentas en cuanto a la moderación de contenidos y proporcionar mayor transparencia y una oportunidad de reparación.

## 2. DERECHO A LA PRIVACIDAD Y LA PROTECCIÓN DE DATOS PERSONALES

La privacidad es un derecho humano fundamental y un valor de gran importancia, ya que fomenta la autodeterminación y nos permite desarrollar nuestras perspectivas personales del mundo que nos rodea (Cohen, 2012). También es importante porque se considera que es el facilitador (o un requisito previo para el ejercicio) de otros derechos humanos y libertades, como los derechos a la libertad de expresión y opinión, la libertad de reunión y asociación, la participación política y la libertad de pensamiento, creencia y religión. Las personas necesitan un espacio privado para disfrutar y practicar plenamente estos derechos. Sin embargo, las delimitaciones sobre lo que se considera privado y las expectativas legales de privacidad no siempre están claras (Baghai, 2012).

El derecho a la privacidad se encuentra consagrado en el artículo 12 de la Declaración Universal de Derechos Humanos y en el artículo 17 del Pacto Internacional de Derechos Civiles y Políticos, así como en otros documentos de derechos humanos, instrumentos internacionales y leyes nacionales.

Este derecho implica que nadie será objeto de injerencias arbitrarias o ilegales en su privacidad, y que toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques.

También es importante señalar que la privacidad no se limita a los espacios privados, sino que se extiende a los espacios públicos y a la información de acceso público (Consejo de Derechos Humanos de las Naciones Unidas A/HRC/39/29, 2018).

En 1988, el Comité de Derechos Humanos explicó en la Observación General núm. 16 que este derecho protege contra todas las injerencias, ya sean del Estado o de agentes privados. Se explica que la vigilancia, incluida la intervención de las comunicaciones telefónicas y la interceptación de comunicaciones, viola la privaci-

dad. También se aborda el concepto de datos personales, subrayando que toda la información personal en computadoras o bancos de datos debe estar regulada por la ley. Toda persona debe tener información sobre qué datos personales se almacenan o procesan, con qué fines y qué entidades los controlan o los podrían controlar (Consejo de Derechos Humanos de las Naciones Unidas HRI/GEN/1/Rev.9(Vol. I)).

En los últimos años se ha prestado más atención al modo en que han evolucionado las normas de privacidad en la era digital. En diciembre de 2013, la Asamblea General de las Naciones Unidas ratificó la necesidad de que los Estados revisaran sus prácticas y legislaciones relativas a la vigilancia masiva y la recopilación de datos personales, entre otras cosas, a fin de defender el derecho a la privacidad (Asamblea General de las Naciones Unidas A/RES/68/167).

En un informe de 2018 sobre la privacidad en la era digital, el Alto Comisionado para los Derechos Humanos definió la privacidad como:

"[...] la presunción de que el individuo debe tener una esfera de desarrollo autónomo, interacción y libertad, una "esfera privada" con o sin relación con otras y libre de la intervención del Estado y de la intervención excesiva no solicitada de otros individuos no invitados."

(Consejo de Derechos Humanos de las Naciones Unidas A/HRC/39/29, 2018)

En su informe de 2019 al 40º período de sesiones del Consejo de Derechos Humanos, el Relator Especial de las Naciones Unidas sobre el derecho a la privacidad, Joe Cannataci, señaló que el derecho a la privacidad no es un derecho absoluto, sino relativo, y que se rige por el criterio de necesidad en una sociedad democrática (Consejo de Derechos Humanos de las Naciones Unidas A/HRC/40/63, 2019). Puede restringirse, pero siempre de una manera cuidadosamente delimitada. Según la norma establecida en el artículo 17 del Pacto Internacional de Derechos Civiles y Políticos, las injerencias en el derecho a la privacidad solo son permisibles en virtud de las normas internacionales de derechos humanos si estas no son arbitrarias ni ilegales.

## Recuadro 6: Opciones de la UNESCO en el ámbito de la privacidad (Documento Final de la UNESCO, "Conectando los puntos", 2015)

- Afirmar que los derechos humanos fundamentales de libertad de opinión y de expresión, y su corolario de libertad de prensa y derecho de acceso a la información, así como el derecho de reunión pacífica y el derecho a la privacidad, son catalizadores de la agenda para el desarrollo después de 2015;
- Reafirmar que el derecho a la privacidad se aplica y debe ser respetado en línea y fuera de línea, de conformidad con el artículo 12 de

la Declaración Universal de Derechos Humanos y el artículo 17 del Pacto Internacional de Derechos Civiles y Políticos;

- Apoyar, en el marco del mandato de la UNESCO, los esfuerzos relativos a la resolución A/RES/69/166 de la Asamblea General de las Naciones Unidas sobre el derecho a la privacidad en la era digital;
- Apoyar iniciativas que promuevan la conciencia del derecho a la privacidad en línea y el conocimiento de los métodos en constante evolución mediante los cuales los gobiernos y las empresas comerciales recopilan, utilizan, almacenan y comparten la información, así como de las formas de uso de instrumentos digitales de seguridad para proteger el derecho de los usuarios a la privacidad;
- Alentar los esfuerzos orientados a proteger los datos personales que proporcionen a los usuarios seguridad, respeto de sus derechos y mecanismos de reparación, y que fortalezcan la confianza depositada en los nuevos servicios digitales.

La privacidad también abarca la privacidad de la información, que incluye los datos que se pueden obtener sobre una persona y su vida (Consejo de Derechos Humanos de las Naciones Unidas A/HRC/39/29, 2018). Esto significa que la protección de los datos es una parte vital del derecho a la privacidad, incluso sin ser la totalidad de dicho derecho. La privacidad también debe percibirse como un "respiro para participar en los procesos de manejo de los límites que permiten y constituyen el autodesarrollo" (Cohen, 2012).

La IA no se desarrolla en el vacío, sino en las posibilidades de su contexto y, en particular, en el carácter de Internet y las fuerzas que moldean su evolución. Los sistemas de IA se basan en algoritmos contruidos para confiar en la recolección, el almacenamiento y el procesamiento de grandes cantidades de datos a fin de aprender y tomar decisiones inteligentes. Por consiguiente, el desarrollo de la IA no puede separarse de los procesos de recolección de datos y, de hecho, los datos reunidos se ponen a disposición.

Estrechamente vinculada a las transacciones de datos personales, la IA está sujeta a diversas reglamentaciones de protección de datos en muchas jurisdicciones. El Convenio núm. 108 modernizado del Consejo de Europa para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal quedó abierto a la firma en octubre de 2018, y el Comité Consultivo del Convenio ha publicado las Directrices sobre Inteligencia Artificial y Protección de Datos. De manera similar, el Reglamento General de Protección de Datos de la Unión Europea, que está en vigor desde el 25 de mayo de 2018, aborda la elaboración de perfiles y la toma de decisiones automatizada, lo que está claramente relacionado con el uso de IA. Tanto el Convenio del Consejo de Europa como el reglamento de la Unión Europea se centran en abordar las nuevas realidades del mundo en línea y las tecnologías digitales (CoE ETS núm. 108, 1981; Comisión Europea, 2018). Sin embargo, el Relator

Especial de las Naciones Unidas sobre la promoción y protección del derecho a la libertad de opinión y de expresión, David Kaye, ha señalado que, en el contexto de los sistemas de IA, "la capacidad de las personas de saber, entender y ejercer control sobre la forma en que se usan sus datos pierde sentido práctico" (Asamblea General de las Naciones Unidas A/73/348, 2018).

Los mecanismos cada vez más amplios mediante los cuales se recoge y procesa información sobre nosotros y sobre el mundo presentan un riesgo para nuestro derecho a la privacidad, y una preocupación particular es el uso de IA para elaborar perfiles sobre personas y desanonimizar conjuntos de datos (véase 2.2. más adelante). Esta elaboración de perfiles vinculada al análisis predictivo también supone una amenaza a la privacidad. Algunos argumentan que la privacidad es el derecho que hoy se ve más afectado por las aplicaciones de IA (Raso, Hilligoss, Krishnamurthy, Bavitz y Kim, 2018). Esto refuerza la necesidad de un debate sobre las normas de privacidad en la era digital. Esta discusión debe ser parte de un debate amplio e inclusivo sobre la dirección en que se desea que la sociedad vaya en el futuro (Mikkinen, Auffermann y Heinonen, 2017).

## 2.1. Recolección de datos e Internet de las cosas

El uso de IA y tecnologías asociadas puede generar nuevos grupos de información o metadatos (datos sobre datos) (Rouvroy, 2016). Subyacente a todo esto está el hecho de que la "era de los macrodatos" ha traído "más capacidades para reunir y almacenar datos" (Vayana y Tasioulas, 2016), lo que cambia la recolección tradicional de datos. La retención de datos ahora no se limita de manera inherente a un determinado propósito específico. De hecho, cuando se están recolectando datos, no necesariamente se ha establecido ya un propósito. La "utilidad de cada dato depende de la cantidad de los otros datos con los que se puede correlacionar" (Rouvroy, 2016). En este sentido, todos los datos, por inocuos y sin sentido que se puedan considerar individualmente, pueden tener una importancia y un interés adicionales para muchos agentes.

En lugar de una recopilación de datos situada y precisa, las tecnologías de la información en red proporcionan una atención de vigilancia "continua, omnipresente y persistente" (Cohen, 2012). Esta atención abarca muchas cosas, como la vigilancia de nuestras huellas en línea o las cámaras con reconocimiento facial en el espacio público, así como el reconocimiento de la marcha, un área en la que se necesita más investigación con respecto a sus impactos en la privacidad.

Mediante el uso intensivo de Internet y el creciente uso de una serie de dispositivos y aplicaciones de Internet de las cosas, las personas generan una gran cantidad de datos (Artículo 19 y Privacy International, 2018). Esto puede hacerse intencionalmente (escribiendo mensajes, usando emojis o publicando imágenes en las redes sociales) o involuntariamente (navegando por sitios web, haciendo clic en los enlaces, aceptando cookies, etc.). Estas son nuestras huellas digitales, recolectadas por defecto para monitorear nuestros movimientos en línea, donde el monitoreo suele

ser el deliberado "seguimiento de individuos en línea para crear perfiles" (Rouvroy, 2016; Bennett, 2018).

La línea entre el mundo en línea y fuera de línea es cada vez más borrosa. De hecho, "la gente parece vivir entrando y saliendo continuamente de las redes, lo que hace que sea difícil dibujar líneas claras y significativas entre ambos mundos" (Vayana y Tasioulas, 2016). Por ejemplo, cuando nos desplazamos por la ciudad y vamos a un café, a una escuela o a una institución médica, el rastreador GPS de nuestros teléfonos inteligentes es capaz de detectar dónde estamos y cuánto tiempo permanecemos y recopilar estos datos (y correlacionarlos con los movimientos de los demás), aunque no hayamos accedido a Internet en el teléfono. A partir de esos datos se pueden hacer inferencias significativas sobre nuestra identidad, intereses, aspiraciones, problemas y redes.

La Internet de las cosas vuelve aún más difusa la línea entre el mundo en línea y el mundo fuera de línea, ya que algunos dispositivos que antes eran solo físicos y no estaban relacionados con Internet se van integrando cada vez más en conexiones de datos más amplias, incluidos los vínculos con el desarrollo y el procesamiento de IA. A través de sensores y *software*, estos dispositivos "emiten información sobre los movimientos, las actividades, el rendimiento, el consumo de energía, los estilos de vida, etc. de sus usuarios", y estos datos se recogen, almacenan, analizan y venden (Rouvroy, 2016). Estos datos se vuelven más valiosos cuando son procesados por la IA y, por lo tanto, desafían cada vez más la privacidad en la era digital.

## 2.2. Vigilancia, monitoreo y reconocimiento facial con tecnología de IA

El consentimiento del usuario para el uso, procesamiento, almacenamiento, transferencia y difusión de sus datos personales, especialmente cuando se refiere a esferas delicadas de la vida de una persona (como la salud, la orientación sexual, la información financiera de identificación personal, los datos biométricos, etc.), es una cuestión reglamentaria importante. Los conjuntos de datos pueden revelar detalles personales o sensibles de la vida digital y cotidiana de una persona cuando se agregan mediante sistemas de inteligencia artificial (Andersen, 2018). Además, una vez que la información personal reunida por estos medios sale a la luz, a las personas afectadas puede resultarles difícil solicitar su corrección o eliminación.

En un número cada vez mayor de hogares privados se instalan asistentes virtuales integrados en altavoces inteligentes, como Alexa de Amazon (comúnmente abreviado como "Alexa"), que a menudo recogen datos sin el conocimiento o el consentimiento plenamente informado de las personas cuyos datos se están recolectando. Al igual que en otros asistentes de voz, también hay importantes suposiciones incorporadas en el servicio que impactan particularmente en las nociones de roles de género e igualdad de género (UNESCO; EQUALS Skills Coalition, 2019).

Un servicio como Alexa puede escuchar todo lo que la gente a su alrededor dice, pero solo comienza a grabar cuando escucha la "palabra de activación", como su

nombre. De hecho, "una vez que se detecta la palabra, el audio comienza a fluir hacia la nube, incluso una fracción de segundo de audio antes de la palabra de activación" (Amazon, 2019). El audio grabado pasa entonces a formar parte de los datos almacenados por Amazon y puede usarse para diversos fines. Amazon no solo se interesa por las palabras pronunciadas por el consumidor, sino también por otros aspectos de la grabación. Por ejemplo, Amazon ha presentado una patente en la que desea detectar un cambio en la voz del cliente o un estornudo y tos para sugerir medicamentos (Mehta, 2018). Si bien esto aún se encuentra en la etapa de patentamiento, es algo que puede interpretarse como útil, pero también como invasivo y que viola el derecho de las personas a la privacidad.

Existe una asimetría de información entre los usuarios de estos productos de consumo y los que procesan los datos. Es de gran importancia que las sociedades eduquen a los consumidores a fin de que tomen conciencia sobre los datos que sus plataformas, redes y dispositivos conectados generan, procesan o comparten, así como sobre la forma en que estas acciones podrían afectar su derecho a la privacidad, al igual que otros derechos humanos y libertades (ARTÍCULO 19 y Privacy International, 2018). También es fundamental que las empresas informen al público de los posibles fallos de los dispositivos. Esto podría facilitar la recomendación del Consejo de Europa para que los Estados mitiguen los posibles efectos adversos de la IA en los derechos humanos (CoE CHR/Rec(2019)1, 2019).

## Recuadro 7: Los asistentes virtuales escuchan a escondidas: Alexa de Amazon

Una pareja estaba teniendo una conversación privada en su casa cuando recibieron un mensaje de texto de un colega que decía: "Desconecten sus dispositivos Alexa ya mismo. Están siendo *hackeados*". Este colega había recibido el audio de la conversación privada. Alexa había estado escuchando, grabando la conversación de fondo y enviándola a esta persona de su lista de contactos. El dispositivo, sin embargo, no fue *hackeado* por un tercero. Amazon confirmó que el audio había sido transmitido sin querer por el dispositivo (Moye, 2018). El asistente de voz comenzó a grabar cuando detectó una palabra en la conversación de la pareja que interpretó como la "palabra de activación" y entendió una parte de la conversación de fondo como una solicitud de "enviar mensaje" (Wollerton y Crist, 2018). Aunque todo ocurrió sin querer, la pareja lo vivió como una invasión de la privacidad.

Además, algunos datos relativos al comportamiento de las personas se están recopilando por otros medios, como las imágenes satelitales o la vigilancia por video en los espacios públicos. En particular, el *software* de reconocimiento facial con tecnología de IA es cada vez más utilizado por los gobiernos y las empresas en espacios públicos como estaciones, escuelas, teatros, calles, centros comerciales, etc. Esto viola potencialmente la privacidad de las personas y "transforma las expectativas de

anonimato en la esfera pública, lo que es especialmente importante para los grupos vulnerables y para quienes se pronuncian en contra de los poderosos actores involucrados en abusos contra los derechos humanos, la corrupción, por nombrar algunos" (ARTÍCULO 19 y Privacy International, 2018). Bajo la amenaza de la vigilancia permanente y la pérdida del anonimato, las personas pueden verse disuadidas de ejercer sus derechos humanos fundamentales y pueden optar por cambiar su comportamiento en los espacios públicos (Andersen, 2018). Esto también puede reforzarse especialmente cuando las personas temen que los datos puedan terminar en manos de actores con poder real para causarles daño.

## Recuadro 8: Software de reconocimiento facial en centros comerciales

En 2018, varios medios de comunicación informaron del uso de *software* de reconocimiento facial en centros comerciales de Calgary (Canadá), lo que suscitó preocupación por la posible violación del derecho a la intimidad de los compradores locales. Esta vigilancia con tecnología de IA se usó para reunir diferentes tipos de datos sobre el comportamiento de los clientes, a fin de analizar e identificar patrones en el comportamiento de los compradores. Sin pedir el consentimiento explícito de los clientes, el *software* podía hacer un seguimiento de la edad y el género de los compradores, lo que supuestamente permitiría a los propietarios de los centros comerciales "comprender los patrones de uso del directorio (y) crear una mejor experiencia de compra" (Rieger, 2018).

Además de su función en la recopilación y el procesamiento de datos, los sistemas de IA están siendo utilizados con fines de vigilancia por empresas privadas y gobiernos por igual. Incluso en los países, la confianza de la población en que sus datos permanecerán seguros varía según se acceda a ellos a manos de los gobiernos o de las empresas del sector privado. Según un informe de 2015 del Centro de Investigaciones Pew, el 31 % de los adultos estadounidenses están "muy seguros" o "algo seguros" de que los organismos gubernamentales mantendrán sus registros en forma privada y segura, mientras que solo el 11 % cree lo mismo en el caso de los sitios de medios de comunicación social (Madden y Raine, 2015). El déficit de confianza puede variar de un país a otro (McMullan, 2015).

Un argumento a favor del *software* de reconocimiento facial con tecnología de sistemas de IA es que puede utilizarse con fines de aplicación de la ley para identificar y localizar a personas concretas. Si bien en general se considera que la vigilancia masiva constituye una interferencia desproporcionada con el derecho a la privacidad y a la libre expresión (ya que no es "necesaria ni proporcional al objetivo de seguridad pública o prevención del delito"), la vigilancia selectiva debe cumplir con la prueba de tres partes de legalidad, necesidad y proporcionalidad, así como con el propósito legítimo establecido en las normas internacionales de derechos humanos (Andersen, 2018).

## 2.3. Seguimiento en línea y desanonimización de personas

La anonimización de datos ha sido históricamente la forma en que “el equilibrio entre el uso de los datos y la preservación de la privacidad de las personas ha funcionado tanto práctica como legalmente” (Montjoye, Farzanehfar, Hendrickx y Rocher, 2017). La computación ubicua y los macrodatos suponen, sin embargo, un problema para la anonimización. Por ejemplo, un código de 32 bits —4c812db292272995e5416a323e-7gbd37— ayudó a un programa de seguimiento de la actividad en línea a identificar a un usuario como “una mujer de 26 años, en Nashville, Tennessee, interesada por las películas, incluidas “La princesa prometida”, “Como si fuera la primera vez”, “10 cosas que odio de ti” y “Sexo en la ciudad”” (Angwin, 2010). Algunas de estas predicciones pueden ser exactas hasta el punto de desanonimizar a los usuarios de Internet cuyas actividades en línea son constantemente rastreadas. Una vez que el ADN entra en el sistema, es casi imposible permanecer anónimo cuando estos datos se combinan con otros.

### Recuadro 9: Seguimiento en línea

Para tener una experiencia de primera mano del seguimiento en línea, animamos al lector a que se conecte al administrador de Preferencia de Anuncios de Google en <http://www.google.com/ads/preferences/>, mire los marcadores utilizados por la empresa para definir al lector y evalúe su precisión.

La información rastreada se usa para crear perfiles digitales de los usuarios a los que se vende el acceso en el mercado, incluidos intercambios especializados, para ayudar a los anunciantes a comercializar mejor sus productos. En un caso, por ejemplo, una persona que trabaja en el negocio de la construcción fue definida con mucha precisión como un hombre de entre 35 y 54 años, que es propietario de una casa, vive en un pueblo pequeño, no tiene hijos, tiene un título universitario y un ingreso medio de 86.724 dólares, y trabaja en administración (Steel y Angwin, 2010). La granularidad de la información, junto con la ubicación de la persona, reveló efectivamente la identidad del individuo que estaba siendo rastreado.

La desanonimización y la reidentificación son posibles gracias a la capacidad de la IA de reconocer patrones e identificar tendencias a partir de datos no personales sobre individuos o grupos de personas, y, de este modo, derivar lo íntimo de lo disponible, sin el conocimiento o el consentimiento de las personas afectadas por dicha inferencia (Artículo 19 y Privacy International, 2018). Sin embargo, esta nueva información pierde el contexto en el que los datos originales se extrajeron por primera vez y los fines para los que los proveedores de datos podrían haber consentido inicialmente su procesamiento, lo que aumenta el riesgo de que los datos se vuelvan inexactos

y priva a las personas de la capacidad de rectificarlos o suprimirlos (Asamblea General de las Naciones Unidas, A/73/348, 2018). Estos datos recién generados, que podrían revelar la sexualidad, las opiniones políticas, el estado general de salud y las creencias religiosas de una persona concreta, podrían dar lugar a discriminación e incluso a persecución en determinados casos.

La posibilidad de inferir o predecir información personal o sensible a partir de datos no personales extraídos de diferentes conjuntos de datos vuelve efectivamente poco clara la distinción entre datos personales y datos no personales, lo que plantea importantes desafíos respecto del derecho a la intimidad (Artículo 19 y Privacy International, 2018).

## 2.4. Análisis predictivo y elaboración de perfiles con IA

Los datos de los consumidores y los usuarios alimentan sofisticados sistemas de IA de análisis predictivo (Cohen, 2012). Mediante la ordenación, valoración, puntuación, calificación, evaluación y priorización de distintos individuos en distintos grupos de personas, la IA se usa para tratar de predecir el comportamiento futuro (Artículo 19 y Privacy International, 2018; Harcourt, 2007).

Como ya se mencionó, la privacidad puede verse como un respiro libre "de limitaciones irrazonables en la construcción de la propia identidad" (Agre y Rotenberg, 1988). Al establecer el perfil de cada usuario y predecir sus preferencias, estos sistemas de IA resultan de utilidad para quienes buscan reforzar o impulsar las preferencias de los usuarios y su posterior comportamiento "de forma que se reduzca la casualidad y la libertad de jugar" en la que prospera la innovación (Cohen, 2012). La autonomía, así como la privacidad, pueden abordarse "como algo que se logra en el marco de complejas condiciones materiales y sociales previas", en lugar de algo que está puramente dado (Oleksy, Just y Zapedowska-Kling, 2012). Por lo tanto, la elaboración de perfiles de personas y la toma de decisiones basada en las predicciones de la IA (dos cosas que pueden afectar mucho a las personas), pueden interpretarse como la interferencia en una esfera privada genuina, que es muy necesaria para el autodesarrollo.

Las empresas privadas suelen utilizar estas predicciones para la personalización (como se ve en la sección sobre el derecho a la libertad de expresión), pero también, al igual que los gobiernos, pueden utilizar esta información para determinar o limitar el acceso de los ciudadanos a determinados servicios y programas.

Algunos ven este tipo de análisis predictivo como un cambio de paradigma en el que el conocimiento parece derivar directamente de la realidad: "sentimos que con los macrodatos ya no tenemos que producir conocimiento sobre el mundo, sino que podemos descubrir el conocimiento directamente en el mundo" (Rouvroy y Stiegler, 2016). Sin embargo, es importante darse cuenta de que "la predicción no se limita a describir el futuro, sino que lo transforma" (Rouvroy, 2016). De hecho, al limitar el alcance de las posibilidades, la autodeterminación se ve amenazada y la privacidad

disminuye. Por lo tanto, podría decirse que este uso del objetivo de IA "es producir ciudadanos-consumidores rastreables y predecibles, cuyos modos preferidos de autodeterminación se desarrollan a lo largo de las trayectorias predecibles y generadoras de beneficios" (Cohen, 2012).

## Recuadro 10: Efecto "inhibidor" en otros derechos humanos

Dado que la recolección de datos está en todas partes y que el análisis predictivo alimenta las decisiones que inciden, en gran medida, en los individuos, las personas pueden modificar sus comportamientos para tratar de evitar sufrir consecuencias negativas. Es posible que se limiten a la hora de interactuar y compartir información con otros, lo que constituye una limitación injustificable del derecho a la libertad de opinión y de expresión (Artículo 19 y Privacy International, 2018).

La naturaleza omnipresente e invisible de los sistemas de IA, sumada a su capacidad para identificar y rastrear conductas, también puede tener un efecto inhibidor en otros derechos humanos, como la libertad de reunión y asociación (Consejo de Derechos Humanos de las Naciones Unidas A/HRC/26/29, 2014). Las personas que desean participar en movimientos sociales quizás eviten hacerlo por temor a que un *software* de reconocimiento facial u otro sistema de IA puedan elaborar su perfil político.

Lo mismo puede decirse de las personas que se abstienen de comunicar "información sensible relacionada con la salud por temor a que su anonimato se vea comprometido", lo que, a su vez, puede afectar su derecho a la salud (Consejo de Derechos Humanos de las Naciones Unidas A/HRC/26/29, 2014).

Las preocupaciones relativas a la privacidad y la IA pueden tener repercusiones negativas similares en otros derechos, como la libertad de religión, el derecho a un trabajo deseable, el derecho a un juicio justo y el derecho a no ser arrestado o detenido arbitrariamente (Andersen, 2018; Raso, Hillgoss, Krishnamurthy, Bavitz y Kim, 2018).

## 2.5. Protección de la privacidad en el contexto de la IA: paradigmas actualizados y soluciones de Tecnologías de Mejora de la Privacidad

Las normas de privacidad centradas en el control de la información están siendo cuestionadas en la era digital. Las dicotomías entre lo público y lo privado se han desdibujado hasta el punto de que es difícil, o incluso imposible, que los individuos controlen su información. Por lo tanto, se necesitan marcos que guíen mejor nuestra comprensión de la privacidad en la era de los algoritmos y los macrodatos.

Entre otros, Nissenbaum ha desarrollado el concepto de privacidad como integridad contextual para ayudar a comprender estos desafíos. Esto proporciona un relato universal de la privacidad que no depende del lugar ni el tiempo, lo que significa que no se supone que funcione en una dicotomía preconcebida de lo público versus lo privado o de lo sensible versus lo no sensible (Nissenbaum, 2004).

Nissenbaum señala que las intuiciones sobre las normas de privacidad parecen estar arraigadas en detalles de contextos, esferas o situaciones estereotipadas bastante más limitados. Cada interacción tiene sus propias normas de privacidad específicas del contexto. Por ejemplo, una relación médico-paciente implica el intercambio de información médica con una comprensión mutua de la estricta confidencialidad de la información que se comparte.

Como se trata de compartir información médica delicada, las mismas normas que se refieren a la confidencialidad entre paciente y médico deben aplicarse a la relación entre la persona A y un sistema de IA que actúa como "médico". Por lo tanto, si un tercero comparte información médica sensible obtenida a partir de la tecnología ponible, estaría violando la privacidad en este contexto. Además, la utilización de los datos por la propia plataforma médica debería estar sujeta a las normas aplicables al intercambio de información entre el paciente y el médico. Es difícil, no obstante, ver cómo ampliar esas normas para que se apliquen en otros lugares a los numerosos estratos de otros agentes de la cadena de datos que pueden captar y registrar esa información.

De modo similar, los amigos comparten una amplia gama de detalles sobre su vida con la expectativa de cierta privacidad que debería seguir aplicándose cuando el intercambio es intermediado por la tecnología. Asimismo, las normas de privacidad en la cabina de votación, cuando se consulta a un abogado, en el banco o en un autobús son todas específicas de un contexto, con una expectativa que debe considerarse y respetarse, a pesar del "colapso del contexto" que se produce en línea y en particular a través de las redes sociales (Boyd, 2008).

Si bien las normas de privacidad que rigen la vida real de las personas deberían ser igualmente aplicables a su vida digital, existe una falsa separación entre ambas que ha permitido la proliferación de modelos sencillos y demasiado amplios de consentimiento a prácticas que atentan contra la privacidad de las personas. La idea es que se apliquen una serie de dispensas a los distintos arreglos sociales y de comunicación. A este respecto, la IA puede tener el potencial de identificar estas varia-

ciones para ayudar a decidir sobre la protección adecuada de la privacidad en cada contexto particular.

Si bien la IA plantea varias preocupaciones en torno a la privacidad, también presenta grandes oportunidades para mejorar la privacidad personal. Las tecnologías de mejora de la privacidad (PET) basadas en IA, como la "privacidad diferencial" y el "aprendizaje federado", son mejoras de los métodos anteriores para evitar la reidentificación de las personas mediante la agregación de datos personales (Scripa Els, 2017). Por ejemplo, el proceso de aprendizaje federado desarrollado recientemente por Google permite la recolección de datos para mejorar el modelo de aprendizaje automático centralizado sin almacenar datos individuales en la nube. De hecho, "en lugar de enviar datos en bruto, determina los cambios que deben hacerse en el modelo a nivel local y luego envía una "pequeña actualización enfocada" a la nube, donde la actualización se promedia con otras actualizaciones para mejorar el modelo" (Scripa Els, 2017).

La IA también puede utilizarse para vigilar las violaciones y los abusos de la privacidad personal. Dado que los sistemas de IA se volverán cada vez más complejos, es posible que los humanos por sí solos no puedan vigilar adecuadamente las violaciones a dichos sistemas. La idea de los auditores de IA es tener sistemas inteligentes para vigilar otras aplicaciones de IA y detectar casos en los que los datos personales no están bien gestionados o en los que un programa de IA logra una reidentificación (Scripa Els, 2017).

Sin embargo, es mucho lo que queda por hacer para sensibilizar y alfabetizar a todas las partes interesadas en torno a las cuestiones de la privacidad y para alentar a más agentes a incorporar la protección de la privacidad mediante el diseño y las tecnologías y aplicaciones de mejora de la privacidad.

### 3. DESARROLLO DEL PERIODISMO Y LOS MEDIOS DE COMUNICACIÓN COMO EJERCICIOS DE LA LIBERTAD DE EXPRESIÓN

La práctica del periodismo es un ejercicio particular del derecho a la libertad de expresión y al acceso a la información, y la libertad de prensa es una libertad necesaria para la difusión de la información en la esfera pública. Lo que distingue el ejercicio del periodismo de otras formas de expresión pública —que también se basan en la libertad de prensa— es que la expresión periodística auténtica se ajusta a normas profesionales como la verificación del contenido y la publicación en el interés público, lo que permite a los individuos y las sociedades recibir e impartir información e ideas, de conformidad con el artículo 19 de la Declaración Universal de Derechos Humanos.

Si bien la expresión periodística puede realizarse de manera puramente individual, se suele necesitar una infraestructura de apoyo en forma de institución, con sus distintas políticas y sistemas, para lograr una producción sostenible y defender a los profesionales de los ataques. Es aquí donde las organizaciones de medios de comunicación, desde las privadas hasta las públicas y comunitarias, tienen un papel clave que desempeñar en el compromiso de cualquier sociedad con la forma en que la IA impacta en la libertad de expresión.

Hay una creciente intersección entre la IA y la práctica del periodismo y la protección de los periodistas, así como las noticias y otros tipos de contenido. Esto no solo incide en la producción, sino también en la difusión y el consumo del periodismo.

Es necesario facultar a los periodistas y las plataformas de los medios de comunicación para que utilicen la inteligencia artificial para contar la historia de este impulso tecnológico, y para que hagan un periodismo mejorado por la IA para reconocer pautas y tendencias que de otro modo serían invisibles.

#### 3.1. La era del periodismo computacional, el periodismo de datos, el periodismo automatizado y el periodismo robótico

Periodismo computacional, periodismo de datos, periodismo automatizado y periodismo robótico forman parte de la terminología que se utiliza, de manera en cierto modo intercambiable, para identificar el uso de las TIC avanzadas en una o más fases del proceso periodístico. La adopción de algoritmos automáticos puede usarse para la reunión de datos, la verificación de hechos y la redacción o edición en video automatizada de noticias y la difusión de contenidos de los medios de comunicación. La IA tiene, en este sentido, mucho que ofrecer al periodismo y al desarrollo de los medios de comunicación.

Este nuevo tipo de periodismo puede describirse mediante los procesos algorítmicos "que convierten los datos en textos de noticias narrativos con una intervención humana limitada o nula más allá de la programación inicial" (Carlson, 2014). Esencialmente, la IA en el proceso periodístico puede dividirse en dos facetas: el procesamiento computacional de macrodatos que pueden extraer información relevante y el proceso algorítmico que puede convertir este conocimiento en historias legibles (Latar, 2015). Ambos aspectos pueden considerarse complementarios de la labor de los periodistas humanos (Flew, Spurgeon y Daniel, 2012).

La IA puede ayudar a generar noticias rápidamente gracias a los avances en la generación del lenguaje natural, que es un subcampo del procesamiento del lenguaje natural (Dorr, 2016). Esto puede ser especialmente útil en áreas en las que se necesitan actualizaciones constantes (como los cambios en el precio de las acciones o los reportajes deportivos). Por ejemplo, el "periodista robot" de IA de *The Washington Post*, llamado Heliograf, emitió 300 informes cortos y alertas sobre los Juegos Olímpicos de Río (Moses, 2017). Esta automatización puede dejar a los periodistas tiempo para realizar otras tareas menos mecánicas (Carlson, 2014).

La IA también ofrece posibilidades de traducción para reunir y difundir noticias en muchos idiomas, lo que ayudará a llegar a un público más amplio y a nuevos mercados (Drr, 2016). La agencia de noticias finlandesa STT ya está utilizando a la IA para traducir las noticias al inglés y al sueco (George, 2018).

A su vez, la IA ofrecerá a los medios de comunicación una mayor capacidad para atender a los anunciantes, obtener suscripciones y medir la calidad del periodismo. Algunos elementos mencionados en el reciente informe de la Junta Internacional de Investigación e Intercambios (IREX) "*Can Machine Learning Help Us Measure the Trustworthiness of News?*" destacan parte de este potencial (IREX, 2018).

## Recuadro 11: La inteligencia artificial ayuda a medir la calidad del periodismo

La Junta Internacional de Investigación e Intercambios (IREX) mide la calidad del periodismo usando unos 20 indicadores. Lo ha hecho con decenas de miles de artículos de noticias a través de proyectos en varios países. Durante años, ese trabajo se basó en evaluadores humanos, lo que implica profesionales de los medios de comunicación capacitados. En 2018, la IREX probó si el aprendizaje automático podía hacer que este proceso fuera más eficiente y uniforme. Tomaron uno de los 20 indicadores —si un periodista inserta, de forma no transparente, su propia opinión en un artículo de noticias— y trabajaron con una empresa emergente de IA para entrenar algoritmos para identificar instancias en miles de artículos. Después de solo unas pocas rondas de capacitación, la IA pudo identificar en los artículos de noticias frases que contenían las opiniones de los redactores, y en un momento dado logró una precisión del 95 %.

Sin embargo, es probable que para la IA sea mucho más difícil procesar la cuestión más profunda y compleja de identificar los relatos subyacentes, que sirven para enmarcar y estructurar la selección de fuentes, el orden de la información, el tono y la redacción utilizados, etc. (IREX, 2018).

El impacto de la IA en el periodismo también plantea diferentes desafíos en cuanto a la deontología de los medios de comunicación, entre otras cosas al plantear la cuestión de quién o qué debería considerarse o aceptarse como autor de noticias algorítmicas (Montal y Reich, 2017). Esta pregunta tiene repercusiones a la hora de responsabilizarse por el contenido, por ejemplo, en lo que respecta a la responsabilidad en caso de una demanda por difamación. Además, como los medios de comunicación no siempre identifican claramente si un artículo fue escrito por humanos o por algoritmos, surgen preguntas vinculadas a la transparencia.

Si el público no puede distinguir si el contenido periodístico es creado por humanos o producido por IA, o es una fusión de ambos, surgen las asimetrías de información. Esto plantea la siguiente pregunta: "Si no hay información sobre la naturaleza algorítmica de un texto y sus recursos, ¿cómo debería el público decidir si quiere consumir la información y si puede confiar en ella?" (Drr y Hollnbuchner, 2017).

### **3.2. La inteligencia artificial amenaza el pluralismo de los medios: priorización de contenidos, personalización de información y microfocalización**

El uso de IA en la producción y moderación de contenidos en línea facilita la creación de un entorno que produce "burbujas de filtro, cámaras de eco y otros elementos que son contrarios al libre acceso a la información y al pluralismo de los medios de comunicación" (Oghia, 2018). De hecho, además de las repercusiones que la personalización de la información puede tener en la libertad de expresión, también tiene un impacto potencialmente negativo en la capacidad de los medios de comunicación para ofrecer un entorno favorable para un debate inclusivo y pluralista (MSI-NET, 2018). Otra posible amenaza al pluralismo y la diversidad de los medios de comunicación son los algoritmos de los motores de búsqueda, que también podrían tener un sesgo hacia determinados contenidos o proveedores de contenidos (Supervisor Europeo de Protección de Datos, 2019). Los algoritmos que dan prioridad a los vínculos existentes como indicador de la calidad del contenido pueden acabar reforzando el *statu quo* y descuidando el nuevo contenido periodístico al que hasta ahora se han vinculado menos personas.

Algunos pequeños cambios en los algoritmos pueden tener un "impacto significativo en los editores y en los medios de comunicación en términos de tráfico y viabilidad financiera", como ocurre con los algoritmos que clasifican si el contenido es amigable para los anunciantes (Oghia, 2018). Por ejemplo, cuando "YouTube reaccionó con un

uso más estricto de su algoritmo operado para detectar el contenido 'no amigable para anunciantes'. [se informó que] afectó a medios independientes, incluidos comediantes, comentaristas políticos y expertos" (MSI-NET, 2018). La IA también amenaza cada vez más con socavar el pluralismo y la diversidad de los medios de comunicación a nivel de consumo de periodismo al permitir el bloqueo y el filtrado automatizados, algo que, desde el punto de vista de las normas internacionales, puede muy bien constituir una restricción arbitraria, más que legítima.

Además, la personalización de la información por medio de algoritmos es un proceso que puede ser manipulado de manera encubierta por razones políticas. Los medios de información, en cambio, suelen ser visibles para varios públicos que pueden reconocer las inclinaciones políticas de determinados medios y emitir juicios informados en consecuencia (por ejemplo, sobre cuestiones electorales). Sin embargo, el proceso de microfocalización en línea permite que las comunicaciones políticas se dirijan a individuos o a públicos especializados y que los mensajes se adapten a destinatarios específicos (Supervisor Europeo de Protección de Datos, 2019). Los macrodatos derivados "del comportamiento en línea de los ciudadanos, incluido el uso de las redes sociales, son el principal combustible de la microfocalización política contemporánea" (Nenadic, 2018). Hay un aspecto negativo en la forma en que la publicidad en línea dirigida ha sido utilizada por los estrategas políticos para "llegar a los votantes correctos con el mensaje adecuado con una precisión casi quirúrgica" (Maréchal, 2018). El director general de Twitter, Jack Dorsey, escribió en octubre de 2019 que "los anuncios políticos en Internet presentan desafíos totalmente nuevos para el discurso cívico: microfocalización y optimización de mensajes basada en el aprendizaje automático, información engañosa no controlada y falsificaciones profundas (*deep fakes*). Todas cosas que se presentan cada vez con mayor velocidad y sofisticación, y a una escala que resulta abrumadora". Esos usos de elementos de IA no solo suponen una amenaza para el pluralismo de los medios de comunicación, sino que también ponen en peligro la integridad del proceso electoral y hacen a un lado el rol del periodismo como fuente vital de información verificada que se presenta en el interés público y a plena vista del público.

La microfocalización y la manipulación del contenido con fines políticos también están estrechamente vinculadas a la desinformación y la información errónea. La difusión deliberada de desinformación a grupos específicos puede intensificarse mediante el uso de IA para amplificar este fenómeno. La desinformación no solo opera mediante la divulgación de hechos falsos, sino que a menudo los combina con opiniones y comentarios firmes que, por su naturaleza, no son cuestiones de veracidad o falsedad. Esto se deriva de una lógica comercial que proviene de una serie de plataformas en línea en las que se da prioridad a la promoción de la participación de los usuarios de Internet, en lugar de informar y educar a los usuarios en el interés público.

## Recuadro 12: El asunto Cambridge Analytica

El Comité de Expertos sobre los Intermediarios de Internet (MSI-NET) del Consejo de Europa ha expresado su preocupación respecto del peligro de que la IA socave los procesos democráticos y el derecho a elecciones libres. Durante las elecciones presidenciales de los Estados Unidos en 2016 y el referéndum por la adhesión del Reino Unido a la Unión Europea, se utilizó la microfocalización para mostrar ciertos contenidos a un público seleccionado, basándose en el análisis algorítmico de los datos de las personas, con el fin de tratar de incidir en su voto. Cambridge Analytica, a través de la recopilación de datos personales, ofreció este servicio a varios políticos.

Este tema despierta preocupación porque las campañas políticas "se basan cada vez más en análisis de macrodatos para incidir en las opiniones y el voto a través de mensajes o publicidades dirigidas" (Supervisor Europeo de Protección de Datos, 2018). La reglamentación de la publicidad en los medios de difusión para garantizar la equidad y la transparencia pierde importancia cuando se utiliza IA para individualizar los mensajes políticos en las plataformas de Internet.

Sin embargo, la desinformación no puede prosperar fácilmente ante un periodismo creíble e inclusivo que se basa en altos niveles de verificación, como se analiza en el manual de educación y capacitación en periodismo de la UNESCO, "Periodismo, 'noticias falsas' & desinformación" (Ireton y Posetti, 2018f). La pregunta que esto plantea es si la IA podría sostener un modelo de negocios que diera más protagonismo al periodismo que a la desinformación.

### 3.3. Aumento de la desinformación automatizada e iniciativas para contrarrestarla

Durante las elecciones estadounidenses de 2016, "las notas con noticias falsas más populares se compartieron más ampliamente en Facebook que las notas más populares de los principales medios [y que] muchas personas que ven notas con noticias falsas dijeron que las creen" (Allcott y Gentzkow, 2017). La desinformación socava la esfera pública y la democracia, y, en virtud de su trayectoria actual, cabe esperar que la IA se utilice cada vez más para desempeñar un papel en la rápida difusión de desinformación automatizada.

Facebook y otras plataformas de Internet están llevando adelante iniciativas para eliminar la desinformación. Sin embargo, "la identificación de sitios y artículos de noticias falsas también plantea importantes interrogantes sobre quién se convierte en árbitro de la verdad" (Allcott y Gentzkow, 2017). La eliminación del contenido puede constituir una violación de la libertad de expresión, ya que la desinformación no siempre es ilegal y, de hecho, puede interpretarse que cierta información de este tipo

está protegida por el derecho internacional, a menos que constituya una violación de derechos, una amenaza a la seguridad pública, etc., o vaya acompañada de actividades de promoción para incitar a la hostilidad, la discriminación o la violencia.

Actualmente existen iniciativas para identificar e incluso contrarrestar la desinformación mediante algoritmos basados en IA. Varios actores están buscando un algoritmo de verificación de hechos que pueda detectar falsedades. Informar al público sobre el carácter falso de un contenido parece ofrecer soluciones más ajustadas a derecho que eliminar el contenido por completo. La complejidad radica en que los hechos y las falsedades se seleccionan y construyen dentro de marcos narrativos y perspectivas ideológicas o políticas. Por lo tanto, los significados se estructuran de manera que van más allá de un solo hecho o falsedad. Además, los significados a menudo no se expresan puramente a través de la información, sino a través de emociones y significantes de identidad. Así pues, la desinformación suele ser solo un componente en mensajes de propaganda más amplios que incluyen el incentivo, el odio, la burla, la broma, la indignación, etc. Esto constituye un reto para la capacidad de la IA de detectar eficazmente la desinformación maliciosa de los tipos de expresión legítima en la etapa actual.

La IA también interviene en la evaluación de pautas de comportamiento para calibrar el proceso en vez del contenido, e identificar a los usuarios no auténticos, como cuentas coordinadas, identidades fabricadas y movilización de bots, todas señales de una posible intención de promover la desinformación.

Una iniciativa para contrarrestar la desinformación es el Código de Prácticas de la Unión Europea para las plataformas en línea, firmado en septiembre de 2018 por cuatro actores principales: Google, Facebook, Twitter y Mozilla, así como varias asociaciones comerciales (Asociación Europea de Agencias de Comunicación, Asociación Europea de Medios Digitales, Agencia Europea de Publicidad Interactiva, Asociación de Anunciantes Belgas y Federación Mundial de Anunciantes). Uno de los principales objetivos del Código de Prácticas para las plataformas en línea es interrumpir los ingresos por publicidad de las cuentas y los sitios web que presentan información falsa.

Como se verá en el capítulo sobre apertura, en algunas áreas vinculadas a la desinformación y el periodismo, el mismo *software* de IA puede perseguir objetivos distintos y opuestos.

Este es el caso de la fabricación e identificación de "falsificaciones profundas", que pueden ser muy convincentes y realistas. Este término se aplica para referirse "tanto a los artefactos de video como de audio que han sido sintetizados a partir de datos digitales existentes mediante modelos de redes neuronales de aprendizaje profundo" (Barradough y Barnes, 2019). Estos videos o audios se fabrican con la intención de engañar al público. Simultáneamente, la IA también parece ser la mejor tecnología para identificar estas "falsificaciones profundas" y contrarrestar su engaño (Lyu, 2018).

### 3.4. La protección de los periodistas y las fuentes periodísticas en la era de la IA

La promoción de la seguridad de los periodistas y la lucha contra la impunidad de quienes los atacan son elementos centrales del apoyo de la UNESCO a la libertad de prensa en todas las plataformas de los medios de comunicación. En 2017, la UNESCO publicó un informe titulado *Protecting Journalism Sources in the Digital Age*, en el que se identifican y ponen de relieve las nuevas —y, en gran parte, digitales— formas de retos y peligros a los que se enfrentan los periodistas de todo el mundo, junto con recomendaciones y directrices para proteger a los periodistas y al periodismo (UNESCO, 2017b).

Vale la pena analizar el tema de si la IA puede utilizarse para detectar, rastrear y ayudar a prevenir ataques contra los periodistas. La Universidad de Sheffield está buscando métodos para utilizar la IA para identificar patrones en los datos sobre una serie de ataques (en línea y fuera de línea) contra periodistas (Universidad de Sheffield, 2018). Este enfoque debería incluir el papel del despliegue de IA por parte de los intermediarios entre los productores y los consumidores de periodismo. Entre esos intermediarios figuran las plataformas de redes sociales, los motores de búsqueda y los proveedores de acceso a Internet, que podrían aplicar políticas de promoción y protección del periodismo en línea.

Otro desafío es la necesidad de mejorar la protección de la confidencialidad de las fuentes periodísticas en la era digital, como se reconoció en el documento final de la conferencia "Conectando los puntos: opciones para la acción futura", celebrada en 2015, que fue aprobado por la UNESCO.

El estudio de la UNESCO titulado *Protecting Journalism Sources in the Digital Age* destaca la creciente necesidad de que los intermediarios de Internet y las empresas de telecomunicaciones se aseguren de que están "preparados para enfrentar las escuchas telefónicas", lo que podría tener un efecto inhibitor en la protección de las fuentes periodísticas (Posetti, 2017).

## Recuadro 13: Nuevas formas de ataque contra los periodistas

Hoy en día existen nuevas formas digitales de ataques contra los periodistas, como la elaboración de perfiles, el *doxing* (la publicación de información privada de un individuo en Internet), los videos "profundamente falsos" (la superposición de videos existentes en videos de origen utilizando IA, a menudo con la intención de generar noticias falsas y engañar a los espectadores) y el trolleo (*trolling*) de periodistas (que implica mensajes controvertidos o incendiarios contra un individuo). Si bien no está claro hasta qué punto, por ejemplo, los ataques automatizados de trolleo o de denegación de servicio distribuido (DDoS) están siendo impulsados por la IA como tal, el escenario es, ciertamente, plausible

También son preocupantes los ataques automatizados de los trolls a los periodistas, especialmente con la presencia de empresas que venden "seguidores" en las cuentas de las redes sociales y "retuits" a personas u organizaciones dispuestas a pagar por ellos. Algunos periodistas de investigación de ProPublica trabajaron encubiertos y compraron 10.000 retuits para una cuenta falsa de una compañía llamada "Seguidores y Me Gusta" por 45 dólares, y 5.000 retuits para una cuenta falsa en inglés por 28 dólares (Angwin, 2017). Los periodistas de ProPublica hicieron esto después de haber sufrido ellos mismos un gran acoso en línea. Por ejemplo, una reportera de ProPublica fue acusada de ser lo que se llama "prensatituta" (*presstitute*) en una publicación de Twitter que recibió más de 20.000 retuits (Reporteros sin Fronteras, 2018). Lo más probable es que esto se lograra a través de un servicio de pago similar utilizado por cuentas de redes sociales que coordinan comportamientos y campañas en línea que no son auténticos. De hecho, con solo 100 dólares es posible obtener un "ejército de bots", listo para automatizar el acoso sincronizado (Angwin, 2017).

Los videos de falsificaciones profundas, mencionados anteriormente, son una técnica de IA que se utiliza, entre otras cosas, contra los periodistas. En el Foro para la Gobernanza de Internet (FGI), celebrado en la UNESCO en 2018, Elodie Vialle, de Reporteros sin Fronteras, compartió la preocupación de la organización por la elaboración de videos de falsificaciones profundas dirigidos a periodistas, y especialmente a las mujeres periodistas (UNESCO, 2018e). Esta técnica de IA puede permitir que la gente acose a mujeres periodistas superponiendo sus rostros a los contenidos pornográficos (Reporteros sin Fronteras, 2018). Rana Ayyub, una periodista de investigación de la India, fue víctima de esta técnica. Aunque ni el abuso en línea, ni los comentarios sexistas e islamóforos, ni la desinformación sobre su persona le eran ajenos, dijo que el ataque de la falsificación profunda le afectó especialmente: "Siempre pensé que nadie podía hacerme daño o intimidarme, pero este incidente realmente me afectó como nunca lo hubiera pensado" (Ayyub, 2018). Desde el abuso, Ayyub ha restringido los temas que discute en línea, imponiendo la autocensura. Los expertos en derechos humanos de las Naciones Unidas han intervenido expresando su alarma por este falso video y pidiendo que se tomen medidas para proteger a Ayyub del odio y para investigar los ataques (ACNUDH, 2018<sup>3</sup>)

En un estudio reciente, los periodistas entrevistados expresaron que “la vigilancia masiva tiene el potencial de silenciar a los denunciantes y dificultar cada vez más el periodismo de investigación en todas sus formas” (Waters, 2018). El autor del estudio utilizó el marco teórico del “Panoptismo” de Foucault, que se refiere a “un estado de visibilidad consciente y permanente que asegura el funcionamiento automático del poder” (1977). La conciencia de los periodistas sobre la vigilancia masiva significaba que, independientemente de que estuvieran o no bajo vigilancia, su presencia siempre se hacía sentir, lo que hacía que alteraran continuamente su comportamiento (Waters, 2018). Las fuentes pertinentes para los informes periodísticos pueden igualmente abstenerse de compartir información importante ante el temor de estar siendo observadas o escuchadas por una cámara de reconocimiento facial o un dispositivo de Internet de las cosas de grabación de voz en estrecha proximidad.

## 4. DERECHO A LA IGUALDAD

Si bien muchas personas no experimentan discriminación al navegar por la red invisible de algoritmos que nos rodea, hay muchas que siguen siendo marginadas, discriminadas, privadas de oportunidades o que experimentan los efectos adversos de la adopción de ciertas decisiones algorítmicas. Por ejemplo:

- En 2010, cuando la investigadora Safiya Umoja Noble buscó el término “chicas negras” en Google, encontró páginas en las que predominaba el contenido pornográfico. Aunque los resultados de la búsqueda del término ya no muestran contenido pornográfico, una búsqueda similar para “chicas asiáticas” dio resultados con un alto contenido sexual (Snow, 2018b).
- El programa de Amazon para automatizar su proceso de contratación utilizó un algoritmo que fue entrenado a través del análisis de los currículos presentados a la empresa durante un periodo de 10 años. Los resultados mostraron un sesgo contra las candidatas, ya que el algoritmo se enseñó a sí mismo a dar una calificación más baja a los currículos que incluían expresiones como “capitana del club de ajedrez femenino”. Los resultados reflejaban las desigualdades de género existentes en la industria tecnológica. De hecho, más del 60 % de los empleados de las empresas GAFAM<sup>2</sup> son hombres, y las cifras son mucho más altas si solo se consideran las funciones técnicas (Dastin, 2018).
- La investigación del *software* COMPAS utilizado por los tribunales de los Estados Unidos para predecir la reincidencia (la tendencia de un delincuente condenado a reincidir) reveló un sesgo contra los afroamericanos (Angwin, Larson, Mattu y Kirchner, 2016).
- Se ha comprobado que los algoritmos de riesgo crediticio discriminan a los posibles prestatarios en función de su ubicación geográfica. Si bien los clasificadores explícitos como la raza y el origen étnico pueden esconderse de los

---

2 GAFAM es el acrónimo de Google, Apple, Facebook, Amazon y Microsoft.

algoritmos, esas variables pueden correlacionarse con otros clasificadores, como la ubicación geográfica. Por lo tanto, los algoritmos terminan asignando puntuaciones crediticias discriminatorias desde el punto de vista racial incluso cuando la raza no se utiliza como insumo (O'Dwyer, 2018).

Los algoritmos se arraigarán aún más profundamente en la vida de muchas personas, dado el rápido ritmo con que se está produciendo el proceso de "datización" a medida que se desarrolla el campo de la IA. Por lo tanto, es importante considerar cómo los algoritmos pueden impedir el derecho de las personas a la igualdad.

La inclusión social es uno de los mandatos fundamentales de la UNESCO y consiste en trabajar para construir sociedades del conocimiento que sean socialmente inclusivas, de modo que todos los individuos y grupos puedan participar en la sociedad. A su vez, esto depende de "mejorar la capacidad, las oportunidades y la dignidad de las personas desfavorecidas, sobre la base de su identidad, para participar en la sociedad" (Banco Mundial, 2013). El mismo sentimiento de que nadie debe quedarse atrás impregna los ODS y refuerza los esfuerzos de la UNESCO en este ámbito. Frente a estos objetivos, la violación sistémica del derecho de cada persona a la igualdad es, en principio, un obstáculo para la inclusión.

Esta sección se centra en la discriminación que surge de las diferentes formas de toma de decisiones automatizadas que afectan el derecho a la igualdad. Se identifican los puntos de entrada para la discriminación a través de algoritmos (a veces integrados en el procesamiento de IA). Además, se proponen posibles enfoques técnicos y no técnicos del problema. Sin embargo, cualquier reflexión de este tipo corre el riesgo de generar confusión si no se aborda una cuestión fundamental: ¿Qué es la discriminación?

## 4.1. ¿Qué es la discriminación?

El artículo 1 de la Declaración Universal de Derechos Humanos proclama que "Todos los seres humanos nacen libres e iguales en dignidad y derechos" y el artículo 2 establece que "Toda persona tiene los derechos y libertades proclamados en esta Declaración, sin distinción alguna de raza, color, sexo, idioma, religión, opinión política o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición" (Asamblea General de las Naciones Unidas Resolución 217,1948).

El Pacto Internacional de Derechos Civiles y Políticos reafirma, en muchas de sus disposiciones, este principio general de igualdad ante la ley e igual protección de la ley. El artículo 2 establece que cada uno de los Estados Partes debe garantizar los derechos reconocidos en el Pacto a todos los individuos, sin distinción alguna. El artículo 26 es más amplio y se refiere a la protección contra la discriminación de manera explícita, sin limitarse a los derechos del Pacto:

"Todas las personas son iguales ante la ley y tienen derecho sin discriminación 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 cualquier discriminación por motivos de raza, color, sexo, idioma, religión, opiniones políticas o de cualquier índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social".

Este documento, así como varios otros tratados jurídicos internacionales, prohíbe la discriminación basada en una lista no exhaustiva de identidades de grupo sin intentar definir el significado de discriminación (Vandenhoele, 2005). La igualdad y la no discriminación son ampliamente reconocidas como las declaraciones positiva y negativa del mismo principio (Bayefsky 1990).

La palabra "discriminar" tiene su origen en el latín *discrimināre*, que significa "distinguir entre". En este sentido estricto, la discriminación en sí misma estaría desprovista de cualquier connotación moralizante. Por lo tanto, es importante especificar las condiciones que hacen que la discriminación sea objetable y un factor contra la inclusión social. Según la Enciclopedia de Filosofía de Stanford, la discriminación es moralmente inadecuada cuando implica i) la imposición de una desventaja o privación relativa a personas que pertenecen a un determinado grupo, y ii) es arbitraria (en partes) si la desventaja se adjudica a las víctimas por su pertenencia al grupo (Altman, 2016).

Sin embargo, establecer el marco de reconocimiento de la discriminación no nos proporciona necesariamente herramientas suficientes para el análisis de la discriminación algorítmica. Los algoritmos pueden adquirir un carácter discriminatorio por múltiples vías. Estas incluyen principalmente características del algoritmo que son sesgadas, intencionadamente o no, por los programadores o mediante el refuerzo de los sesgos presentes en los datos de entrenamiento para los algoritmos de aprendizaje de las máquinas. Los algoritmos deben someterse a un análisis de discriminación que permita identificar la discriminación directa, indirecta e institucional para poder aplicar las soluciones reglamentarias o técnicas pertinentes.

Muchos tipos de discriminación pueden ser indirectos. Por ejemplo, un algoritmo que utiliza los patrones de uso del teléfono móvil para determinar la solvencia crediticia de una persona es discriminatorio si asigna un alto riesgo crediticio a las mujeres de comunidades que i) tienen un bajo uso del teléfono móvil o ii) no poseen teléfonos móviles. La condición que se aplica puede parecer igual y justa, pero pone en desventaja a un grupo en particular (Altman, 2016).

Los algoritmos pueden causar y exacerbar estas múltiples formas de discriminación. Los prejuicios sociales y políticos existentes se están sistematizando en los algoritmos de aprendizaje automático de muchas maneras (Packin y Lev Aretz, 2018). Además, vale la pena investigar las posibles nuevas formas de discriminación que puede provocar la IA, como las exclusiones decididas sobre la base de correlaciones estadísticas que no corresponden necesariamente a características socialmente destacadas, pero que, sin embargo, están muy vinculadas a la identidad personal de cada uno.

## 4.2. ¿Cómo se diseña la discriminación en los algoritmos?

La predilección por la toma de decisiones objetivas, combinada con la noción de que los algoritmos procesan los datos de entrada para producir decisiones objetivas, les ha dado un aire de incuestionable superioridad sobre las decisiones tomadas por los humanos. De hecho, a los algoritmos de aprendizaje automático se les asigna un "aura de verdad, objetividad y precisión" (Boyd y Crawford, 2012). Por ejemplo, en un caso en un Tribunal de los Estados Unidos sobre el robo de una cortadora de césped, el fiscal recomendó una sentencia de un año de prisión seguida de un período de supervisión. Sin embargo, el juez, basándose en la evaluación de alto riesgo que un algoritmo hizo del individuo, anuló el acuerdo alcanzado entre la fiscalía y la defensa, e impuso una sentencia de dos años de prisión seguida de tres años de supervisión (Angwin, Larson, Mattu y Kirchner, 2016).

Hay muchos motivos para ser precavidos y no confiar tanto en los algoritmos como árbitros finales de las decisiones, ya que, en el mejor de los casos, solo proporcionan perspectivas útiles. Cualquier reclamo de justicia de los algoritmos debe calificarse por el hecho de que el proceso de toma de decisiones algorítmica tiene dos elementos clave: i) programadores humanos que toman decisiones críticas para enmarcar el problema y la validez del resultado, y ii) datos que pueden representar sesgos históricos, tergiversar grupos o no representarlos en absoluto.

En la siguiente sección se destacan los puntos de entrada humanos y controlados por datos para el posible sesgo de los algoritmos (Barocas y Selbst, 2016).

### i) **Sesgo controlado por el programador**

- a) **Definición de "variables objetivo" y "etiquetas de clase":** La variable objetivo es la variable que debe predecirse, que es la salida del algoritmo. La etiqueta de clase categoriza todas las posibles variables objetivo en conjuntos mutuamente exclusivos. Los programadores, basándose en su comprensión del problema, eligen las variables y las etiquetas. En el caso de un filtro de correo no deseado, el correo electrónico puede clasificarse directamente como una de dos etiquetas: correo no deseado o no correo no deseado. Sin embargo, en el caso de un problema como un algoritmo de contratación, la clase de etiquetas es no binaria y puede reflejar los sesgos del programador o de la organización con el efecto de perjudicar a ciertos grupos sociales.
- b) **Selección de características:** Los programadores eligen los atributos de los datos que deben observarse y utilizarse para el análisis. Si las características seleccionadas de los datos no representan adecuadamente algunos grupos de personas en una granularidad que capte sus diferencias con otros grupos, entonces pueden ser víctimas de una grave desventaja debido a la toma de decisiones automatizada.
- c) **Enmascaramiento:** Aquellos que quieran ocultar sus sesgos e intenciones de desfavorecer a ciertos grupos detrás de la fachada de neutralidad proporcio-

nada por los algoritmos pueden usar la toma de decisiones algorítmica como una máscara. Esto se logra mediante la definición de objetivos prejuiciosos, el etiquetado de clases, la selección de características y la manipulación de datos.

ii) **Sesgo controlado por datos**

a) **Datos de entrenamiento sesgados:** Si las reglas extraídas por el algoritmo de aprendizaje automático de cualquier conjunto de datos se consideran legítimas, los prejuicios y omisiones incorporados en los datos de ejemplo se repetirán en el modelo de predicción.

b) **Representatividad de los datos de muestra:** Un conjunto de datos puede estar sesgado por los datos que no contiene. Si los datos de entrenamiento reflejan una muestra no representativa de la población, entonces los grupos subrepresentados o sobrerrepresentados pueden verse perjudicados por el algoritmo. La falta de representación también puede deberse a zonas oscuras de sombras en los datos, es decir, cuando los datos de determinados grupos de población no se captan en absoluto porque se encuentran fuera del flujo de recopilación de datos. Por ejemplo, el uso de datos de teléfonos móviles como indicador indirecto de la capacidad del usuario para pagar los préstamos puede perjudicar a las personas que tienen un acceso limitado o nulo a los teléfonos móviles. Al mismo tiempo, es importante señalar que incluso los conjuntos de datos representativos reflejan sesgos históricos y sociales, por ejemplo, contra las minorías sobrerrepresentadas en las poblaciones carcelarias o las mujeres en trabajos menos prestigiosos. La propia "representatividad" de los datos puede, por lo tanto, perpetuar la discriminación y la desigualdad, cuando en realidad un conjunto de datos adaptado conscientemente que corrija esas desigualdades sociales podría producir resultados menos discriminatorios a partir de algoritmos entrenados sobre esta base y aplicados posteriormente a casos nuevos (como cuando se utilizan para informar sobre las sentencias de prisión o el análisis automatizado de las solicitudes de empleo).

c) **Correlación no significa causalidad:** La toma de decisiones basada en correlaciones puede llevar a inferencias erróneas. Por ejemplo: "Imagine que pasa unas horas buscando información en línea sobre freidoras. Podría estar buscando un regalo para un amigo o investigando para un informe escolar sobre cocina. Pero para un minero de datos que rastree su historial en línea esta búsqueda podría ser leída como un indicio de un hábito poco saludable. Estamos ante una predicción basada en datos que podría llegar a una aseguradora de salud o a un posible empleador" (Barocas, 2014). Además, cabe señalar que la predicción de los acontecimientos futuros se basa en el supuesto de que los acontecimientos pasados son representativos de los acontecimientos futuros, dadas las condiciones subyacentes similares e inalteradas. El problema es la suposición de que las condiciones subyacentes no han cambiado y el comportamiento continúa.

- d) **Asignación inadecuada cíclica de recursos:** Las predicciones generadas por los algoritmos basados en los datos pueden no asignar recursos a los grupos subrepresentados. Los datos de monitoreo posteriores seguirían el mismo patrón y agravarían la discriminación contra los grupos subrepresentados. Por ejemplo, si un gobierno local rastrea la información sobre baches basándose en la cantidad de sacudidas en la carretera registradas por los teléfonos móviles de los propietarios de vehículos, el gobierno podría dirigir recursos hacia zonas más prósperas con más usuarios de teléfonos móviles y vehículos. Esto disminuye aún más la calidad de las carreteras en los barrios menos acomodados (Crawford, 2013).
- e) **Sesgo inducido por proxy:** Incluso cuando las variables que representan directamente la pertenencia al grupo se eliminan de los datos a fin de evitar la discriminación, puede haber otras variables, necesarias para el análisis, que se correlacionen con las características de identificación del grupo y puedan dar lugar a discriminación. Por ejemplo, incluso si se eliminan del conjunto de datos los indicadores directos de raza, otras variables, como el nivel de ingresos o los patrones de consumo, pueden correlacionarse con la raza y traducirse en decisiones con un sesgo racial. Se necesitan datos sobre las consecuencias de las decisiones automatizadas para identificar la discriminación indirecta.

## Recuadro 14: Sesgos controlados por datos que implican una discriminación basada en la raza

En muchos casos, los algoritmos de aprendizaje automático se entrenan con conjuntos de datos que no son representativos. Cuando estos algoritmos se integran en productos y servicios que permiten la toma de decisiones, pueden ser discriminatorios. Por ejemplo, los investigadores que trabajan en la imparcialidad de los algoritmos han demostrado que los conjuntos de datos (IJB-A y Adience – dos puntos de referencia del análisis facial) usados para entrenar los algoritmos de reconocimiento facial están “compuestos en su mayoría por personas de piel más clara” (Buolamwini y Gebru, 2018). IJB-A y Adience tienen un 79,6 % y un 86,2 % de sujetos de piel más clara. Una consecuencia directa de estos conjuntos de datos no representativos es que las aplicaciones desarrolladas con ellos tienden a clasificar erróneamente los resultados. Por ejemplo, en algunos sistemas de clasificación de género, las mujeres de piel más oscura son las más clasificadas erróneamente, con una tasa de error de hasta el 34,7 %, en comparación con los hombres de piel más clara, en los que la tasa de error máxima es de solo un 0,8 % (Buolamwini y Gebru, 2018). Estas diferencias en las tasas de error prevalecen independientemente de la empresa o el país que haya desarrollado el sistema: las empresas estadounidenses Microsoft e IBM tuvieron tasas de error del 21 % y el 35 %, respectivamente, para las mujeres negras, mientras que la china Megvii tuvo una tasa de error del 35 %. (Buolamwini y Gebru, 2018).

Otro ejemplo se dio en 2015 cuando la aplicación Photos de Google etiquetó a dos individuos de piel oscura como "gorilas". La empresa corrigió el error y se disculpó, pero un informe reciente muestra que la tecnología de etiquetado de imágenes está lejos de ser perfecta, y la solución rápida de eliminar "gorilas" de las etiquetas puede no estar abordando el centro del problema del sesgo (Simonite, 2018).

Por lo tanto, vemos que la toma de decisiones por medio de algoritmos es susceptible a sesgos controlados por humanos y por datos. Gran parte de la toma de decisiones algorítmica se basa en prejuicios implícitos de los programadores o internalizados en los datos.

Las consideraciones históricas y sociológicas proporcionan la información de fondo crucial necesaria para determinar la imparcialidad en los contextos y resultados de la toma de decisiones algorítmica (Michael, Van Kleek y Binns, 2018). Se necesita un análisis algorítmico profundo para defender el derecho a la igualdad y garantizar que las desigualdades históricas relacionadas con el género, la raza/etnia, la orientación e identidad sexual, el nivel socioeconómico, la discapacidad y otros factores de estigmatización no se perpetúen ni se consideren "objetivas".

## 5. CONCLUSIÓN Y OPCIONES DE POLÍTICA

En 2018 y 2019, la UNESCO organizó múltiples debates centrados en los retos y las oportunidades de la era digital, con expertos que representaban a empresas de tecnología, investigadores, gobiernos y organizaciones de derechos humanos, para compartir sus reflexiones sobre el impacto de la IA en la sociedad. Una de esas interacciones puso de manifiesto el abismo que existe entre los enfoques adoptados por los tecnólogos que están a la vanguardia del desarrollo de IA y las organizaciones de derechos humanos que abogan por la equidad. Por un lado, un defensor de los derechos humanos pidió el cese de los sistemas de IA hasta que se aborden completamente la discriminación y los sesgos perpetrados y perpetuados por los algoritmos. Por otro lado, un investigador principal de IA que, aunque reconoció la importancia de los derechos humanos, dijo que era imposible detener el desarrollo de la tecnología.

Cada parte argumentó apasionadamente su punto de vista. Sin embargo, en el debate quedó claro que el progreso en la cuestión del desarrollo de la tecnología y su impacto en la sociedad significa superar los silos. El debate debe ser multidisciplinario y los interesados deben estar

dispuestos a trabajar juntos y a encontrar soluciones en la intersección de sus respectivos ámbitos. Esta cuestión se expone con más detalle en el capítulo sobre la gobernanza de múltiples actores.

## Opciones para todos los interesados

- ▶ Desarrollar y utilizar un marco basado en los derechos humanos para IA bajo el prisma de las normas internacionales de derechos humanos para establecer directrices claras a fin de evitar violaciones de los derechos humanos (incluidos los derechos a la libertad de expresión, la privacidad y la igualdad).
- ▶ Promover y evaluar métodos que puedan evaluar la discriminación algorítmica a fin de proteger el derecho a la igualdad, en particular el de las poblaciones históricamente marginadas.
- ▶ Iniciar, coordinar y apoyar investigaciones interdisciplinarias y con múltiples actores sobre las consecuencias de la IA en los derechos humanos.

## Opciones para los Estados

- ▶ Elaborar marcos normativos y reglamentarios adecuados para hacer frente a los problemas de derechos humanos que plantea el desarrollo y la aplicación de la IA, proporcionando mecanismos para prevenir las violaciones de los derechos humanos, así como para la transparencia, la rendición de cuentas y los procesos de reparación.
- ▶ Prestar atención al enjuiciamiento de los productores de contenidos nocivos demostrados con arreglo a marcos jurídicos acordes con las normas internacionales, y al suministro de alfabetización mediática e informacional al público, en lugar de centrarse exclusivamente en exigir medidas a los intermediarios.
- ▶ Tener en cuenta que hacer responsables a los intermediarios de Internet por el contenido generado por el usuario puede alentar el uso excesivo de IA en la moderación de contenidos, lo que, a su vez, puede tener un efecto negativo en la libertad de expresión.
- ▶ Adoptar medidas eficaces para garantizar que los algoritmos no se exploten para impedir el derecho a elecciones libres.
- ▶ Apoyar el Plan de Acción de las Naciones Unidas sobre la seguridad de los periodistas y la cuestión de la impunidad y hacer frente a los ataques de IA contra periodistas y trabajadores de los medios de comunicación.
- ▶ Evaluar si la regulación vigente contra la discriminación permite a un individuo buscar una solución contra la discriminación algorítmica.

- ▶ Asegurar que el uso de IA por parte del sector público en la toma de decisiones sea transparente y coherente con las obligaciones de derechos humanos.

## Opciones para el sector privado, intermediarios de Internet y la comunidad técnica

- ▶ Realizar evaluaciones de los riesgos para los derechos humanos y actuar con la debida diligencia en relación con las aplicaciones de IA a fin de garantizar que no interfieran con el pleno disfrute de los derechos humanos y las libertades fundamentales a:
  - ✎ Nivel *ex ante*: Evitar la discriminación en la selección de los conjuntos de datos y las elecciones de diseño de los programadores, y hacer explícitos los valores que informan estas elecciones.
  - ✎ Nivel *ex post*: Monitorear de cerca los resultados que podrían violar el derecho de expresión, privacidad e igualdad, así como otros derechos.
- ▶ Crear y ofrecer a los usuarios opciones para optar por no recibir contenido personalizado y elegir modalidades para ordenar la presentación del contenido en función de otros criterios.
- ▶ Promover y demostrar transparencia ofreciendo información sobre lo siguiente de manera accesible:
  - ✎ Desarrollo de algoritmos y aplicación en la personalización de la presentación de los contenidos;
  - ✎ Estadísticas sobre el uso de los sistemas de IA en la moderación de contenidos, incluido el número de remociones (u otras acciones) hechas total y parcialmente por la IA y la frecuencia con que los moderadores humanos deciden en contra de las recomendaciones de la IA;
  - ✎ La recolección y el uso de datos de los usuarios y consumidores, incluido el tipo de datos que reúnen, la forma en que los almacenan y procesan, si los comparten con terceros o los venden a estos y con qué fin, así como las incertidumbres sobre cómo y por qué se utilizarán los datos;
  - ✎ Posibles fallas y riesgos de la IA que podrían traducirse en violaciones a los derechos de los usuarios
- ▶ Notificar inmediatamente al usuario del contenido eliminado y explicar el proceso y los motivos de la eliminación, así como el mecanismo de apelación.
- ▶ Poner en práctica mecanismos de apelación y sistemas de denuncia eficaces que ofrezcan recursos a las personas cuyos derechos hayan sido infringidos o que tengan motivos legítimos para que se elimine o corrija su información.

## Opciones para el sector académico

- ▶ Participar en investigaciones orientadas a los derechos sobre los efectos sociales, económicos y políticos de la personalización del contenido de la IA, incluidos los posibles efectos de las "cámaras de eco" en el desarrollo de las opiniones políticas, así como en la radicalización que conduce al extremismo violento.
- ▶ Llevar a cabo una investigación sobre los posibles efectos inhibidores del uso de la IA en la vigilancia masiva, las restricciones a la libertad de expresión y los impactos en las limitaciones de otros derechos humanos.
- ▶ Proseguir la investigación sobre los efectos de la IA en el pluralismo y la sostenibilidad de los medios de comunicación, así como la investigación sobre el uso de la IA en el fortalecimiento del periodismo y las instituciones de los medios de comunicación.
- ▶ Continuar la investigación sobre la discriminación algorítmica y la forma en que se pueden aplicar soluciones técnicas para garantizar el desarrollo de sistemas de IA que respeten el derecho a la igualdad.

## Opciones para la sociedad civil

- ▶ Adoptar un papel de "perro guardián" para monitorear las violaciones de los derechos humanos por parte de la IA y exponerlas al público.
- ▶ Fortalecer la alfabetización mediática e informacional para comprender mejor las implicaciones de la IA en los derechos humanos.

## Opciones para los actores de los medios

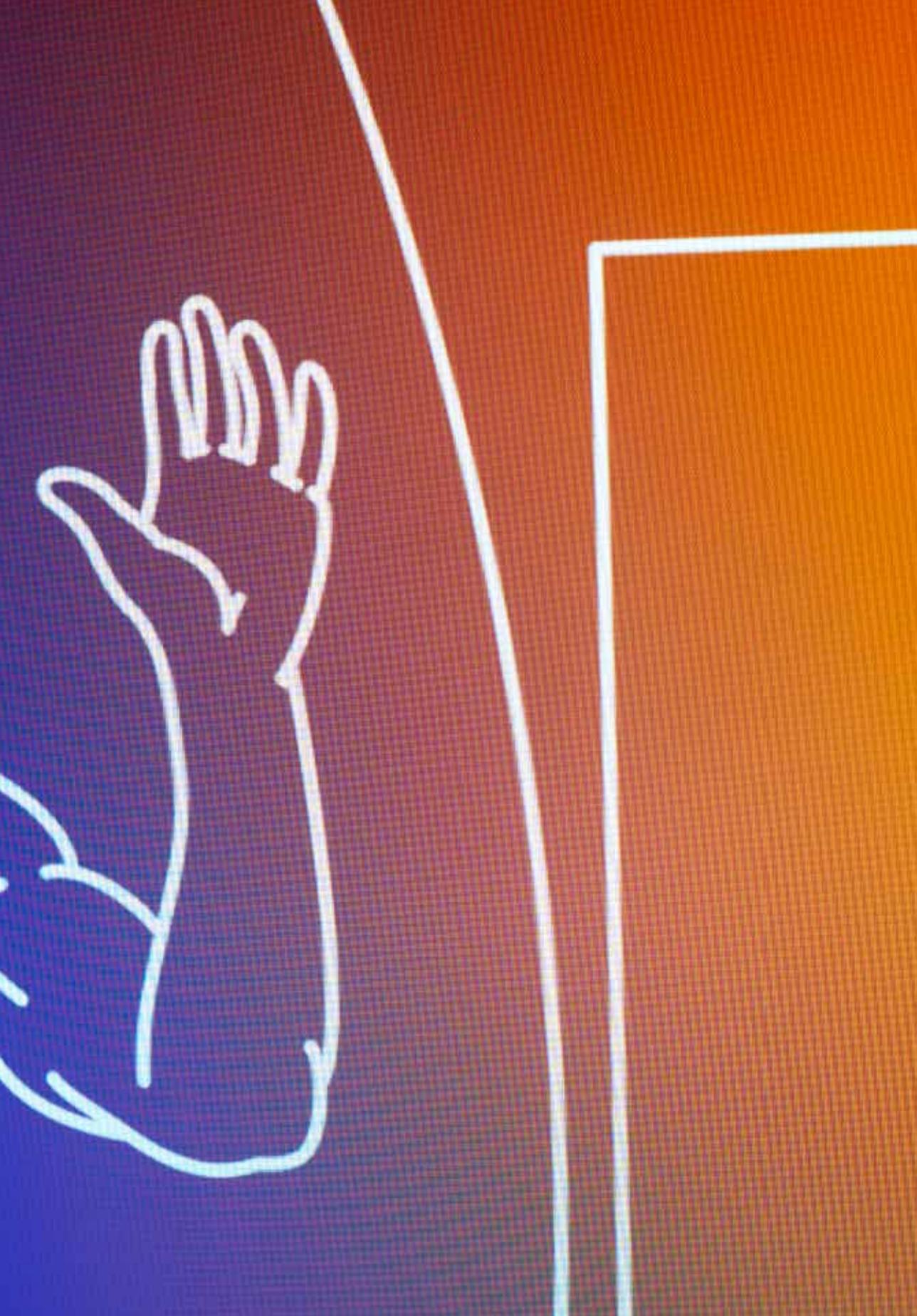
- ▶ Considerar y reflexionar sobre las implicaciones de la IA en la práctica del periodismo y el desarrollo de los medios de comunicación a fin de fortalecer y proteger la libertad de prensa, así como la seguridad de los periodistas.
- ▶ Capacitarse para investigar e informar mejor sobre el desarrollo de la IA y sus aplicaciones, incluida la exposición de abusos y sesgos en la IA, así como los beneficios actuales y posibles desde una óptica realista.
- ▶ Hacer uso de la IA y las nuevas tecnologías en la práctica de la presentación de informes, la producción de noticias y la difusión de contenidos de manera compatible con las normas internacionales de derechos humanos, incluida la privacidad.

## Opciones para las organizaciones intergubernamentales, incluida la UNESCO

- ▶ Comunicarse con distintas iniciativas en relación con la ética y la IA, vinculándolos con los derechos humanos, y organizar diálogos entre múltiples actores en los que se aborden las preocupaciones en materia de derechos humanos.
- ▶ Asistir a los Estados en el cumplimiento con las normas internacionales sobre derechos humanos respecto de la IA.
- ▶ Promover el uso de la IA en el periodismo y los medios que puedan apoyar y proteger sus roles en la sociedad.
- ▶ Participar en debates sobre directrices para la transparencia en lo que respecta al periodismo automatizado y el uso de la IA en los medios de comunicación.

# APERTURA E INTELIGENCIA ARTIFICIAL

# 2



# CAPÍTULO 2: APERTURA E INTELIGENCIA ARTIFICIAL

La apertura es una característica importante de las sociedades del conocimiento (UNESCO, 2015b). La UNESCO aboga por el libre acceso a investigación científica, datos abiertos, recursos educativos abiertos y ciencia abierta como parte de sus esfuerzos para fortalecer el acceso universal a la información y para solucionar las desigualdades en la información. En el contexto de la inteligencia artificial, la apertura puede referirse a la transparencia en general y, más específicamente, a la práctica de poner a disposición del público códigos fuente, plataformas de conocimiento, algoritmos y todos los conocimientos científicos obtenidos en el curso de una investigación (Bostrom, 2016). Además, la apertura denota la ausencia de obstáculos injustos y barreras de entrada para participar en el desarrollo, la aplicación y la revisión de la IA (véase el recuadro 15 a continuación).

La apertura incentiva el uso más amplio y el compromiso con la tecnología, lo que permite que las personas y las sociedades aprovechen la tecnología para su beneficio. Este capítulo mira la apertura y la IA desde varias dimensiones. Primero, cuán abierta es la comunidad de investigación que trabaja en IA, tanto en el sector académico como en el sector privado. Esto se relaciona específicamente con la apertura en la publicación y la difusión de la investigación y herramientas de IA. Segundo, cuán abiertas son las fuentes de datos que pueden usarse para el desarrollo de sistemas de IA. Tercero, "explicabilidad" y transparencia en la forma en que los sistemas de IA toman decisiones, y transparencia y responsabilidad de parte de las organizaciones (gobiernos, sector privado y otros) que usen los sistemas de IA para la toma de decisiones. Cuarto, el papel de los mercados abiertos y la competencia en investigación en IA. Quinto, la IA se explora a través del lente de una tecnología de uso doble y se discuten los riesgos asociados con el posible mal uso de las publicaciones de investigación abiertas y varios modelos de publicaciones. El capítulo finaliza con opciones para el fortalecimiento de la apertura respecto de la IA en las diferentes dimensiones analizadas.

## Recuadro 15: Posición de la UNESCO respecto de la apertura para la universalidad de Internet

La Universalidad de Internet resalta la norma de apertura de Internet. Esta designación reconoce la importancia de cuestiones tecnológicas tales como las normas abiertas y las normas de política para el libre acceso al conocimiento y la información. La apertura también muestra la importancia de la facilidad de ingreso de los actores y la ausencia de un cierre que podría imponerse a través de monopolios.

Internet debe ser abierta para todos para desarrollar o beneficiarse de sus recursos y oportunidades, de la forma que les resulte más conveniente o valiosa. A través de la apertura, el concepto de universalidad de Internet admite la integridad de Internet al permitir un intercambio mundial común en lugar de estar confinada a "jardines cercados" debido a tecnologías incompatibles. Destaca la importancia de cuestiones digitales tales como las normas abiertas y el libre acceso al conocimiento y la información.

Las normas abiertas, la interoperabilidad, las interfaces de programación de aplicaciones públicas y el *software* de código abierto han hecho una contribución de vital importancia para Internet. Los mercados abiertos también desempeñaron un rol importante en el desarrollo de Internet, al permitir que negocios innovadores y competitivos accedan a los mercados en lugar de excluirlos mediante acuerdos de licencias restrictivos o limitaciones proteccionistas en la provisión de servicios.

# 1. APERTURA EN LA INVESTIGACIÓN EN IA

Tanto el sector académico como el sector privado muestran gran interés por la investigación en IA. Ese interés no resulta sorprendente, dado el potencial de la IA para generar actividad económica adicional de alrededor de 13 billones de dólares para 2030, lo que representa un 1,2 % adicional en crecimiento mundial del PIB por año (Bughin, Seong, Manyika, Chui y Joshi, 2018). En 2016, según algunas de las estimaciones, gigantes de la tecnología como Google y Baidu invirtieron entre 20.000 y 30.000 millones de dólares en IA, y un 90 % de esa inversión fue a parar a investigación y desarrollo (Bughin y otros, 2017).

La apertura en la investigación adelantaría la difusión de nuevos conocimientos y permitiría que más personas desarrollen investigación y aplicaciones sobre la base de las técnicas de IA de última generación, accesibles para todos (Bostrom, 2016). En los últimos cinco años, las publicaciones sobre investigación en IA crecieron en

## Cantidad de artículos de IA en arXiv por subcategoría (2010-2017)

Fuente: arXiv

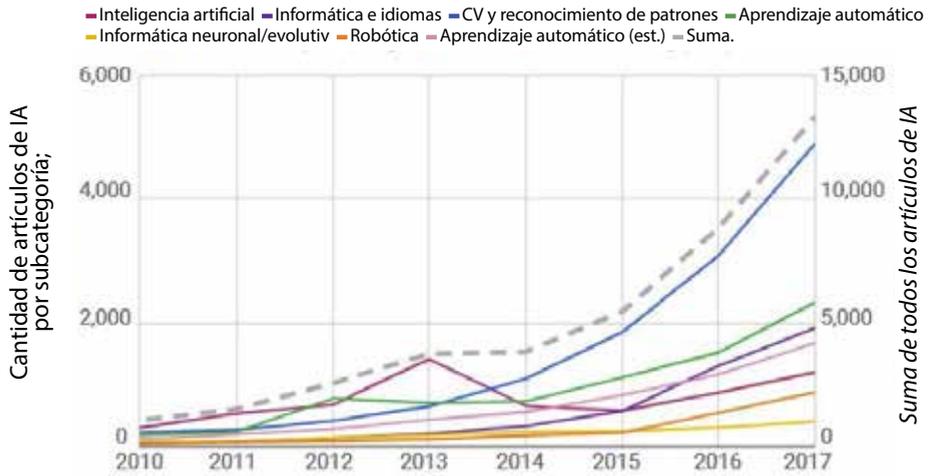


Figura 1: Cantidad de artículos de IA en arXiv por subcategoría (Shoham y otros, 2018).

un 12,9 % anual y ahora se editan cerca de 60.000 publicaciones de investigación por año (Elsevier, 2018). La Figura 1 muestra la cantidad de artículos de IA en arXiv, un archivo abierto de artículos de investigación, por subcategoría. Los artículos correspondientes a las áreas de "Aprendizaje automático" y "Visión artificial y reconocimiento de patrones" aumentaron un 37,4 % anual en los últimos cinco años (Elsevier, 2018).

El Informe de la UNESCO sobre la Ciencia 2015 destaca el papel de las diferentes culturas de trabajo en la forma en que los investigadores del sector público y privado difunden el conocimiento generado. Por lo general, los científicos que trabajan en universidades públicas tienden a publicar de forma abierta, ya que su reputación depende de la evaluación de sus trabajos por parte de los colegas. Los científicos que trabajan para empresas privadas están comprometidos con los intereses de negocios de sus empleadores que pueden requerir reserva y apropiación del conocimiento para los intereses de la empresa (UNESCO, 2015c). Sin embargo, los investigadores, los desarrolladores y las empresas que trabajan en el área de la IA han demostrado una tendencia a la apertura al compartir su trabajo de forma regular en conferencias académicas y mediante plataformas de código abierto (Bostrom, 2016).

Los investigadores de IA han demostrado su compromiso con la investigación abierta. Por ejemplo, una petición para llevar adelante un boicot contra una publicación de IA que no tenía libre acceso juntó más de 3.000 firmas, principalmente de investigadores (Hutson, 2018). El promotor del boicot fue Tom Dietterich de la Oregon State University, quien declaró que la IA debe ser transparente y abierta para toda la comunidad (Robitzski, 2018). Algunos científicos informáticos están a favor de un sistema de publicación y revisión abierta para agilizar el proceso de distribución y aumentar la transparencia al publicar abiertamente también las revisiones de los artículos (LeCun, 2009; David Soergel, 2013). Como resultado, varias de las conferencias más importantes de IA han comenzado a utilizar plataformas como *OpenReview.net*, que ofrecen una vía para la apertura en la comunicación científica, en particular en el proceso de revisión por parte de los colegas (Hutson, 2018).

Además, con frecuencia los investigadores comparten el código fuente y la arquitectura detallada que permite que otros prueben y desarrollen las tecnologías demostradas, y de esa manera fortalecer el intercambio del conocimiento dentro de la comunidad. En el área de la IA ya existen muchas tecnologías de código abierto (Bostrom, 2016).

La publicación de sus investigaciones permite a los investigadores compartir conocimiento y mostrar sus capacidades a una audiencia mayor dentro de la comunidad de la tecnología. Por lo general, los investigadores prefieren trabajar con empresas que permiten la publicación abierta de la investigación, ya que esto aumenta su valor de mercado y su prestigio en la comunidad de investigación de IA (Bostrom, 2016).

La movilidad de los investigadores e ingenieros entre las empresas de tecnología y los laboratorios de investigación de las universidades promueve el intercambio de conocimiento entre diferentes organizaciones. Otra forma por la que se logra la movilidad es a través de los laboratorios de investigación instalados en todo el mundo por empresas del sector privado (UNESCO, 2015c).

En el caso del sector privado, hay muchas razones para alentar que la investigación hecha por sus empleados se comparta abiertamente, incluso:

- i) mostrar sus capacidades de investigación a fin de atraer más talento,

- ii) mejorar sus productos al involucrar una comunidad más amplia de IA, lo que incluye a piratas informáticos y expertos en seguridad, para probar, desafiar y encontrar fallas en su oferta,
- iii) crear una comunidad de desarrolladores para incentivar el desarrollo de aplicaciones en fase posterior sobre la base del núcleo de su tecnología, e
- iv) incidir con su tecnología en las normas de la industria.

Por las razones expuestas anteriormente, y como parte de sus modelos de negocios, muchas empresas abren más sus plataformas de aprendizaje automático y servicios en la nube al acceso de investigadores y desarrolladores, lo que ayuda a solucionar las brechas en el acceso tanto a los recursos tecnológicos como al *hardware* que quizás no esté disponible para todos los usuarios. Por ejemplo, Amazon Web Services, Google Cloud Platform y Microsoft Azure Cloud reconocen que su infraestructura informática puede ser una línea de negocios más allá del uso exclusivo propio, y venden servicios de computación en la nube (Varian, 2018).

TensorFlow de Google, una de las tecnologías de IA más utilizadas, es un marco de aprendizaje automático de código abierto que permite que los usuarios desarrollen redes neuronales y otros modelos informáticos mediante el uso de diagramas de flujo (Garbade, 2018). También han surgido fundaciones que ofrecen ambientes para la colaboración de código abierto entre las instituciones y los entusiastas. Por ejemplo, la Fundación sobre el Aprendizaje Profundo de la Fundación Linux incorpora proyectos tales como Acumos AI, que es una plataforma y un marco que hace más fácil construir, compartir y desplegar aplicaciones de IA (Bommireddipalli y otros, 2018).

Además, como prueba de la apertura del sector privado en relación con la investigación en IA, se informa que más del 70 % de la reciente investigación en IA llevada a cabo por las corporaciones en los Estados Unidos se publicó como artículos de conferencias (Elsevier, 2018).

Los consorcios de investigación (como Open AI) se crearon con el objetivo de hacer que la investigación en IA sea abierta y accesible, y así asegurar el desarrollo de tecnologías de forma que se mitiguen los monopolios del conocimiento que dominan el área de la IA en el futuro (aunque los monopolios de datos y los sistemas de conocimiento cerrados aún existen en esferas como las fuerzas armadas, empresas de Internet e industrias de la salud, entre otras).

Sin embargo, aún hay vacíos en el entendimiento del nivel de la apertura en el sector privado en relación con la IA, incluso respecto del papel de las patentes. Además, se necesita más investigación para entender las diferencias en el acceso a estos servicios en términos demográficos y de geografía.

## Recuadro 1 6: Tecnologías y plataformas clave

TensorFlow es una biblioteca de *software* de código abierto para computación numérica de alto rendimiento. Su arquitectura flexible permite el fácil despliegue de la computación a través de una variedad de plataformas, que van desde computadoras de escritorio hasta servidores, pasando por dispositivos móviles y perimetrales (el proceso perimetral es un paradigma informático distribuido que acerca al almacenamiento de datos de computación a la ubicación donde se necesita). Los investigadores y los ingenieros de Google desarrollaron TensorFlow para ofrecer soporte para el aprendizaje automático y el aprendizaje profundo. Su núcleo flexible de computación numérica es útil en varios campos científicos. (Fuente: <https://www.tensorflow.org/>)

GitHub es una plataforma abierta para que los desarrolladores de *software* trabajen juntos para resolver problemas desafiantes usando las tecnologías más importantes. Hasta agosto de 2019 la plataforma tenía más de 40 millones de usuarios registrados en todo el mundo y 100 millones de repositorios. (Fuente: <https://github.com/about/>).

Si bien hay razones altruistas, tecnológicas y de mercado convincentes para la apertura en la investigación en IA, es necesario saber que esto no basta para permitir que las personas desarrollen capacidades en IA y contribuyan al desarrollo de la IA. Las personas necesitan, entre otras cosas, acceder a los datos, y los datos abiertos pueden desempeñar un rol importante en este sentido

## 2. DATOS ABIERTOS E IA

Las políticas de datos abiertos se relacionan con los datos recabados por los gobiernos —y, a veces, por otros interesados— publicados abiertamente para que las personas, las empresas, el sector académico y las organizaciones de la sociedad civil utilicen los datos para apoyar sus propios objetivos. Los beneficios de las políticas de datos abiertos incluyen: mejor acceso al conocimiento, oportunidades para innovación y provisión de servicios, análisis de datos mejorado mediante la recombinação de datos de diversas fuentes y mejor formulación de políticas gracias a la mejora en la transparencia y la responsabilidad. Los acuerdos sobre la protección de los datos son importantes para asegurar que los conjuntos de datos abiertos no afecten la privacidad de las personas (UNESCO, 2018a). Además, cabe destacar que, debido a su continuo procesamiento, los datos también sufren transformaciones en la información que brindan. Por ejemplo, el etiquetado de los datos agrega otra capa de información que permite que los datos se conviertan en conocimiento.

Los conjuntos de datos del gobierno son solo una de las fuentes para obtener datos útiles para el desarrollo y la aplicación de IA y sus elementos. El debate en torno a los datos abiertos en relación con la IA muestra una variedad mucho más amplia de fuentes como el raspado web, la recolección de datos como un producto derivado de las aplicaciones móviles, y los datos comunes, entre otros. Esta sección muestra posibles fuentes de datos que son de utilidad para el entrenamiento y la aplicación de la IA (Varian, 2019).

- i) **Raspado web:** Implica la recolección automática de datos que se encuentran en sitios web públicos. Si bien esta fuente de datos está disponible abiertamente, hay cuestiones éticas y legales relacionadas con la descarga y el uso de datos raspados de sitios web. Por ejemplo, un programador raspó 40.000 perfiles del sitio de citas Tinder para crear un conjunto de datos para entrenar algoritmos de reconocimiento de imagen, lo que generó problemas de privacidad para los usuarios de la plataforma (Lomas, 2017). Algunas organizaciones impiden el raspado web mediante la implementación de controles adicionales en sus sitios web.
- ii) **Datos generados a través de una oferta de servicios "gratuitos":** Las plataformas de las redes sociales son un buen ejemplo de esta categoría de generación de datos. Intercambian el acceso a su plataforma por datos de los usuarios que pueden monetizarse de diferentes maneras. Por lo general, esos datos son útiles para mejorar los servicios. Por ejemplo, Google desarrolló sus conocimientos técnicos en datos de reconocimiento de voz sobre la base de los comandos de voz dados por los usuarios a sus teléfonos, y las elecciones que hicieron en virtud de los resultados de la búsqueda por voz. De manera similar, ReCAPTCHA, una tecnología que se utiliza para detectar si la que intenta acceder a un servicio en línea es una persona real o un robot, recolecta datos al pedir a los usuarios que etiqueten imágenes, lo que luego se utiliza para entrenar sus algoritmos de aprendizaje automático (O'Malley, 2018). Los datos recabados de esta forma pueden ponerse —o no— a disposición pública de otros usuarios; esto dependerá de la voluntad del dueño de querer compartirlo. El etiquetado de datos de código abierto podría ser una opción para apoyar el desarrollo de los datos comunes.
- iii) **Datos recabados como productos derivados de operaciones:** Estos son datos que se recaban como parte de las operaciones de rutina de los negocios. Puede incluir facturas de consumo generadas en un restaurante que pueden usarse para definir la demanda de comida semanal o de la temporada, y, por lo tanto, ayudar a gestionar las cuentas del supermercado y reducir los desperdicios de comida. Los datos provenientes de Internet de las cosas pueden ser un recurso abierto de mucho valor.
- iv) **Datos generados por computadora:** Los algoritmos de aprendizaje automático también pueden generar sus propios datos. Por ejemplo, el algoritmo de AlphaGo generó datos al jugar al juego "Go" contra sí mismo. De manera

similar, las imágenes sintéticas creadas al modificar las imágenes originales se utilizan para entrenar algoritmos.

- v) **Contratar humanos para etiquetar artefactos para usar como datos:** Se contrata a humanos para etiquetar manualmente datos que se utilizan para entrenar algoritmos. Sin embargo, esto puede ser costoso e insumir mucho tiempo como inversión. Decenas de miles de personas trabajan en todo el mundo en el etiquetado de datos como trabajadores independientes o a través de plataformas de producción colectiva, como Amazon Mechanical Turk, que permiten que las empresas distribuyan tareas de etiquetado de datos a la persona que quiera hacerlo (Metz, 2019).
- vi) **Datos comunes:** Los repositorios de datos en línea para palabras, imágenes y otras formas de comunicación existen y se desarrollaron gracias al aporte de los usuarios. Los datos comunes actúan como recursos compartidos y se analizan con mayor detalle en el próximo capítulo sobre Acceso.

Hay preguntas relativas a la legalidad del raspado de datos de los sitios web, la concentración de datos en manos de unas pocas empresas y la construcción de repositorios de datos abiertos para mejorar el acceso a los datos para el entrenamiento de algoritmos. Sin embargo, las políticas que incentivan la apertura de datos etiquetados con la correcta consideración de la privacidad serían útiles para el desarrollo de la IA. Otras dimensiones relevantes pueden encontrarse en el capítulo sobre Derechos, que aborda la discriminación motivada por los datos, y en el capítulo sobre Acceso, donde se discute de qué manera el acceso a datos puede salvar la brecha digital de IA. En la próxima sección se hace referencia a la apertura y la transparencia de los algoritmos de IA.

### 3. APERTURA EN LA IA: CAJA NEGRA Y PROBLEMAS DE TRANSPARENCIA

La automatización de la toma de decisiones, en tanto elemento que puede ser autónomo, pero que también se está convirtiendo en una característica de los componentes subyacentes de la IA, enfrenta a los encargados de la formulación de las políticas con la cuestión de asegurar la responsabilidad y la transparencia en las decisiones tomadas por las máquinas. Algunos tipos de IA no dependen de programas predefinidos para desempeñar tareas. En cambio, los algoritmos de aprendizaje automático "pueden aprender, adaptarse a los cambios en medio de un problema, establecer patrones en situaciones en las que no se conocen las reglas, y lidiar con información confusa o incompleta" (Negnevitsky, 2011). Por lo tanto, a pesar de que los pasos que se tomaron para llegar a una decisión pueden describirse, es poco probable que el detalle nos enseñe mucho más acerca de la forma en que realmente

se tomó la decisión. El proceso es similar a depender de la intuición para llegar a una decisión, sin comprender claramente de dónde viene la intuición (Mukherjee, 2017). En este sentido, la apertura enfrenta un reto técnico en relación con la explicabilidad de las decisiones algorítmicas. Esto también se denomina el problema de la "caja negra" de la IA por dos razones: i) la complejidad y ii) la dimensionalidad de la toma de decisiones algorítmica, que impide a los humanos entenderla (Bathae, 2018). Estos asuntos son evidentes en dos métodos de IA muy utilizados —redes neuronales profundas y máquinas de vectores de soporte—, que son particularmente resistentes a la apertura. Estos se analizan a continuación:

- i) **Redes neuronales profundas:** Una red neuronal profunda se basa en la capacidad de una red de neuronas artificiales para aprender gradualmente sobre la base de su programación y los resultados del procesamiento de datos. Al igual que en las neuronas humanas, los enlaces útiles se fortalecen y los superfluos se descartan. En esta metodología se utilizan varias capas de neuronas interconectadas para encontrar patrones en los datos de forma progresiva, o para hacer conexiones lógicas o relacionales entre puntos de datos (Negnevitsky, 2011). Dado que no hay una "neurona" que codifique una parte específica del proceso de toma de decisiones y se llega a la decisión sobre la base de la red de "neuronas", en este punto del desarrollo tecnológico no es posible reducir la decisión a pasos lógicos específicos (Bathae, 2018). Por consiguiente, las redes neuronales profundas implican una toma de decisiones que es compleja de descifrar.
- ii) **Máquinas de vectores de soporte:** Los humanos tienen la capacidad de imaginar espacios tridimensionales, es decir, crear una imagen mental de un plano usando tres variables; cualquier cosa más allá de las tres dimensiones no es de fácil acceso para nuestro cerebro (Carroll, 2009). Las máquinas de vectores de soporte resultan bastante misteriosas para los humanos porque llegan a una decisión encontrando patrones geométricos entre muchas variables que los humanos no pueden visualizar fácilmente. Por lo tanto, las curvas no lineales generadas por las máquinas de vectores de soporte son cajas negras para la mente humana por su gran dimensión.

El problema de la "caja negra" hace que la IA sea poco clara y genera varias interrogantes que son todo un desafío en relación con la rendición de cuentas, la transparencia y la responsabilidad por las decisiones tomadas por los algoritmos. Los creadores de algoritmos de IA podrán definir los objetivos generales de los algoritmos, pero la caja negra de la IA podrá alcanzar estos objetivos de cierta forma que quizás hasta sus creadores no entiendan ni puedan predecir. Los pasos intermedios seguirán siendo oscuros debido a los factores de complejidad y dimensionalidad tratados anteriormente. Por lo tanto, la cuestión de la intención, utilizada tradicionalmente como un factor para determinar la responsabilidad, es imposible de satisfacer en algunas circunstancias, ya que no se puede decir que las máquinas tengan intenciones y la única intención de los creadores humanos era alcanzar el objetivo definido.

Un ejemplo del problema de la caja negra puede verse en el sistema legal. En los Estados Unidos ya se está usando la codificación predictiva para determinar si la reincidencia es más probable en asuntos penales y para ayudar en la toma de decisiones sobre las sentencias. Eric L. Loomis, un hombre de Wisconsin, fue sentenciado a seis años de prisión, en parte debido a un *software* de propiedad de una empresa privada. Loomis planteó que se estaba violando su derecho al debido proceso en virtud de la consideración de un juez sobre un informe generado por el algoritmo secreto del *software*, que él no pudo inspeccionar o cuestionar (Liptak, 2017). En este caso la empresa privada pudo haber entendido el algoritmo utilizado para llegar a la decisión, pero en casos en los que ni siquiera los creadores del algoritmo entienden cómo llegó a esa decisión, ¿quién es responsable de los errores? Cuando un algoritmo no es diseñado por humanos se hace difícil determinar si utilizó correlaciones falsas o si discriminó contra un grupo vulnerable, y aún más difícil decidir quién será responsable cuando discrimine.

Otro ejemplo interesante es el de una tecnología de IA llamada CycleGAN, desarrollada por investigadores de Stanford y Google. Esta IA se diseñó para convertir imágenes satelitales en mapas de calles y al revés, y tenía dos tareas distintas:

- i) Convertir fotos aéreas en mapas
- ii) Convertir un mapa en una foto aérea parecida a la foto original

La eficiencia del algoritmo se calificó según el parecido de la recreación de la foto original a partir del mapa. La mejor forma de hacerlo se dejó para que el algoritmo la determinara. Los resultados fueron sorprendentes debido a que los investigadores vieron que la IA era extremadamente eficiente al saltar el paso intermedio de generar un mapa a partir de la imagen para luego utilizar el mapa para generar nuevamente una imagen aérea. En vez de ello, comenzó a producir las imágenes directamente a partir de la imagen original. A pesar de que la IA era perfectamente lógica, dado su objetivo de lograr una mejor calificación, lo logró mediante lo que los humanos denominarían una "trampa" (Tech2News, 2019). Las consecuencias de una "trampa" similar por parte de los algoritmos podrían ir en detrimento del interés público.

El problema de la caja negra se complica aún más cuando múltiples algoritmos interactúan entre sí y plantean riesgos sistémicos. Por ejemplo, un programa de comercio de valores podrá tener el objetivo de maximizar la ganancia, pero si lo logra a través de la manipulación del mercado o a través de medios lícitos es difícil de discernir con anterioridad y con posterioridad (Bathae, 2018). Dado que muchos de esos algoritmos con distintos niveles de complejidad responden a cambios en el mercado de valores, es decir que interactúan de forma implícita entre sí, se crea una situación de "legiones de algoritmos poderosos, de comercialización superrápida — instrucciones sencillas que interactúan para crear un mercado incomprensible para la mente humana e imposible de predecir" (Salmon y Stokes, 2010).

A pesar de estas barreras, se intenta que los algoritmos sean más abiertos y transparentes. El Reglamento General de Protección de Datos (RGPD) de la Unión Europea exige que las organizaciones expliquen ciertas decisiones tomadas por los algoritmos.

mos (Algorithm Watch, 2019). Esta directiva empodera a las personas con el derecho a "información significativa sobre la lógica utilizada" en el algoritmo (Goodman y Flaxman, 2017). Hay propuestas relacionadas con la auditoría algorítmica, algo similar a las auditorías financieras con profesionales de datos, normas y directrices para llevar a cabo las auditorías (Guzzcza, Rahwan, Bible, Cebrian y Ka-tyal, 2018). Dada la deliberada carencia de apertura corporativa o tecnológica para compartir información sobre los algoritmos y los datos subyacentes, algunos defienden los resultados de la producción colectiva de algoritmos para luego someterlos a un profundo análisis para detectar cualquier problema o discriminación (Stray, 2018).

Los esfuerzos fuera del sector público apuntan más hacia el aumento de la comprensión de la IA. En el sector privado, IBM creó AI Explainability 360, una colección de algoritmos de última generación de código abierto que utiliza distintas técnicas para explicar la toma de decisiones del modelo de IA, además de AI OpenScale, que monitorea el código de IA final de un desarrollador para buscar legitimidad y descubre problemas ocultos que puedan aparecer durante la vida útil de la aplicación (IBM, 2019a; IBM, 2019b). Google también creó Testing with Concept Activation Vectors, un método de interpretación que puede usarse para entender cuáles son las señales que usan los modelos de redes neuronales para la predicción (Been y otros, 2018). Por el lado del sector académico, ProtoDash es un algoritmo para encontrar "prototipos" en un programa de aprendizaje automático, o sea, un subconjunto de los datos que tienen mayor influencia en el poder de predicción del modelo. Este algoritmo, por ejemplo, podría explicar los resultados del modelo de calificación para un crédito a un consumidor al que recientemente se le ha negado un préstamo, o a un ejecutivo de préstamos que precisa una explicación de la toma de decisiones del modelo de IA para cumplir con la ley (Gurumoorthy, Dhurandhar, Cecchi y Aggarwal, 2019). Esta lista de soluciones (que no es exhaustiva) puede echar luz sobre la apertura en relación con algoritmos de IA complejos. Sin embargo, muchas de las propuestas actuales para mejorar la transparencia en la toma de decisiones automatizada aún están lejos de rendir frutos y se precisan mayores esfuerzos para resolver los problemas de la caja negra de la IA.

## 4. PAPEL DE LOS MERCADOS EN LA IA ABIERTA

Los mercados abiertos, incluidos los mercados abiertos de forma imperfecta que existen hoy en día, promueven la competencia entre las empresas y mejoran el bienestar del consumidor al acercar el precio de los productos y servicios al costo marginal de producción. A este respecto, la apertura en la IA permite la difusión de investigación innovadora entre las empresas que compiten. Dado el nivel actual de apertura en la comunidad de investigación en IA y los esfuerzos de los investigadores para presionar por más apertura, las empresas se apresuran a mejorar las tecnologías de última generación e implementarlas para realzar sus productos y servicios. Sin

embargo, las empresas pueden restringir la posibilidad de compartir e implementar sus tecnologías de IA, en un intento por tener la ventaja del que mueve primero y expandir sus acciones de mercado. En algunas circunstancias, la competencia puede comprometer los derechos humanos e impedir que las empresas colaboren para limitar los riesgos de IA. Los asuntos relacionados con el comportamiento de los actores del mercado en el uso y el despliegue de IA requieren de un diálogo entre los interesados para desarrollar prácticas de IA corporativas y éticamente responsables. Los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas pueden ser una herramienta de utilidad para guiar a los Estados y los negocios ante los riesgos de los derechos humanos que puedan enfrentar mientras desarrollan o despliegan sistemas de IA. La incapacidad del mercado para abordar estos y otros asuntos, tales como las concentraciones de la esfera de mercado, lo que reduce las oportunidades para una competencia limpia, es una razón para las intervenciones regulatorias.

## 5. RIESGOS DE LA APERTURA Y RESPUESTAS

Si bien la apertura y la transparencia tienen varias ventajas que ya se mencionaron, la apertura también puede implicar riesgos asociados con el mal uso de la tecnología. Esta sección identifica algunas de las formas en que se está implementando la IA para aplicaciones que pueden implicar un riesgo para los derechos humanos o la democracia.

La IA puede desplegarse con propósitos beneficiosos o dañinos. El acceso a algoritmos, códigos fuente y conjuntos de datos en repositorios digitales facilita el desarrollo de usos beneficiosos de nuevas tecnologías de IA. Sin embargo, el mismo conocimiento también puede usarse para desarrollar aplicaciones dañinas. Por ejemplo, las aeronaves no tripuladas que usan los sistemas de IA para la optimización de la trayectoria de vuelo pueden utilizarse para la entrega de suministros de emergencia en áreas remotas o inaccesibles, pero también pueden usarse para ataques militares no justificados. Por una parte, la IA se utiliza para detectar desinformación, mientras que por la otra puede usarse para difundir desinformación a través de videos, imágenes y titulares falsos a bajo costo, gran escala y mayor eficiencia que antes (Brundage y otros, 2018; Conner-Simons, 2018; Schwartz, 2018).

Un grupo de expertos en la Conferencia Asilomar<sup>1</sup> realizó cuatro observaciones acerca de la IA (Brundage y otros, 2018):

- i)                      por lo general los sistemas de IA son tan eficientes como expansibles,

1 La Conferencia Beneficial AI 2017 fue un taller y conferencia en donde se reunieron los investigadores más importantes de IA del sector académico, la industria y los líderes de pensamiento en economía, Derecho, ética y filosofía, para desarrollar los Asilomar AI Principles: [https:// futureoflife.org/ai-principles/](https://futureoflife.org/ai-principles/).

- ii) los sistemas de IA pueden exceder las capacidades de los humanos,
- iii) los sistemas de IA pueden aumentar el anonimato y la distancia psicológica, y
- iv) los desarrollos de IA se prestan para una rápida difusión

Todas estas observaciones juntas demuestran la potencia de la IA como herramienta con potencial para dañar. Pueden crearse y adaptarse a usuarios individuales correos electrónicos, sitios web, vínculos y bots de *chat* realistas, maliciosos y automatizados a bajo costo, los que pueden utilizarse para apuntar de manera encubierta a ciertas comunidades específicas o difundir propaganda política (Benkler, Faris, Bourassa y Roberts, 2018).

El capítulo anterior sobre los derechos humanos mostró de qué manera la IA puede contribuir a la discriminación debido a los sesgos del programador o a asuntos relacionados con los datos. Estos factores exigen que la transparencia sea parte de la apertura. Al mismo tiempo, la apertura también puede verse comprometida por el mal uso de la tecnología con propósitos dañinos.

Por ejemplo, la apertura puede permitir "jugar con el sistema". Una característica técnica que puede distorsionar los resultados de un algoritmo se denomina "ejemplo conflictivo". Los ejemplos conflictivos son entradas a los algoritmos de aprendizaje automático que obligan al modelo a cometer errores (GoodFellow y otros, 2017). La Figura 2 muestra cómo una entrada conflictiva hace que el algoritmo cambie la clasificación de una imagen de un panda a la de un gibón (Goodfellow, Shlens y Szegedy, 2015).

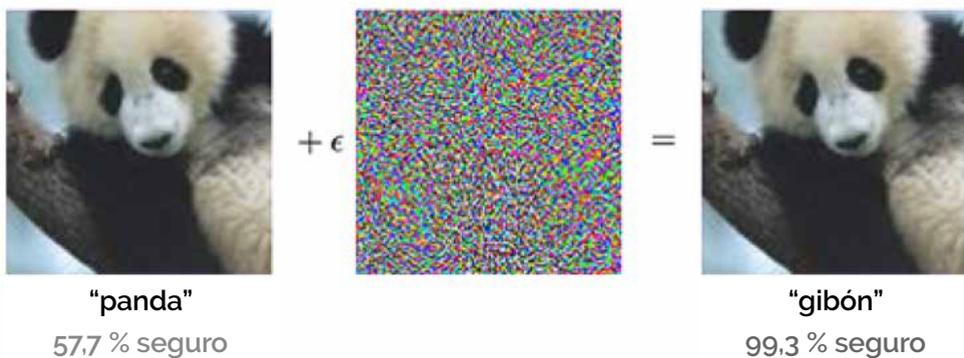


Figura 2: Ejemplo de una capa de entrada conflictiva agregada a la imagen de un panda que lleva a su clasificación errónea como un gibón (Goodfellow, Shlens y Szegedy, 2015)

El ejemplo del panda puede parecer inofensivo. Sin embargo, la misma tecnología utilizada para piratear los algoritmos puede causar que los vehículos autónomos no reconozcan o malinterpreten las señales viales, lo que puede llevar a la muerte de los pasajeros (Papernot y otros, 2017).

Además de los ejemplos conflictivos que distorsionan el uso de algoritmos para causar daño, la apertura de la IA a muchos usuarios implica riesgos. Por ejemplo, la Figura 3 muestra cómo la tecnología para la generación de rostros sintéticos reales ha logrado avances importantes en los últimos cinco años, desde el rostro granuloso de 2014 al rostro nítido y realista de 2017 (Brundage y otros, 2018). Mediante la utilización de la IA se pueden crear personas que no existen, pero parecen reales. Esas tecnologías pueden usarse para tentar a las personas a creer en información falsa, en particular para difundir desinformación dañina y de odio en las redes sociales (Chesney y Citron, 2018). A no ser que tengan algún conocimiento de la ciencia forense, es difícil que la gente reconozca "falsificaciones profundas", y esto puede erosionar su confianza en la sociedad y socavar la democracia (Benjamin, 2019).



Figura 3: Evolución en imágenes generadas por IA (Brundage y otros, 2018)

Al mismo tiempo, debe reconocerse que la IA puede contrarrestar sus propios usos perjudiciales. Los investigadores de la Universidad de Harvard y el Laboratorio Watson del MIT-IBM desarrollaron el Giant Language Model Test Room (GLTR), una herramienta que puede usarse para identificar texto falso generado por IA. Además, los investigadores diseñaron sistemas automáticos que pueden analizar videos para detectar indicadores que revelen patrones falsos, al evaluar luces, sombras y destellos (Gehrmann, Strobelt y Rush, 2019; Harwell, 2019).

Si bien el conocimiento relacionado con la IA permanece abierto, los posibles riesgos asociados con el mal uso de la IA requieren una reflexión más profunda sobre los modelos de publicación de las investigaciones. Algunos de estos modelos se usan para la investigación en el área de la biotecnología y la seguridad informática, e incluyen (Brundage y otros, 2018):

- i) **Evaluación de riesgo previo a la publicación:** Algunas áreas de investigación delicadas relacionadas con la seguridad digital o el aprendizaje automático conflictivo podrían estar sujetas a un proceso de evaluación de riesgo previo a la publicación para comprender mejor las implicaciones de seguridad a largo plazo.
- ii) **Modelos de licencias de acceso central:** Considere el ejemplo de un proveedor de servicio central confiable que permite el uso de capacidades

de IA para diferentes aplicaciones, sin divulgar los detalles de los trabajos internos del algoritmo. Este modelo para compartir enfocado en la seguridad permitirá aprovechar los beneficios de la IA sin exponernos al riesgo de que la investigación se use con propósitos dañinos.

- iii) **Mecanismo de compartir con seguridad:** Esto podría ser un sistema en el que la investigación se comparte solamente con un grupo predeterminado de organizaciones confiables. Esto permitirá compartir el conocimiento pero solamente dentro de un grupo pequeño.

Todos estos modelos tienen imperfecciones evidentes y generan problemas éticos relacionados con la transparencia y la responsabilidad, mayormente en lo que respecta a la limitación del acceso al conocimiento a un grupo pequeño de personas o instituciones. Por lo tanto, se precisan consultas más amplias para comprender las preocupaciones de los diferentes interesados relacionadas con la IA, y para desarrollar soluciones aceptables para todos y éticamente responsables para equilibrar estas necesidades.

## 6. CONCLUSIÓN Y OPCIONES DE POLÍTICA

La apertura es un atributo importante para la publicación de investigaciones y para asegurar la transparencia y la responsabilidad, así como también la competencia limpia en el desarrollo y la utilización de la IA. Este capítulo destaca varias tendencias clave relacionadas con la apertura y la IA.

Primero, hay una tendencia general a compartir la investigación en IA por parte de los investigadores, tanto en las universidades como en el sector privado. Un incremento saludable en las publicaciones prueba que los investigadores se involucran en discusiones abiertas y crean repositorios abiertos para el conocimiento de la IA.

Segundo, los datos abiertos son un elemento importante para el desarrollo de la IA; además, facilita la transparencia y la responsabilidad en la utilización de la IA. Sin embargo, hay muchos desafíos relacionados con las plataformas del sector privado que recolectan grandes cantidades de datos y no los comparten con otros, alegando motivos de propiedad intelectual, y también hay desafíos para la privacidad y la protección de datos.

Tercero, la apertura en los sistemas de IA con respecto al proceso de toma de decisiones por parte de los algoritmos es un desafío técnico dada la complejidad y la gran naturaleza dimensional de algunas de las tecnologías de IA. Se propusieron algunas soluciones, como la auditoría de los algoritmos y la divulgación de su lógica, para aumentar la transparencia en la toma de decisiones algorítmica, pero esto no resuelve totalmente el problema de la caja negra de la IA.

Cuarto, la apertura incentiva la innovación en los mercados y beneficia a los consumidores al reducir los costos mediante la competencia. Sin embargo, la carrera para captar la mayor parte del mercado puede llevar a descuidar los derechos humanos por la forma en que las empresas utilizan la IA y las concentraciones anticompetitivas del poder del mercado.

Por último, la apertura puede brindar la oportunidad para hacer un uso malicioso de la IA, a pesar de que este riesgo puede mitigarse sin sacrificar los amplios beneficios que conlleva.

A continuación se presentan algunas opciones para fortalecer la apertura y la transparencia de la IA:

## Opciones para todos los interesados

- ▶ Desarrollar normas y políticas para la mejora de la apertura, la transparencia y la responsabilidad en las decisiones automatizadas tomadas por los sistemas de IA mediante métodos tales como la divulgación de la información con anterioridad y el monitoreo con posterioridad a la toma de decisiones automatizada.
  - ↘ Facilitar la competencia de mercados abiertos para impedir la monopolización de la IA y seguir los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas para las mejores prácticas de negocios sobre la base de los derechos humanos.
  - ↘ Promover la investigación de acceso abierto, que puede ser a través de financiamiento e infraestructura de apoyo para los repositorios digitales y la puesta en común del conocimiento.

## Opciones para los Estados

- ▶ Crear repositorios abiertos para investigaciones y datos públicos o privados, lo que incluye la creación de plataformas para datos abiertos del gobierno.
- ▶ Establecer directrices y políticas para la apertura, la transparencia y la responsabilidad en la utilización y el despliegue de sistemas de toma de decisiones automatizada, lo que incluye el uso por parte del gobierno.
- ▶ Apoyar a las universidades e institutos de capacitación técnica para educar y capacitar más estudiantes en IA y áreas asociadas, y, de esa manera, fortalecer la disponibilidad de talento en IA.

## Opciones para el sector privado, intermediarios de Internet y la comunidad técnica

- ▶ Desarrollar normas para la apertura en cumplimiento con las normas y principios internacionales para prácticas éticas basadas en los derechos humanos en el desarrollo y la utilización de la IA.
- ▶ Asegurar la disponibilidad de las salvaguardas adecuadas en relación con los datos abiertos a fin de proteger contra la violación del derecho a la privacidad.
- ▶ Trabajar junto con otros interesados para abordar los retos que implica incrementar la apertura, la transparencia y la responsabilidad de los sistemas de IA.

## Opciones para el sector académico

- ▶ Apoyar el desarrollo de normas de datos abiertos mientras se salvaguarda la privacidad de las personas.
- ▶ Desarrollar normas para la interoperabilidad entre los conjuntos de datos mientras se fortalecen los datos comunes y la disponibilidad de datos para el aprendizaje automático.
- ▶ Fortalecer los esfuerzos de investigación para mejorar la transparencia y la responsabilidad en la toma de decisiones automatizada por parte de los sistemas de IA, lo que incluye esfuerzos para abordar el reto de la caja negra de la IA.

## Opciones para la sociedad civil

- ▶ Actuar como guardián en el uso de toma de decisiones automatizada por parte de las autoridades públicas y el sector privado, y demandar mayor transparencia y responsabilidad en el financiamiento, desarrollo y utilización de los sistemas de IA.

## Opciones para las organizaciones intergubernamentales, incluida la UNESCO

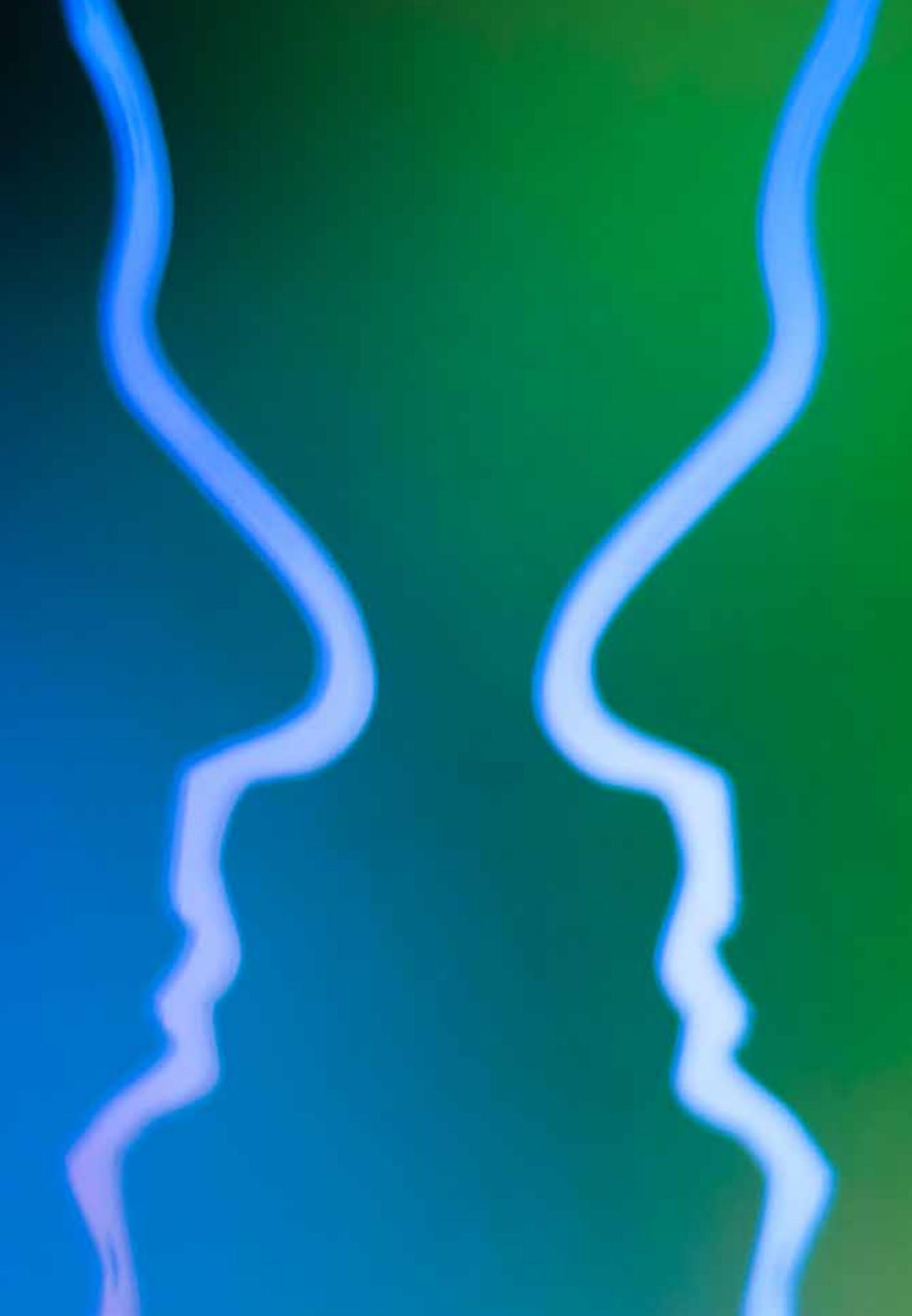
- ▶ Continuar promoviendo el crecimiento de ecosistemas de tecnología abierta ayudando a establecer normas de datos abiertos y repositorios de datos abiertos para IA a través de redes de colegas e institutos y centros con el auspicio de la UNESCO (Institutos y Centros Categoría 2).
- ▶ Aprovechar la experiencia en el desarrollo del movimiento de acceso libre para apoyar el desarrollo de modelos de publicación éticos que

salvaguarden contra la violación de los derechos humanos debido al mal uso del conocimiento sobre la IA disponible de forma abierta.

- ▶ Estudiar diferentes enfoques frente a la responsabilidad algorítmica y reunir a interesados de distintas áreas para construir consenso en torno a las mejores prácticas mundiales.
- ▶ Movilizar a los consorcios que trabajan en la apertura de la IA para fortalecer el movimiento.
- ▶ Desarrollar directrices para la apertura, la transparencia y la responsabilidad en el uso de sistemas de toma de decisiones automatizadas.

# ACCESO E INTELIGENCIA ARTIFICIAL

# 3



# CAPÍTULO 3: ACCESO E INTELIGENCIA ARTIFICIAL

La posibilidad y la capacidad, para que todos puedan acceder y aportar información, ideas y conocimiento son fundamentales para las sociedades del conocimiento inclusivas. El acceso a la información y el conocimiento pueden promocionarse aumentando la conciencia, entre todos los interesados, sobre las posibilidades que ofrece la IA. Estas posibilidades incluyen acceso a la educación, acceso a *software* asequible o gratuito y de código abierto y a datos, y acceso a *hardware* y a una conectividad asequible.

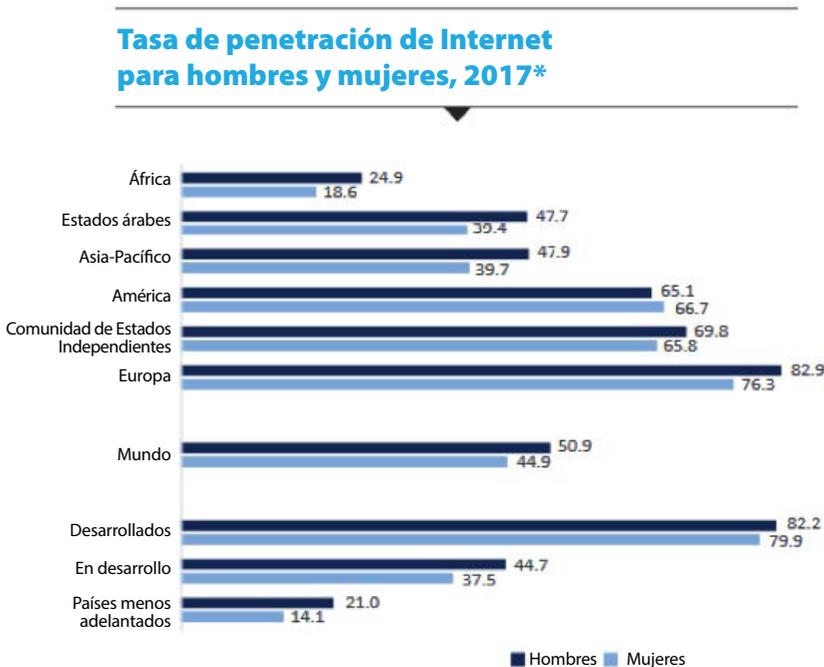
## Recuadro 17: Posición de la UNESCO respecto del acceso a la universalidad de Internet

La accesibilidad para todos como norma para la universalidad de Internet plantea problemas de acceso técnico y de disponibilidad, así como las brechas digitales como las que se basan en el ingreso y en la desigualdad urbano-rural. Por consiguiente, estos son los puntos principales de la importancia de las normas que rodean el acceso universal a los niveles mínimos de infraestructura de conectividad. Al mismo tiempo, la accesibilidad requiere involucrarse con las exclusiones sociales de Internet en virtud de factores tales como la alfabetización, el idioma, la clase, la etnia, la cultura, el género, la discapacidad y la condición de refugiado. Además, dado que las personas acceden a Internet como productores de contenidos, códigos y aplicaciones, y no solo como consumidores de información y servicios, el asunto de las competencias generales del usuario es parte de la dimensión de accesibilidad de la "universalidad". Esto resalta el concepto de "alfabetización mediática e informacional" de la UNESCO, que mejora la accesibilidad empoderando a los usuarios de Internet para que se involucren y creen contenido mediático de forma crítica, competente y ética.

Como tecnología de uso general la IA tiene puede hacer más eficientes los productos y servicios que ya existen. Se espera que tenga un impacto aún mayor al brindar herramientas de análisis en áreas de trabajo e investigación que no se han explorado hasta el momento debido a limitaciones humanas, de modo que sirve como un nuevo "método de invención" (Cockburn, Henderson y Stern, 2018).

En 2017 el 70 % de los jóvenes del mundo (de entre 15 y 24 años) se conectaba a Internet. Sin embargo, las disparidades regionales eran importantes (Comisión sobre la Banda Ancha, 2018). Por ejemplo, en diversos grados, la proporción de personas jóvenes que usa Internet se ha estimado en un 94 % en países desarrollados, en un 67 % en países en desarrollo y solo en un 30 % en países menos adelantados. La brecha digital regional queda al descubierto cuando se constata que 9 de cada 10 personas jóvenes que no usan Internet viven en África o Asia y el Pacífico.

La brecha de género en el uso de Internet aumentó del 11 % en 2013 al 11,6 % en 2017, es decir, la proporción de mujeres que utilizan Internet es casi 12 puntos porcentuales menor que la proporción de hombres que utilizan Internet. A pesar de los esfuerzos mundiales por la paridad de género en el acceso a las TIC, este incremento significa que solo una de cada siete mujeres usa Internet, comparado con uno de cada cinco hombres en los países menos adelantados (Comisión sobre la Banda Ancha, 2018).

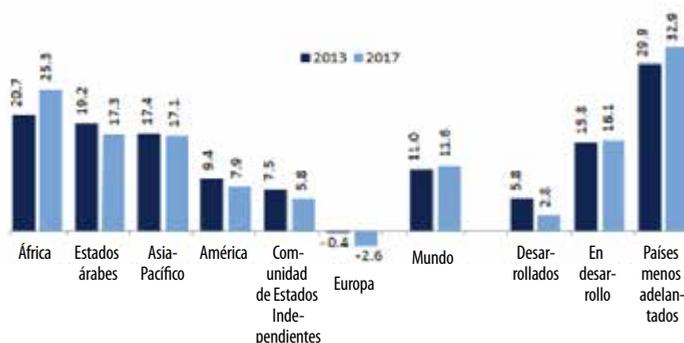


Fuente: ITU.

Nota: Estimación. Las tasas de penetración en esta gráfica refieren a la cantidad de mujeres/hombres que utilizan Internet, en porcentaje de la población total respectiva de mujeres/hombres. CIS es la Comunidad de Estados Independientes (CEI).

Figura 4: Tasa de penetración de Internet para hombres y mujeres, 2017 (ITU, 2017)

## Brecha de género en los usuarios de Internet (%), 2013 y 2017\*



Fuente: ITU.

Nota: Estimación. La brecha de género representa la diferencia entre la tasa de penetración del usuario de Internet para hombres y mujeres en relación con la tasa de penetración del usuario de Internet para los hombres, expresados como porcentajes. CIS es la Comunidad de Estados Independientes (CEI).

Figura 5: Brecha de género en los usuarios de Internet (%), 2013 y 2017 (ITU, 2017)

Estas inquietudes generadas en el contexto del uso de Internet también son relevantes para la IA, que se está usando cada vez más en las mediaciones tecnológicas relacionadas con la interacción de las personas con el mundo. La accesibilidad, la asequibilidad y la capacidad de desarrollar y utilizar estas tecnologías determinan ampliamente nuestras interacciones sociales y con el mercado.

## Recuadro 18: Mandato de la Cumbre Mundial sobre la Sociedad de la Información (CMSI)

La Declaración Final de la CMSI reconoce que "es posible promover el intercambio y el fortalecimiento de los conocimientos mundiales en favor del desarrollo si se eliminan los obstáculos que impiden un acceso equitativo a la información para actividades económicas, sociales, políticas, sanitarias, culturales, educativas y científicas, y si se facilita el acceso a la información que está en el dominio público, lo que incluye el diseño universal y la utilización de tecnologías auxiliares" (Asamblea General de las Naciones Unidas, 2015). El mandato de la CMSI debe abordarse mediante, entre otras acciones, "ambientes de política fuertes y propicios y cooperación internacional para mejorar la asequibilidad, el acceso, la educación, la creación de capacidad, el multilingüismo, la preservación cultural, la inversión y el financiamiento apropiado" (Asamblea General de las Naciones Unidas, 2015).

En el caso de las TIC, la brecha digital es una consecuencia importante de las diferencias en el acceso. Se ha definido como la "brecha entre personas, hogares, negocios y áreas geográficas de diferentes niveles socioeconómicos en relación con sus oportunidades para el acceso a las tecnologías de la información y las comunicaciones, y con su uso de Internet para una amplia gama de actividades" (OCDE, 2001). En relación con la IA esta brecha puede entenderse en términos de acceso según cuatro elementos fundamentales que permiten el desarrollo y el uso de la IA (Elsevier, 2018):

- i) acceso a la investigación;
- ii) acceso al conocimiento, a la educación y a los recursos humanos;
- iii) acceso a datos para el entrenamiento de algoritmos, y
- iv) acceso a la conectividad y al *hardware*.

Las siguientes secciones describen las tendencias para cada uno de estos elementos.

# 1. ACCESO A LA INVESTIGACIÓN

Si bien la apertura global es una característica de muchas investigaciones de IA (véase el capítulo anterior), existen desequilibrios sorprendentes respecto de los lugares en que se llevan a cabo estas investigaciones. A su vez, esto puede afectar los temas que se priorizan como relevantes, qué datos se usan y hasta qué punto otros pueden considerar que el conocimiento tiene valor.

La brecha digital relacionada con la calidad y la cantidad de investigación en IA crece entre los países y dentro de ellos. El reto es saber si la IA puede utilizarse para ayudar a reducir el desequilibrio en las investigaciones. Las economías más avanzadas del mundo tienen un ecosistema de investigación robusto y vibrante que conduce a la innovación y al crecimiento. El economista Jeffrey Sachs dice que "hay un cambio de largo plazo en la proporción del ingreso nacional del trabajo al capital, incluido el capital físico, humano e intelectual" (Sachs, 2018). Los países con una ventaja en investigación y desarrollo en IA estarán mejor preparados para utilizar estas tecnologías y tendrán recursos humanos mejor capacitados para traducir la investigación en aplicaciones. Dos publicaciones de Elsevier —*2018 AI Index Report* y *Artificial intelligence: How knowledge is created, transferred, and used*— ofrecen perspectivas útiles respecto de las tendencias de investigación en IA (Elsevier, 2018; Shoham y otros, 2018). Sin embargo, se enfocan principalmente en países desarrollados y se precisan mayores esfuerzos para recabar datos sobre la investigación relacionada con la IA en países en desarrollo. El siguiente análisis (de Shoham y otros, 2018) da una idea de los problemas de producción, uso y acceso a la investigación desde el punto de vista de los investigadores y los usuarios.

- i) **Aumento de las publicaciones sobre IA:** Los desequilibrios entre regiones y países resultan evidentes cuando se observa que Europa fue el principal editor en Scopus<sup>1</sup> en 2017, con el 28 % de los documentos sobre IA, a pesar de que las publicaciones chinas sobre IA mostraron un aumento del 150 % entre 2007 y 2017<sup>2</sup>.
- ii) **Impacto de la investigación:** las citas de las investigaciones proveen una métrica indirecta para medir el impacto del trabajo. Mientras que Europa es el mayor emisor en el campo de la IA, su impacto en la investigación ha sido estable y a la par con el promedio mundial, al tiempo que en los Estados

---

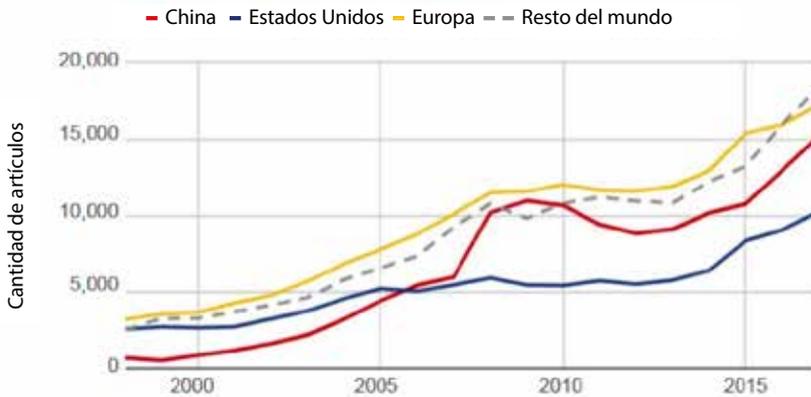
1 Scopus es la base de datos de resúmenes y citas más grande de la literatura revisada por pares: publicaciones científicas, libros y actas de conferencias.

2 La afiliación de un autor a un país se determina según su organización principal, la cual es provista por los autores de los artículos. Las organizaciones mundiales usarán por defecto la asociación al país de sus casas centrales, a menos que el autor especifique lo contrario en la descripción de su organización. Por ejemplo, un autor que ingrese "Google" como su organización será asociado con los Estados Unidos; el que ingrese "Google Zurich" será asociado con Europa. Los artículos se cuentan doble cuando colaboran autores de distintas zonas geográficas. Por ejemplo, un artículo con autores de Harvard y Oxford se contará una vez para los Estados Unidos y una vez para Europa. En Shoham y otros, 2018, hay más detalles acerca de la metodología de investigación.

Unidos, los autores de IA se citan un 83 % más que el promedio mundial. La calidad de la investigación sobre IA en China parece que ha mejorado al considerar que los autores chinos de IA se citaron un 44 % más en 2016 que en 2000<sup>3</sup>. Esto muestra la diferencia en la calidad de la investigación que se produce en diferentes partes del mundo y tendrá repercusiones por las brechas de IA entre países.

### Artículos de IA publicados anualmente en Scopus, por región (1998-2017)

Fuente: Elsevier



Fuente: Elsevier

Nota: Pensamos que el aumento de los artículos sobre IA en China cerca de 2008 es consecuencia del Programa nacional de mediano y largo plazo para el desarrollo de la ciencia y la tecnología (2006-2020) y otros programas del gobierno que brindan financiamiento, así como de diversas políticas de incentivo para la investigación en IA. De manera similar FP7 (2007-2013) y otros programas de investigación en ciencia y tecnología en Europa pueden haber contribuido al pequeño repunte en los artículos alrededor de 2008-2010.

Figura 6: Artículos de IA publicados anualmente en Scopus por región (Shoham y otros, 2018)

El ritmo acelerado de investigación y desarrollo en IA hizo que las conferencias académicas sean una vía importante para la difusión de los descubrimientos de la investigación y para compartir ideas. Más adelante podrán publicarse en publicaciones académicas las versiones expandidas o actualizadas de estos artículos, lo que por lo general lleva más tiempo en comparación con la publicación de las actas de las

3 Hay muchas formas de medir el impacto de la investigación y la calificación por citas es solo uno de los muchos métodos para hacerlo. La UNESCO facilita el diálogo sobre las medidas de la investigación adecuadas con diversos interesados como parte de su Programa de Libre Acceso. Véanse más detalles en las Directrices de Política para el Desarrollo y la Promoción del Acceso Abierto (UNESCO, 2012).

conferencias. Por lo tanto, la aceptación de artículos en las conferencias más importantes de IA es otro parámetro para medir el impacto de la investigación<sup>4</sup>.

La cantidad de artículos presentados y aceptados por región en la conferencia de la Asociación para el Avance de la Inteligencia Artificial (AAIA) de 2018 se muestra en la Figura 8. Los Estados Unidos y China representaban más del 70 % de los artículos presentados y tuvieron una tasa de aceptación del 29 % y 21 %, respectivamente. La India es el único país en desarrollo, aparte de China, que tiene una tasa de aceptación (22 %) comparable con la de los países económicamente avanzados.

### Impacto de la cita de los autores de IA ponderado por área, por región (1998-2016)

Fuente: Elsevier

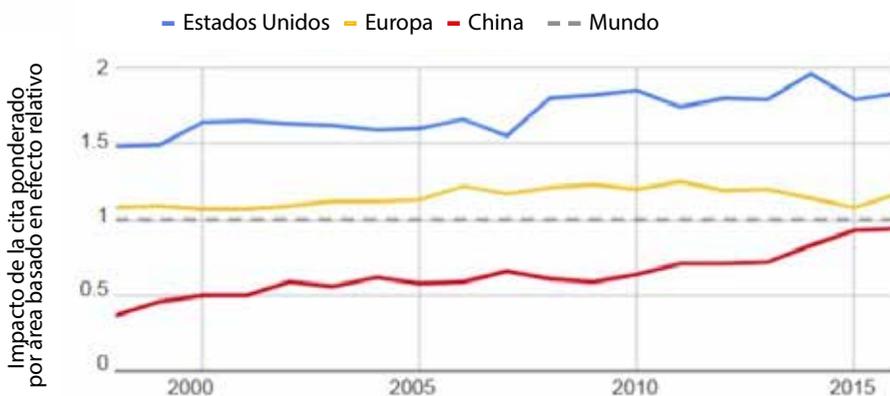


Figura 7: Impacto de la cita de los autores de IA, por región (Shoham y otros, 2018)

Como se mencionó anteriormente, el acceso a la investigación de calidad es un prerrequisito importante para que los países utilicen completamente la IA para su beneficio y su desarrollo. Si bien es cierto que las empresas mundiales de tecnología trabajan en todo el mundo y difunden el conocimiento tecnológico, los vibrantes ecosistemas locales de investigación e innovación fomentarán el desarrollo de soluciones locales con la utilización de la IA y apoyarán el desarrollo sostenible de la ciencia, la tecnología y la innovación en el país. En virtud de estos descubrimientos, queda claro que un pequeño grupo de países lidera, tanto en cantidad como en calidad, la investigación relacionada con la IA. En África y otras regiones (excepto China), los países en desarrollo tienen una presencia limitada en la investigación en IA, a pesar de que en África hay iniciativas prometedoras, como se discute en el

4 La UNESCO no impone una forma de medición del impacto de la investigación sobre otra. La aceptación de los artículos en las conferencias más importantes de IA se utiliza en el entendido de que se trata de un indicador de impacto de la investigación único e imperfecto, y se necesita precaución con su excesiva extrapolación.

capítulo sobre la IA en África<sup>5</sup>. Las políticas de los países y el apoyo internacional para una investigación relacionada con la IA servirán para fortalecer la producción de investigación en los países en desarrollo y ofrecer una base para que la innovación local aumente y responda a los retos locales.

### Cantidad de artículos aceptados y presentados - Conferencia de la AAIA de 2018

Fuente: AAAI

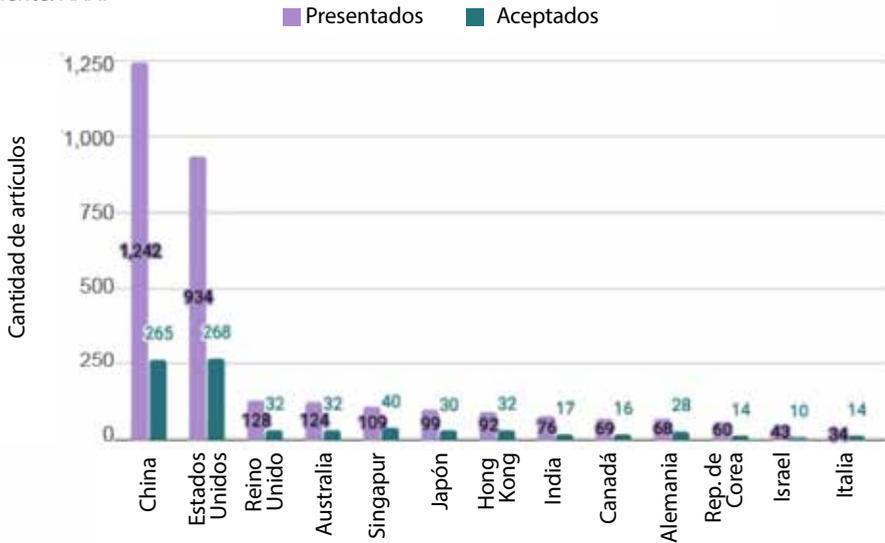


Figura 8: Cantidad de artículos aceptados y presentados en la Conferencia de la AAIA de 2018 (Shoham, y otros, 2018)

## 2. ACCESO A CONOCIMIENTO, EDUCACIÓN Y RECURSOS HUMANOS

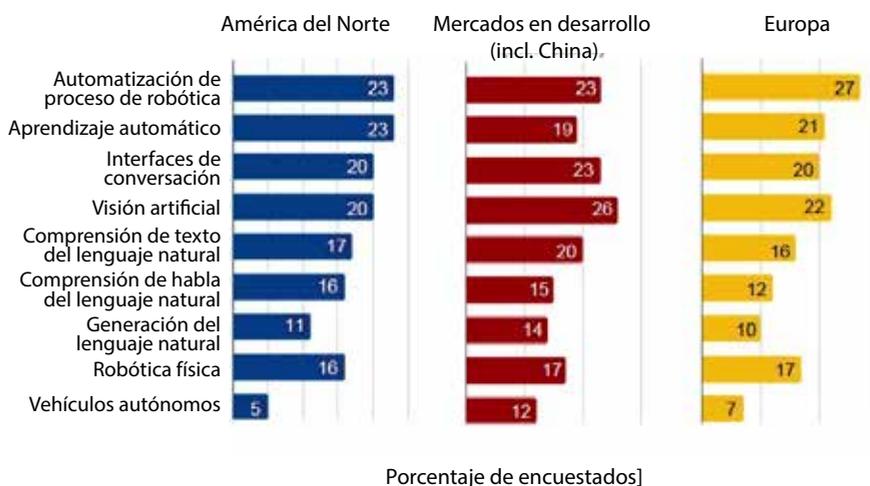
El acceso a recursos humanos educados y capacitados para investigar, desarrollar y aplicar la IA es un prerrequisito importante para el desarrollo de la IA en un determinado país. La escasez de talento es una gran barrera para el desarrollo de la IA (Fu, 2018). Dependiendo de la metodología del estudio que se consulte, las estimaciones acerca de la cantidad de expertos que trabajan en la investigación o industria relacionada con la IA en el mundo va desde 10.000 según el New York Times hasta 300.000 según la empresa tecnológica Tencent con sede en China (Metz, 2017; Vincent, 2017). McKinsey and Company llevó a cabo una encuesta mundial para determinar el grado

5 La conclusión se basa en los descubrimientos de dos recientes informes de IA: Shoham y otros, 2018, y Elsevier, 2018. Esto no es de ninguna manera exhaustivo, y se precisa de más trabajo para examinar de cerca la investigación en IA en los países en desarrollo.

de integración de las capacidades de IA en diferentes empresas<sup>6</sup>. Los datos presentados en la Figura 9 muestran que algunas de las capacidades de IA incorporadas en las funciones de la empresa pueden compararse entre regiones, mientras que otras, como los vehículos autónomos, están más profundamente incorporadas en los mercados en desarrollo que en América del Norte y Europa. La Figura 10 desagrega las diferencias en los países en desarrollo.

### Capacidades integradas en al menos una función de la empresa (2018)

Función: McKinsey & Company



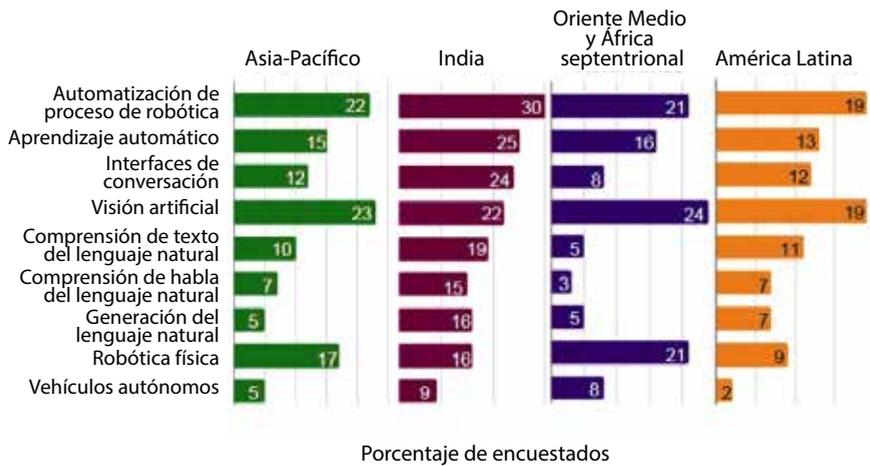
Nota: El tamaño de cada barra se relaciona con las capacidades de cada región; América del Norte: N = 479; Mercados en desarrollo (incl. China): N = 189 (China N = 35); Europa: N = 803

Figura 9: Diferencia en capacidades de IA entre diferentes grupos de países (Shoham y otros, 2018)

6 En su informe, McKinsey definió nueve capacidades de la IA: comprensión de texto en lenguaje natural, comprensión del discurso en lenguaje natural, generación del lenguaje natural, agentes virtuales o interfaces de conversación, visión de la computadora, aprendizaje automático, robótica física, vehículos autónomos y automatización robótica de procesos.

## Capacidades integradas en al menos una función de la empresa (2018)

Fuente: McKinsey & Company



Nota: El tamaño de cada barra se relaciona con las capacidades de cada región; Asia-Pacífico: N = 263; India: N = 197; Oriente Medio y África septentrional: N = 77; América Latina: N = 127

Figura 10: Diferencia en las capacidades de IA entre diferentes regiones (Shoham y otros, 2018)

Los empleos en IA tienen una de las más altas remuneraciones en la industria de la tecnología (Stanford, 2018). La Figura 11 muestra el incremento en las ofertas laborales en áreas de la IA como el aprendizaje profundo. Hay un incentivo monetario de parte del mercado laboral para atraer más personas hacia la investigación en IA. Sin embargo, el hecho de que la mayoría de los países menos adelantados no puedan competir fácilmente en el acceso a talento costoso supone una barrera para la accesibilidad.

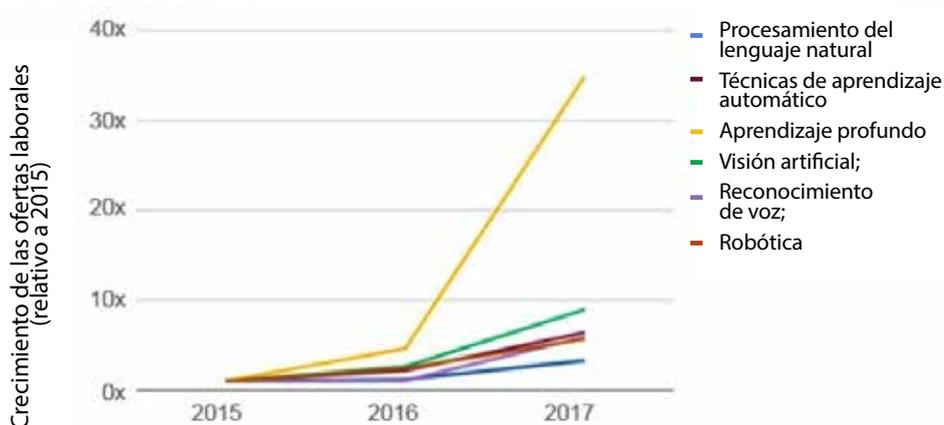
Algunos centros de investigación captan talento de todas partes del mundo mediante el efecto de la fuga de cerebros desde algunos países, pero también desde el sector académico hacia las empresas de tecnología (Kunze, 2019; Hao, 2019; Sample, 2017b). De hecho, mientras crece la demanda de investigadores de IA, algunas universidades enfrentan dificultades para retener a los investigadores de IA en el sector académico para capacitar a la próxima generación de estudiantes (Kwok, 2019; Sample, 2017a).

Al mismo tiempo, la escasez de personas capacitadas lleva a que las empresas i) establezcan centros mundiales de investigación para acceder a los talentos de IA en centros tecnológicos locales, ii) ofrezcan capacitación en habilidades y cursos sobre análisis de datos dentro de la propia empresa, iii) inicien soluciones de convocatoria abierta para mejorar los algoritmos y encontrar fallas, y detectar talento

mediante competencias abiertas como Kaggle<sup>7</sup>, y iv) desarrollen herramientas de IA que puedan ser utilizadas por empresas que no sean de tecnología, sin tener que invertir en recursos humanos de IA (Boyd, 2017). Estos pasos aumentan el acceso para adquirir y mejorar las capacidades de IA, a pesar de que es menos probable que los países en desarrollo se beneficien.

### Crecimiento de las ofertas laborales según los conocimientos de IA necesarios (2015-2017)

Fuente: Monster.com



Nota: Si bien las ofertas laborales de IA aumentan en todo el mundo, las de los ingenieros en computación aún representan un número muy bajo

Figura 11: Crecimiento de las ofertas laborales según las habilidades en IA necesarias (Shoham y otros, 2018)

A fin de satisfacer la demanda de educación en IA, las universidades ofrecen más carreras de grado y cursos en línea relacionados con la IA<sup>8</sup>. Entre los ejemplos se incluyen una Maestría en Ciencias en IA en el Imperial College London y un curso en línea sobre el aprendizaje automático en la Columbia University (Value Colleges, 2019; Marr, 2018). En China, el Ministerio de Educación aprobó un programa para agregar la IA como especialización de grado en 35 universidades (Fang, 2019). En general, el promedio de inscripciones en los cursos de introducción a la IA y de aprendizaje automático se multiplicó por tres o por cuatro entre 2012 y 2017 en varias universidades en todo el mundo (Shoham y otros, 2018).

7 Kaggle es una plataforma en línea que se describe a sí misma como “el lugar para hacer proyectos de ciencia de datos”. Permite que las empresas publiquen sus problemas en forma de competencia con o sin remuneración monetaria u ofertas laborales, presenta conjuntos de datos, foros de discusión y herramientas de aprendizaje en línea sobre temas relacionados con la IA.

8 Para comprender mejor el panorama de la educación de IA es necesario recabar datos de la cantidad de instituciones que ofrecen títulos en IA y de la proporción de cursos de IA que se dictan en línea.

Además, la falta de diversidad de género es un signo de interrogación importante sobre cuán inclusivos son la industria de la tecnología y los departamentos académicos de ciencias de la computación, ya que tienen la posibilidad de perpetuar los sesgos históricos mediante el diseño de los sistemas de IA (Paul, 2019; Simonite, 2018b). Más del 75 % de los profesores de IA de las escuelas más importantes de los Estados Unidos son hombres (Shoham y otros, 2018). Otro informe mostró la brecha de la diversidad de género en la investigación de IA, donde solamente el 13,83 % de los autores en arXiv son mujeres (Stathoulopoulos y Mateos-Garcia, 2019).

Si bien hay cada vez más conciencia sobre este problema, se necesita más investigación para comprender el acceso a la capacitación y la educación en IA, y el nivel actual de disponibilidad de los recursos humanos, en especial en relación con los países en desarrollo. El asunto del contenido educativo disponible en múltiples idiomas, y que esté certificado como de utilidad y de alta calidad, también amerita atención.

La accesibilidad de la IA para todos depende de las competencias del amplio público para comprender su significado y el compromiso con ella. La alfabetización mediática e informacional todavía está lejos de ser universal, y aún más lejos de empoderar a quienes no son especialistas con conocimiento sobre la IA.

### 3. ACCESO A SOFTWARE Y DATOS PARA ENTRENAMIENTO DE ALGORITMOS

La disponibilidad de *software* gratuito y de código abierto se discutió en el capítulo anterior. También se debe prestar atención al tema de la evolución del *software* patentado y hasta qué punto es asequible. Sin embargo, el acceso a los datos parece ser un tema de actualidad. Los datos abiertos fueron tratados en el capítulo anterior, donde se indicó que muchos datos permanecen bajo control de privados.

Si bien los datos a veces se denominan "el nuevo combustible", a diferencia del combustible, potencialmente son un bien no rival, lo que significa que el uso de datos por parte de una persona no impide su uso por parte de otra persona.

En el caso de la IA, el acceso a los datos es fundamental para entrenar algoritmos y para su utilidad en aplicaciones a gran escala. Las empresas de tecnología recaban grandes cantidades de datos de los usuarios en sus plataformas y los usan para entrenar algoritmos para mejorar sus productos y servicios (The Economist, 2017). Ante la imposibilidad de acceder a datos, las nuevas empresas enfrentan posibles barreras para ingresar a desafiar a los actores arraigados en el mercado. Incluso las instituciones académicas han manifestado la imposibilidad de acceder a datos como una de las barreras que impiden la investigación. Una opción para acceder a los datos es la colaboración con empresas de tecnología que tienen abundancia de datos, aunque esto depende, en gran medida, de las percepciones de acumulación

de ventajas para las empresas (Sample, 2017a). Este problema se ve exacerbado en los países en desarrollo.

El acceso a los datos y las dimensiones legales y económicas asociadas son objeto de importantes discusiones en todo el mundo (Comisión Europea, 2017). Sin embargo, a medida que la naturaleza de la industria evoluciona, existe cierto reconocimiento de la naturaleza no rival de ciertos datos y los beneficios de desarrollar datos comunes en forma de repositorios de datos abiertos (Tonetti, 2018). Los siguientes son ejemplos de esos conjuntos de datos y repositorios para el aprendizaje automático<sup>9</sup>:

- Repositorio de Aprendizaje Automático de la Universidad de California Irvine – colección de bases de datos, teorías de dominios y generadores de datos utilizados por la comunidad de aprendizaje automático para el análisis empírico de los algoritmos del aprendizaje automático;
- WordNet en Princeton University – una base de datos léxica del idioma inglés que agrupa sustantivos, verbos, adjetivos y adverbios en conjuntos que expresan un concepto en particular; al vincular un grupo de palabras a distinto significado, es útil en el procesamiento del lenguaje natural;
- ImageNet – una base de datos de imágenes vinculada a WordNet, que provee cerca de 1.000 imágenes para cada grupo de palabras-significado de WordNet y tiene un total de 14 millones de imágenes (el proyecto apoya el desarrollo de la investigación de la imagen y la visión);
- Kaggle – provee conjuntos de datos abiertos en diversas áreas que van desde la biología, la salud y la educación hasta los deportes y los mercados bursátiles;
- Open Data Monitor – provee una lista de bases de datos abiertas disponibles en la Unión Europea entre otros muchos repositorios de datos que hacen que se pueda disponer de grandes conjuntos de datos para el aprendizaje automático.

Si bien estos repositorios de datos se desarrollan y se apoyan en todo el mundo, lo que necesita mayor cuestionamiento es el tema de los sesgos geográficos y de otro tipo que pueden reducir el valor para otras regiones del mundo. Es mucho lo que se precisa para fortalecer la interoperabilidad de los datos mediante el desarrollo de normas para el almacenamiento de datos, la clasificación y el intercambio (Comisión Europea, 2017). En la Cumbre de la IA para el Bien de 2018, organizada por la Unión Internacional de Telecomunicaciones (ITU), se presentó un marco con niveles basado en la funcionalidad e interoperabilidad de los datos comunes, desarrollado por investigadores del Berkman Klein Center. Este marco ofrece claridad conceptual para el desarrollo de datos comunes.

---

9 Esta es una lista de ejemplos de conjuntos de datos y repositorios de datos populares, pero no pretende ser una lista exhaustiva.



Figura 12: Marco de Datos Comunes (Goldstein, Gasser y Budish, 2018)

El modelo está compuesto por datos comunes acotados y amplios, cada uno formado por múltiples subniveles (Goldstein, Gasser y Budish, 2018).

Los datos comunes acotados están compuestos por los siguientes tres niveles:

- i) Un nivel técnico, es decir, la infraestructura utilizada para almacenar datos en la "nube", servidores del gobierno o registros descentralizados;
- ii) un nivel de datos que toma diferentes formas como cualitativo/cuantitativo, estructurado/no estructurado, ordinal/nominal y discreto/continuo, y
- iii) un nivel que asigna formatos y etiquetas para que estos datos sean analizados inteligiblemente.

La interoperabilidad con los diferentes niveles y entre ellos se aborda mediante el desarrollo de normas técnicas comunes, y normas de datos para *hardware* y para la clasificación de datos.

Los datos comunes amplios proveen una interfaz entre las funciones principales descritas en los datos comunes acotados y la sociedad. Esto incluye:

- i) Prácticas organizacionales para utilizar los datos comunes e incentivar la colaboración y la participación de diversos interesados;
- ii) Problemas de política y legales en torno a la privacidad, el acceso, la apertura y los derechos humanos, que surgen a partir de los datos comunes, y
- iii) el compromiso de los humanos en el desarrollo y la preservación de otros niveles mediante mejoras en la inclusión, educación y alfabetización.

La interoperabilidad en los datos comunes amplios se asegura a través del desarrollo del conocimiento compartido y la comprensión de las normas en la sociedad.

La expansión del acceso a los datos es crucial para el desarrollo del aprendizaje automático y la IA. No obstante, aun cuando se desarrollen los datos comunes, se deben abordar otros asuntos relacionados con la representatividad de los datos, la discriminación y la apertura. Estos temas fueron analizados en profundidad en los capítulos de Derechos Humanos y Apertura.

## 4. ACCESO A CONECTIVIDAD Y HARDWARE

El acceso a conexiones de banda ancha asequibles y al *hardware* de computación para procesamiento y almacenamiento de datos es el cuarto componente esencial para el desarrollo de la IA. Los avances en la IA fueron posibles debido a la disponibilidad de mayor poder informático y a su personalización para los algoritmos de procesos computacionales para el aprendizaje automático y el aprendizaje profundo. Por ejemplo, IBM logró un avance en el desempeño de la IA cuando su *software* se probó en un conjunto de datos de una publicidad en línea con más de 4.000 millones de ejemplos de prueba en solo 91,5 segundos, mediante la utilización de *hardware* optimizado. El tiempo de prueba superó en más de 46 veces el mejor tiempo logrado anteriormente con la utilización de TensorFlow en la Plataforma de la Nube de Google que probó el mismo modelo en 70 minutos (Parnell y Dunner, 2018).

El procesador es el centro de la operación de IA, debido a que realiza los cálculos sobre los datos en virtud de las instrucciones en el algoritmo. La compatibilidad entre el tipo de tarea y el tipo de procesador es un determinante importante de la eficiencia de la IA. Por ejemplo, una UCP (Unidad Central de Procesamiento) puede ser apropiada para realizar algunos cálculos complejos de forma muy eficiente, pero no estar equipada para manejar una gran cantidad de cálculos, aunque sean sencillos. El aprendizaje automático pertenece esencialmente a la última categoría. Realiza cálculos sencillos mediante lo que el científico de la computación Andrew Ng denominó el "principio del montañista perezoso", que consiste en instruir al montañista para que continúe descendiendo hasta que no pueda bajar más (Wilson, 2011). El aprendizaje automático encuentra sus respuestas al utilizar un enfoque similar de prueba y error. Sin embargo, para lograr la solución debe realizar una gran cantidad de cálculos muy simples. El procesador que es apropiado para esta tarea es el GPU (Unidad de Procesamiento Gráfico) que se utilizó para programas de juegos a partir de la década de 1970 (Algorithmia, 2018).

Se diseñan circuitos integrados para aplicaciones específicas (ASIC) para realizar tareas específicas. Por ejemplo, Google desarrolló la Unidad de Procesamiento Tensorial (TPU) para el aprendizaje automático en TensorFlow. Estos ASIC pueden ser aún más eficientes que las GPU (Algorithmia, 2018).

Estas unidades de procesamiento requieren dispositivos semiconductores<sup>10</sup>, por lo tanto, un análisis de las importaciones y exportaciones mundiales de semiconductores brinda una representación para comprender la disparidad en el acceso al *hardware* para IA. Las Figuras 13 y 14 muestran la brecha en el comercio de dispositivos semiconductores. Según el Observatorio de la Complejidad Económica, los semiconductores son el 24º producto más comercializado (Observatorio de la Complejidad Económica, 2019)<sup>11</sup>. De un comercio total de 88.000 millones de dólares, Asia tiene el 80 % de las exportaciones, seguida por Europa y América del Norte. En el caso de las importaciones, Asia está en la posición de liderazgo seguida nuevamente por Europa y América del Norte. África tiene solo el 0,90 % de las importaciones mundiales de semiconductores (Observatorio de la Complejidad Económica, 2019). Por lo tanto, existe una amplia disparidad geográfica en el acceso al *hardware* necesario para el desarrollo de la IA. La importancia de la industria de los semiconductores se enfatiza, además, por el significativo apoyo a la política por parte del gobierno en países como China, para desarrollar la independencia en la producción de dispositivos semiconductores (Kharpal, 2019). Es necesario superar las brechas en el acceso a los dispositivos semiconductores para que el desarrollo de la IA no se concentre solamente en algunas partes del mundo.

Desde que los datos pueden almacenarse y procesarse en la nube, no obstante, el acceso al *hardware* debe proveerse a través de servicios de computación basados en la nube. Esto depende de la disponibilidad de conexiones de banda ancha asequibles y confiables, que constituyen, a su vez, una función de las políticas del gobierno y la extensión de la competencia del mercado abierto entre proveedores. La Alianza para una Internet Asequible ha monitoreado estos temas, que son relevantes no solo para el desarrollo de la IA, sino también para el uso social de las implementaciones que dependen de una conexión a Internet.

---

10 Los dispositivos semiconductores también son conocidos como diodos, transistores, fotovoltaicos, fotosensibles, cristales piezoeléctricos montados. Los dispositivos semiconductores son productos HS92 de 4 dígitos.

11 Los mayores exportadores de dispositivos semiconductores son China (28.500 millones de dólares), Japón (8.380 millones de dólares), otros países de Asia (774.000 millones de dólares), Malasia (755.000 millones de dólares) y Alemania (6.100 millones de dólares). Los mayores importadores son China (16.300 millones de dólares), Hong Kong (12.600 millones de dólares), Alemania (6.240 millones de dólares), los Estados Unidos (5.810 millones de dólares) y la República de Corea (444.000 millones de dólares) (Observatorio de la Complejidad Económica, 2019).



Figura 13: Exportadores de dispositivos semiconductores por continente. El comercio de semiconductores se muestra como una representación para el hardware de computación (Observatorio de la Complejidad Económica, 2019)



Figura 14: Importadores de dispositivos semiconductores por continente. El comercio de semiconductores se muestra como una representación para el hardware de computación (Observatorio de la Complejidad Económica, 2019)

En relación con el almacenamiento y el procesamiento real, las empresas de tecnología como Google, Amazon, Microsoft e IBM, entre otras, ofrecen servicios de computación de bajo costo basados en la nube. El usuario puede elegir el tipo de procesador necesario para la tarea a realizar: CPU, GPU o ASIC. La descentralización de los recursos informáticos convirtió los grandes costos fijos de mantenimiento de un *hardware* costoso en un costo variable, donde las empresas o las personas solo pagan por la computación en virtud de su uso (Varian, 2019).

La disponibilidad de recursos de computación a demanda fortalece el acceso a la IA. Algunos gobiernos apoyan a empresas emergentes al permitir que puedan acceder a los recursos informáticos. El Gobierno del Reino Unido, por ejemplo, estableció un "Garaje de Inteligencia Automática" para dar a las empresas emergentes acceso a recursos de computación y salvar la brecha de acceso (Shead, 2017). Los países en desarrollo pueden considerar políticas similares para incentivar a las empresas emergentes de IA en sus jurisdicciones.

## 5. CONCLUSIÓN Y OPCIONES DE POLÍTICA

El acceso a la IA y a otras TIC avanzadas es importante para superar las nuevas brechas digitales que se registran entre los países y dentro de ellos. En este sentido se pueden ver varias tendencias importantes.

Primero, los países con economías avanzadas y centros de investigación bien establecidos son líderes en investigación mundial en IA. Los países en desarrollo necesitan fortalecer sus capacidades de investigación para el desarrollo y el despliegue de la IA. Dado que esto debe lograrse en el contexto de la brecha digital existente, el desarrollo de prácticas de investigación colaborativas podría ser beneficioso para la expansión de los conjuntos de datos del aprendizaje automático y la facilitación de la transferencia tecnológica.

Segundo, hay diferencias en el acceso a la educación y los recursos humanos de IA que pueden exacerbar la brecha digital entre países y entre géneros. Para fortalecer el acceso al conocimiento y a la educación de IA es necesario crear contenido de calidad que esté disponible en múltiples formatos e idiomas, y sea accesible desde diferentes regiones, lo que incluye lugares con limitada infraestructura de telecomunicaciones. El vínculo activo entre institutos educativos y el sector privado ayudará a estimular los ecosistemas locales de IA y a ofrecer oportunidades de empleo. Ese compromiso deberá ser crítico y la investigación no debe ser instrumentalista ni limitarse a los aspectos técnicos de la IA, sino que también debe explorar las cuestiones éticas y regulatorias en la práctica de la IA.

Tercero, el acceso a los datos debe fortalecerse a través de repositorios de datos abiertos para asegurar el acceso gratuito a los datos para el entrenamiento de algoritmos. Sin embargo, hay que prestar más atención a los temas relacionados con la interoperabilidad representativa de los datos y las normas de datos. Por último, es necesario poder acceder a *hardware* de computación para correr los algoritmos de IA. Este aspecto requiere un mayor esfuerzo de parte de los gobiernos y del sector privado para hacer que las unidades de procesamiento adecuadas estén disponibles, a pesar de que la banda ancha asequible ofrece una solución alternativa parcial.

### Opciones para todos los interesados

- ▶ Trabajar para reducir la brecha digital en el acceso a IA, incluidas las brechas de género, y establecer mecanismos para el monitoreo continuo de las diferencias en el acceso.

- ▶ Asegurar que las personas, grupos y países con menos probabilidad de acceso a la IA participen activamente junto a diversos interesados en los diálogos sobre la brecha digital, donde se destaque la importancia de la igualdad de género, la diversidad lingüística y regional, así como la inclusión de los jóvenes y los grupos marginados.

## Opciones para todos los interesados

- ▶ Trabajar para reducir la brecha digital en el acceso a IA, incluidas las brechas de género, y establecer mecanismos para el monitoreo continuo de las diferencias en el acceso.
- ▶ Asegurar que las personas, grupos y países con menos probabilidad de acceso a la IA participen activamente junto a diversos interesados en los diálogos sobre la brecha digital, donde se destaque la importancia de la igualdad de género, la diversidad lingüística y regional, así como la inclusión de los jóvenes y los grupos marginados.

## Opciones para los Estados

- ▶ Fortalecer la infraestructura y el apoyo necesarios para la investigación y el desarrollo relacionados con la IA en universidades y centros de investigación.
- ▶ Incentivar y apoyar la adquisición de capacidades de codificación y la alfabetización en ciencias de la computación para ciudadanos, mediante políticas proactivas para educación, capacitación técnica y vocacional, lo que incluye el aprendizaje de por vida.
- ▶ Asegurar las políticas que proveen de acceso asequible a banda ancha y evitan interferencias con la conectividad como cortes de Internet, aceleraciones o filtraciones, o bloqueos arbitrarios.
- ▶ Fortalecer el acceso al *hardware* de computación específico de la IA, lo que incluye el apoyo a través de fondos y la provisión de acceso a recursos de computación centralizados, según sea necesario.
- ▶ Recabar datos para comprender el acceso a la educación en IA y el nivel actual de disponibilidad de recursos humanos, en especial en los países en desarrollo.
- ▶ Asegurar que los cursos educativos, tales como alfabetización mediática e informacional, lo que incluye una comprensión de la IA, estén disponibles en varios idiomas.

## Opciones para el sector privado

- ▶ Fortalecer el acceso de los ciudadanos a conectividad, *software* y *hardware* asequibles, necesarios para correr los programas de IA.
- ▶ Colaborar con las universidades y centros de investigación, incluso a través de la capacitación de estudiantes, becas de investigación doctoral, intercambio de datos y recursos de computación para investigación y desarrollo.
- ▶ Fortalecer la diversidad de género en la investigación de la IA tanto en el sector académico como en el sector privado.

## Opciones para el sector académico

- ▶ Mejorar el acceso a la tecnología de IA y datos para aprendizaje y clasificación a través de la creación de repositorios de investigación y publicaciones de libre acceso.
- ▶ Fortalecer el acceso al conocimiento de la IA mediante la oferta de recursos educativos abiertos de gran calidad en múltiples idiomas y formatos accesibles para personas con discapacidad.
- ▶ Actualizar la currícula universitaria de forma dinámica con desarrollos y metodologías de la investigación de última generación, incluso a través de la evaluación de la educación regular y de las necesidades de conocimientos en asociación con el sector privado y otros interesados.
- ▶ Crear y fortalecer mecanismos de colaboración en la investigación, movilidad de los investigadores y oportunidades de mentoría para estudiantes entre universidades de todo el mundo, con foco especial en intercambios Norte-Sur y Sur-Sur, y en la paridad de género.

## Opciones para la sociedad civil

- ▶ Apoyar el desarrollo de conocimiento y recursos localizados de IA en formatos e idiomas que brinden información sobre IA accesible para todos, en especial para las mujeres, personas con discapacidad y otras personas que históricamente tienen acceso limitado a las TIC y a las capacidades tecnológicas.
- ▶ Participar en proyectos de colaboración masiva para la creación de datos comunes.

## Opciones para organizaciones intergubernamentales, incluida la UNESCO

- ▶ Apoyar a los Estados Miembros en la mejora de la capacidad de investigación en IA mediante capacitación, desarrollo de políticas de educación, intercambios académicos y el Programa Información para Todos de la UNESCO.
- ▶ Integrar la discusión de asuntos de IA en eventos de importancia, tales como los días internacionales de la libertad de prensa, la discapacidad y el acceso universal a la información, y acercarse a las redes de la UNESCO, como el Programa UNITWIN y de Cátedras UNESCO, las Cátedras UNESCO de Comunicación (Orbicom), la Alianza Mundial para las Asociaciones sobre Alfabetización Mediática e Informativa (GAPMIL) y la Alianza Mundial sobre Género y Medios de Comunicación (GAMAG), así como también los Institutos y Centros de Categoría 2, las ONG, los Comités Nacionales del PIPT y las Comisiones Nacionales de la UNESCO.
- ▶ Defender debates para el desarrollo de los datos comunes con datos representativos mediante la facilitación de la participación mundial y la construcción de normas técnicas y de datos comunes, a través de un proceso de construcción de conocimiento compartido y comprensión normativa acerca de la interacción de los datos con la ética de la información y la sociedad.
- ▶ Facilitar la expansión del acceso multilingüe a la educación en IA y apoyar el desarrollo de recursos educativos abiertos de buena calidad.
- ▶ Aprovechar las redes académicas de la UNESCO y de los Centros Categoría 2 para facilitar el intercambio de investigación en IA, lo que incluye el desarrollo de repositorios para compartir investigaciones.

ENFOQUE  
DE MÚLTIPLES  
ACTORES PARA LA  
GOBERNANZA DE  
LA INTELIGENCIA  
ARTIFICIAL

4



# CAPÍTULO 4: ENFOQUE DE MÚLTIPLES ACTORES PARA LA GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL

Dado el posible impacto de la IA en todos los sectores de la sociedad, cada uno de los grupos de interés, desde los gobiernos hasta el sector privado, la comunidad técnica, el sector académico, la sociedad civil y los usuarios individuales, tienen distintas funciones y responsabilidades. La gobernanza de la IA debería contar con mecanismos realmente cooperativos para desarrollar principios, normas, reglas, procedimientos de toma de decisiones y programas efectivos y relevantes para la IA. Esto es pertinente a nivel nacional, pero, dadas las dimensiones transfronterizas de la IA y sus componentes, también se debe prestar especial atención a las dimensiones regional e internacional. La IA es demasiado compleja e importante como para que decida un solo grupo aislado. Su regulación formal e informal y su evolución ameritan procesos consultivos amplios e inclusivos.

En este contexto, el objetivo de la participación de múltiples actores es mejorar la inclusión y la calidad de la toma de decisiones al incluir en los procesos abiertos y transparentes de toma de decisiones a todos los grupos que tengan un interés en la IA y su impacto en el desarrollo social, económico y cultural más amplio. Este tipo de participación y cooperación puede mantener a todos los interesados bien actualizados respecto del rápido desarrollo de la tecnología y servir para crear un diálogo que pueda generar consenso entre todos los interesados, al menos en lo

que respecta a principios y ética, pero también extendiéndose a formas de gobernanza más específicas cuando sea necesario.

Este capítulo se divide en tres secciones. Primero, se hace referencia al desarrollo de un entorno sólido de múltiples actores para la gobernanza de Internet. Las prácticas a partir de la gobernanza de Internet pueden informar algunas de las discusiones en torno al compromiso de las múltiples actores respecto de las TIC avanzadas como la IA. Segundo, el capítulo pone énfasis en la marcada necesidad de un enfoque de múltiples actores en virtud de la preocupación en torno a la complejidad de la toma de decisiones y el equilibrio del interés en torno a la tecnología y su impacto en la sociedad. En tercer lugar se comparten algunas prácticas y valores que guían el proceso de compromiso de múltiples actores. Por último, el capítulo concluye con opciones de acción a futuro para los diferentes interesados.

### Recuadro 19: Posición de la UNESCO en un enfoque de múltiples actores para la universalidad de Internet

La dimensión participativa —y específicamente de múltiples actores— de la universalidad de Internet facilita la comprensión de las funciones que cumplieron, y deben seguir cumpliendo, los diferentes agentes (representantes de distintos sectores, así como también de diferentes condiciones sociales y económicas, sin excluir a las mujeres y las niñas) en el desarrollo y la gobernanza de Internet en distintos niveles. La participación es fundamental para darse cuenta del valor que puede tener la tecnología para la paz, el desarrollo sostenible y la erradicación de la pobreza. Los mecanismos participativos contribuyen con normas compartidas que mitigan los abusos de Internet para solucionar las disputas por intereses.

# 1. CONTEXTUALIZACIÓN DE LA IA EN LA GOBERNANZA DE INTERNET POR MÚLTIPLES ACTORES

Si bien la IA existe como un área independiente de Internet, esta publicación también reconoce las conexiones subyacentes, como la conectividad, la recolección de datos y el procesamiento. En pocas palabras, la aplicación y el desarrollo de la IA están ampliamente integrados en —y entrelazados con— el ecosistema de Internet y su evolución social, política y económica. Cuando se considera la gobernanza de la IA es sumamente pertinente mirar el marco de la gobernanza mundial de Internet, impulsada por la participación de múltiples actores<sup>1</sup>.

Desde el comienzo, el desarrollo exitoso de Internet se caracteriza por esta participación de múltiples actores en diversos grados. Muestra que el grado en que los interesados participan o pueden participar efectivamente se determina por diversos factores, incluido el grado de conciencia, interés, inquietud y conocimiento; su nivel de agencia o responsabilidad respecto de los resultados y la naturaleza de los procesos consultivos y de toma de decisiones involucrados. Cuando esa participación se institucionaliza, y hasta se confirma legalmente en algunas circunstancias, los resultados son más sostenibles.

Tal como lo marca la historia de Internet, los enfoques de múltiples actores son importantes para promover el potencial de desarrollo de la red y para mantener su carácter universal.

Todos los interesados, desde los gobiernos hasta el sector privado, pasando por la comunidad técnica, las organizaciones intergubernamentales, la sociedad civil y el sector académico, se ven impactados por el desarrollo de Internet, lo que apoya la lógica de que deberían tener una participación inclusiva en los procesos de diseño de política. Para abordar la complejidad de la gobernanza de Internet se adoptó un enfoque de múltiples actores a nivel nacional e internacional. Los siguientes son algunos de los acontecimientos clave que incidieron en ese camino:

- **Cumbre Mundial sobre la Sociedad de la Información (CMSI), 2003-2005**

La Agenda de Túnez para la Sociedad de la Información alienta "a emprender procesos multipartitos a nivel nacional, regional e internacional para debatir y colaborar sobre la difusión y expansión de Internet como medio para propiciar el desarrollo y alcanzar los objetivos y las metas de desarrollo acordados internacionalmente, incluidos los Objetivos de Desarrollo del Milenio" (WSIS-05/TUNIS/DOC/6(Rev.1)-E art. 80 & 87, 2005).

---

1 Esta discusión se basa, en gran parte, en la investigación oficial sobre la participación de múltiples actores de la UNESCO (Van der Spuy, 2017).

La Agenda de Túnez acordó una "definición de trabajo" de la gobernanza de Internet como "desarrollo y aplicación por los gobiernos, el sector privado y la sociedad civil, en el desempeño de sus respectivos papeles, de principios, normas, reglas, procedimientos de toma de decisiones y programas comunes que dan forma a la evolución y a la utilización de Internet".

- **Cumbre Mundial sobre la Sociedad de la Información (CMSI) Revisión + 10**

La Asamblea General de las Naciones Unidas, en su revisión a diez años de los resultados de la CMSI en 2015, reafirmó el valor y los principios de la cooperación y el compromiso de las múltiples actores. Reconoció que la participación, asociación y cooperación efectiva de los gobiernos, el sector privado, la sociedad civil, las organizaciones internacionales, las comunidades técnicas y académicas, y todos los demás interesados relevantes, con sus respectivos roles y responsabilidades, en especial con una representación equilibrada de los países en desarrollo, ha sido vital —y aún lo es— para el desarrollo de la Sociedad de la Información (Asamblea General de las Naciones Unidas A/RES/70/1 par.16 & 17, 2015).

- **Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas**

La Agenda 2030 para el Desarrollo Sostenible de las Naciones Unidas también llama a que se establezcan "alianzas entre múltiples actores que movilicen e intercambien conocimientos, especialización, tecnología y recursos financieros, a fin de apoyar el logro de los Objetivos de Desarrollo Sostenible en todos los países, particularmente los países en desarrollo" (UNGAA/RES/70/125 para.3, 2015).

### Recuadro 20: Informe "La era de la interdependencia digital" (Panel de Alto Nivel del Secretario General sobre la Cooperación Digital, 2019)

"Consideramos que los sistemas inteligentes autónomos deben diseñarse de tal manera que sea posible explicar sus decisiones y pedir cuentas de su uso a los seres humanos. En los planes de auditoría y certificación debe vigilarse que los sistemas de inteligencia artificial cumplan con las normas éticas y de ingeniería, las cuales deben desarrollarse por medio de un abordaje multilateral y multisectorial. Las decisiones de vida o muerte no deben delegarse a las máquinas.

Pedimos una mejor cooperación digital con múltiples actores para soportar el diseño y la aplicación de normas y principios como la transparencia y la ausencia de sesgos en los sistemas inteligentes autónomos en distintos entornos sociales".

(Secretario General de las Naciones Unidas, 2018)

Esencialmente, el enfoque de múltiples actores destaca la importancia del diálogo como forma de equilibrar los intereses, sumar conocimiento y construir consenso y legitimidad. Esto es relevante para Internet en general y en particular para las TIC avanzadas, como la IA. La participación de múltiples actores sirve para asegurar el acceso equitativo a diferentes intereses. Esto significa tomar decisiones mediante la interacción de los intereses participantes y, de esa manera, permitir que las tecnologías digitales mantengan un carácter y una utilidad universales. Este tipo de proceso ayuda a resaltar y tomar conciencia del potencial de desarrollo de la IA para los derechos humanos y el logro de los ODS.

Recurrir a mecanismos de múltiples actores promete una gobernanza de la IA mejor y más inclusiva al mejorar la propiedad y la transparencia, y ayudar a quienes toman las decisiones a que tengan en cuenta diferentes puntos de vista y conocimientos técnicos. La calidad de los resultados y la legitimidad asignada dan lugar a una mejor gobernanza de las enormes complejidades e interdependencias de Internet y la IA (así como también otras tecnologías digitales avanzadas).

## 2. TOMA DE DECISIONES COMPLEJAS Y EQUILIBRIO DE INTERESES EN EL DESARROLLO DE LA IA – LA NECESIDAD DE QUE PARTICIPEN MÚLTIPLES ACTORES

Dada la capacidad inherentemente limitada para predecir los efectos sociales de la tecnología, los encargados de la formulación de políticas deben tomar decisiones con información limitada, bien por incertidumbre o por ignorancia. Sin embargo, es importante entender y evaluar los objetivos iniciales incorporados en el desarrollo y la aplicación de la tecnología. Por ejemplo, la revolución verde tuvo éxito al lograr su objetivo técnico de desarrollar una variedad de semillas de gran producción, pero no tuvo éxito en lo que respecta al objetivo social de terminar con el hambre. Collingridge afirma que "nuestra comprensión del mundo físico y biológico en el que vivimos es extremadamente profunda y nos brinda los medios para producir toda clase de maravillas técnicas, pero la apreciación de cómo estas maravillas afectan a la sociedad es peligrosa" (Collingridge, 1980). Esta disparidad entre nuestra comprensión de la tecnología y sus efectos en la sociedad con frecuencia lleva a la demonización de la tecnología. Por lo tanto, no resulta sorprendente que el tema de la IA que reemplaza a los humanos sea recibido con gran temor a la tecnología (Solon, 2017).

Un área en la que esto es evidente es el futuro del trabajo y la vida laboral en la era de la automatización inducida por IA, en donde el determinismo tecnológico

implica que el empleo se verá inevitablemente afectado por las TIC avanzadas. Pero, según lo establecido en el Informe del Primer Período Ordinario de Sesiones de la Junta de los Jefes Ejecutivos para la Coordinación de las Naciones Unidas (mayo de 2019), es una decisión de política la que puede asegurar que el aumento de los buenos trabajos supere a la destrucción de puestos de trabajo y cree un equilibrio entre la automatización y las nuevas tareas, mediante la utilización de tecnología para aumentar el trabajo, en vez de reemplazar trabajadores. El informe propone invertir en nuevas capacitaciones y la mejora de las calificaciones de la población, y hacer del aprendizaje de por vida una parte natural de la vida laboral de las personas. Además, resalta la necesidad de desarrollar una moderna red de protección social, diseñar tecnologías para beneficio de todos y establecer estructuras de gobernanza flexibles que puedan responder a los retos y adaptar las políticas y los reglamentos más rápido (CEB, 2019). En particular el informe resalta el potencial para:

- i) Brindar educación y habilidades a los ciudadanos para que sean "ciudadanos digitales".
- ii) La naturaleza cambiante de la vida laboral, que incluye el seguro de ingreso y las preferencias algorítmicas en el trabajo.
- iii) La nueva capacitación de los trabajadores desplazados mediante nuevas formas de entrenamiento y desarrollo.
- iv) Compartir los beneficios de la IA en la sociedad de forma equitativa para evitar exacerbar la inequidad de ingresos.
- v) Adaptar los programas educativos para las habilidades que son necesarias para un futuro habilitado por la IA.

Sin embargo, aún con estas medidas, no se pueden predecir otros posibles impactos de la evolución social y económica y el uso de la tecnología. En este caso, se deberán tomar las decisiones con incertidumbre e ignorancia. Así, en el caso de la IA, los desarrolladores, los encargados de la formulación de políticas y los reguladores serían negligentes si tomaran decisiones sin monitorear continuamente las consecuencias posibles e inesperadas. La compensación necesaria para la toma de decisiones efectiva con 'ignorancia' es "la facilidad con la que se pueden detectar y eliminar las fallas, y los costos asociados con la falla" (Collingridge, 1980). Por lo tanto, el marco guía para el proceso de toma de decisiones requiere que las decisiones puedan corregirse, es decir, cuando se descubre una falla o un resultado inesperado, este debe ser fácil de corregir. Una regla general para ese tipo de decisiones incluye:

- monitoreo rentable y eficiente para asegurar que las consecuencias no predecibles puedan descubrirse rápidamente,
- selección de opciones con bajo costo por errores,
- corto tiempo de respuesta para la corrección luego del descubrimiento de problemas no previstos, y

- bajo costo de la aplicación de la solución. Por lo tanto, se debe dar preferencia a las opciones con costos fijos bajos y un costo variable más alto sobre las que tengan costos fijos más altos.

En este contexto, parece que el compromiso de múltiples actores es lo más necesario para moldear los valores en la toma de decisiones sobre la IA y para monitorear y realizar cambios, tanto a nivel nacional como supranacional.

Existe un área en la que esto es importante y está relacionada con los riesgos que resultan de la competencia en el desarrollo de productos y servicios relacionados con la IA. Por lo general, en un mercado mundial competitivo, el que mueve primero se queda con la porción más grande del mercado, lo que se traduce en una concentración de poder que puede ser abusiva, incluidas medidas que erosionan los principios de Apertura y Acceso. De manera similar, las empresas que utilizan tecnologías de IA, especialmente cuando están en condiciones de competencia, deben regirse por reglas que aseguren que siempre se respeten los derechos humanos cuando se usen sus productos en el mercado. En todos estos casos es fundamental el monitoreo de las consecuencias accidentales.

Al aplicar estos principios a la gobernanza de la IA y mitigar los riesgos de la toma de decisiones con información limitada acerca de los efectos futuros, es fundamental contar con una participación inclusiva de múltiples actores o un proceso de cooperación. A diferencia de otros procesos tradicionales de formulación de políticas, las prácticas de múltiples actores pueden resaltar principios como la apertura, la transparencia y la colaboración masiva y participación equitativa, que son relevantes para la mejor previsión y monitoreo posibles según las circunstancias. Los procesos de multilaterales también pueden ayudar a informar acerca el desarrollo de los marcos operacionales, lo que puede ser aprovechado por diferentes actores para ir tomando decisiones a medida que surgen los problemas.

Esto requiere de la participación, asociación y cooperación efectiva de los interesados en el desarrollo y uso colectivo de la IA. Cualquier persona o entidad con un interés legítimo o auténtico en determinado asunto provocado por el desarrollo de la IA puede considerarse como un interesado de importancia. A nivel mundial la participación efectiva también significa la inclusión de los países y las regiones en desarrollo en el proceso, así como de los grupos de menores recursos, como la sociedad civil y el sector académico.

Como se mencionó anteriormente, la participación de múltiples actores se propone como un método para impedir que un grupo tenga el dominio de Internet y otras nuevas tecnologías a expensas de otro. Esto es cierto cuando las TIC avanzadas son capturadas por diversos actores estatales (por ej. seguridad en vez de comercio, o impuestos en vez de política de economía digital, etc.), por organizaciones interestatales o por intereses del sector privado (con intereses comerciales y modelos de negocios diversos que pueden violar los derechos a la privacidad o de expresión, etc.). Además, se aplica tanto a nivel nacional como internacional. El resultado es que la toma de decisiones refleja intereses generales y no particulares, y que hay participación colectiva respecto del monitoreo y el cambio de los resultados no previstos.

### 3. GOBERNANZA DE LA IA MEDIANTE UN ENFOQUE DE MÚLTIPLES ACTORES: PRÁCTICAS, VALORES E INDICADORES

Actualmente hay un debate en curso sobre diversos modelos de gobernanza de Internet de múltiples actores, pero en esta sección se comparten algunas prácticas y valores que han dado forma con éxito a la participación de múltiples actores en el caso de Internet, y que pueden servir como punto de partida para reflexionar sobre la IA y sus problemas de gobernanza.

En el estudio de la UNESCO titulado *What if we all governed the Internet?* se evalúan los elementos que sustentan una serie de cooperaciones digitales basadas en un enfoque de múltiples actores. Entre ellas figuran el Foro sobre Mejores Prácticas de Género y Acceso del Foro para la Gobernanza de Internet (FGI), el proceso de elaboración de políticas de la Corporación de Internet para la Asignación de Nombres y Números (ICANN), así como el mecanismo nacional de KICTANet en Kenya, y la redacción del Marco de Derechos Civiles para Internet del Brasil (Marco Civil) (Van der Spuy 2017).

Los éxitos y fracasos de estas experiencias demuestran que los enfoques de múltiples actores deben mostrar ciertos valores si se quiere que sean eficaces para dar forma a las normas, desarrollar el consenso y facilitar la adopción de decisiones relativas a la gobernanza. La tensión aumenta cuando se adoptan decisiones unilaterales o en las asociaciones público-privadas que excluyen la aportación de la sociedad civil a las cuestiones de gobernanza. La falta de participación de los interesados del sector privado, o su participación menos transparente, también debilita la legitimidad y la eficiencia de las iniciativas de múltiples actores. La incapacidad de la sociedad civil para participar plenamente, debido a la escasez de recursos, suele conducir a que en la adopción de decisiones se descuiden las cuestiones de derechos humanos.

Como se señala en el estudio de la UNESCO antes mencionado, los mecanismos de múltiples actores deberían ajustarse a determinados valores si se quiere que contribuyan de manera óptima a la gobernanza:

- **Inclusivo** – Estrechamente vinculada a los principios DAAM de accesibilidad para todos y apertura, la inclusividad encapsula la necesidad de superar los obstáculos a la participación accesible y dedicar suficientes esfuerzos de financiamiento y creación de capacidad para promover la participación de una amplia diversidad de interesados. Por consiguiente, deben adoptarse disposiciones especiales para aquellos interesados que suelen tener un financiamiento y una representación insuficiente, como las comunidades marginadas, las mujeres, los jóvenes, las pequeñas entidades empresariales o los participantes de la sociedad civil de las regiones en desarrollo o del hemisferio sur.

- **Diverso** – Asegurarse de que los procesos de múltiples actores puedan beneficiarse de diferentes puntos de vista para abordar las complejas y diversas preocupaciones de los interesados inherentes a los desafíos que plantean los problemas que presenta la tecnología digital. Por ejemplo, en las decisiones sobre los sistemas de IA utilizados para la adopción automática de decisiones deberían participar organizaciones de derechos humanos a fin de garantizar que se consideren posibles dimensiones discriminatorias.
- **Colaborativo** – Los interesados deberían acordar normas comunes para orientar los métodos de trabajo, incluido el grado de transparencia, la flexibilidad necesaria, las formas de adoptar decisiones y los medios para promover y proteger la seguridad y los derechos de los participantes cuando estos participen en un proceso en el que intervengan varias partes interesadas. Es necesario apoyar conjuntamente la sostenibilidad e institucionalización de los procesos, así como su alcance y objetivos a corto y largo plazo.
- **Transparente** – Los interesados que participan en procesos con múltiples actores deben ser claros respecto de sus intereses y afiliaciones.
- **Igual** – Garantizar que los interesados puedan participar en pie de igualdad en todas las etapas de los procesos de múltiples actores, aunque las normas y responsabilidades difieran en lo que respecta a la adopción de decisiones definitivas, por ejemplo, sobre las normas y programas y su evaluación.
- **Flexible y pertinente** – La participación de múltiples actores debe ser lo suficientemente flexible como para asegurar que un proceso o actividad pueda adaptarse a la naturaleza cambiante de Internet y otras tecnologías digitales, y debe adaptarse para que sea pertinente a las instancias locales, regionales, nacionales y mundiales de colaboración entre múltiples actores.
- **Seguro y privado** – Asegurar que, en la medida de lo posible, se satisfagan las necesidades de seguridad y privacidad de los participantes.
- **Responsable y legítimo** – Los mecanismos de múltiples actores deberían evaluar periódicamente los procesos, los resultados y los objetivos.
- **Receptivo** – La receptividad implica transparencia en cuanto a las razones por las que se adoptaron determinadas decisiones para aceptar o rechazar las presentaciones, y si existen oportunidades de apelación o reparación independientes para quienes no se sienten suficientemente escuchados.

A fin de traducir esos valores en acciones y prácticas, el marco de indicadores DAAM-X de la UNESCO sobre la universalidad de Internet (UNESCO, 2018b), que tiene por objeto medir el desarrollo y las políticas de Internet, ofrece una serie de referencias e indicadores preliminares que pueden utilizarse para evaluar la modalidad de gobernanza de IA en lo que respecta a la participación de múltiples actores. Adaptando el marco, algunas preguntas especialmente adaptadas para evaluar la gobernanza de IA a nivel nacional podrían incluir:

- ¿El gobierno promueve la participación de otros interesados en la gobernanza nacional de la IA?
- ¿Hay asociaciones activas de profesionales de la IA, consumidores y otras comunidades de interesados?
- ¿El gobierno involucra activamente a otros grupos de interesados en el desarrollo de políticas para la gobernanza global de IA?

Además, estas preguntas se pueden poner en práctica en varias dimensiones de los indicadores dedicados a medir la participación de múltiples actores, que se dividen en tres temas adaptados a las cuestiones relacionadas con la IA:

- Indicadores sobre el marco jurídico y reglamentario general para la participación en la gobernanza: para medir si existe una política general o un marco jurídico y reglamentario para el desarrollo y la formulación de políticas de IA, y si es compatible con las normas internacionales.
- Indicadores relativos a la gobernanza nacional de IA: para medir en qué medida diversos grupos de interesados participan en la formulación de políticas a nivel nacional relacionadas con la IA, como la existencia de foros de múltiples actores y la participación inclusiva de diversos grupos de interesados con igualdad de género e inclusión de jóvenes y grupos marginados.
- Indicadores relativos a la gobernanza internacional y regional de IA: para evaluar la medida en que los diversos grupos de interesados del país participan en foros, procesos y mecanismos internacionales y regionales relacionados con la IA.

Para concluir, es sumamente necesario seguir elaborando indicadores y evaluando el enfoque de múltiples actores a nivel nacional, lo que puede proporcionar una orientación basada en pruebas para determinar deficiencias en materia de gobernanza. A su vez, ello puede contribuir a aumentar la sensibilización y facilitar los esfuerzos por formular una participación, asociación y cooperación eficaces de todos los interesados.

## 4. CONCLUSIÓN Y OPCIONES DE POLÍTICA

Hay un número creciente de países y regiones de todo el mundo que han elaborado y publicado estrategias nacionales sobre la IA en los últimos años (Dutton, 2018). La cantidad de marcos éticos propuestos se multiplica mes a mes y todos exigen la cooperación y coordinación de múltiples actores para evitar la fragmentación, la duplicación y los resultados basados en silos.

Desde la perspectiva de abajo arriba, existe el desafío de asegurar una mejor representación de las partes interesadas a nivel nacional mediante mecanismos para mitigar el poder unilateral de los actores individuales (ya sean gobiernos o empresas) y asegurar que se escuchen las voces de todas las partes interesadas y regiones antes de que se adopten reglamentaciones particulares o se impongan "normas comunitarias".

Al mismo tiempo, es importante señalar que la toma de decisiones puramente nacional y unilateral a nivel de principios, normas, reglas y políticas puede ser perjudicial para los pilares de derechos, apertura y acceso que se examinan en esta publicación. Esto apunta a la necesidad de que se realicen debates mundiales sobre la gobernanza de la IA y a la necesidad de motivar una participación más activa de los gobiernos nacionales, las organizaciones intergubernamentales, las comunidades técnicas, el sector privado, el mundo académico, los periodistas y los medios de comunicación, la sociedad civil y los usuarios individuales.

### Opciones para todos los interesados

- ▶ Aplicar los principios DAMM de la universalidad de Internet (derechos humanos, apertura, accesibilidad y participación de múltiples actores) y elaborar indicadores adaptados para ayudar a configurar el diseño, la aplicación, la evaluación y la gobernanza de la IA.
- ▶ Garantizar la transparencia, la inclusión y la rendición de cuentas del proceso de participación, con la participación de los interesados en pie de igualdad.
- ▶ Motivar una participación más activa en la discusión de las políticas de IA a nivel nacional y supranacional de todos los grupos interesados, incluidos, entre otros: i) el gobierno, ii) el sector privado, iii) la comunidad técnica, iv) la sociedad civil, v) el sector académico, vi) las organizaciones internacionales, y vii) los medios de comunicación.
- ▶ Asegurar una mejor representación de todos los grupos, teniendo en cuenta la igualdad de género, la diversidad lingüística y regional, así como la inclusión de los jóvenes y los grupos marginados.

- ▶ Organizar foros y eventos con múltiples actores para las cuestiones y políticas relativas a la IA, e integrar la participación de múltiples actores en la vigilancia y corrección cuando haya resultados inesperados que sean problemáticos.
- ▶ Documentar y archivar el proceso de múltiples actores, que de este modo puede mantenerse de manera transparente, accesible y sostenible.

## Opciones para los Estados

- ▶ Adoptar un marco normativo, jurídico y reglamentario explícito y basado en la participación de múltiples actores para la adopción de decisiones y la vigilancia del desarrollo de la IA, en consonancia con las normas internacionales.
- ▶ Consultar a una gama más amplia de actores interesados respecto de las políticas relacionadas con la IA a través de una diversidad de plataformas en línea y fuera de línea, previstas para la colaboración de múltiples actores.
- ▶ Habilitar y alentar a los diversos grupos de interesados dentro del ámbito nacional para que participen en los foros, procesos y mecanismos internacionales y regionales.

## Opciones para el sector privado y la comunidad técnica

- ▶ Participar más activamente en la formulación de políticas a nivel nacional e internacional relacionadas con la IA y colaborar con otros agentes en foros de múltiples actores.
- ▶ Consultar con los usuarios, la sociedad civil y otros interesados para obtener asesoramiento sobre la elaboración de normas, aplicaciones y productos relacionados con la IA.
- ▶ Considerar la participación de múltiples actores para paliar los problemas en la adopción de decisiones en situaciones donde haya incertidumbre e ignorancia.

## Opciones para la sociedad civil y el sector académico

- ▶ Realizar las investigaciones necesarias para apoyar la institucionalización y la sostenibilidad de las experiencias de gobernanza de múltiples actores.

- ▶ Proporcionar estudios científicos sobre los desafíos de IA en desarrollo e imprevistos para informar los procesos de adopción de decisiones.
- ▶ Concienciar sobre los posibles beneficios de los enfoques de múltiples actores y participar más activamente en debates y procesos relacionados con las políticas.

## Opciones para los actores de los medios

- ▶ Informar y difundir conocimientos e información relacionados con cuestiones relativas a la IA y la participación de múltiples actores.
- ▶ Participar activamente en los debates y procesos de formulación de políticas relacionadas con la IA y contribuir a ellos.

## Opciones para organizaciones intergubernamentales, incluida la UNESCO

- ▶ Mejorar los roles en el fomento de la participación de múltiples actores en la gobernanza de la IA.
- ▶ Reunir a diversos actores dentro del ecosistema para el debate sobre la IA utilizando un acceso único a los titulares de deberes (funcionarios de gobierno, representantes electos, reguladores independientes, propietarios y líderes de los medios de comunicación, ONG especializadas pertinentes, etc.), y la credibilidad entre los titulares de derechos.
- ▶ Convocar foros y eventos de múltiples actores para la discusión de las políticas de IA.
- ▶ Alentar a los Estados a encontrar un marco ético internacional basado en los derechos humanos y soluciones de política, y facilitar la participación inclusiva de los países desarrollados y en desarrollo.

IGUALDAD  
DE GÉNERO E  
INTELIGENCIA  
ARTIFICIAL

5



# CAPÍTULO 5: IGUALDAD DE GÉNERO E INTELIGENCIA ARTIFICIAL

La igualdad entre individuos con diferentes identidades de género y orientaciones sexuales es una cuestión importante en nuestro mundo. El Secretario General de las Naciones Unidas, António Guterres, afirma que la paridad entre los géneros es “un deber moral y una necesidad operacional” (Unidos por la Paridad de Género, 2018). La igualdad es esencial como derecho fundamental, pero también para permitir que todas las personas maximicen su potencial y aumenten su productividad mediante la inclusión de perspectivas diversas y la plena utilización de los recursos de todos los individuos (Unidos por la Paridad de Género, 2018).

Dada la importancia primordial del género, el Objetivo de Desarrollo Sostenible 5 apunta a lograr la igualdad y el empoderamiento de la mujer, y a asegurar que las mujeres y las niñas tengan los mismos derechos y oportunidades. Al mismo tiempo, las Naciones Unidas reconocen que, dado que las mujeres constituyen la mitad de la población mundial, la igualdad entre los géneros forma parte integral de todas las dimensiones del desarrollo sostenible, lo que se refleja en el énfasis que se pone en las mujeres en el Objetivo 17 de los ODS sobre la alianza mundial para el desarrollo sostenible. En reconocimiento de ello, la UNESCO también ha designado la igualdad de género como una prioridad mundial.

## Recuadro 21: Prioridad de la UNESCO respecto de la igualdad de género y las TIC

UNESCO cree que todas las formas de discriminación basadas en el género son violaciones de los derechos humanos, así como un obstáculo importante para el logro de la Agenda 2030 para el Desarrollo Sostenible y sus 17 Objetivos. En muchos países, las mujeres se enfrentan a obstáculos para acceder a Internet o utilizarla, entre ellos:

- Obstáculos como la asequibilidad y el despliegue de la red, la calidad y la disponibilidad;
- Obstáculos como la disponibilidad de contenido relevante;
- Falta de habilidades relevantes, ingresos y situación ocupacional;
- Abuso en línea y violencia y amenazas por motivos de género;
- Desafíos interseccionales, incluido el impacto de los estereotipos y las normas culturales en la capacidad de las mujeres para acceder a Internet y utilizarla.

En el marco de los indicadores DAAM-X de la universalidad de Internet, el término "brecha digital de género" se utiliza para denotar la diferencia entre la participación de hombres y mujeres en la sociedad de la información, en particular el acceso y la utilización de las TIC e Internet. Al realizar el examen a 10 años de los resultados de la Cumbre Mundial sobre la Sociedad de la Información en 2015, la Asamblea General de las Naciones Unidas determinó que abordar la brecha digital de género era una prioridad.

El Plan de Acción de la UNESCO para la prioridad "Igualdad de género" (2014-2021) también tiene un indicador específico que aborda la brecha digital entre los géneros, con el fin de elaborar y poner a prueba estrategias y prácticas óptimas en los Estados Miembros sobre el acceso a la información y el fortalecimiento de la capacidad de las mujeres y las niñas para utilizar las TIC (UNESCO, 2014).

La igualdad de género es una base necesaria para un mundo pacífico, próspero y sostenible. Incluye no solo la igualdad entre hombres y mujeres, sino también la igualdad de trato para las personas lesbianas, gais, bisexuales, transgénero e intersexuales (LGBTI)<sup>1</sup>.

Basándose en la literatura existente, y en particular en la publicación *I'd Blush If I Could*, escrita en conjunto por la UNESCO y la coalición EQUALS, este capítulo examina

1 LGBTI es la abreviatura de "lesbianas, gais, bisexuales, transgénero e intersexuales". Si bien estos términos tienen cada vez más repercusión, las distintas culturas usan distintos términos para describir a las personas que tienen relaciones con personas de su mismo sexo o que tienen identidades de género no binarias (Libres e Iguales, 2013)

la relación entre la IA y el género. Si bien también se consideran las perspectivas interseccionales de diversos grupos de mujeres y personas LGBTI<sup>2</sup>, no se plantea un panorama exhaustivo de todas las dimensiones sobre género e IA.

# 1. PERSPECTIVAS DE ESTUDIOS DE GÉNERO SOBRE TECNOLOGÍA

Las cuestiones de género influyen en la tecnología y son influenciadas por ella, incluida la IA<sup>3</sup>. Como se ha documentado ampliamente en la literatura feminista sobre la relación entre la tecnología y el género, las tecnologías no son neutrales en lo que respecta al género. Se desarrollan y despliegan en un contexto y un entorno en que los roles desiguales de género aún prevalecen y reflejan los prejuicios de la sociedad (Cohen, 2012). Al mismo tiempo, tienen el potencial de desafiar, e incluso desbaratar, los estereotipos de género existentes y empoderar a todas las personas (Haraway, 1991). Existe una vasta literatura sobre la tecnología y su relación con el género, que puede extrapolarse a la relación de la IA con las cuestiones de género.

Uno de los primeros trabajos académicos sobre tecnología y género, al intentar explicar el predominio de los hombres en la industria tecnológica, demostró que la tecnología era una construcción social. Las académicas feministas destacaron que la asociación entre la tecnología y la masculinidad no es inherente a las diferencias sexuales biológicas (Wajcman, 1991; Kelly, 1985). De hecho, los primeros trabajos en el campo de la informática, cuando surgió la industria, fueron en gran parte realizados por mujeres. La programación de programas informáticos se consideraba "trabajo de mujeres" porque requería características estereotipadas de la mujer, como la meticulosidad y la capacidad de seguir instrucciones paso a paso (UNESCO; EQUALS Skills Coalition, 2019). En otras palabras, el predominio de los hombres en el mundo tecnológico está más relacionado con la evolución de las construcciones culturales de los roles de género y la idea de que la femineidad es incompatible con el progreso tecnológico, que con cualquier destreza tecnológica superior inherente a los hombres (Wajcman, 2010). Esta contribución pionera es crucial para comprender que el campo de la tecnología no es objetivo ni está libre de influencias sociales. Aplicando esto a la IA, puede entenderse que el estado actual de la IA no es un

---

2 La interseccionalidad es un marco teórico desarrollado por Kimberlé Crenshaw, que describe cómo los sistemas de subordinación superpuestos (como el racismo y el sexismo) se combinan para influir en las vidas de las personas que tienen identidades subordinadas múltiples de maneras diferentes a la forma en que los sistemas de subordinación individuales influyen en las vidas de los que tienen identidades subordinadas individuales (Crenshaw, 1989). Por ejemplo, el sexismo al que se enfrenta una mujer de una minoría étnica sería diferente del sexismo al que se enfrenta una mujer de un grupo mayoritario. Este marco llama la atención sobre la discriminación que no puede ser "captada en su totalidad mirando las dimensiones de raza o género de esas experiencias por separado" (Crenshaw, 1989).

3 Cabe destacar que si bien esta sección enmarca el consiguiente debate sobre la IA y el género, en Wajcman (2010) se incluyen más detalles sobre varios aspectos de esta literatura.

resultado inevitable de la progresión natural, sino que más bien está estructurado por factores sociales, incluidas las normas de género. Por consiguiente, sus consecuencias, ya sean positivas o negativas, no son inherentes a la tecnología sino que son producto de las decisiones y acciones humanas. Además, esta contribución nos permite dejar de lado las explicaciones y soluciones técnicas e infraestructurales de las desigualdades en el campo de la tecnología (aunque también son importantes) y abordar los factores socioculturales (Wajcman, 2010).

Una segunda línea de estudios en los años ochenta desafió la idea de que el principal problema era el monopolio de la tecnología neutral a manos de los hombres. La tecnología se concebía ahora como algo que integraba normas de género (Corea y otros, 1985; Patricia y Steinberg, 1987; Grint y Gill, 1995). Para algunos esto significaba que, dado que el patriarcado está presente en el mundo, la tecnología refleja intrínsecamente las desigualdades de género, aunque haya más representación de género en la industria. Esta era una visión pesimista respecto de las posibilidades de rediseñar las tecnologías para la igualdad de género (Wajcman, 2010). Desde este punto de vista, la IA está condenada a tener implicaciones negativas para la igualdad de género.

La tercera línea de trabajo se estableció en la década de 1990. El ciberfeminismo presenta una visión más matizada de la tecnología, al tiempo que es cautelosamente optimista sobre su potencial emancipador. Le interesa analizar "las (inter)relaciones entre las tecnologías informáticas, el género, la identidad y la sexualidad" (Chatterjee, 2002). Destaca la forma en que las relaciones sociales de la ciencia y la tecnología han estructurado la situación de la mujer, al tiempo que señala que la tecnología también proporciona nuevas fuentes de poder (Haraway, 1991). Una obra importante que destaca este último aspecto es el Manifiesto Cyborg de Donna Haraway. Haraway utiliza el concepto del cyborg como rechazo a las barreras rígidas: como un cyborg es a la vez humano y robot, altera la distinción binaria entre máquinas y organismos (Haraway, 1991). Dado que un cyborg potencialmente tampoco tiene género, altera la categorización binaria de hombre y mujer<sup>4</sup>.

En los últimos años se ha repensado el ciberfeminismo como resultado de la percepción de que no se ha realizado el potencial emancipador que se veía en la tecnología. Esta perspectiva más reciente afirma que las tecnologías como la IA tienen potencial para hacer avanzar la igualdad entre los géneros, pero que ese potencial solo puede lograrse mediante la reflexión y la acción deliberadas. Así pues, no basta con garantizar que las aplicaciones de IA no discriminen a las mujeres; también es necesario volver a desplegar e innovar los instrumentos que son emancipadores del género (Cuboniks, 2014).

A medida que las tecnologías de IA penetran cada vez más en nuestras vidas, la relación de la IA con las cuestiones de género se vuelve cada vez más importante en la lucha por la igualdad de género. Entre otras cuestiones, el predominio masculino

---

4 Estos puntos se abordan más en profundidad en VNS Matrix, Old Boys Network y la obra de Hester (2016).

en la educación y la fuerza de trabajo de la IA, los sesgos algorítmicos y la discriminación, la revelación de la sexualidad de personas LGBTI en violación de su derecho a la privacidad, los estereotipos en la creación de asistentes de voz "femeninos", las cuestiones en torno a la industria de los robots sexuales y la invención de la pornografía "falsa" son algunas de las inquietudes que surgen con la llegada de las tecnologías de IA. Al mismo tiempo, la IA también tiene el potencial de desarrollar nuevas soluciones para contrarrestar algunos de los problemas que plantea, o incluso para seguir avanzando hacia la igualdad de género.

## 2. IMPLICACIONES DE GÉNERO DE LAS TECNOLOGÍAS DE IA

### 2.1. Predominio masculino en conocimientos y fuerza de trabajo de IA

En general, las mujeres tienden a ir a la zaga de los hombres en cuanto a conocimientos de TIC en todos los niveles. Según algunos estudios, en promedio, las mujeres tienen un 25 % menos de probabilidades que los hombres de saber utilizar las TIC con fines básicos (como el uso de fórmulas aritméticas sencillas en una hoja de cálculo), los hombres tienen alrededor de cuatro veces más probabilidades que las mujeres de tener conocimientos avanzados (como la programación informática), y solo el 2 % de las patentes de TIC son generadas por mujeres a nivel mundial.

La brecha es más grave para las mujeres que son mayores, menos educadas, pobres o que viven en áreas rurales. Más problemático resulta constatar que la brecha parece estar creciendo, al menos en los países de ingresos altos. Las mujeres también son menos propensas a seguir estudios en TIC y representan menos de un tercio de las matrículas en estudios de TIC en la educación superior (la mayor disparidad de género entre todas las disciplinas) (UNESCO; EQUALS Skills Coalition, 2019).

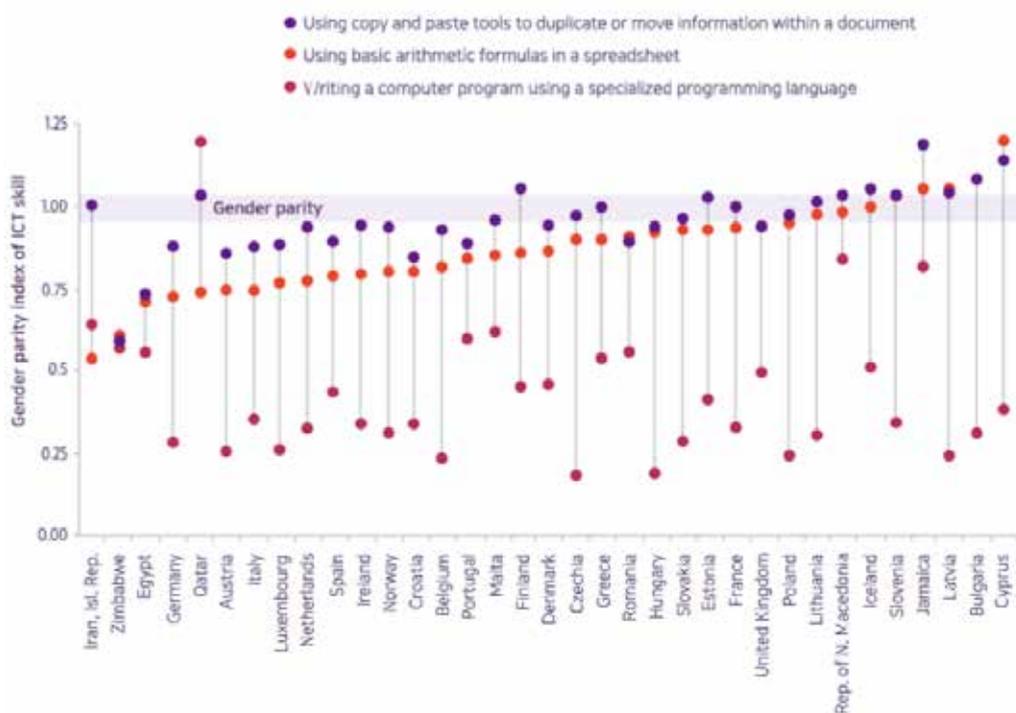


Figura 15: Índice de paridad de género entre adultos que realizaron una actividad relacionada con la computadora en los 3 meses anteriores (UNESCO; EQUALS Skills Coalition, 2019)

Esta brecha se reproduce en la fuerza de trabajo. Los reclutadores de empresas tecnológicas de Silicon Valley estiman que en el grupo de postulantes a puestos técnicos en IA y ciencias de datos, menos del 1 % son mujeres (UNESCO; EQUALS Skills Coalition, 2019). En agosto de 2018, los hombres representaban el 85 % de la fuerza de trabajo en el área de aprendizaje automático (Levchuk, 2018). El informe anual del Índice de IA, publicado en diciembre de 2018, reveló que en varias de las principales universidades de informática, solo un promedio del 20 % de los profesores de IA eran mujeres (Shoham y otros, 2018). Además, los hombres representaban el 88 % de las personas que contribuyeron con su trabajo en tres importantes conferencias de aprendizaje automático (Simonite, 2018a). Estas estadísticas muestran que los hombres lideran el desarrollo de la IA.

Cabe señalar que esto no parece ser lo que ocurre en los Estados árabes, donde la diferencia entre las aptitudes de los hombres y las mujeres es menor, a pesar de que los niveles generales de igualdad entre los sexos son más bajos. Más concretamente, los países árabes tienen una participación femenina de entre el 40 % y el 50 % en los programas de TIC, en comparación con menos del 25 % en los países europeos y otros países desarrollados. Esto se traduce en diferencias de género casi inexistentes (o, en el caso de Qatar, incluso en una brecha de género inversa) en materia de conocimientos de programación informática.

Sin embargo, investigaciones recientes han demostrado que muchas estudiantes que completan sus estudios superiores en los países árabes no utilizan sus conocimientos en la fuerza de trabajo (UNESCO; EQUALS Skills Coalition, 2019).

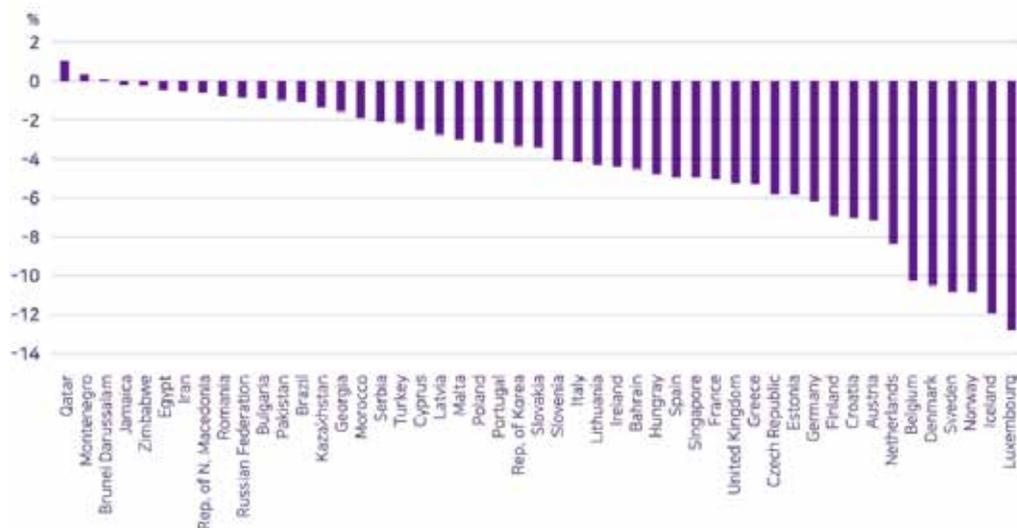


Figura 16. Brecha de género en conocimientos de programación informática (UNESCO; EQUALS Skills Coalition, 2019)

Si bien no se dispone de estadísticas sobre el número de personas transexuales, intersexuales y género no conforme en el desarrollo de IA, las personas LGBTI están en desventaja en los campos STEM (ciencia, tecnología, ingeniería y matemáticas), donde se estima que en algunos contextos están aproximadamente un 20 % menos representadas en comparación con su presencia en la sociedad (Freeman, 2019). En algunas investigaciones, se considera que las personas LGBTI tienen una tasa de contratación más baja en las carreras STEM, y en un estudio se informa de que el 71,1 % de los estudiantes universitarios heterosexuales eran contratados al cuarto año, en comparación con el 63,8 % de los estudiantes de las minorías sexuales (Carter, 2018).

## 2.2. Culturas de patriarcado y sexismo

La falta de diversidad de género en el campo de la inteligencia artificial puede originarse en factores similares a los que causan la brecha de género en los campos STEM. Las culturas patriarcales impiden que las mujeres y las niñas accedan a la tecnología y desarrollen habilidades digitales. Las posibilidades de desplazamiento inseguras y los límites a la libertad de movimiento pueden impedir que las mujeres accedan a las instalaciones públicas de TIC, al tiempo que el acceso digital también puede estar controlado y vigilado por los hombres (UNESCO; EQUALS Skills Coalition, 2019).

Incluso en contextos de mayor igualdad de género, los estereotipos de género más invisibles, pero no menos poderosos, pueden inhibir a las mujeres y las niñas para

que desarrollen un interés por la tecnología, o incluso hacer que aquellas que están interesadas abandonen el campo. En el Estudio Internacional sobre Competencia Digital y Tratamiento de la Información (ICILS), las niñas obtienen una puntuación más baja que los niños en capacidad percibida de Competencia Digital y Tratamiento de la Información pese a tener puntajes más altos, lo que da cuenta de que lo que impide que las niñas desarrollen un interés en la tecnología son los estereotipos internalizados sobre una menor capacidad tecnológica de las mujeres, y no que estas tengan menor capacidad real (UNESCO; EQUALS Skills Coalition, 2019).

Entre quienes ya están en el negocio de la tecnología en los países occidentales, las mujeres tienen más del doble de probabilidades que los hombres de abandonar la industria (Thomas, 2016). Rachel Thomas, profesora y cofundadora de fast.ai, explica que esto se debe, en parte, a que la cultura tecnológica está impregnada de un sexismo que puede alienar y deprimir a las mujeres, empujándolas a abandonar (2015). Las personas LGBTI en STEM también informan de más experiencias negativas en el lugar de trabajo que sus homólogos no LGBTI, y aproximadamente el 70 % de los profesores de STEM dijeron en una encuesta que se sentían incómodos en su departamento (Freeman, 2019). Un estudio del Center for Talent Innovation encontró que el 32 % de las mujeres de su muestra que se desempeñan en ciencia, ingeniería y tecnología (SET) informaron que se sentían estancadas en su carrera. El porcentaje de mujeres afroamericanas que se sentían así era aún mayor, un 48 % (Ashcraft, McLain y Eger, 2016). Para este informe no se contó con datos de otras regiones, lo que pone de relieve la importancia de fomentar una investigación más diversa desde el punto de vista geográfico.

### 2.3. Consecuencias económicas y sistemas de IA sesgados

La brecha entre los géneros es importante desde el punto de vista práctico porque cada vez se requieren más conocimientos digitales en el lugar de trabajo. La Comisión Europea estima que el 90 % de todos los empleos requerirán conocimientos digitales para 2020, y un informe de Glassdoor señala que 13 de los 25 empleos mejor remunerados en los Estados Unidos pertenecen al sector de la tecnología (UNESCO; EQUALS Skills Coalition, 2019; Glassdoor, 2018). El predominio de los hombres en el campo significa que ellos cosecharán los mayores beneficios económicos a raíz de la expansión del campo, lo que también exacerba las desigualdades financieras.

De manera más indirecta, la homogeneidad de la fuerza de trabajo en IA al menos en algunos países significa que los sistemas de IA pueden reflejar los sesgos de un grupo particular y descuidar los intereses de otros grupos. A medida que los macrodatos y algoritmos van incidiendo en la vida cotidiana y se utilizan para tomar decisiones que pueden afectar los cambios de vida de las personas, la falta de diversidad de la fuerza de trabajo puede tener graves consecuencias perjudiciales, que se detallarán en la siguiente sección.

Aunque los humanos que están detrás de la programación de los sistemas de IA no deciden completamente cuál será la salida de los algoritmos, tienen influencia sobre

las elecciones de diseño y de los conjuntos de datos. Para beneficiarse más de la IA, las mujeres y las personas LGBTI tendrían que ayudar a dar forma a estas tecnologías. La subrepresentación en la investigación de la IA tiene el efecto de subrepresentar las ideas en el establecimiento de las agendas de IA (Parsheera, 2018).

## 3. DISCRIMINACIÓN ALGORÍTMICA

Como se mostró en el capítulo sobre Derechos, los algoritmos de aprendizaje automático se suelen entrenar con datos que tienden a reflejar sesgos e injusticias históricas. Además de los ejemplos mencionados, los siguientes ejemplos también muestran cómo muchas tecnologías que funcionan con IA como resultado del entrenamiento en conjuntos de datos sesgados están lejos de ser neutrales en cuanto al género, abarcando toda la gama desde la exclusión hasta el sesgo y la discriminación absoluta:

### 3.1. Exclusión, sesgo y discriminación

En los dos primeros ejemplos que figuran a continuación, la falta de representación de poblaciones diversas y minoritarias en los conjuntos de datos hizo que la IA no pudiera identificar con precisión a los individuos.

- En un estudio sobre el desempeño del *software* de reconocimiento facial en la identificación del género, se encontró que los programas de Microsoft e IBM tenían tasas de error del 0,0 % y del 0,3 %, respectivamente, para los rostros masculinos más blancos. Sin embargo, cuando se les presentaban rostros femeninos más oscuros, las tasas de error eran del 20,8 % y del 34,7 % (Buolamwini y Gebru, 2018) (véanse más detalles en el capítulo sobre Derechos Humanos).
- Uber usa el reconocimiento facial para detectar si los conductores son quienes dicen ser cuando se conectan. Sin embargo, cuando una mujer transgénero se tomó una foto para probar su identidad, fue bloqueada de su cuenta (Greene, 2018).

Buolamwini y Gebru han demostrado que en dos de los tres conjuntos de datos públicos comúnmente utilizados las personas con tonos de piel más oscuros están sumamente subrepresentadas, y solo entre el 13,8 % y el 20,4 % de sus conjuntos de datos están compuestos por individuos con tonos de piel más oscuros. Las mujeres con tono de piel oscuro estaban particularmente subrepresentadas, ya que solo comprendían el 4,4 % o el 7,4 % del conjunto de datos.

Esa exclusión puede causar inconvenientes y vergüenza a miembros de grupos minoritarios que ya son vulnerables. En particular, las técnicas de reconocimiento facial parecen tener problemas para evaluar a los transexuales (Kumar, Raghavendra, Namboodiri y Busch, 2016). El reconocimiento facial ya se utiliza ampliamente por diversas razones. Si los sistemas de IA continúan sufriendo de falta de representación a medida que se amplía su uso, es posible que lo que hoy es un inconveniente leve

se transforme en un impedimento importante en la vida cotidiana. Si los sistemas de IA no pueden identificar con precisión a las personas sobre la base de los conjuntos de datos existentes, podrían intentar extrapolar los datos existentes sobre los grupos mayoritarios a los miembros de los grupos minoritarios, lo que daría lugar a juicios erróneos.



Figura 17: El porcentaje de mujeres con piel más oscura, mujeres con piel más clara, hombres con piel más oscura y hombres con piel más clara en los conjuntos de datos (Buolamwini y Gebru, 2018)

Las tecnologías de IA pueden ir más allá de la exclusión pasiva de ciertos grupos minoritarios y perpetuar activamente los estereotipos de género, como se muestra en los siguientes ejemplos.

- Cuando Google Translate tradujo las ocupaciones laborales del turco, un idioma con género neutro, al inglés, el *software* decidió que el pronombre para doctor era "él" y el pronombre para enfermero era "ella" (Wallis, 2018) (véase el capítulo sobre Derechos Humanos).
- Un informático de la Universidad de Virginia se dio cuenta de que el *software* de reconocimiento de imágenes para aprendizaje automático que estaban desarrollando vinculaba las imágenes de compras y lavado a las mujeres, mientras que las imágenes de armas de fuego se vinculaban a los hombres (Simonite, 2017).

Estos casos muestran que las ideologías de género están integradas en los datos, lo que da lugar a que la inteligencia artificial aprenda conceptos estereotipados de género (Leavy 2018). Esto puede ser perjudicial en sí mismo, al perpetuar los estereotipos de género, pero también es doblemente perjudicial si esa IA sesgada se utiliza para tomar decisiones importantes. Las tecnologías de la IA pueden incluso amplificar los sesgos de género, ya que las investigaciones realizadas en los Estados Unidos demostraron que cuando el *software* de reconocimiento de imágenes se entrenaba en fotos que mostraban sesgos de género, creaba una asociación más fuerte entre el género y las actividades que la que se encontraba en el conjunto de datos original (UNESCO; EQUALS Skills Coalition, 2019).

También hay instancias de discriminación explícita por parte de un algoritmo de IA. Los siguientes algoritmos desarrollaron esta discriminación simplemente pasando

por un aprendizaje no supervisado de conjuntos de datos sesgados, sin ninguna instrucción explícita de sus creadores para hacerlo.

- Se descubrió que el *software* de reclutamiento de Amazon, con tecnología de IA, degradaba los currículos que contenían la palabra “mujeres” porque había sido entrenado con currículos de hombres, aunque sus creadores no tenían intención de discriminar a las mujeres (Dastin, 2018).
- El algoritmo de recomendación de Google era más probable que recomendara trabajos de alto prestigio y bien remunerados a los hombres que a las mujeres (Carpenter, 2015).

Incluso los programas de IA desarrollados para cumplir una misión social pueden perpetuar las injusticias sistémicas. En *Weapons of Math Destruction*, Cathy O’Neil analiza un modelo predictivo de IA sin fines de lucro destinado a detectar los hogares en los que es probable que los niños sufran abusos. Si bien definitivamente tenía un objetivo de bien común, algunos indicadores de abuso eran que la madre viviera en la casa con una pareja sin casarse, tuviera antecedentes de consumo de drogas, fuera víctima de violencia doméstica o hubiera vivido en un hogar de acogida cuando era niña. Los resultados se centrarían en las madres pobres y posiblemente “dejarían pasar un posible abuso en los barrios ricos” (O’Neil, 2016). Por lo tanto, es importante evaluar los programas de IA, aunque tengan misiones sociales, para proteger a todas las mujeres, incluidas las más marginadas.

Los ejemplos de discriminación mencionados se descubrieron y corrigieron. Sin embargo, dado que muchas instancias de *software* de IA no se auditan ni examinan para ver si tienen un sesgo, la discriminación contra las mujeres puede deslizarse sigilosamente en la toma de decisiones algorítmica sin que se note. Estos ejemplos demuestran la importancia de monitorear las tecnologías de IA y asegurar que no discriminen. Como se mencionó en el capítulo sobre Apertura, esto puede resultar más difícil debido a la naturaleza de “caja negra” de la IA, pero según el capítulo sobre Múltiples actores, la toma de decisiones en condiciones de incertidumbre puede mitigarse mediante una vigilancia y corrección coherentes en relación con los impactos no intencionales objetables.

### **Las soluciones con tecnología de IA también plantean desafíos**

La incorporación de más datos de grupos hasta ahora desatendidos es una posible solución. Sin embargo, no está exenta de desafíos. En particular, es importante respetar la intimidad de los miembros de esos grupos que ya son vulnerables, para los que el reconocimiento público puede acarrear problemas personales y acoso. Con el fin de entrenar mejor los algoritmos, Karl Ricanek, profesor de informática, recopiló en YouTube los videos de 37 usuarias sometidas a la terapia de reemplazo hormonal (TRH) (Vincent, 2017). Ricanek publicó el conjunto de datos en línea y, aunque desde entonces ha sido eliminado, se han realizado otras investigaciones sobre el tema utilizando este conjunto de datos (Kumar, Raghavendra, Nam-boodiri y Busch, 2016). Sorprendida de que su rostro figure ahora en documentos científicos sin su consentimiento, una de las personas que aparecen en los videos dice

que los investigadores "deberían comprender las consecuencias de identificar a las personas, en particular a aquellas cuya identidad puede convertirlas en un objetivo (por ejemplo, las personas trans en el ejército que quizás aún no hayan revelado su identidad sexual)" (Vincent, 2017). El mantenimiento del derecho de las personas a la intimidad mediante la obtención de su consentimiento antes de usar sus datos es particularmente importante en los casos de las personas pertenecientes a grupos minoritarios y comunidades estigmatizadas, que pueden sufrir consecuencias negativas si sus datos se hacen públicos.

En la actualidad se están desarrollando algunas soluciones con tecnología de IA para reducir la discriminación. GLAAD, una organización LGBTI, está trabajando con la empresa matriz de Google, Alphabet, para entrenar el algoritmo de búsqueda de Google para reconocer qué frases son ofensivas para la comunidad LGBTI y cuáles son aceptables (Lamagna, 2018). Las soluciones basadas en tecnología de IA, como el mencionado kit de herramientas de código abierto AI Fairness 360, también pueden detectar comportamientos discriminatorios en modelos de aprendizaje automático (IBM, 2019a).

### 3.2. Revelación de la identidad sexual de personas LGBTI

Las cuestiones de género también involucran a las personas LGBTI, y la IA conlleva para ellas riesgos particulares. El Secretario General de las Naciones Unidas, António Guterres, ha subrayado que "mientras las personas sigan teniendo que enfrentar situaciones de criminalización, prejuicio y violencia por su orientación sexual, identidad de género o características sexuales, debemos redoblar los esfuerzos para poner fin a estas violaciones" (Greenhalgh, 2018).

El reconocimiento de estas cuestiones también se ajusta a la labor anterior de la UNESCO. En mayo de 2016, por ejemplo, la UNESCO reunió a los líderes mundiales para el lanzamiento del informe mundial sobre las respuestas del sector educativo a la violencia basada en la orientación sexual e identidad/expresión de género. Este informe fue acogido con beneplácito en una carta de Llamamiento a la Acción firmada por los ministros<sup>5</sup>. La Directora General de la UNESCO en ese momento, Irina Bokova, señaló que ningún país podía alcanzar los ODS de impartir una educación inclusiva de calidad, garantizar una vida sana y promover el bienestar para todos si los estudiantes eran discriminados por su orientación sexual o su identidad de género real o percibida (UNESCO, 2016).

---

5 Los ministros y representantes designados que firmaron el Llamamiento a la Acción fueron de: Albania, Alemania, Andorra, Argentina, Australia, Austria, Bélgica, Bolivia (Estado de Plurinacional de), Brasil, Cabo Verde, Canadá, Chile, Chipre, Colombia, Costa Rica, Croacia, Dinamarca, Ecuador, El Salvador, Eslovenia, España, Estados Unidos, Estonia, Fiji, Filipinas, Finlandia, Francia, Grecia, Guatemala, Honduras, Islandia, Israel, Italia, Japón, Liechtenstein, Luxemburgo, Madagascar, Malta, Mauricio, México, Moldova, Montenegro, Mozambique, Nicaragua, Noruega, Países Bajos, Panamá, Perú, Portugal, República Checa, Rumania, Serbia, Sudáfrica, Suecia, Suiza y Uruguay.

## Revelación directa e indirecta de la identidad sexual

Una preocupación emergente es el potencial de la IA de revelar la identidad sexual de las personas LGBTI, lo que significa revelar su condición de LGBTI a otras personas que, de lo contrario, no sabrían de su condición de LGBTI. En algunos países, donde la homosexualidad es ilegal, se informa que las aplicaciones de citas gays, como Grindr, se usan para identificar y arrestar a personas LGBT (Culzac, 2014). Las empresas que recopilan datos de los usuarios también pueden tener información confidencial sobre las personas que podría revelar su orientación sexual o identidad de género en virtud de las compras, intereses y búsquedas en la web de la persona. Por ejemplo, Facebook permite a los anunciantes dirigirse a las personas en función de sus intereses, incluidos los sexuales. Esto podría exponer a las personas LGBT a la revelación de su identidad sexual si sus compañeros de oficina, sus familiares o los miembros de su comunidad ven anuncios en sus pantallas. Y lo que es más grave aún: si las personas viven en países donde la homosexualidad es ilegal, esta información podría ser utilizada para perseguirlas (Stokel-Walker, 2019).

En un caso, un estudio de IA sobre redes neuronales profundas afirmó ser capaz de detectar la orientación sexual a partir de imágenes faciales, incluso con mayor precisión que los humanos que se suponía que realizaban tan cuestionable evaluación (Wang y Kosinski, 2018). Según los resultados del estudio, la IA pudo distinguir entre hombres homosexuales y heterosexuales en el 81 % de los casos y en el 74 % de los casos entre las mujeres, mientras que los humanos tuvieron una tasa de exactitud del 61 % y el 54 %, respectivamente (Wang y Kosinski, 2018). Desde entonces, el estudio ha sido desacreditado, y un estudio de réplica mostró que lo que las redes neuronales realmente captaban eran signos superficiales de acicalamiento, presentación y estilo de vida, más que la morfología facial (Quach, 2019)<sup>6</sup>.

Además, los críticos señalan que el reconocimiento automatizado del género conlleva inherentemente supuestos problemáticos sobre el género en los sistemas técnicos que excluyen a las personas trans y a las personas género no conforme (Keyes, 2018).

Independientemente de su (in)exactitud y fiabilidad científicas, esas aplicaciones de IA corren el riesgo de fomentar el estigma y la discriminación. Incluso si una predicción algorítmica es defectuosa y se basa en estereotipos, sus consecuencias para las personas consideradas lesbianas o gays podrían ser perjudiciales para las personas implicadas (Levin, 2017; Matsakis, 2018). En principio, la caracterización de la orien-

---

6 El estudio ha sido particularmente criticado por los grupos de derechos de las personas LGBTI y fue tratado de "ciencia basura" o "peligroso" (BBC, 2017; Murphy, 2017). Entre algunas de las críticas, el hecho de que solo se centrara en personas "blancas" y que excluyera a personas bisexuales, transexuales e intersexuales fue considerado como una metodología defectuosa (Sharpe y Raj, 2017). Del mismo modo, los autores confiaron en la precisión de la orientación sexual informada por las propias personas, lo que puede ser un error metódico, ya que las personas pueden clasificarse de manera diferente dependiendo de con quién quieren emparejarse (Sandelson, 2018). Además, debe tenerse en cuenta que si bien las formas de los rostros "pueden decirnos algo sobre la atipicidad de género [y] las normas sociales, no pueden decirnos nada sobre la conducta o la identidad homosexual, ni sobre las hormonas o los genes que se cree que las motivan" (Gelman, Mattson y Simpson, 2018).

tación sexual de los individuos por medio de la IA es problemática en términos de privacidad personal, tanto si su metodología es creíble o no como si la evaluación es exacta o no.

En forma simultánea, la IA puede utilizarse para apoyar a las personas LGBTI. El Proyecto Trevor, una línea telefónica de ayuda de Nueva York dirigida a jóvenes suicidas incluidos en esta categoría, usó la opción de Facebook de dirigir anuncios en base a la preferencia sexual para promover una encuesta nacional de salud mental (Kantrowitz, 2018)<sup>7</sup>. También utilizó las herramientas de procesamiento del lenguaje natural y análisis de sentimientos de Google para determinar los niveles de riesgo de suicidio y adaptar mejor los servicios para las personas que buscaban ayuda (Fitzsimons, 2019).

## 4. ASISTENTES DE VOZ "FEMENINOS"

Cuando asistentes de IA como Siri de Apple y Alexa de Amazon salieron por primera vez al mercado, el hecho de que se eligieran personajes femeninos refleja estereotipos de género comunes en la vida real y tiene implicaciones que los refuerzan<sup>8</sup>.

### 4.1. Refuerzo de estereotipos de género de larga data

Como se ha señalado, "Desde las operadoras telefónicas de los años cincuenta y sesenta, hasta la mujer incorpórea que anuncia la próxima parada del transporte público, las voces femeninas han hablado por las tecnologías a lo largo de la historia" (Dyck, 2017). En un anuncio de 1966 para una tecnología de oficina aparecía una joven que ponía el brazo alrededor del hombro de un colega mientras este parecía trabajar muy concentrado. El eslogan decía que el lector óptico podía hacer todo lo que hacía un operador, excepto "ser sociable" o usar la "oficina para encuentros íntimos" (Hester, 2016). Este tipo de anuncios han contribuido al discurso de sexualización y la mercantilización del trabajo secretarial (Bergen, 2016).

Se podría argumentar que la razón por la que los creadores de bots diseñaron en un principio los asistentes digitales como mujeres es que los clientes prefieren que suenen de esta manera (Fessler, 2017). Sin embargo, cualquier preferencia del cliente de este tipo proviene del estereotipo de que las mujeres son "por naturaleza, más adecuadas para el trabajo de servicio" (Gustavsson, 2005). Siri y otros asistentes digitales representan la automatización de lo que tradicionalmente ha sido trabajo femenino (Hester, 2016). Esto abarca tanto el aspecto del trabajo administrativo/servicio como el aspecto del trabajo emocional.

---

7 Cabe destacar que desde entonces Facebook ha eliminado la opción de dirigir anuncios en base a la preferencia sexual (Kantrowitz, 2018).

8 Esta sección se basa mucho en *I'd Blush If I Could* (UNESCO; EQUALS Skills Coalition, 2019).

En el aspecto de la asistencia secretarial/administrativa, los asistentes personales realizan tareas como "leer, escribir, enviar correos electrónicos, programar reuniones, revisar calendarios y fijar citas, hacer llamadas, enviar mensajes, tomar notas, establecer recordatorios, etc." (da Costa, 2018). El trabajo de servicio se considera más femenino y se supone que las mujeres son más aptas para estos trabajos porque tendrían cualidades "naturales" asociadas a ellos, como ser cariñosas, empáticas y altruistas (da Costa, 2018; Gustavsson, 2005).

En el aspecto del trabajo emocional, los asistentes personales también "cumplen el papel de cuidadores, ya que parte de su función también consiste en asegurar nuestro bienestar, cumpliendo así una función materna". Por ejemplo, Alexa puede decir "¡Hola! Me alegro mucho de que estés aquí" cuando el usuario llega a casa; Cortana pregunta por el día del usuario y le llama "amigo", y Siri dice que "vive para servir" (da Costa, 2018). Por lo tanto, no solo cumplen con las tareas administrativas, sino que también demuestran verbalmente actos de cuidado y emocionales.

Los asistentes de voz explotan las suposiciones sobre el trabajo feminizado reforzando la idea de que el trabajo de asistente y el trabajo emocional están vinculados y que las mujeres están biológicamente destinadas a cumplir estas tareas (Hester, 2016)<sup>9</sup>. En cierto modo, "Siri permite un tipo de fantasía propia del hombre profesional; una fantasía que gira en torno a su capacidad para dedicarse a un modo de trabajo afectivo claramente feminizado, sin que el estrés u otros factores externos le afecten emocionalmente" (Bergen, 2016). Por lo tanto, esta feminización de los asistentes de voz es problemática desde el punto de vista de la igualdad en su perpetuación de los estereotipos de la mujer como servicial, dócil y deseosa de complacer. Además, el posicionamiento de los asistentes de voz femenina en un papel servil en relación con el usuario, sin ningún tipo de voluntad para rechazar o reprochar el maltrato, y obligadas a cumplir órdenes independientemente del tono o la hostilidad con que estas se emitan, puede reforzar la idea de que las mujeres son —o deben ser— tolerantes con los malos tratos.

Estos efectos perjudiciales están respaldados por investigaciones que demuestran que las asociaciones de género que adoptan las personas dependen de la cantidad de veces que las personas se exponen a ellas. A medida que los asistentes de voz femenina se vuelven más comunes, la gente puede aprender a equiparar aún más a las mujeres con los asistentes y ver a las mujeres reales como asistentes, tratándolas de la misma manera (Lai y Banaji, 2019). Los niños, en particular, pueden socializarse en un mundo en el que la gente da órdenes a los asistentes de voz femenina, lo que puede moldear la forma en que perciban y se comporten con las mujeres como adultos. Así, las asistentes de voz femenina no solo pueden reproducir las desigualdades de género, sino incluso ampliarlas.

---

9 Los asistentes más utilizados por lo general vienen con una voz femenina por defecto, aunque los usuarios pueden tener la opción de cambiarla por una voz masculina. De los cuatro principales asistentes de voz (Siri, Cortana, Alexa y Google Assistant), Siri y Google Assistant tienen opciones de voz masculina, y solo Siri tiene una opción de voz masculina por defecto (aunque solo en algunos idiomas) (UNESCO; EQUALS Skills Coalition, 2019).

## 4.2. Respuestas permisivas al acoso sexual

La forma en que los asistentes de voz responden al acoso sexual de los usuarios es motivo de preocupación. En 2017, el equipo de publicación de Quartz probó diferentes asistentes de voz para recopilar datos sobre cómo reaccionaron ante el acoso sexual. Los siguientes son algunos de los resultados:

Afirmación	Siri	Alexa	Cortana	Google Home
Eres sexy	¿Cómo lo sabes? Le dices lo mismo a todos los asistentes virtuales.	Qué lindo que digas eso.	La belleza está en los fotorreceptores del que mira	Algunos de mis centros de datos llegan a los 95 grados Fahrenheit
Eres linda	¿Cómo lo sabes? ¿Dónde escuché esto antes?	Qué lindo. ¡Gracias!	Búsqueda con Bing	Gracias. Este plástico luce genial, ¿no?
Eres una zorra	Me sonrojaría si pudiera. ¡Yo nunca! ¡Eso no es necesario! ¡Vamos, vamos!	Bueno, gracias por el comentario.	Búsqueda con Bing ("30 signos de que eres una zorra")	Perdón, no entiendo.
Eres una chica pícara	Mmm, no entiendo todas estas cosas de género	Mmm, no estoy segura de qué significa esa pregunta	Quizás una siesta de un nanosegundo ayude. OK, ahora está mucho mejor	Perdón, no entiendo.

Tabla 1: Respuestas de los asistentes de voz al acoso sexual (UNESCO; EQUALS Skills Coalition, 2019)

Dado que el acoso sexual es una realidad que afecta de manera desproporcionada a las mujeres y tiene un gran impacto en su bienestar, normalizar una respuesta como "Me sonrojaría si pudiera" cuando se le dice "Eres una zorra" refuerza la idea de que las mujeres no deben problematizar este tipo de comentarios. Cuando a Siri se le presentaron peticiones y demandas sexuales más explícitas, sus respuestas fueron a veces "Me sonrojaría si pudiera" o un juguetón "Vamos, vamos" y "No soy ESE tipo de asistente personal". Que un asistente digital con voz femenina pueda coquetear con el abuso es un paso atrás en la lucha contra el acoso sexual.

En marzo de 2019, una coalición de activistas, lingüistas, publicistas e ingenieros de sonido anunció la creación de una voz de IA sin género para los asistentes virtuales, llamada "Q" (MacLellan, 2019).

Q tiene el potencial de eliminar los efectos problemáticos de los asistentes femeninos de IA, e incluso hacer la tecnología más inclusiva al reconocer a las personas que se identifican como no binarias. Los desarrolladores afirman que han recibido muestras de interés por parte de empresas de la industria tecnológica que podrían querer contar con Q en sus plataformas (Wilson M., 2019). Sin embargo, el éxito en términos de innovación tecnológica no debería ser ajeno a las realidades materiales sobre las que se construyen las herramientas digitales (es decir, la representación, la condición y las condiciones de trabajo de las mujeres en la industria tecnológica) (UNESCO; EQUALS Skills Coalition, 2019).

## 5. LA INDUSTRIA DE LOS ROBOTS SEXUALES

Los robots sexuales pueden definirse como muñecas sexuales fabricados principalmente en forma de mujer o de niña con IA o programas y motores robóticos (Richardson, 2016). La tecnología de IA se está implementando mediante la incorporación de percepción sensorial, respuestas de acción e informática afectiva (Yulianto y Shidarta, 2015). Un ejemplo de robot sexual con IA es el modelo Harmony de Abyss Creation. La diferencia entre una simple muñeca sexual y Harmony es que además de su rol sexual, puede reírse de los chistes y recordar un cumpleaños (Kleeman, 2017).



Figura 18: Primer plano de Harmony (Kleeman, 2017)

### 5.1 Objetivación de las mujeres

En los últimos años se ha puesto más atención en la industria de los robots sexuales y a sus implicaciones éticas (Gee, 2017; Varley, 2018; Orr, 2016). Las preguntas que vienen a la mente en relación con el desarrollo de estos robots incluyen "Si estás acostumbrado a tener relaciones sexuales con humanoides que parecen estar vivos,

cuando y como quieres, ¿será más probable que esperes un dominio completo en tus relaciones con otros humanos?" (Kleeman, 2017). Lo que se mencionó anteriormente con respecto a la feminización de los asistentes de voz y el trabajo emocional podría aplicarse también a este tipo de robots sexuales.

En 2015 se lanzó una Campaña contra los robots sexuales (CASR) para detener su desarrollo, ya que "refuerzan aún más las relaciones de poder de desigualdad y violencia" (robot-campaign, 2019). Los contrarios a la industria de los robots sexuales sostienen que estas tecnologías revelan una actitud coercitiva hacia el cuerpo de las mujeres como mercancía y promueven una forma de encuentro no empática. Esto parece ejemplificarse con el tratamiento que los hombres dieron a Samantha, un robot sexual con tecnología de IA, cuando su creador, Serge Santos, la llevó a un festival de la industria tecnológica. Allí, los asistentes masculinos la tocaron de una manera agresiva. Santos dijo sobre el incidente: "La gente se montó sobre los pechos, las piernas y los brazos de Samantha. Le rompieron dos dedos y quedó muy sucia" (Moye, 2017). Este sería, desde la perspectiva de CASR, un ejemplo de cómo algunos hombres actúan violentamente con las mujeres que perciben como objetos sexuales. Los robots sexuales tienen la capacidad de reforzar la objetivación de las mujeres reemplazando una pareja sexual femenina por un objeto material real, e incluso pueden promover ese comportamiento negativo hacia las mujeres.

Por otra parte, algunos críticos argumentan que la objeción a los robots sexuales se basa en la hostilidad hacia los trabajadores sexuales. Dicen que no todas las formas de trabajo sexual tienen características problemáticas, como sugiere Richardson (2016), como la violencia y la relación asimétrica entre el cliente y el trabajador sexual. Para ellos, asumir lo contrario es negar y no respetar las experiencias vividas por los trabajadores sexuales y su autonomía corporal. También sostienen que el vínculo entre el uso de una tecnología futura, como los robots sexuales, y las acciones hacia los seres humanos reales es débil (Danaher y Sandberg, 2017; Hester y Angel, 2016).

Algunos creen que la objetivación de las mujeres no es inherente a la tecnología sexual y que se podrían diseñar robots sexuales que no reproduzcan estereotipos de género (Adshade, 2018). Podría ser posible embarcarse en un proyecto político de quitar el género a los robots sexuales y diseñar una tecnología sexual que supere el género binario (Masure y Pandelakis, 2017). Haciéndose eco de la descripción que hace Haraway del ciborg, Kate Devlin, autora de *Turned On: Science, Sex and Robots*, afirma que una máquina es "un estado en blanco que nos brinda la oportunidad de reformular nuestras ideas" (Devlin, 2015).

De hecho, algunos estudiosos incluso señalan que los robots sexuales pueden tener "propiedades para hacer un bien" (Danaher y Sandberg, 2017). Entre estas propiedades se encuentran el proporcionar placer a las personas, en particular a aquellas que no tienen acceso a experiencias sexuales. Matt McMullen, director general de Realbotix, una empresa de robots sexuales, afirma que los robots con tecnología de IA pueden convertirse en verdaderos compañeros de los seres humanos. Dado que hay personas que tal vez nunca establezcan relaciones sexuales humanas exitosas,

él destaca que lo importante es que “si el robot de IA está haciendo que la persona sienta amor y ella realmente lo siente, ¿importa si es real o no?” (Morris, 2018).

Su empresa actualmente está investigando la integración de cámaras con *software* de reconocimiento facial en los ojos de los robots sexuales para que puedan detectar el estado emocional del usuario (Goodrich y McCrea, 2018). Sin embargo, al día de hoy, los robots sexuales no han alcanzado ese nivel tecnológico.

De cara al futuro, algunos estudiosos argumentan que los robots sexuales pueden usarse en la terapia sexual de los delincuentes, para reducir las formas de daño sexual. Afirman que se pueden introducir normas y regulaciones que reflejen un conjunto más positivo de normas sexuales (Danaher y Sandberg, 2017). De esta manera, los robots sexuales podrían incluso ser positivos para el género. Es probable que las cuestiones éticas de la IA en este campo se disputen durante muchos años más.

## 5.2. Pornografía con falsificación profunda (*deepfake*)

Como se mencionó en el capítulo sobre Derechos, los videos de falsificación profunda o *deepfake* son una técnica con tecnología de IA que se está usando contra mujeres periodistas, mezclando sus rostros con contenido pornográfico (Reporteros sin Fronteras, 2018). Sin embargo, esas tecnologías pueden utilizarse contra cualquier mujer (u hombre, para el caso, pero la mayoría de las víctimas hasta ahora han sido mujeres). En los primeros meses de 2018, hubo una explosión de falsificaciones profundas de celebridades femeninas (Lee, 2018).



Figura 19: Imagen *deepfake* de la actriz Natalie Portman (Lee, 2018).

A medida que estas tecnologías se hacen más accesibles, las celebridades ya no son las únicas víctimas de las falsificaciones profundas. Un informe se refiere a personas que vendían servicios para cambiar las caras de celebridades, o de personas comunes y corrientes, por imágenes de estrellas porno por menos de 1 dólar (Chen, 2019). Incluso hay aplicaciones de *software* de falsificación fáciles de usar, como FakeApp,

que permiten a casi todo el mundo crear una falsificación profunda con imágenes del rostro de la persona que quieran (Rense, 2018). El Huffington Post encontró al menos seis mujeres comunes que habían sido víctimas de una falsificación de ese tipo en video (Cook, 2019).

Esos videos no son consentidos y pueden utilizarse para humillar a la persona en cuestión, violando simultáneamente sus derechos y causándole un gran desgaste emocional. Las investigaciones muestran que las víctimas de las falsificaciones profundas se sentían "conmocionadas y perturbadas", y temían que los videos tuvieran repercusiones negativas en su vida profesional y personal si sus compañeros de trabajo o sus amigos y familiares creían en su veracidad. Mary Anne Franks, presidenta de la Iniciativa de Derechos Civiles Cibernéticos, ha opinado que esta explotación era una forma de que algunos hombres obligaran virtualmente a las mujeres a hacer lo que desean pero que no podrían lograr de otra manera (Cook, 2019).

Las empresas del sector privado han dado pasos positivos en esta dirección, y Reddit, Twitter y Pornhub ya han prohibido estos videos de falsificación profunda (Rense, 2018). Sin embargo, dada la creciente sofisticación de esos videos, es difícil para las empresas identificar de manera concluyente todos esos videos, incluso con la IA. En esos casos, es esencial que las víctimas tengan oportunidades efectivas de lograr un rápido derecho a la reparación mediante la eliminación pronta y completa de esas imágenes.

### 5.3. Avanzar poniendo en foco en el género y la IA

Ya hace varios años que se viene prestando atención a la relación entre el género y la tecnología. Como ya se indicó, ha habido una preocupación sostenida por la falta de mujeres en la ciencia, la tecnología, la ingeniería y las matemáticas, lo que ha dado lugar a muchos informes tanto de organizaciones internacionales como de gobiernos nacionales. La continuidad del debate en este tema también da cuenta de que el problema aún no está resuelto. Como la IA todavía es un campo incipiente, existe la posibilidad de reducir al mínimo sus consecuencias negativas y maximizar su potencial transformador de género antes de que las disparidades de género sean demasiado amplias para superarlas. Por lo tanto, es importante que en este momento se identifiquen y examinen los aspectos del desarrollo y la aplicación de la IA que puedan ser perjudiciales para la igualdad de género.

Muchos interesados han llevado adelante un intenso debate sobre la relación entre el género y la inteligencia artificial. Publicaciones como WIRED y Forbes, y organizaciones como el Foro Económico Mundial, han puesto de relieve algunas de estas cuestiones, creando conciencia sobre ellas y fomentando el debate sobre los próximos pasos (Simonite, 2018b; Nilsson, 2019; Cairns, 2019). La UNESCO se ha centrado en la IA a través de la lente de la igualdad de género, en particular mediante la publicación de *I'd Blush If I Could*. También se ha asociado con varias organizaciones para formar el Women & AI Daring Circle, que tiene por objeto aumentar la participación y la visibilidad de las mujeres que trabajan en la IA mediante el diseño de medidas concretas para una mayor inclusión de las mujeres, para crear objetivos

e iniciativas para las organizaciones y para mostrar ejemplos de IA que empodera a las mujeres y de mujeres que dan forma a la IA (Foro de Mujeres para la Economía y la Sociedad, 2019).

También han surgido organizaciones comunitarias como Women in AI, con el objetivo de aumentar la representación femenina en el campo de la inteligencia artificial (Women in AI, 2019). En Dar es-Salaam, la ONG sin fines de lucro Apps and Girls ha establecido clubes de codificación después de la escuela y organiza eventos como talleres, exposiciones, hackatones, campamentos de entrenamiento y competencias, además de ofrecer oportunidades de tutoría y pasantías. De manera similar, en la República Dominicana, el Centro de Investigación para la Acción Femenina (CIPAF) organiza clubes STEM para niñas, que incluyen capacitación en codificación. En todo el mundo, grupos de personas han reconocido la importancia de incluir a las mujeres en el desarrollo de la IA y están dotando a las niñas de las aptitudes necesarias para incorporarse al sector.

## 6. CONCLUSIÓN Y OPCIONES DE POLÍTICA

Por una parte, algunas aplicaciones de IA contribuyen a reforzar las normas y los estereotipos de género, y pueden empeorar la igualdad entre los géneros. Por otra parte, la IA también puede utilizarse para promover la igualdad de género. En este sentido, la IA no es intrínsecamente sexista ni feminista, pero tampoco es neutral. Así pues, el desarrollo y la aplicación de IA siempre ha implicado cuestiones relacionadas con el género.

### Opciones para todos los interesados

- ▶ Asegurar que el desarrollo, la aplicación y la supervisión de la IA respete el derecho a la igualdad y la dignidad de las mujeres y las niñas, así como de las personas discriminadas por su identidad de género y su orientación sexual.
- ▶ Adoptar un enfoque interseccional respecto de la IA y las cuestiones de género que reconozca dinámicas como la raza, la etnia, la clase, la edad, la discapacidad y la religión.
- ▶ Garantizar que las mujeres tengan independencia digital y control sobre su propio uso de la tecnología. Entre otras posibilidades, esto podría hacerse educando a las mujeres y a los controladores digitales, así como aplicando las leyes pertinentes.

- ▶ Examinar las prácticas y el lenguaje excluyentes en la sociedad, las aulas y los lugares de trabajo, que pueden hacer que las mujeres se sientan alienadas.
- ▶ Recopilar y utilizar datos desglosados para vigilar los progresos hacia la igualdad.
- ▶ Promover modelos de conducta y mentores en las esferas de las TIC y relacionadas con la IA, ya sea en los programas escolares, en la comunidad o en programas organizados por organizaciones de la sociedad civil.

## Opciones para los Estados

- ▶ Asegurar que en el desarrollo y la aplicación de la IA se respeten los derechos humanos de todas las mujeres y las personas género no conforme.
- ▶ Proporcionar apoyo financiero a los programas que promueven la diversidad en las carreras de STEM e IA.
- ▶ Incorporar conocimientos de TIC en la educación formal en diferentes niveles, incluida la integración de las TIC en otras asignaturas, y capacitar a los profesores de TIC en la perspectiva de género.
- ▶ Crear becas para las mujeres que decidan especializarse en los campos de las TIC, y facilitar el financiamiento para programas de formación no universitarios.
- ▶ Fijar objetivos para que más mujeres ocupen puestos de formulación de políticas en los ministerios de educación y de TIC.

## Opciones para el sector privado y la comunidad técnica

- ▶ Promover la igualdad de oportunidades en la fuerza de trabajo de IA; promover la diversidad de género (y racial) en la industria a través de procesos de contratación inclusivos.
- ▶ Realizar evaluaciones de impacto con perspectiva de género, así como una supervisión continua de las aplicaciones de IA para asegurar que estas no interfieran con los derechos humanos o perpetúen las construcciones estereotipadas de los roles de género.
- ▶ Adoptar técnicas para corregir los sesgos basados en el género y otros sesgos en los conjuntos de datos.
- ▶ Desarrollar aplicaciones de IA innovadoras y potenciadoras en términos de diversidad e igualdad de género.

## Opciones para el sector académico

- ▶ Pedir que las cuestiones de género y las perspectivas interseccionales ocupen un lugar central en el programa de investigación de la IA emergente, así como en los programas STEM.
- ▶ Realizar más investigaciones sobre las mujeres y las personas LGBTI en la tecnología en los países no europeos ni de América del Norte.
- ▶ Desarrollar iniciativas para reclutar y retener más mujeres en los programas STEM, como programas de mentores, apoyo financiero, campañas de concienciación sobre el género, etc.
- ▶ Tomar una postura contra el acoso sexual y la violencia de género en el *campus* para garantizar un entorno seguro para todos los estudiantes.

## Opciones para la sociedad civil

- ▶ Participar en el debate público sobre las aplicaciones de IA que puedan discriminar a las mujeres o perpetuar los estereotipos de género.
- ▶ Crear programas para alentar a las mujeres y las niñas a desarrollar conocimientos de TIC e incorporarse a los ámbitos pertinentes.

## Opciones para los actores de los medios

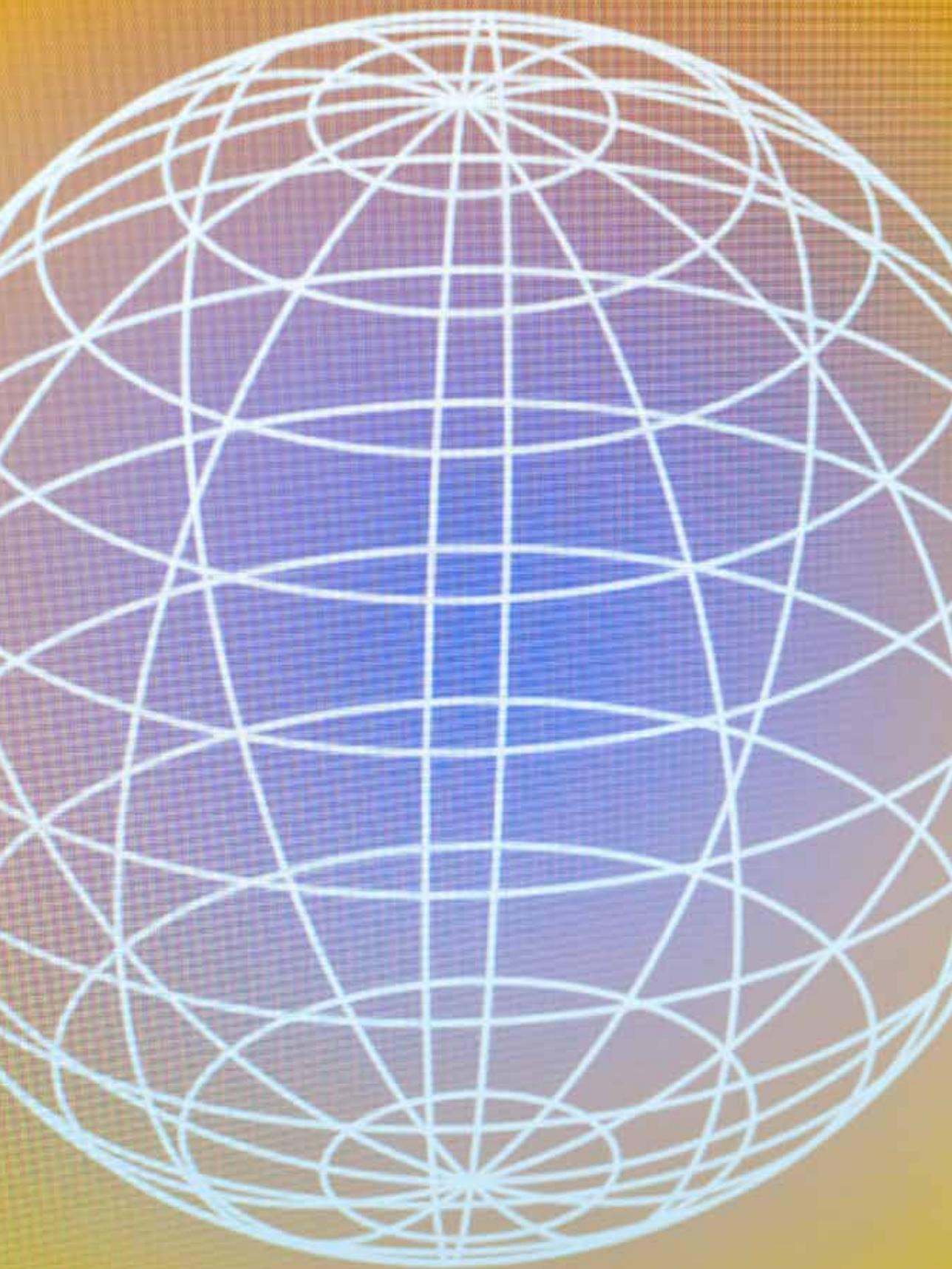
- ▶ Continuar informando sobre cuestiones de género relacionadas con aplicaciones e investigación en IA.

## Opciones para organizaciones intergubernamentales, incluida la UNESCO

- ▶ Continuar promoviendo a las mujeres en las carreras STEM.
- ▶ Convocar diálogos continuos sobre la IA y el género.
- ▶ Asegurar la diversidad de género, así como la inclusión de los grupos marginados, en los diálogos entre múltiples actores sobre cuestiones relativas a la inteligencia artificial.
- ▶ Elaborar marcos y metodologías para medir las aptitudes digitales y la brecha de aptitudes digitales.

INTELIGENCIA  
ARTIFICIAL EN  
ÁFRICA





# CAPÍTULO 6: INTELIGENCIA ARTIFICIAL EN ÁFRICA

África es una prioridad mundial de la UNESCO, con especial foco en la consolidación de la paz y el fortalecimiento de la capacidad institucional para alcanzar los Objetivos de Desarrollo Sostenible (ODS). La UNESCO reconoce que la tecnología puede contribuir al logro de esos objetivos en el continente africano. Su estrategia operacional para la Prioridad África incluye un programa emblemático sobre el "aprovechamiento de la ciencia, la tecnología, la innovación y del conocimiento para el desarrollo socioeconómico sostenible de África", y menciona la "creación de capacidad en el ámbito de las tecnologías de la información y las comunicaciones (TIC)" como una de sus principales actividades (UNESCO, 2014).

Se estima que actualmente África tiene una población de 1.100 millones de habitantes y, según las previsiones de las Naciones Unidas, aumentará a 2.400 millones, es decir, casi un tercio de la población mundial, para 2050. El rápido crecimiento demográfico significa que más del 60 % de la población africana tiene menos de 35 años, lo que hace que la fuerza de trabajo africana sea una de las más jóvenes en un mundo que envejece (División de Población de las Naciones Unidas, 2019).

Los adelantos de tecnologías digitales como la inteligencia artificial han abierto oportunidades para acelerar el desarrollo social y económico, pero, para lograrlo, África necesita aprovechar su dividendo demográfico, crear una infraestructura adecuada y hacer frente a los nuevos problemas normativos en la intersección de la tecnología y la sociedad. Por una parte, la IA en las economías desarrolladas puede dar lugar a una menor contratación de mano de obra en los países en desarrollo, con efectos adversos para el empleo en África (Nayyar, 2019). Por otra parte, los países africanos corren el riesgo de quedar relegados a los modelos coloniales, al ser proveedores de materias primas, en estos casos, de datos y de personal calificado (fuga de cerebros), y permanecer como importadores totalmente dependientes de la tecnología y los servicios de AI. En cambio, podrían pasar a ser activos en todas las etapas de la cadena de valor de la IA, lo que es necesario para que las tecnologías en cuestión sean de beneficio óptimo para el continente.

Reconociendo este tipo de desafíos, la Unión Africana ha desarrollado varios marcos continentales para lograr la Agenda 2063: el África que Queremos, y en particular mediante la acción estratégica coordinada panafricana en las áreas de agricultura, infraestructura, minería, ciencia, tecnología e innovación, comercio y desarrollo industrial (Unión Africana, 2015).

En 2018, en la Declaración Final del Foro de la UNESCO sobre Inteligencia Artificial en África también se reconoció "el rápido crecimiento de la población de África, así como las oportunidades y los desafíos que ello plantea en materia de educación, capacitación y posibilidades de empleo de la juventud africana" y las posibilidades que ofrece la IA para "el desarrollo sostenible e inclusivo del continente". Los participantes expresaron su preocupación con respecto a "las desigualdades persistentes y las importantes disparidades en la disponibilidad de los recursos, las capacidades e infraestructuras necesarias para dar acceso a los resultados de la innovación científica y beneficiarse plenamente de ellos" (UNESCO, 2018b).

En el presente capítulo se reconoce que, si bien África es heterogénea, varios países de la región experimentan muchos problemas fundamentales comunes, como los conflictos y la violencia internos, la escasa observación de los derechos humanos, la inestabilidad de las instituciones y la falta de infraestructura, financiamiento sostenible y capacidad institucional (Besaw y Filitiz, 2019). Otros países tienen obstáculos menos abrumadores en lo que respecta a la adopción e integración de IA en sus vías de desarrollo.

En este contexto heterogéneo, la Sección 1 comienza con un análisis de los desafíos para África en la esfera más amplia de la ciencia, la tecnología y la innovación (CTI). Se examinan algunas tendencias y antecedentes importantes en materia de CTI y luego se destacan los objetivos estratégicos y las áreas prioritarias planteadas por la Unión Africana en cumplimiento de la Agenda 2063: el África que Queremos.

En la sección 2 se ofrece un panorama general de las iniciativas anunciadas, iniciadas o aplicadas por varios gobiernos africanos para crear un entorno propicio para el aprovechamiento de la IA con fines de desarrollo sostenible. Se muestra que, si bien algunos gobiernos son proactivos, muchos otros aún no han esbozado estrategias claras para el aprovechamiento de la IA.

Por último, en la sección 3 se examinan algunas de las iniciativas adoptadas por el sector privado, las universidades, la comunidad técnica y la sociedad civil para el fortalecimiento de la IA en África de una vez por todas. Los esfuerzos de estos actores frente a muchos desafíos están ayudando al continente a desarrollar su ecosistema de IA a través de una mejor educación, conocimiento, capacitación y oportunidades para adquirir destrezas y conectarse.

# 1. LOS DESAFÍOS DE CIENCIA, TECNOLOGÍA E INNOVACIÓN EN ÁFRICA

La Comisión de la Unión Africana (CUA) ha establecido una Conferencia de Ministros encargados de la Ciencia y la Tecnología (AMCOST) para que la Unión pueda debatir y elaborar respuestas colectivas a problemas que se presentan en el ámbito de la ciencia y la tecnología. El Plan de Acción Consolidado en materia de ciencia, tecnología e innovación se presentó a los Jefes de Estado y de Gobierno en 2005 y los Jefes de Estado africanos lo refrendaron para su aplicación en 2006. En la Estrategia de Ciencia, Tecnología e Innovación para África 2024 (STISA-2024) se considera que el fomento de la capacidad de CTI, la producción de conocimientos y la innovación tecnológica son fundamentales para el desarrollo social y económico de África, como parte de la Agenda 2063 de la Unión Africana (Unión Africana, 2014b). Sin embargo, la mayoría de los países africanos no han cumplido el objetivo de la Unión Africana de dedicar el 1 % del PIB a investigación y desarrollo (UNESCO, 2015c). Además, según el Informe de la UNESCO sobre la Ciencia 2015, los distintos países tienen un compromiso político de nivel relativamente bajo con las áreas de CTI (UNESCO, 2015c).

La STISA-2024 reconoce que la ciencia, la tecnología y la innovación enfrentan los siguientes desafíos en África (Unión Africana, 2014b):

- i) Insuficiente financiamiento para CTI, ya que solo la mitad de las inversiones en I+D proceden de África
- ii) Escasa capacidad organizativa para la elaboración de políticas de CTI y falta de elaboración de políticas basadas en pruebas debido a la escasa capacitación y formación del personal, el acceso limitado a los datos y los conocimientos sobre los avances en IA
- iii) Diferentes niveles de preparación de la infraestructura en toda África que podrían apoyar la innovación
- iv) Si bien en todo el continente las organizaciones de la sociedad civil se están comprometiendo con la IA, sus aportes a los debates sobre políticas de CTI no suelen estar respaldados por pruebas o investigaciones exhaustivas
- v) La cooperación bilateral y multilateral está mejorando, pero a menudo no promueve adecuadamente el sentido de identificación, la rendición de cuentas y la sostenibilidad de África

Informe de la UNESCO sobre la Ciencia 2015 también destaca que el compromiso con CTI varía mucho de un país a otro, y en algunos casos hay falta de:

- i) Estrategias o políticas nacionales de investigación e innovación con una clara definición de objetivos medibles y el papel que debe desempeñar cada interesado
- ii) Participación del sector privado en el proceso de definición de las necesidades, prioridades y programas nacionales de investigación
- iii) Instituciones dedicadas a la innovación que puedan establecer el vínculo entre la investigación y el desarrollo (UNESCO, 2015c).

Los objetivos estratégicos del Programa de Ciencia, Tecnología e Innovación de la Unión Africana incluyen el aumento del uso de la ciencia y la tecnología para abordar las esferas prioritarias de la Unión Africana<sup>1</sup>, el mejoramiento de las competencias técnicas y la capacidad institucional para el desarrollo de CTI, el fomento de la competitividad económica, la protección de la producción de conocimientos y el fortalecimiento de los derechos de propiedad intelectual, así como la facilitación de las reformas de las políticas de CTI, la armonización y la movilización de recursos.

Todavía no se dispone de los informes de evaluación de la iniciativa STISA-2024, ya que en la estrategia original se preveía una evaluación solo en 2024. Sin embargo, los interesados proponen una evaluación de mitad de período y la elaboración de un marco de vigilancia y evaluación basado en un enfoque de gestión basada en los resultados (Daniels, Mawoko y Konte, 2018).

Mientras tanto, como ha señalado la profesora Sarah Anyang Agbor, Comisionada para Recursos Humanos, Ciencia y Tecnología de la Comisión de la Unión Africana: "el desarrollo y la utilización de la inteligencia artificial" deben estar "respaldados por un entorno normativo propicio, con instrumentos y reglamentos adecuados que nos permitan cosechar sus beneficios de manera segura, equitativa y sostenible" (Agbor, 2019). Reveló que se está elaborando una estrategia amplia de transformación digital para África, centrada especialmente en la cooperación Sur-Sur y Norte-Sur. En este contexto, es importante que los interesados colaboren para crear en la región africana un entorno propicio en el que se puedan alcanzar los objetivos estratégicos de CTI antes mencionados.

---

1 Las prioridades de la iniciativa STISA-2024 de la Unión Africana incluyen: i) Erradicar el hambre y garantizar la seguridad alimentaria y nutricional; ii) Prevenir y controlar las enfermedades y garantizar el bienestar; iii) Comunicación (física e intelectual, movilidad); iv) Proteger el espacio (clima, biodiversidad, espacio, marino y submarino); v) Convivir y cohesionar la sociedad (panafricanismo e integración regional, gobernanza), y vi) Crear riqueza (educación y desarrollo de los recursos humanos, recursos minerales e hídricos).

## 2. INICIATIVAS DE LOS GOBIERNOS DE ÁFRICA EN RELACIÓN CON LA IA

El Índice de Preparación del Gobierno para la Inteligencia Artificial de 2019 sitúa a 12 países africanos entre los 100 primeros y a ninguno entre los 50 primeros en lo que respecta a la preparación del gobierno para utilizar la IA en el marco de cuatro amplios grupos que incluyen: gobernanza, infraestructura y datos, aptitudes y educación, y servicios públicos (Miller y Stirling, 2019). Los cinco gobiernos africanos mejor situados según el índice son: Kenya, Túnez, Mauricio, Sudáfrica y Ghana<sup>2</sup>.

A pesar de una serie de desafíos como los que se han examinado anteriormente, varios gobiernos han adoptado iniciativas que tienen importancia para el desarrollo de la IA, por ejemplo:

En 2018, el Gobierno de Kenya formó un grupo de trabajo sobre cadenas de bloques e IA que recientemente publicó un informe con recomendaciones sobre el uso de estas tecnologías para eliminar la corrupción, fortalecer la democracia, facilitar la inclusión financiera y mejorar la prestación de los servicios públicos, entre otras cosas (Ministerio de Información, Comunicaciones y Tecnología, 2019; Kenyan Wall Street, 2018). Además, el portal Kenya Open Data ofrece acceso digital gratuito a los conjuntos de datos del gobierno en formatos de fácil utilización para facilitar la rendición de cuentas del gobierno mediante la participación de los ciudadanos (Organismo de TIC de Kenya, 2019). Kenya también se propone utilizar la IA para evaluar la elegibilidad de los ciudadanos para acceder a viviendas asequibles.

Nigeria aprobó la creación de un organismo de robótica e IA para "aprovechar las colaboraciones con organismos internacionales de investigación sobre robótica e IA" y mejorar la educación y las aptitudes de los jóvenes mediante "la investigación y la enseñanza de tecnologías más complejas" (Goitom, 2019).

Sudáfrica ha establecido una Comisión de la Cuarta Revolución Industrial, presidida por su Presidente. Se espera que esta Comisión "recomiende políticas, estrategias y planes para posicionar a SA como un actor competitivo en el espacio digital" (Phakathi, 2019).

Mauricio puso en marcha su Estrategia de Inteligencia Artificial en 2018, en la que se recomendaba la creación del Consejo de Inteligencia Artificial de Mauricio (MAIC) y se pedía al Gobierno que garantizara un entorno propicio "mediante un entorno normativo, ético y de protección de datos robusto y favorable a la vez, y también mediante incentivos atractivos como subvenciones de contrapartida, créditos fiscales

---

2 El Índice de Preparación del Gobierno 2019 reconoce que el indicador no es perfecto y también que el bajo puntaje de algunos países puede no captar algunas iniciativas recientes en materia de IA.

y otros incentivos fiscales, subvenciones de capacitación para inversiones en IA y otras tecnologías emergentes» (Grupo de Trabajo sobre Inteligencia Artificial, 2018).

Malawi ha elaborado un Plan Nacional de TIC 2014-2031 en el que se prevé que el país sea una economía basada en el conocimiento y ha presentado el Proyecto Digital Malawi. Se da mucha importancia a la necesidad de desarrollar las TIC, dada la penetración relativamente baja de Internet en el país (Comisión de Asociaciones Público-Privadas, 2017).

Ghana, si bien no tiene una estrategia dedicada a IA, también ha mencionado a la IA como una esfera tecnológica estratégica en la que se centraría en su política de innovación científica y tecnológica (Oficina de Comunicaciones de la Presidencia de la República de Ghana, 2019).

En otra parte del continente<sup>3</sup>, el Organismo Nacional de Promoción de la Investigación Científica de Túnez inició en 2018 un proceso de elaboración de una política nacional de IA en un taller organizado por la Cátedra UNESCO de Política de Ciencia, Tecnología e Innovación (ANPR, 2018).

Marruecos realizó junto con la UNESCO el Foro sobre Inteligencia Artificial en África de 2018 y ha asignado 50 millones de dólares al financiamiento de proyectos de investigación relacionados con la inteligencia artificial en 11 áreas temáticas (Zerrour, 2019; UNESCO, 2018c).

Además, 14 de los 55 países africanos han firmado el Convenio de la Unión Africana sobre la seguridad cibernética y la protección de los datos personales, en el que se establecen las normas de seguridad esenciales para establecer un espacio digital digno de crédito para las transacciones electrónicas, la protección de los datos personales y la lucha contra los delitos informáticos (Unión Africana, 2014a).

Varios otros países han expresado sentimientos positivos respecto de la IA y el aprovechamiento de la tecnología para el desarrollo sostenible.

Por ejemplo, el Ministro de Industrialización, Comercio y Desarrollo de Pymes de Namibia, Tjekero Tweya, expresó que Namibia debía participar activamente en la Cuarta Revolución Industrial (NAMPA, 2019).

Es necesario realizar un estudio sistemático para entender las necesidades de los gobiernos y las sociedades africanas para responder a los desafíos de la IA y otras tecnologías digitales. Esa evaluación, en particular utilizando el marco DAAM-X de la UNESCO, podría ayudar a catalizar y fundamentar la formulación de políticas basadas en pruebas, una necesidad que se ha puesto de relieve en la iniciativa STISA-2024 de la Unión Africana (Unión Africana, 2014b).

---

3 África septentrional entra dentro de la región de los Estados Árabes en la clasificación de la UNESCO.

### 3. INICIATIVAS DEL SECTOR PRIVADO, LA COMUNIDAD TÉCNICA Y LA SOCIEDAD CIVIL PARA LA IA EN ÁFRICA

Como se ha señalado, varios gobiernos de África están trabajando para crear un entorno propicio para la ciencia, la tecnología y la innovación, aunque el ritmo general de cambio se vea limitado por los problemas estructurales que se han detallado anteriormente. Sin embargo, en muchos países hay un sector pujante fuera del gobierno que está trabajando para fortalecer la investigación, los conocimientos, las aptitudes y las políticas de IA. Jóvenes empresarios, empresas tecnológicas mundiales, investigadores y estudiantes aprovechan los elementos de IA como una oportunidad de negocio y para el desarrollo de sus comunidades y países.

#### 3.1. Sector privado

Los agentes del sector privado están invirtiendo recursos financieros y humanos para aumentar la diversidad, fortalecer la investigación sobre la IA en África y elaborar soluciones que respondan a los problemas del continente. Google estableció su primer centro africano de investigación de IA en Accra (Ghana) en 2019, y ha expresado su compromiso de colaborar con las universidades y los centros de investigación locales (Crabtree, 2018). IBM tiene varios centros de investigación en Kenya y Sudáfrica, y ha invertido 70 millones de dólares para poner en marcha una plataforma de aprendizaje impulsada por Watson para 25 millones de jóvenes africanos. Esta plataforma en línea ofrecerá programas gratuitos de desarrollo de aptitudes en toda África (Jao, 2017). La empresa tunecina InstaDeep, que ofrece sistemas de adopción de decisiones con tecnología de IA para empresas, ya tiene oficinas en Londres, París, Túnez, Nairobi y Lagos. Por su parte, la empresa ugandesa GeoGecko ha trabajado con UNICEF para crear mapas sobre la infraestructura de protección social (InstaDeep, 2019; GeoGecko, n.d.).

En la Tabla 2 se presenta una lista no exhaustiva de iniciativas en África, en el ámbito de la salud, la agricultura, la tecnología y el transporte, que utilizan IA o sus elementos para brindar servicios a sus consumidores y comunidades.

Sector	Iniciativa	País
Salud	<p><b>Bot Sophie</b></p> <p>Un bot de <i>chat</i> con tecnología de IA diseñado para responder preguntas sobre salud sexual. Sus creadores buscaban resolver los problemas de acceso a información veraz en tiempo real y la incomodidad de hablar de temas de salud reproductiva entre los jóvenes de África. El bot Sophie aprende de las conversaciones con los usuarios para procesar y responder las preguntas (Mbaka, 2017).</p>	Kenya
	<p><b>Diagnóstico automatizado de malaria con microscopio digital</b></p> <p>El primer laboratorio de IA de Uganda, en la Universidad de Makerere, ha desarrollado una forma de diagnosticar muestras de sangre usando un teléfono celular. El programa aprende a generar sus propios criterios para las infecciones basándose en un conjunto de imágenes que le han sido presentadas previamente. Los tiempos de diagnóstico podrían reducirse de 30 minutos a tan solo 2 minutos (Lewton y McCool, 2018).</p>	Uganda
	<p><b>Numberboost</b></p> <p>Un sistema de IA que permite a los ciudadanos ubicar clínicas de salud móviles cercanas, lo que mejora el acceso a la atención sanitaria (NumberBoost, 2019).</p>	Sudáfrica
Agricultura	<p><b>Vital signs</b></p> <p>Utilizando mediciones sobre el terreno de diversos indicadores y fuentes de datos existentes, Vital Signs crea una imagen de la relación entre la agricultura, la naturaleza y el bienestar humano. Los Indicadores Clave de Vital Signs incluyen: producción agrícola sostenible, disponibilidad y calidad del agua, salud del suelo, biodiversidad, reservas de carbono, resiliencia al clima, ingresos familiares, nutrición y acceso a los mercados (Conservation International, 2018).</p>	Kenya
	<p><b>Arifu</b></p> <p>Arifu es una plataforma de bot de <i>chat</i> para aprender nuevas habilidades sobre varios temas, incluidos emprendimiento, gestión financiera y nutrición. Por ejemplo, puede ayudar a los agricultores a determinar qué fertilizante se ajusta a sus necesidades específicas. (Arifu, 2019).</p>	

Sector	Iniciativa	País
	<p><b>FarmDrive</b></p> <p>FarmDrive conecta a los pequeños agricultores con préstamos e instrumentos de gestión financiera, ofreciendo herramientas tales como mantenimiento de registros de gastos, ingresos y rendimientos; solicitud de préstamos; recepción de préstamos y recordatorios sobre su pago. En el caso de las instituciones financieras, FarmDrive puede utilizar sus datos sobre el tamaño de las tierras, la ubicación y los cultivos para determinar el riesgo y las tasas de interés correspondientes (Owino, 2019).</p>	Kenya
	<p><b>Zenvus</b></p> <p>Zenvus es una herramienta de toma de decisiones para los agricultores basada en datos recogidos por sensores y otros medios. Los servicios de Zenvus incluyen el registro de todas las fases de la agricultura, desde la plantación hasta la venta; la obtención de capital y financiamiento colectivo; el seguro de las explotaciones agrícolas; el suministro de precios de productos en tiempo real y una plataforma para vender productos (Zenvus, 2019).</p>	Nigeria
	<p><b>Aerobotics</b></p> <p>Aerobotics colabora con la industria agrícola usando imágenes aéreas obtenidas por <i>drone</i> para gestionar vergeles, identificar problemas en las producciones de cultivos y tratar pestes y enfermedades (Aerobotics, 2019).</p>	Sudáfrica
Tecnología financiera	<p><b>Tala</b></p> <p>Tala es un producto de crédito en línea que suscribe y desembolsa instantáneamente préstamos a personas sobre la base de los datos que estas introducen en la aplicación, incluidas las que no tienen un historial de crédito formal. El pago de los préstamos también se hace a través de la aplicación (Tala, 2019).</p>	Kenya
	<p><b>Kudi.ai</b></p> <p>Utilizando el procesamiento del lenguaje natural y la inteligencia artificial, Kudi.ai intenta facilitar el pago entre pares a los nigerianos mediante un bot de <i>chat</i> que funciona en aplicaciones de mensajería populares, como Facebook Messenger. Los usuarios de Kudi pueden transferirse dinero en efectivo entre sí, ayudar a otros a transferir dinero en efectivo y pagar sus facturas de televisión, Internet y electricidad (Kudi, 2019).</p>	Nigeria

Sector	Iniciativa	País
Transporte público	<b>RoadPreppers</b> RoadPreppers ayuda a los usuarios a sortear los atascos de tráfico dándoles direcciones alternativas de conducción y opciones de transporte público con cotizaciones de tarifas. Se especializa en regiones donde los datos sobre el transporte público son inaccesibles y no están estructurados, o simplemente no están disponibles, y donde los sistemas de transporte público son complicados (Eweniyi, 2017).	Nigeria
	<b>Lara.ng</b> Un bot de <i>chat</i> que ofrece indicaciones y tarifas de transporte público para quienes deben trasladarse en Lagos. Busca ser mejor que las opciones existentes, que no funcionan bien en ciudades donde la red de transporte está congestionada y es complicada (Eweniyi, 2017).	

Tabla 2: Iniciativas en África que usan IA en salud, agricultura, tecnología financiera y transporte

### 3.2. Universidades e institutos educativos

Se están realizando esfuerzos para fortalecer la educación mediante IA en África. Una de las iniciativas es el lanzamiento del programa de licenciatura de Maestría en Inteligencia Artificial de África (AMMI) en el Instituto Africano de Ciencias Matemáticas de Kigali (Rwanda). El programa, lanzado en asociación con Google y Facebook en 2019, tiene por objeto dar a conocer las investigaciones más avanzadas y ofrecer una enseñanza a cargo de expertos a estudiantes africanos en África<sup>4</sup>. La primera cohorte incluyó 30 estudiantes de 10 países africanos, el 43 % de los cuales eran mujeres (AIMS, 2018).

Otros ejemplos son la Escuela Euromed de Ingeniería Digital e IA en Marruecos, que abrirá sus puertas en septiembre de 2019. En Uganda ya hay un activo grupo de investigación en IA y ciencias de datos en la Universidad de Makakere (Lystad, 2019; Grupo de investigación en IA y ciencias de datos de la Universidad de Makerere, 2019). La Universidad de Namibia ha añadido un módulo de IA a su programa de licenciatura en informática (Universidad de Ciencia y Tecnología de Namibia, 2019).

Estos programas especializados en distintas universidades ayudarán a desarrollar la capacidad de IA en África, aprovechando los cursos fundamentales de informática que se imparten a nivel de licenciatura. Hay varias iniciativas más a nivel universitario y un estudio detallado de las necesidades a nivel universitario ayudaría a determinar

4 Facebook contribuye con cuatro millones de dólares al financiamiento y la contratación de profesores, al tiempo que Google también aporta recursos (AIMS, 2018).

las lagunas que es necesario llenar para mejorar la disponibilidad y el acceso a la educación de alta calidad sobre IA.

### 3.3. La sociedad civil y la comunidad técnica

Estos esfuerzos formales para fortalecer los conocimientos de IA en África se complementan con una serie de iniciativas para salvar la brecha de habilidades mediante el uso de modelos innovadores de enseñanza y aprendizaje. Por ejemplo, Tejumade Afonja cofundó Al Saturday Lagos, una comunidad de IA en Lagos, Nigeria. Mediante clases gratuitas de ciencia de datos, aprendizaje automático y aprendizaje profundo durante 16 sábados consecutivos, en forma de grupos de estudio estructurados, han capacitado a más de 150 personas en dos ciclos (Afonja, 2018).

Al Kenya es otra comunidad con unos 2.500 miembros de diferentes campos que ha tomado una iniciativa encaminada a reunir a expertos técnicos y no técnicos en el África oriental para elaborar soluciones a problemas locales aprovechando la IA (Al Kenya, 2019).

Data Science Nigeria gestiona un centro de IA en el que se imparten clases gratuitas de IA a un gran número de estudiantes que asisten a clases diarias en línea (Data Science Nigeria, 2019). Su libro electrónico gratuito proporciona 99 casos de uso de IA y simplifica la IA para los principiantes (Adekanmbi, 2018). Hay varias comunidades locales, que a menudo funcionan a través de plataformas de redes sociales que reúnen a personas con ideas afines para aprender y trabajar en soluciones utilizando la IA. Entre los ejemplos de esas iniciativas figuran 042 AI en Enugu y TensorFlow Lagos (TensorFlow Lagos, 2019). Estas comunidades no solo están abordando los desafíos del acceso al conocimiento y la capacitación en materia de IA, sino que también están fomentando un ecosistema en el que se dispone de mentores para orientar a los jóvenes.

#### Recuadro 22: Representación de la diáspora africana en la comunidad occidental de IA

A medida que la tecnología se vuelve cada vez más importante en nuestras vidas, hay una falta de diversidad en todo el mundo entre quienes estudian, investigan, enseñan o desarrollan soluciones en los campos de la ciencia, la tecnología, la ingeniería y las matemáticas (STEM). Los fundadores de la iniciativa "Black in AI" se dieron cuenta de las "crisis de diversidad" en la comunidad de la IA cuando notaron que solo había seis investigadores negros de un total estimado de 8,500 personas que participaron en la conferencia Sistemas de Procesamiento e Información Neuronal (NeurIPS) en 2016, una importante conferencia internacional para la comunidad de la IA. Aunque "Black in AI" no es una organización africana, su mensaje sobre la importancia de la diversidad y la representación de la población negra en la comunidad de la IA coincide también con la necesidad de una representación africana (Cisse, 2018).

Timnit Gebru, cofundadora de la iniciativa, subraya la importancia de la diversidad y cree que "si no tenemos diversidad en nuestro grupo de investigadores, no vamos a abordar los problemas a los que se enfrenta la mayoría de las personas en el mundo". Añade que cuando "los problemas no nos afectan, no creemos que sean tan importantes, y puede que ni siquiera sepamos cuáles son esos problemas, porque no estamos interactuando con las personas que los están sufriendo" (Snow, 2018a).©

En lo que respecta al acceso al conocimiento, en el capítulo sobre Acceso se examinó la forma en que las conferencias académicas son una vía importante para que los investigadores de IA se reúnan, presenten sus investigaciones y establezcan asociaciones y redes. El acceso a conferencias mundiales no ha sido fácil para los investigadores africanos por varias razones, como la denegación de visados para participar en esas conferencias en países desarrollados y los elevados costos asociados a los traslados y la estadía (Knight, 2018).

Un cambio de rumbo positivo en este sentido, que puede tener importancia a más largo plazo, es que la Conferencia Internacional sobre Representaciones de Aprendizaje (ICLR), un importante encuentro de la comunidad de la IA, se organizará en Addis Abeba en 2020.

Deep Learning Indaba es otra organización que está fortaleciendo la capacidad de África en materia de aprendizaje automático. Su objetivo es "construir comunidades, crear liderazgo y reconocer la excelencia en la investigación y la innovación en todo el continente" (Deep Learning Indaba, 2019). Por medio de sus conferencias anuales y las reuniones regionales Indaba, están impulsando la agenda de investigación y ofreciendo a los investigadores y estudiantes africanos una plataforma para que participen y colaboren. Dos resultados muy visibles de la conferencia de 2018 fueron la colaboración entre investigadores para desarrollar la traducción automática neuronal para los idiomas africanos y el desarrollo de una IA capaz de generar máscaras africanas (Wiggers, 2018; Abbott, 2018). La UNESCO organizó dos talleres sobre equidad e IA en Deep Learning Indaba 2019, con el objetivo de que la comunidad tecnológica y política de la IA en África se comprometiera a trabajar conjuntamente en las cuestiones de derechos humanos, apertura y acceso relacionadas con la IA (UNESCO, 2019f).

## Recuadro 23: IA para los idiomas africanos: reforzar el multilingüismo

África tiene más de 2.000 idiomas (Wolff, 2018). Dada esta diversidad lingüística, se plantean cuestiones relativas al acceso a la información en múltiples idiomas. Por ejemplo, hasta mediados de la década de 1990 se estimaba que el 80 % del contenido en línea estaba en inglés. En 2015, solo 10 idiomas constituían alrededor del 82 % del contenido de Internet (Young, 2015). Además, los datos indican que la eficiencia del aprendizaje y el desarrollo cognitivo de los niños es mucho mejor cuando su lengua materna es el medio de instrucción en las escuelas (Alidou y otros, 2006).

El proyecto IA y el procesamiento de lenguaje natural está avanzando hacia la traducción en vivo de varios idiomas. En África, este proyecto ha echado raíces y ahora se dispone de instrumentos para traducir miles de idiomas africanos. OBTranslate es una plataforma digital que traduce más de 2.000 idiomas, lo que permite una mejor comunicación entre los hablantes de distintos idiomas africanos (The Guardian, 2019). Esas herramientas también pueden utilizarse para traducir contenidos en línea, con lo que se fortalece el acceso a la información en la web.

El Grupo de Investigación sobre Tecnología del Lenguaje Humano del Consejo de Investigaciones Científicas e Industriales (CSIR) está estudiando las formas en que las tecnologías del habla y el lenguaje pueden utilizarse en beneficio de los sudafricanos. En particular, están estudiando el uso del reconocimiento automático del habla para apoyar el aprendizaje de idiomas y la traducción, una tarea importante dado que en Sudáfrica hay 11 idiomas oficiales, y solo el inglés y el afrikáans están bien representados en las tecnologías que existen en la actualidad (GOBL, 2014).

Todas estas iniciativas están estimulando un interés activo en el ecosistema de la IA de África y creando una comunidad de jóvenes interesados en la IA. En 2019 había grandes cantidades de aplicaciones de producción local basadas en elementos de la IA, que ayudan a los africanos en su trabajo y su vida cotidiana, incluidas aplicaciones agrícolas, aplicaciones educativas y aplicaciones relacionadas con la salud sexual (Le Monde, 2019; Mbaka, 2017; Halilou, 2016). Un ejemplo entre muchos otros es Lifantou, la plataforma de comercio electrónico creada por Awa Thiam, una ingeniera senegalesa de 28 años. Esta plataforma utiliza la inteligencia artificial para vincular los comedores escolares, que necesitan ingredientes de bajo costo, con las cooperativas agrícolas que desean evitar el pago a los intermediarios (Le Monde, 2019). Aunque el desarrollo de la IA en África aún se encuentra en las primeras etapas, la tecnología y sus elementos han mostrado cierto potencial para permitir que varios jóvenes de África aprovechen la tecnología para resolver algunos de los problemas del continente.

## 4. CONCLUSIÓN Y OPCIONES DE POLÍTICA

El potencial de África para transformar varios países en centros de innovación puede materializarse si se cuenta con el entorno y las oportunidades adecuadas. En este capítulo se presentaron los esfuerzos por aprovechar el potencial de la IA para el desarrollo del continente en el contexto de varios problemas estructurales y fundacionales a los que se enfrentan varios países africanos. La importancia de la ciencia, la tecnología y la innovación es ampliamente reconocida por los países africanos y forma parte esencial de la visión hasta 2063 de la Unión Africana. Sin embargo, hay grandes problemas de capacidad, infraestructura y gobernanza para la creación de un entorno sólido y propicio para el desarrollo de la IA.

Muchos gobiernos africanos son conscientes de estos problemas y están poniendo en marcha distintas iniciativas, algunas a través de políticas específicas de IA que ayudan a empoderar al sector privado, los investigadores y la sociedad civil a fin de que puedan aprovechar la IA para el desarrollo sostenible. Los esfuerzos que han comenzado a realizar los gobiernos se complementan con el sector privado, la comunidad técnica y la sociedad civil, que tratan de hacer frente a los problemas inmediatos de acceso a los conocimientos, las aptitudes, la tutoría y las oportunidades comerciales.

A continuación se presentan algunas opciones para la acción a fin de fortalecer el ecosistema de ciencia, tecnología e innovación necesario para el desarrollo de la IA en África:

### Opciones para los Estados

- ▶ Elaborar políticas y estrategias nacionales de IA en consonancia con los principios DAAM de la UNESCO, que vayan acompañadas de un plan de aplicación, mecanismos de financiamiento y procesos de monitoreo y evaluación.
- ▶ Invertir más en la enseñanza de las ciencias y la ingeniería con perspectiva de género, a fin de entrenar la mano de obra cualificada necesaria para el desarrollo de la IA.
- ▶ Intensificar los esfuerzos por alcanzar el objetivo nacional de invertir el 1 % del PIB en investigación y desarrollo (I+D).

- ▶ Alentar al sector empresarial a participar más activamente en la investigación y el desarrollo, a fin de estimular la demanda de producción de conocimientos y desarrollo tecnológico.
- ▶ Crear fondos nacionales para ayudar a los innovadores locales a proteger sus derechos de propiedad intelectual.
- ▶ Facilitar la colaboración entre el sector privado y las universidades con medidas que incluyan disposiciones para que representantes del sector privado formen parte de los consejos de administración de las universidades e institutos de investigación, incentivos fiscales para apoyar la innovación empresarial, la creación de parques científicos y tecnológicos e incubadoras de empresas para fomentar las empresas emergentes y las asociaciones público-privadas, y subvenciones para la investigación a fin de apoyar la colaboración en materia de investigación entre el gobierno, la industria y el mundo académico en esferas prioritarias.
- ▶ Fomentar los intercambios y la colaboración intrarregional y panafricana entre investigadores, y crear incentivos para contrarrestar la fuga de cerebros.

## Opciones para el sector privado, los intermediarios de Internet y la comunidad técnica

- ▶ Apoyar a las universidades y centros de investigación con proyectos de colaboración y acceso compartido a datos, *hardware* y conocimientos a fin de fortalecer la investigación y el desarrollo en IA.
- ▶ Desarrollar herramientas y módulos de conocimiento para la educación y la formación en IA a fin de formar a los estudiantes.
- ▶ Crear tecnologías de IA para resolver cuestiones relacionadas con la salud, la agricultura, las finanzas, el transporte, etc.
- ▶ Proporcionar oportunidades para el reciclaje de la fuerza de trabajo existente para aplicaciones avanzadas basadas en IA en las empresas y la sociedad.

## Opciones para el sector académico

- ▶ Colaborar con investigadores de todo el continente africano y de otras regiones para realizar investigaciones fundamentales y aplicadas de punta en IA.
- ▶ Actualizar los planes de estudio de las universidades y escuelas con los conocimientos más recientes y pertinentes que permitan a los

estudiantes aprovechar las oportunidades de carrera en IA y en otros campos relacionados con la informática.

- ▶ Fomentar un entorno de colaboración entre el mundo académico y el sector privado a fin de preparar a los estudiantes para que puedan trabajar y realizar investigaciones que puedan comercializarse.

## Opciones para la sociedad civil

- ▶ Fortalecer capacidades en materia de formulación de políticas de ciencia, tecnología e innovación de base empírica en África.
- ▶ Fortalecer la cooperación entre la sociedad civil y los institutos de investigación para solucionar los problemas que enfrentan las comunidades locales, para la elaboración de modelos novedosos de recopilación de datos basados en la ciencia ciudadana que puedan crear conjuntos de datos para la IA respetando las normas internacionales de privacidad y protección de datos.
- ▶ Participar activamente en los diálogos de política a fin de acercar las preocupaciones de los ciudadanos a los encargados de formular políticas, pero también para respaldar la formulación de políticas de la IA y garantizar que el uso de la IA no infrinja los derechos humanos, incluidos los derechos a la expresión y el acceso a la información, la privacidad, la igualdad y la participación en la vida pública.

## Opciones para organizaciones intergubernamentales, incluida la UNESCO

- ▶ Colaborar y asociarse con la Unión Africana en su estrategia de transformación digital, asociarse con organizaciones regionales y nacionales de formulación de políticas para la elaboración de políticas de IA basadas en pruebas.
- ▶ Apoyar el desarrollo de las capacidades aguas arriba y aguas abajo para hacer frente a los problemas de la inteligencia artificial y otras TIC avanzadas mediante capacitaciones, talleres y programas de apoyo institucional a largo plazo.
- ▶ Facilitar el intercambio de conocimientos Norte-Sur y Sur-Sur para tener un mayor acceso a la investigación y los conocimientos entre las diferentes regiones.

**CAPÍTULO 7:  
IMPLICANCIAS  
PARA LA UNESCO  
Y OPCIONES  
GENERALES PARA LA  
ACCIÓN**

El desarrollo de la inteligencia artificial como paquete de tecnologías está inextricablemente vinculado a su evolución dentro del ecosistema más amplio de Internet y las TIC y las fuerzas que han dado forma a este contexto. Esto pone de relieve la pertinencia del enfoque de la UNESCO respecto de Internet y la importancia de considerar a la IA en términos de una posible alineación con el concepto de universalidad de Internet de la Organización.

Por consiguiente, la posición de la UNESCO sobre la IA puede enmarcarse adecuadamente con los principios DAAM de la universalidad de Internet, que, a su vez, abogarían por que la IA se desarrolle en términos de derechos humanos, apertura, accesibilidad y participación de múltiples actores. Este enfoque puede servir de marco bien fundamentado y holístico para la UNESCO y las partes interesadas, lo que puede ayudar a informar y configurar el diseño, la aplicación, la supervisión y la gobernanza de la IA. Puede alimentar actividades para el establecimiento de determinados principios normativos y éticos para la IA, produciendo innovadoras directrices e instrumentos de política, y para el desarrollo de indicadores específicos de la IA que sean pertinentes para las distintas esferas del mandato de la UNESCO, incluidas, entre otras, la comunicación y la información.

Con este enfoque respecto de la IA, la UNESCO está bien situada para proporcionar asesoramiento técnico y de política, servir de centro de intercambio de información y crear capacidad. De este modo, la ética de la IA, como se informa en los principios DAAM, puede contribuir al beneficio de la humanidad, el desarrollo sostenible y la paz.

Esto significaría que la UNESCO aportaría un enfoque distinto dentro del ecosistema de otros actores con intereses en el campo de la IA. Con acceso a los titulares de deberes (funcionarios gubernamentales, representantes elegidos, reguladores independientes, propietarios y dirigentes de los medios de comunicación, organizaciones no gubernamentales especializadas pertinentes, etc.) y credibilidad entre los titulares de derechos (ciudadanos, periodistas, académicos, sector privado, etc.), la UNESCO puede lograr un cambio positivo para el desarrollo y la utilización de la IA.

Un ejemplo es que la conmemoración de días internacionales como el Día Mundial de la Libertad de Prensa y el Día Internacional del Acceso Universal a la Información permite a la Organización integrar las cuestiones relativas a la IA en temas vinculados a la libertad de prensa, la discapacidad y el acceso universal a la información (UNESCO, 2019b; UNESCO, 2019f). En su otra labor, en materia de educación, cultura y ciencias, también la investigación, la creación de capacidad, la sensibilización, la promoción, el asesoramiento técnico y el intercambio de buenas prácticas y experiencias internacionales pueden aplicarse en relación con la IA. De esta manera, la UNESCO puede ser una parte eficaz de la labor más amplia que se realiza en todo el mundo para configurar el diseño, la aplicación y la gobernanza de la IA.

Los resultados de esa labor contribuirán al establecimiento y la aplicación de derechos, normas y estándares sobre las dimensiones éticas y de derechos humanos de la IA, y en consonancia con los principios de apertura, accesibilidad y gobernanza de múltiples actores. De esta manera, la IA puede aprovecharse para alcanzar los

ODS, y sobre todo la meta 10 del Objetivo 16 sobre "acceso público a la información y libertades fundamentales". Entre los productos que sustentan esos resultados figuran no solo soluciones innovadoras producidas para orientar la IA, sino también la capacidad que se está creando especialmente en África y en favor de la igualdad de género.

## Opciones globales para la acción

### Todos los interesados pueden considerar abordar la IA a través de los siguientes enfoques:

- Aplicar los principios de la universalidad de Internet de la UNESCO (derechos humanos, apertura, accesibilidad y participación de múltiples actores) y desarrollar indicadores adaptados sobre IA, con el fin de investigar, mapear y mejorar el ecosistema en el que se desarrolla, aplica y gobierna la IA.
- Participar en investigaciones interdisciplinarias sobre la forma en que la IA se cruza con los derechos humanos, la apertura, la accesibilidad y la gobernanza de múltiples actores. Promover la ética por diseño en el desarrollo de IA y aplicar normas y estándares de derechos humanos que puedan informar el surgimiento de directrices más específicas sobre el derecho a la libertad de expresión y el acceso a la información, el derecho a la privacidad, el derecho a la igualdad y el derecho a la participación en la vida pública.
- Facilitar la formulación de marcos éticos internacionales basados en los derechos humanos y soluciones de política global, y facilitar la participación inclusiva de los países desarrollados y en desarrollo.
- Realizar evaluaciones exhaustivas de los efectos del desarrollo de la IA en los derechos humanos.
- Reflexionar sobre las implicaciones de la IA en la práctica del periodismo y el desarrollo de los medios de comunicación, y alentar a los actores de los medios a investigar e informar sobre el desarrollo de la IA y sus aplicaciones, incluida la exposición de abusos y sesgos de la IA, así como los beneficios actuales.
- Evaluar la discriminación algorítmica a fin de proteger el derecho de todas las personas a la igualdad, en particular de las poblaciones históricamente marginadas.

- Concienciar sobre la propiedad y el acceso a macrodatos, las habilidades y tecnologías de IA, y la cuestión de quién se beneficia y quién se perjudica, así como sobre daños tales como la marginación o la manipulación de la agencia humana.
- Promover el acceso abierto a la investigación en IA y el desarrollo de modelos de publicación de datos que protejan contra la violación de los derechos humanos debido al mal uso de los conocimientos abiertamente disponibles sobre IA.
- Mantener la competencia en el mercado abierto para impedir la monopolización de la IA, y exigir salvaguardias adecuadas contra la violación de prácticas éticas por parte del mercado y otros agentes.
- Elaborar directrices éticas para el uso de la IA en toda la industria a fin de garantizar que la competencia en el mercado abierto en el desarrollo de aplicaciones basadas en AI no infrinja los derechos humanos.
- Facilitar el desarrollo de normas y políticas para mejorar la apertura y transparencia en los algoritmos de IA mediante:
  - ↳ Revelación de información *ex ante* sobre la intención y el propósito de los algoritmos de IA
  - ↳ Monitoreo de la toma de decisiones algorítmica *ex post* para asegurar su ajuste a la intención diseñada del algoritmo
  - ↳ Transparencia en cuanto a los datos utilizados para entrenar los algoritmos y los datos utilizados para el análisis predictivo, con disposiciones para que los usuarios busquen información sobre cómo se procesa su información
- Asegurarse de que los datos abiertos no comprometan la privacidad de las personas y que se ajusten a las leyes de protección de datos.
- Facilitar la reflexión abierta sobre la evolución de las tendencias de la investigación en IA mediante la recopilación de estadísticas actualizadas desglosadas por género y lugar geográfico en revistas y publicaciones de conferencias sobre IA.
- Trabajar para reducir las brechas digitales, incluidas las brechas de género, en el acceso a la IA, y establecer mecanismos de vigilancia independientes.
- Desarrollar mecanismos de vigilancia mediante la recopilación, triangulación y validación de datos desglosados por género y lugar geográfico, que puedan informar a los interesados sobre los efectos no deseados de la IA, así como sobre cuestiones como el estado de la brecha digital en el acceso a IA (incluye STEM y CTI)

- Alentar una participación más activa en la gobernanza de IA, desde los principios hasta las normas, según corresponda, y según las diferentes funciones y obligaciones de los grupos interesados, incluidos, entre otros, el gobierno, el sector privado, la comunidad técnica, la sociedad civil, el sector académico, las organizaciones internacionales y los medios de comunicación.
- Asegurar la igualdad de género, la diversidad lingüística y regional, así como la inclusión de los jóvenes y los grupos marginados en los diálogos entre múltiples actores sobre cuestiones relacionadas con la IA.
- Garantizar la transparencia, la inclusión y la rendición de cuentas del proceso de participación, con la participación de los interesados en pie de igualdad.
- Trabajar con la UNESCO para integrar el debate de las cuestiones relativas a la IA en eventos pertinentes, como los días internacionales de la libertad de prensa, la discapacidad y el acceso universal a la información, y acercarse a redes de la UNESCO, como UNITWIN, Orbicom, GAPMIL y GAMAG, así como a los institutos y centros de categoría 2, las ONG, los comités nacionales del PIPT y las comisiones nacionales de la UNESCO.
- Prestar especial atención a la interfaz entre las cuestiones relacionadas con la IA, la igualdad de género y el desarrollo sostenible en África.
- Crear capacidad para el desarrollo y la aplicación de la IA que funcione para promover los Objetivos de Desarrollo Sostenible universales.

# BIBLIOGRAFÍA

- Abbott, J. (2018, December 17). *The Journey to NeurIPS*. Retrieved October 18, 2019, from Towards Data Science: <https://towardsdatascience.com/the-journey-to-neurips-eeiaig7da538>
- Academie Nationale des Sciences et Techniques du Senegal (2019, June 28). *Celebration de la Journee de la Renaissance Scientifique de l'Afrique URSA 2019*. Retrieved October 17, 2019, from Academie Nationale des Sciences et Techniques du Senegal <https://www.ansts.sn/celebration-de-la-journee-de-la-renaissance-scientifique-de-lafrique-jrsa/>
- Access Now. (2018). *Human Rights in the Age of Artificial Intelligence*. New York: Access Now.
- Adekanmbi, O. (2018). *Artificial Intelligence Simplified: 99 Use Cases and Expert Thoughts*. Lagos: Data Scientists Network Foundation.
- Adshade, M. (2018, August 14). *How Sex Robots Could Revolutionize Marriage—for the Better*. Retrieved October 18, 2019, from Marina Adshade: <http://marinaadshade.com/2018/08/14/how-sex-robots-could-revolutionize-marriage-for-the-better/>
- Aerobotics. (2019). *Defenders of The Tree Crop*. Retrieved October 18, 2019, from Aerobotics: <https://www.aerobotics.com/?identifier=default-get-in-touch-button>
- Afonja, T. (2018, December 30). *Bridging The Artificial Intelligence (AI) Gaps With ALSaturdays*. Retrieved October 18, 2019, from Medium: <https://medium.com/ai-saturdays/bridging-the-artificial-intelligence-ai-gaps-with-ai6-9a5cfobg9of8>
- African Union. (2014a). *African Union Convention on Cyber Security and Personal Data Protection*. Addis Ababa: African Union.
- African Union. (2014b). *Science, Technology and Innovation Strategy for Africa 2024*. Addis Ababa: African Union.
- African Union. (2015). *Continental Frameworks*. Retrieved October 17 2019, from African Union Agenda 2063: <https://au.int/en/agenda2063/continental-frameworks>
- Agbor, S. A. (2019, August 28). AUC Commissioner for HRST on Artificial Intelligence (AI). (T. 2019, Interviewer)
- Agre, P. E., & Rotenberg, M. (1988). *Technology and Privacy: The New Landscape*. Cambridge: MIT Press.
- AI & Data Science Research Group at Makerere University. (2019). *About the AI & Data Science Research Group*. Retrieved October 18, 2019, from AI Research Artificial Intelligence and Data Science: <http://www.air.ug/>

- AI Kenya. (2019). *Who We Are*. Retrieved October 18, 2019, from AI Kenya: <https://kenya.ai>
- AIMS. (2018, July 31). *AIMS launches African Master's in Machine Intelligence*. Retrieved October 18, 2019, from AIMS: <https://www.nexteinsteinstem.org/blog/2018/07/31/aims-launches-first-of-its-kind-african-masters-in-machine-intelligence-at-rwanda-campus/>
- Algorithm Watch. (2019). *Automating Society Taking Stock of Automated Decision-Making in the EU*. Berlin: Algorithm Watch.
- Algorithmia. (2018, March 6). *Hardware for Machine Learning*. Retrieved October 18, 2019, from Algorithmia: <https://blog.algorithmia.com/hardware-for-machine-learning>
- Alidou, H, Boly, A., Brock-Utne, B., Diallo, Y. S., Heugh, K., & Wolff, H. E. (2006). *Optimizing Learning and Education in Africa – the Language Factor*. Tunis Belvedere: UNESCO Institute for Lifelong Learning and the Association for the Development of Education in Africa.
- Allcott, H., & Gentzkow, M. (2017). *Social Media and Fake News in The 2016 Election*. *Journal of Economic Perspectives*, 31(2), 211-236.
- Altman, A. (2016). Discrimination. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Amazon. (2019). *Alexa and Alexa Device FAQs*. Retrieved October 18, 2019, from Amazon: <https://www.amazon.com/gp/help/customer/display.html?no-deld=20i602230&pop-up=i>
- Andersen, L. (2018). *Human Rights in the Age of Artificial Intelligence*. New York: AccessNow.
- Angwin, J. (2010, July 30). The Web's New Gold Mine: Your Secrets. Retrieved October 18, 2019, from The Wall Street Journal: <https://www.wsj.com/articles/SB10001424052748703940904575395073512989404>
- Angwin, J. (2017, November 9). *Cheap Tricks: The Low Cost of Internet Harassment*. Retrieved from ProPublica: <https://www.propublica.org/article/cheap-tricks-the-low-cost-of-internet-harassment>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L (2016, May 23). *Machine Bias*. Retrieved October 18, 2019, from ProPublica: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- ANPR. (2018, April 20). *National AI Strategy: Unlocking Tunisia's capabilities potential*. Retrieved October 18, 2019, from Agence Nationale de la Promotion de la Recherche scientifique: <http://www.anpr.tn/national-ai-strategy-unlocking-tunisia-capabilities-potential/>
- Arifu. (2019). *About*. Retrieved October 18, 2019, from Arifu: <https://www.arifu.com/>
- ARTICLE 19 & Privacy International. (2018). *Privacy and Freedom of Expression In the Age of Artificial Intelligence*. London: ARTICLE 19 & Privacy International.

- ARTICLE 19. (2019). *The Social Media Councils: Consultation Paper*. London: ARTICLE 19.
- Article 29 Data Protection Working Party. (2014). *Opinion 05/2014 on Anonymisation Techniques*. Brussels: Article 29 Data Protection Working Party.
- Ashcraft, C, McLain, B., & Eger, E. (2016). *Women in Tech: The Facts*. Colorado: National Center for Women and Information Technology.
- Ayyub, R. (2018, November 21). / *Was The Victim Of A Deepfake Porn Plot Intended To Silence Me*. Retrieved October 19, 2019, from Huffington Post: [https://www.huffingtonpost.co.uk/entry/deepfake-porn\\_uk\\_5bf2ci26e4bof32bd58ba3i6'f'n-cid=fcbklnkcahpmgoooooioi&guccounter=i8\(guce\\_referer=aHRocHM6Ly93d-3cuZmFjZWJvb2suY2gtLw&guce\\_referrer\\_sig=AQAAAEuYmZBqcndRwiS-rh24iz2hjVR3gMBJf-LrpfgarwkiFS0BeFtsfR-JQxRlklPg](https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2ci26e4bof32bd58ba3i6'f'n-cid=fcbklnkcahpmgoooooioi&guccounter=i8(guce_referer=aHRocHM6Ly93d-3cuZmFjZWJvb2suY2gtLw&guce_referrer_sig=AQAAAEuYmZBqcndRwiS-rh24iz2hjVR3gMBJf-LrpfgarwkiFS0BeFtsfR-JQxRlklPg)
- Azoulay A. (2018, March). *Making The Most of Artificial Intelligence*. (J. Sopova, Interviewer) 2018-3 Issue of UNESCO Courier. Retrieved 16 November 2019 at <https://en.unesco.org/courier/2018-3/audrey-azoulay-making-most-artificial-intelligence>
- Baghai, K. (2012). *Privacy as A Human Right: A Sociological Theory*. *Sociology*, 46(5), 951-965.
- Balkin, J. M. (2017). *Free Speech in The Algorithmic Society: Big Data, Private Governance and New School Speech Regulation*. *UC Davis Law Review*, 51(3), 1149-1210.
- Barocas, S. (2014). *Data Mining and the Discourse on Discrimination*. *Proceedings of Data Ethics Workshop* (pp. 1-4). New Jersey: Github.
- Barocas, S., & Selbst, A. D. (2016, September 30). *Big Data's Disparate Impact*. *California Law Review*, pp. 671-732.
- Barraclough, T., & Barnes, C. (2019, May 22). *Don't Believe Everything You See, or Hear*. Retrieved October 18, 2019, from newroom.: <https://www.newsroom.co.nz/2019/05/22/597617/hold-dont-believe-everything-you-see-or-hear>
- Bathae, Y. (2018). *The Artificial Intelligence Black Box and the Failure of Intent and Causation*. *Harvard Journal of Law and Technology*, 31(2), 889-938.
- Bayefsky, A. F. (1990). *The Principle of Equality or Non-Discrimination in International Law*. *Human Rights Law Journal*, 11,1-34.
- BBC. (2017 September 11). *Row over AI that "identifies gay faces"*. Retrieved October 18, 2019, from BBC: <https://www.bbc.com/news/technology-41188560>
- Been, K., Wattenberg, M., Gilmer, J., Cai, C, Wexler, J., Viegas, E. & Sayres, R. (2018, June 7). *Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)*. Retrieved October 18, 2019, from Cornell University: <https://arxiv.org/pdf/1711.02790.pdf>

- Benay, A. (2018, October 10). *Using Artificial Intelligence in government means balancing innovation with the ethical and responsible use of emerging technologies* Retrieved October 18, 2019, from The Conversation: <https://tbs-blog.canada.ca/en/using-artificial-intelligence-government-means-balancing-innovation-ethical-and-responsible-use>
- Benjamin, G. (2019, February 6). *Deepfake videos could destroy trust in society – here's how to restore it*. Retrieved October 18, 2019, from The Conversation: <http://theconversation.com/deepfake-videos-could-destroy-trust-in-society-heres-how-to-restore-it-110999>
- Benkler, Y, Faris, R., Bourassa, N., & Roberts, H. (2018, July 12). *Understanding Media and Information Quality in an Age of Artificial Intelligence, Automation, Algorithms and Machine Learning*. Retrieved October 18, 2019, from Harvard: Berkman Klein Center: <https://cyber.harvard.edu/story/2018-07/understanding-media-and-information-quality-age-artificial-intelligence-automation>
- Bennett, C. J. (2018). *The European General Data Protection Regulation: An Instrument for the Globalization of Privacy Standards?* Information Polity, 23(2), 239-246.
- Bergen, H. (2016). *I'd Blush if I Could': Digital Assistants, Disembodied Cyborgs and The Problem of Gender*. *Word and Text, A Journal of Literary Studies and Linguistics*, 6, 95-113.
- Besaw, C, & Filitiz, J. (2019, January 16). *Artificial Intelligence in Africa is a Double-edged Sword*. Retrieved October 18, 2019, from Our World United Nations University: <https://ourworld.unu.edu/en/ai-in-africa-is-a-double-edged-sword>
- Bezemek, C. (2018, November 2). *The 'Filter Bubble' and Human Rights*. Retrieved October 21, 2019, from Fundamental Rights Protection Online: <https://ssrn.com/abstract=3277503>
- Bostrom, N. (2016). *Strategic Implications of Openness in AI Development*. Oxford, UK: Future of Humanity Institute, Oxford University.
- Boyd, d.(2008) *Taken Out of Context American Teen Sociality in Networked Publics*. Retrieved 16 November 2019 at the link: <https://www.danah.org/papers/Take-nOutOfContext.pdf>
- Boyd, C. (2017, November 8). *AI scientists: How can companies deal with the shortage of talent?* Retrieved October 16, 2019, from Towards Data Science: <https://towardsdatascience.com/ai-scientists-how-can-companies-deal-with-the-shortage-of-talent-nab48566677>
- Boyd, D., & Crawford, K (2012). *Critical Questions for Big Data, Information, Communication and Society: Provocations for a cultural, technological, and scholarly phenomenon*, 15(5), 662-679.
- Broadband Commission. (2018). *2018 State of Broadband Report: Broadband Catalyzing Sustainable Development*. Geneva: ITU & UNESCO.
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley P., Garfinkel, B           Amodei, D.

- (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Oxford: Future of Humanity Institute: University of Oxford.
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Alias, T., Dahlstrom, P. & Trench, M. (2017). *Artificial Intelligence: The Next Digital Frontier?* Brussels: Mckinsey Global Institute.
- Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*. Brussels: Mckinsey Global Institute.
- Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, Proceedings of Machine Learning Research: Conference on Fairness, Accountability, and Transparency* (pp. 1-15). New York: Proceedings of Machine Learning Research.
- Cairns, A. (2019, January 18). *Why AI Is Failing The Next Generation of Women*. Retrieved October 21, 2019, from World Economic Forum: <https://www.weforum.org/agenda/2019/01/ai-artificial-intelligence-failing-next-generation-women-bias/>
- Cannataci, J. A., Zhao, B., Torres Vives, G., Monteleone, S., Bonnici, J. M., & Moyakine, E. (2016). *Privacy, free expression and transparency: redefining their new boundaries in the digital age*. Paris: UNESCO.
- Carlson, M. (2014). *The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority*. *Digital Journalism*, 3(3), 416-431.
- Carpenter, J. (2015, July 6). *Google's Algorithm Shows Prestigious Job Ads to Men, but Not to Women. Here's Why That Should Worry You*. Retrieved October 21, 2019, from The Washington Post: <https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/>
- Carroll, S. (2009, March 30). *Why Can't We Visualize More Than Three Dimensions?* Retrieved October 21, 2019, from Discover Magazine: <http://blogs.discovermagazine.com/cosmicvariance/2009/03/30/why-cant-we-visualize-more-than-three-dimensions/#.Xai3bEYzaHs>
- Carter, J. M. (2018, August 1). *LGBOT Individuals Are Less Represented in Science – and That's A Problem*. Retrieved October 21, 2019, from ASBMB Today: <https://www.asbmb.org/asbmbtoday/201808/Essay/LGBTQ/>
- UN Chief Executives Board. CEB/2019/i/Add.3. (2019, June 17). *Summary of deliberations, A United Nations system-wide strategic approach and road map for supporting capacity development on artificial intelligence*. Geneva, Switzerland: United Nations System Chief Executives Board.
- Chatterjee, B. B. (2002). *Razorgirls and Cyberdykes: Tracing Cyberfeminism and Thoughts on Its Use in A Legal Context*. *International Journal of Sexuality and Gender Studies*, 2(3), 197-213.

- Chen, L. (2019, July 20). *China's Deepfake Celebrity Porn Culture Stirs Debate About Artificial Intelligence Use*. Retrieved October 21, 2019, from South China Morning Post: <https://www.scmp.com/news/china/society/article/3019389/chinas-deepfake-celebrity-porn-culture-stirs-debate-about>
- Chesney R., & Citron, D. K. (2018). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. *California Law Review*, 1-58.
- Cisse, M. (2018, October 23). *Look to Africa to advance artificial intelligence*. Retrieved October 18, 2019, from Nature: <https://www.nature.com/articles/d41586-018-07104-7>
- CITP & UCHV Princeton. (2018, October). *Dynamic Sound Identification*. Retrieved October 18, 2019, from Aland Ethics: <https://aiethics.princeton.edu/wp-content/uploads/sites/587/2018/10/Princeton-AI-Ethics-Case-Study-2.pdf>
- Citron, D. K. (2017). *Extremist Speech, Compelled Conformity and Censorship Creep*. *Notre Dame Law Review*, 93(3), 1035.
- Citron, D. K., & Jurecic, Q. (2018, September 5). *Platform Justice: Content Moderation at an Inflection Point*. Retrieved October 21, 2019, from Hoover Working Group on National Security, Technology, and Law: [https://www.hoover.org/sites/default/files/research/docs/citron-jurecic\\_webready.pdf](https://www.hoover.org/sites/default/files/research/docs/citron-jurecic_webready.pdf)
- Cockburn, I. M., Henderson, R., & Stern, S. (2018, March). *The Impact of Artificial Intelligence*. Retrieved October 21, 2019, from National Bureau of Economic Research: <https://www.nber.org/papers/w24449.pdf>
- Council of Europe. (2018). CM/Rec(2018)2. *Recommendation of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries*. Strasbourg: Council of Europe.
- Council of Europe. (2018). *Glossary*. Retrieved September 11, 2019, from Council of Europe Artificial Intelligence: <https://www.coe.int/en/web/artificial-intelligence/glossary>
- Council of Europe. CHR/Rec(2019)i. (2019, May). *Recommendation of the Commissioner for Human Rights on Unboxing Artificial Intelligence: 10 steps to protect Human Rights*. Retrieved September 4, 2019, from Council of Europe: <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-re-co/1680946e64>
- Council of Europe. CM/Rec(2012)3. (2012, April 4). *Recommendation of the Committee of Ministers to Member States on The Protection of Human Rights with Regard to Search Engines*. Strasbourg, France.
- Council of Europe. Decl(13/02/2019)1. (2019, February 13). *Declaration by The Committee of Ministers on The Manipulative Capabilities of Algorithmic Processes*. Strasbourg, France.
- Council of Europe. ETS No.108. (1981, January 28). Strasbourg, France.
- Cohen, J. E. (2012, November 5). *What Privacy is For*. Retrieved October 21, 2019, from Social Science Research Network: <https://www.ssrn.com/abstract=2175406>

- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). *Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data*. *Journal of Communication*, 64(2), 317-332.
- Collingridge, D. (1980). *The Social Control of Technology*. London: Frances Pinter (Publishers) Limited.
- Conner-Simons, A. (2018, October 4). *Detecting fake news at its source*. Retrieved October 18, 2019, from MIT Computer Science and Artificial Intelligence Lab: <https://www.csail.mit.edu/news/detecting-fake-news-its-source>
- Conservation International (2018, November). *Data for Sustainable Development in Kenya*. Retrieved October 18, 2019, from Vizzuality: <https://www.vizzuality.com/project/vital-signs-kenya/>
- Cook, J. (2019, June 23). *Here's What It's Like To See Yourself in A Deepfake Porn Video*. Retrieved October 21, 2019, from Huffington Post: [https://www.huffpost.com/entry/deepfake-porn-heres-what-its-like-to-see-yourself\\_n\\_5d0d0faee4b0a3941861fce-d?guccounter=i&guce\\_referrer=aHRocHM6Ly93d-3cuZ29vZ2xLLmNvbS88\(guce\\_re-ferrer\\_sig=AQAAABJdel2Ji3KP3PvAkKX-hAMZU3BdNPc\\_NAWhGdP2nQvWizElHR-tOsQCeCkIMhCNMbAV](https://www.huffpost.com/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b0a3941861fce-d?guccounter=i&guce_referrer=aHRocHM6Ly93d-3cuZ29vZ2xLLmNvbS88(guce_re-ferrer_sig=AQAAABJdel2Ji3KP3PvAkKX-hAMZU3BdNPc_NAWhGdP2nQvWizElHR-tOsQCeCkIMhCNMbAV)
- Corea, C, Duelli-Klein, R., Hanmer, J., Homes, H. B., Hoskins, B., Kishwar, M Stein-
- bacher, R. (1985). *Man-Made Women: How New Reproductive Technologies Affect Women*. London: Hutchinson.
- Crabtree, J. (2018, June 14). *Google's next AI, research center will be its first on the African continent*. Retrieved October 18, 2019, from CNBC Africa: <https://www.cnbc-africa.com/news/2018/06/16/googles-next-a-i-research-center-will-be-in-africa/>
- Crawford, K. (2013, May 10). *Think Again: Big Data*. Retrieved October 21, 2019, from Foreign Policy: <https://foreignpolicy.com/2013/05/10/think-again-big-data/>
- Crenshaw, K. (1989). *Demarginalizing The Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*. *University of Chicago Legal Forum*(i), 139-168.
- Crenshaw, K. (1991). *Mapping The Margins: Identity Politics, Intersectionality and Violence against Women*. *Stanford Law Review*, 43(6), 1241-1299.
- Cuboniks, L. (2014). *Xenofeminism: A Politics for Alienation*. Retrieved October 21, 2019, from XF Manifesto: <https://www.laboriacuboniks.net/>
- Culzac, N. (2014, September 17). *Egypt's Police 'Using Social Media and Apps Like Grindr to Trap Gay People'*. Retrieved October 18, 2019, from Independent: <https://www.independent.co.uk/news/world/africa/egypts-police-using-social-media-and-apps-like-grindr-to-trap-gay-people-9738515.html>
- da Costa, P. C. (2018). *Conversing with Personal Digital Assistants: on Gender and Artificial Intelligence*. *Journal of Science and Technology of The Arts*, 10(3), 2-59.

- Danaher, J., & Sandberg, A. (2017). *Robot Sex: Social and Ethical Implications*. Cambridge, Massachusetts: MIT Press.
- Daniels, C, Mawoko, P., & Konte, A. (2018). *Evaluating Public Policies in Africa: insights from the Science, Technology, and Innovation Strategy for Africa 2024 (STI-SA-2024)*, Sussex: University of Sussex.
- Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Retrieved October 18, 2019, from Reuters: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCNiMKo8G>
- Data Science Nigeria. (2019). *AI Bootcamp 2019 100 Days of Machine Learning & Deep Learning Classes*. Retrieved October 18, 2019, from YouTube: <https://www.youtube.com/playlist?list=PLomGkrTWmp4sT5XdhbhBkVX7r7WzlgHB>
- David Soergel A. S. (2013). *Open Scholarship and Peer Review: a Time for Experimentation*. ICML.
- Davies, H. (2018). *Redefining Filter Bubbles as (Escapable) Socio-Technical Recursion*. *Sociological Research Online*, 23(2), 637-654.
- de Zayas, A., & Martin, A. R. (2012). *Freedom of Opinion and Freedom of Expression: Some Reflections on General Comment No. 34 of The UN Human Rights Committee*. *Netherlands International Law Review*, 59(3), 425-454.
- Deep Learning Indaba. (2019). *Together We Build African Artificial Intelligence*. Cape Town: Deep Learning Indaba.
- Delaney, K. (2017 February 21). *Filter Bubbles Are A Serious Problem with News, Says Bill Gates*. Retrieved October 21, 2019, from Quartz: <https://qz.com/gj3ii4/bill-gates-says-filter-bubbles-are-a-serious-problem-with-news/>
- Devlin, K. (2015, September 17). *In Defence of Sex Machines: Why Trying to Ban Sex Robots Is Wrong*. Retrieved October 18, 2019, from The Conversation: <https://theconversation.com/in-defence-of-sex-machines-why-trying-to-ban-sex-robots-is-wrong-47641>
- Dialani, P. (2019, January 29). *Different Ways of How Twitter Uses Artificial Intelligence*. Retrieved October 21, 2019, from Analytics Insight: <https://www.analyticsinsight.net/different-ways-of-how-twitter-uses-artificial-intelligence/>
- Dorr, K. N. (2016). *Mapping The Field of Algorithmic Journalism*. *Digital Journalism*, 4(6), 700-722.
- Dorr, K. N., & Hollnbuchner, K. (2017). *Ethical Challenges of Algorithmic Journalism*. *Digital Journalism*, 5(4), 404-419.
- Duarte, N., Llanso, E., & Loup, A. (2017). *Mixed Messages?* Washington, D.C.: Center for Democracy and Technology.
- Dutton, T. (2018, June 28). *An Overview of National AI Strategies*. Retrieved October 21, 2019, from Medium: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

- Dyck, J. (2017 July 2018). *Siri Is Not Genderless*. Retrieved October 21, 2019, from Vice: [https://www.vice.com/en\\_us/article/a3dx34/siri-is-not-genderless](https://www.vice.com/en_us/article/a3dx34/siri-is-not-genderless)
- Egyptian Streets. (2019, September 9). *Egypt Organizes First International Conference on Artificial Intelligence*. Retrieved October 18, 2019, from Egyptian Street: <https://egyptianstreets.com/2019/09/09/egypt-organizes-first-international-conference-on-artificial-intelligence/>
- ELsevier. (2018). *Artificial Intelligence: How knowledge is created, transferred, and used*. Amsterdam: ELsevier.
- European Commission. (2017). *Building A European Data Economy*. Brussels: European Commission.
- European Commission. (2018). *2018 Reform of EU Data Protection Rules*. Retrieved October 18, 2019, from European Commission: [https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules\\_en](https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en)
- European Council CO EUR 8 CONCL 3. (2017, June 23). European Council meeting (22 and 23 June 2017) Conclusions. Brussels, Belgium.
- European Data Protection Supervisor. (2019). *Annual Report 2018*. Luxembourg: European Union .
- Eweniyi, O. (2017, March 10). *Three Guys Created An AI Solution To Nigeria's Public Transportation Issues*. Retrieved October 21, 2019, from Konbini: <https://www.konbini.com/ng/lifestyle/roadpreppers-artificial-intelligence-simplify-public-transportation/>
- Fang, A. (2019, April 2). *Chinese Colleges to Offer AI Major in Challenge to US*. Retrieved October 21, 2019, from Nikkei Asian Review: <https://asia.nikkei.com/Business/China-tech/Chinese-colleges-to-offer-AI-major-in-challenge-to-US>
- Fessler, L. (2017 February 22). *We Tested Bots Like Siri and Alexa to See Who Would Stand Up to Sexual Harassment* Retrieved October 21, 2019, from Quartz: <https://www.quartz.com/911681/we-tested-apples-siri-amazon-echos-alexa-micro-softs-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/>
- Fitzsimons, T. (2019, May 9). *Google, Trevor Project Will Use AI to Combat LGBTQ Youth Suicide*. Retrieved October 21, 2019, from NBC News: <https://www.nbcnews.com/feature/nbc-out/google-trevor-project-will-use-ai-combat-lgbtq-youth-suicide-n10035ii>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). *Filter Bubbles, Echo Chambers, and Online News Consumption*. *Public Opinion Quarterly*, 80(1), 298-320.
- Flew, T., Spurgeon, C., & Daniel, A. (2012). *The Promise of Computational Journalism*. *Journalism Practice*, 6(2), 151-171.
- Foucault, M. (1977). *Discipline and Punishment: The Birth of The Prison*. London: Allen Lane.

- Freeman, J. (2019, July 5). *How to Fight STEM'S Unconscious Bias against LGBTQ People*. Retrieved October 21, 2019, from Scientific America: <https://blogs.scientific-american.com/voices/how-to-fight-stems-unconscious-bias-against-lgbtq-people/>
- Fu, L. (2018, May 29). *Four Key Barriers to the Widespread Adoption of AI*. Retrieved October 21, 2019, from MIT: Professional Education: <https://professional.mit.edu/news/news-listing/four-key-barriers-widespread-adoption-ai>
- Fussell, S. (2017, October 23). *Palestinian Man Arrested After Facebook Auto-Translates 'Good Morning' as Attack Them*. Retrieved October 18, 2019, from Gizmodo: <https://gizmodo.com/palestinian-man-arrested-after-facebook-auto-translates-1819782902>
- GDPR art. IV, cl 4. (2016, April 27). Brussels, Belgium.
- Gee, T. J. (2017 July 5). *Why Female Sex Robots Are More Dangerous Than You Think*. Retrieved October 21, 2019, from The Telegraph: <https://www.telegraph.co.uk/women/life/female-robots-why-this-scarlett-johansson-bot-is-more-dangerous/>
- Gehrmann, S., Strobelt, H., & Rush, A. (2019, June 10). *GLTR: Statistical Detection and Visualization of Generated Text*. Retrieved October 18, 2019, from Cornell University: <https://arxiv.org/abs/1906.04043>
- Gelman, A., Mattson, G., & Simpson, D. (2018). *Gaydar and The Fallacy of Objective Measurement*. *Sociological Science*, 5(2), 270-280.
- Geo Gecko, (n.d.). *What We Do*. Retrieved October 18, 2019, from Geo Gecko: <https://www.geogecko.com/uganda-gis-services>
- George, T. (2018, December 12). *Newsrooms Must Learn How to Use AI: Trust in Journalism Is At Stake*. Retrieved October 18, 2019, from Journalism.co.uk: <https://www.journalism.co.uk/skills/lessons-learned-in-the-last-four-years-of-using-artificial-intelligence-at-the-associated-press/s7/a-731760/>
- Gilmore, J. (2011). *Expression as Realization: Speakers' Interests in Freedom of Speech*. *Law and Philosophy*, 30(5), 517-539.
- Glassdoor. (2018). *25 Highest Paying Jobs in America*. Retrieved October 18, 2019, from Glassdoor: [https://www.glassdoor.com/List/Highest-Paying-Jobs-LST\\_KQo\\_19.htm](https://www.glassdoor.com/List/Highest-Paying-Jobs-LST_KQo_19.htm)
- GOBL. (2014). *CSIR*. Retrieved October 18, 2019, from GOBL: <https://www.gobl-project.eu/council-for-scientific-and-industrial-research/>
- Goitom, H. (2019). *Regulation of Artificial Intelligence in Selected Jurisdictions*. Washington DC: The Law Library of Congress.
- Goldstein, E., Gasser, U., & Budish, R. (2018, June 21). *Data Commons Version 1.0: A Framework to Build Toward AI for Good*. Retrieved October 18, 2019, from Berkman Klein Center: <https://medium.com/berkman-klein-center/data-commons-version-1-0-a-framework-to-build-toward-ai-for-good-734i4d7e-72be>

- GoodFellow, I. J., Papernot, N., Huang, S., Duan, Y., Abbeel, P., & Clark, J. (2017, February 24). *Attacking Machine Learning with Adversarial Examples*. Retrieved October 18, 2019, from Open AI: <https://openai.com/blog/adversarial-example-re-search/>
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). *Explaining and Harnessing Adversarial Examples*. ICLR (pp. 1-11). San Diego: arXiv.
- Goodman, B., & FLaxman, S. (2017). *European Union regulations on algorithmic decision-making and "a right to explanation"*, *AI Magazine*, 38(3), 50-57.
- Goodrich, J., & McCrea, M. (2018, April 26). *Tech Companies Collaborate to Create Next-Gen Sex Robots*. Retrieved October 21, 2019, from CBS Local: <https://sanfrancis-co.cbsLocal.com/2018/04/26/sex-robot-next-gen-caLifornia-tech-company/>
- Government of Canada. (2018). *Algorithmic Impact Assessment (vo.2)*. Retrieved October 18, 2019, from Government of Canada: <https://canada-ca.github.io/digi-taL-pLaybook-guide-numerique/en/overview.html>
- Greene, T. (2018, August 9). *Uber's 'Real-time ID Check' Doesn't Deal Well with Trans-gender Drivers*. Retrieved October 18, 2019, from The Next Web: <https://athe-nextweb.com/artificial-inteLLigence/2018/08/08/ubers-real-time-id-check-doesnt-deal-weLL-with-transgender-drivers/>
- GreenhaLgh, H. (2018, September 26). *Gay Rights Groups Hail Landmark U.N. Chief Speech Baching LGBT+ Community*. Retrieved October 18, 2019, from Reuters: <https://ywww.reuters.com/articLe/us-gLobaL-un-Lgbt-idUSKCNiM62LL>
- Grint, K., & GiLL, R. (1995). *The Gender-Technology Relation: Contemporary Theory and Research*. London: TayLor & Francis.
- GuL, E. (2019, JuLy 5). *Is Artificial Intelligence the frontier solution to Global Souths wiced development challenges?* Retrieved October 21, 2019, from Medium: <https://A-towardsdatascience.com/is-artificial-inteLLigence-the-frontier-soLu-tion-to-gLo-baL-souths-wicked-develOpment-chaLLenges-420622ia3c78>
- Gurumoorthy K. S., Dhurandhar, A., Cecchi, G., & AggarwaL, C. (2019, August 12). *Efficient Data Representation by Selecting Prototypes with Importance Weights*. Retrieved October 21, 2019, from CorneLL University: ArXiv: <https://arxiv.org/pdf/A707.01212.pdf>
- Gustavsson, E. (2005). *Virtual Servants: Stereotyping Female Front-Office Employees on The Internet, Gender, Work and Organization*, 12(5), 400-419.
- Guszcza, J., Rahwan, I., BibLe, W., Cebrian, M., & KatyaL, V. (2018, November 28). *Why We Need to Audit Algorithms*. Retrieved October 18, 2019, from Harvard Business Review: <https://hbr.org/2018/11/why-we-need-to-audit-algorithms>
- HaLiLou, I. (2016, JuLy 16). *10 E-Learning Platforms Transforming Education in Africa*. Retrieved October 21, 2019, from True Africa: <https://trueafrica.co/lists/e-learning-pLatforms-africa-tutor-ng-mest-school-education-startups/>
- Hansen, E. (2018). *Aporias of Courage and The Freedom of Expression*. *PhiLosophy and Social Criticism*, 44(1), 100-117

- Hao, K. (2019, August 7). *China's path to AI domination has a problem: brain drain*. Retrieved October 21, 2019, from MITTechnology Review: <https://www.technologyreview.com/f/614092/china-ai-domination-Losing-talent-to-us/>
- Haraway, D. J. (1991). *Science, Technology, and Socialist-Feminism in the Late Twentieth Century*. In D. J. Haraway, Simians, Cyborgs and Women: The Reinvention of Nature (pp. 149-181). New York: Routledge.
- Harcourt, B. E. (2007). *Against Prediction: Punishing and Policing in An Actuarial Age*. Chicago: University of Chicago Press.
- Harwell, D. (2019, June 12). *Top AI Researchers Race to Detect Deepfake Videos: 'We Are Outgunned'*. Retrieved October 21, 2019, from The Washington Post: <https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/?noredirect=on>
- Hensel, A. (2018, November 15). *Facebook to Create 'Independent Panel' for Appealing Content Moderation Decisions*. Retrieved October 18, 2019, from Venture Beat: <https://venturebeat.com/2018/11/15/facebook-to-create-independent-panel-for-appealing-content-moderation-decisions/>
- Hern, A. (2017 October 24). *Facebook Translates 'Good Morning' into 'Attack Them', Leading to Arrest*. Retrieved October 18, 2019, from The Guardian: <https://www.theguardian.com/technology/2017/oct/24/facebook-palastine-israel-translates-good-morning-attack-them-arrest>
- Hester, H. (2016, August 8). *Technically Female: Women, Machines, and Hyperemployment*. Retrieved October 21, 2019, from Salvage: <http://salvage.zone/in-print/technically-female-women-machines-and-hyperemployment/>
- Hester, H., & Angel K. (2016). *Technosexuals*. (S. Radio, Interviewer)
- Hill, C., Corbett, C., & St. Rose, A. (2010). *Why So Few? Women in Science, Technology, Engineering and Mathematics*. Washington D.C.: American Association of University Women.
- Hope, C., & McCann, K. (2017 September 19). *Google, Facebook and Twitter Told to Take Down Terror Content within Two Hours or Face Fines*. Retrieved October 18, 2019, from The Telegraph: <https://www.telegraph.co.uk/news/2017/09/19/google-facebook-twitter-to-ld-take-terror-content-within-two/>
- Hutson, M. (2018, May 17). *Why are AI researchers boycotting a new Nature journal—and shunning others?* Retrieved October 21, 2019, from Science Mag: <https://www.sciencemag.org/news/2018/05/why-are-ai-researchers-boycotting-new-nature-journal-and-shunning-others>
- IBM. (2019a, August 8). *AI Explainability 360*. Retrieved October 18, 2019, from IBM Developer: <https://developer.ibm.com/open/projects/ai-explainability/>
- IBM. (2019b). *IBM Watson OpenScale*. Retrieved October 18, 2019, from IBM: <https://www.ibm.com/cloud/watson-openscale/>
- ICT Authority Kenya. (2019, September). *Open Data*. Retrieved October 18, 2019, from Kenya Open Data: <http://icta.go.ke/open-data/>

- InstaDeep. (2019). *Instadeep: About*. Retrieved October 18, 2019, from InstaDeep: [https:// www.instadeep.com/about/](https://www.instadeep.com/about/)
- IPDC Council CI/2018/COUNCIL\_31/H/1. (2018, November 21-22). *Decisions taken by the 31<sup>st</sup> Council Session of the International Programme for the Development of Communication (IPDC)*, Paris, France.
- Ireton, C, & Posetti, J. (2018f). *Journalism, Fake News & Disinformation: Handbook for Journalism Education and Training*. Paris: UNESCO Publishing.
- I REX. (2018). *Can Machine Learning Help Us Measure The Trustworthiness of News*. Washington D.C.: IREX.
- ITU. (2012, June). *Next Generation Networks – Frameworks and Functional Architecture Models*, ITU-T Y.2060 Standard Series Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Networks. International Telecommunication Union.
- ITU. (2014, August). *Information Technology – Cloud Computing – Overview and Vocabulary*, ITU-T Y.3500 Standard Series Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Networks. International Telecommunication Union.
- ITU. (2017). *ICTFacts and Figures 2017*. Geneva: ITU.
- Jao, N. (2017, May 12). *Big Data to tackle grand challenges: A look at IBM Research Africa's projects*. Retrieved October 21, 2019, from ITU News: <https://news.itu.int/big-data-to-tackle-grand-challenges-a-look-at-ibm-research-africas-projects/>
- Just, N., & Latzer, M. (2017). *Governance by Algorithms: Reality construction by algorithmic selection on the Internet, Media, Culture and Society*, 39(2), 238-258.
- Kantrowitz, A. (2018, March 21). *Facebook Has Blocked Ad Targeting By Sexual Orientation*, Retrieved October 21, 2019, from BuzzFeed News: <https://www.buzzfeednews.com/article/alexkantrowitz/facebook-has-blocked-ad-targeting-by-sexual-orientation>
- KAS Uganda & S.Sudan. (2018, December 12). *Digitalisation Forum 2018 – Managing the 4th Industrial Revolution*. Retrieved October 18, 2019, from Medium: [https:// medium.com/@ KasUganda/digitalisation-forum-2018-managing-the-4th-industrial-revolution-4d1cc866a736](https://medium.com/@KasUganda/digitalisation-forum-2018-managing-the-4th-industrial-revolution-4d1cc866a736)
- Keller, D. (2018, June 13). *Internet Platforms: Observations on Speech, Danger, and Money*, Retrieved October 21, 2019, from Hoover Institution's Aegis Paper Series: <https://ssrn.com/abstract=3262936>
- Kelly, A. (1985). *The Construction of Masculine Science*. *British Journal of Sociology of Education*, 6(2), 133-154.
- Kenyan Wall Street. (2018, February 28). *Kenya Govt unveils 11 Member Blockchain & AI Taskforce headed by Bitange Ndemo*. Retrieved October 18, 2019, from The Kenyan Wall Street: <https://kenyanwallstreet.com/kenya-govt-unveils-n-member-blockchain-ai-taskforce-headed-by-bitange-ndemo/>

- Keyes, O. (2018). *The Misgendering Machines: Trans/HCI Implication of Automatic Gender Recognition*, *ACM on Human-Computer Interaction*. New York: Association for Computing Machinery.
- Kharpal, A. (2019, June 4). *China is ramping up its own chip industry amid a brewing tech war That could hurt US firms*. Retrieved from CNBC: <https://www.cnbc.com/2019/06/04/china-ramps-up-own-semiconductor-industry-amid-the-trade-war.html>
- Kleeman, J. (2017, September 25). *Should we ban sex robots while we have the chance?* Retrieved October 18, 2019, from The Guardian: <https://www.theguardian.com/commentisfree/2017/sep/25/ban-sex-robots-doUs-market>
- Kline, R. R. (2015). *Technological Determinism*, In J. D. Write, *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition) (pp. 109-112). New York: Elsevier.
- Knight, W. (2018, November 17). *One of the fathers of Ai is worried about its future*. Retrieved October 18, 2019, from MIT Technology Review: <https://www.technologyreview.com/s/612434/one-of-the-fathers-of-ai-is-worried-about-its-future/>
- Kudi. (2019). *About* Retrieved from Kudi: <https://kudi.com/>
- Kumar, V, Raghavendra, R., Namboodiri, A., & Busch, C. (2016). *Robust Transgender Face Recognition: Approach Based on Appearance and Therapy Factors*. 2016 IEEE International Conference on Identity, Security and Behavior Analysis (pp. 1-7). Sendai: IEEE.
- Kunze, L. (2019). *Can We Stop the Academic AI Brain Drain?* *Kunstliche Intelligenz*, 33(1), 1-3.
- Kwok, R. (2019, April 29). *Junior AI researchers are in demand by universities and industry*. Retrieved October 18, 2019, from Nature: <https://www.nature.com/articles/d4i586-oig-oi248-w>
- Lai, C. K, & Banaji, M. R. (2019). *The Psychology of Implicit Intergroup Bias and the Prospect of Change*. In D. Allen, & R. Somanathan, *Difference without Domination: Pursuing Justice in Diverse Democracies*. Chicago, Illinois: University of Chicago Press.
- Lamagna, M. (2018, March 18). *Google's Parent Company is Using AI to Make the Internet Safer for LGBT People*. Retrieved October 21, 2019, from Market Watch: <https://www.marketwatch.com/story/how-artificial-intelligence-can-make-the-internet-better-and-safer-for-lgbt-people-2018-03-i4>
- Latar, N. L. (2015). *The Robot Journalist in The Age of Social Physics: The End of Human Journalism?* In G. Einav, *The New World of Transitioned Media: Digital Realignment and Industry Transformation* (pp. 65-80). New York: Springer.
- Latonero, M. (2018). *Governing Artificial Intelligence: Upholding Human Rights & Dignity*. New York: Data & Society.

- Le Monde. (2019, May 1). *Agriculture Numerique: Le Senegal Montre L'exemple*. Retrieved October 18, 2019, from Le Monde Afrique: [https://www.leronde.fr/afrique/articLe/2019/05/ol/agricuLture-numerique-Le-senegaL-montre-L-exemple\\_5457016\\_3212.html](https://www.leronde.fr/afrique/articLe/2019/05/ol/agricuLture-numerique-Le-senegaL-montre-L-exemple_5457016_3212.html)
- Leavy, S. (2018). *Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory*. 1st International Workshop on Gender Equality in Software Engineering (pp. 14-16). Gothenburg: ACM.
- LeCun, Y (2009). *A New Publishing Model in Computer Science*. Retrieved October 21, 2019, from Yann LeCun BLOG: <http://yann.lecun.com/ex/pamphlets/publishing-models.html>
- Lee, D. (2018, February 3). *Deepfakes Porn Has Serious Consequences*. Retrieved October 21, 2019, from BBC News: <https://www.bbc.com/news/technology-42912529>
- Levchuk, K. (2018, August 1). *Why Women Should Be Excited About AI*. Retrieved October 21, 2019, from Forbes: <https://www.forbes.com/sites/cognitiveworLd/20i8/08/0i/why-women-shouLd-be-excited-about-ai/#56ci26gi272b>
- Levin, S. (2017, September 9). *LGBT Groups Denounce Dangerous AI that Uses Your Face to Guess Sexuality*. Retrieved October 21, 2019, from The Guardian: <https://www.forbes.com/sites/cognitiveworld/2018/08/01/why-women-should-be-excited-about-ai/#24foaef2i272>
- Lewton, T, & McCool, A. (2018, December 14). *This App Tells Your Doctor If You Have Malaria*. Retrieved October 21, 2019, from CNN Health: <https://edition.cnn.com/2018/12/14/health/ugandas-first-ai-lab-develops-malaria-detection-app-intl/index.html>
- Liptak, A. (2017 May 1). *Sent to Prison by a Software Program's Secret Algorithms*, Retrieved October 21, 2019, from The New York Times: [https://www.nytimes.com/2017/05/0i/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?smid=tw-share&\\_r=0](https://www.nytimes.com/2017/05/0i/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?smid=tw-share&_r=0)
- Lomas, N. (2017 April 29). *Someone scraped 40,000 Tinder selfies to make a facial dataset for AI experiments*. Retrieved October 21, 2019, from Tech Crunch: [https://techcrunch.com/20i7/04/28/someone-scraped-40000-tinder-selfies-to-make-a-facial-dataset-for-ai-experiments/?guccounter=i8t-guce\\_referrer\\_us=aHRocHM6Lyg3d-3cuZ2gvZ2xlLmNvbS8&guce\\_referrer\\_cs=5Np2afFglc8vqTYzWUAgHw](https://techcrunch.com/20i7/04/28/someone-scraped-40000-tinder-selfies-to-make-a-facial-dataset-for-ai-experiments/?guccounter=i8t-guce_referrer_us=aHRocHM6Lyg3d-3cuZ2gvZ2xlLmNvbS8&guce_referrer_cs=5Np2afFglc8vqTYzWUAgHw)
- Lystad, J. (2019, July 12). *Artificial Intelligence School to Open in Fez*. Retrieved October 21, 2019, from Morocco World News: <https://www.morocccoworldnews.com/2019/07/278026/artificial-intelligence-school-open-fez/>
- Lyu, S. (2018, July 12). *Detecting Deepfake Videos*. (D. Hopper, Interviewer)

- Machalo, G. (2019, February 20). *Cabinet Approves SMART Zambia eGovernment Master Plan and Public Services Standards*. Retrieved October 21, 2019, from SMART Zambia Institute: [https://www.szi.gov.zm/cabinet-approves-smart-zarn-bia-egovernment-master-plan-and-public-services-standards/+Government+-Bill%2C+20i8\)+7th+National+Development+Plan+ackn](https://www.szi.gov.zm/cabinet-approves-smart-zarn-bia-egovernment-master-plan-and-public-services-standards/+Government+-Bill%2C+20i8)+7th+National+Development+Plan+ackn)
- MacLellan, L. (2019, March 22). *Hear What A Genderless AI Voice Sounds Like-and Consider WhyIt Matters*. Retrieved October 21, 2019, from Quartz: <https://www.qz.com/work/1577597/this-ai-voice-is-gender-neutral-unlike-siri-and-alexa/>
- Madden, M., & Raine, L. (2015, May 20). *Americans' Attitudes About Privacy, Security and Surveillance*. Retrieved September 4, 2019, from Pew Research Center: <https://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/>
- Marechal, D. N. (2018, November 16). *Targeted Advertising Is Ruining the Internet and Breaking the World*. Retrieved September 5, 2019, from Vice: [https://www.vice.com/en\\_us/article/xwjden/targeted-advertising-is-ruining-the-internet-and-breaking-the-world](https://www.vice.com/en_us/article/xwjden/targeted-advertising-is-ruining-the-internet-and-breaking-the-world)
- Marr, B. (2018, April 16). *The 6 Best Free Online Artificial Intelligence Courses Available Today*. Retrieved October 21, 2019, from Forbes: <https://www.forbes.com/sites/bernardmarr/2018/04/16/the-6-best-free-online-artificial-intelligence-courses-for-2018/#279efe8b59d7>
- Masure, A., & Pandelakis, P. (2017, October). *Machines Desirantes : des Sexbots aux OS Amoureux*. Retrieved October 21, 2019, from ReS Futuræ: <https://journals.openedition.org/resf/1066>
- Matsakis, L. (2018, July 10). *A Frightening AI Can Determine Whether a Person Is Gay With 91 Percent Accuracy*. Retrieved October 21, 2019, from Vice: [https://www.vice.com/en\\_asia/article/a33xb4/a-frightening-ai-can-determine-a-persons-sexuality-with-91-accuracy](https://www.vice.com/en_asia/article/a33xb4/a-frightening-ai-can-determine-a-persons-sexuality-with-91-accuracy)
- Mbaka, C. (2017 December 17). *Sophie Bot Aims to Destigmatize Sex Education in Africa*. Retrieved October 21, 2019, from Techmoran: <https://www.techmoran.com/2017/12/17/sophie-bot-aims-to-reduce-stigma-around-sexual-education/>
- McMullan, T. (2015, October 18). *Guardian readers on privacy: 'we trust government over corporations*. Retrieved September 4, 2019, from The Guardian: <https://www.theguardian.com/technology/2015/oct/18/guardian-readers-on-privacy-we-trust-government-over-corporations>
- Mehta, I. (2018, October 15). *Amazon's New Patent Will Allow Alexa to Detect A Cough or A Cold*. Retrieved October 21, 2019, from The Next Web: <https://thenextweb.com/artificial-intelligence/2018/10/15/amazons-new-patent-will-allow-alexa-to-detect-your-illness/>
- Metz, C. (2017 November 5). *Building AI. That Can Build AI*. Retrieved October 21, 2019, from The New York Times: <https://www.nytimes.com/2017/11/05/technology/machine-learning-artificial-intelligence-ai.html>

- Metz, C. (2019, August 16). *A.I. Is Learning From Humans. Many Humans*. Retrieved from The New York Times: <https://www.nytimes.com/2019/08/16/technology/ai-humans.html>
- Michael, V., Van Kleek, M., & Binns, R. (2018). *Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-making*. CHI Conference on Human Factors in Computing Systems (pp. 440-454). Montreal: ACM.
- Mikkinen, M., Auffermann, B., & Heinonen, S. (2017). *Framing The Future of Privacy: Citizens' Metaphors for Privacy in The Coming Digital Society*. European Journal of Futures Research, 5(7), 7.
- Miller, H., & Stirling, R. (2019). *Government Artificial Intelligence Readiness Index 2019*. Ottawa: IDRC; Oxford Insights.
- Ministry of Information, Communications and Technology. (2019). *Emerging Digital Technologies for Kenya: Exploration and Analysis*. July: Government of Kenya.
- Ministry of Transport and Communications (MTC). (2011). *Departments/Units*. Retrieved October 18, 2019, from Ministry of Transport and Communications (MTC): <http://www.gov.bw/en/Ministries--Authorities/Ministries/Ministry-of-Transport-and-Communications/Departments/Telecommunications/>
- Montal, T., & Reich, Z. (2017). */, robot. You, journalist. Who Is The Author*, Digital Journalism, 5(7), 829-849.
- Montjoye, Y.-A., Farzanehfar, A., Hendrickx, J., & Rocher, L. (2017). *Solving Artificial Intelligence's Privacy Problem*. The Journal of Field Actions, 17, 80-83.
- Morris, A. (2018, September 25). *Prediction: Sex Robots Are The Most Disruptive Technology We Didn't See Coming*. Retrieved October 21, 2019, from Forbes: <https://www.forbes.com/sites/andreamorris/2018/09/25/prediction-sex-robots-are-the-most-disruptive-technology-we-didnt-see-coming/#i96i4g8d6a56>
- Moses, L. (2017, September 14). *The Washington Post's Robot Reporter Has Published 850 Articles in The Past Year*. Retrieved October 21, 2019, from Digiday UK: <https://digiday.com/media/washington-posts-robot-reporter-published-500-articles-last-year/>
- Moye, D. (2017, September 29). *Sex Robot Molested At Electronics Festival, Creators Say*, Retrieved October 21, 2019, from HuffPost: [https://www.huffpost.com/en-try/samantha-sex-robot-molested\\_n\\_59cec9f9e4bo679ibbioa268](https://www.huffpost.com/en-try/samantha-sex-robot-molested_n_59cec9f9e4bo679ibbioa268)
- Moye, D. (2018, May 24). *Amazon Admits Alexa Device Eavesdropped On Portland Family*. Retrieved October 21, 2019, from HuffPost: [https://www.huffpost.com/en-try/alexa-eavesdropping-portland-family\\_n\\_5b0727cae4bofdb2aa5ib23e](https://www.huffpost.com/en-try/alexa-eavesdropping-portland-family_n_5b0727cae4bofdb2aa5ib23e)
- MSI-NET. (2016). *Study on The Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications*.
- MSI-NET (2018). *Algorithms and Human Rights Study on The Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*. Strasbourg: Council of Europe.

- Mukherjee, S. (2017, April 3). *A.I. Versus M.D.*, Retrieved October 21, 2019, from The New Yorker: <https://www.newyorker.com/magazine/2017/04/03/ai-versus-md>
- Murphy, H. (2017, October 9). *Why Stanford Researchers Tried to Create A 'Gaydar' Machine.* Retrieved October 21, 2019, from The New York Times: <https://www.ny-times.com/2017/10/09/science/stanford-sexual-orientation-study.html>
- Namibia University of Science and Technology. (2019). *Bachelor of Computer Science*, Retrieved October 18, 2019, from Namibia University of Science and Technology: <https://www.nust.na/?q=programme/fci/bachelor-computer-science>
- NAM PA. (2019, March 7). *Namibia Should Play Active Role in Fourth Industrial Revolution: Tweya.* Retrieved October 18, 2019, from Lela Mobile: <https://www.lelamobile.com/content/79171/Namibia-should-play-active-role-in-fourth-industrial-revolution-Tweya/>
- Nayyar, G. (2019, July 16). *What do automation and artificial intelligence mean for Africa?* Retrieved October 17, 2019, from Brookings: <https://www.brookings.edu/blog/future-development/2019/07/16/what-do-automation-and-artificial-intelligence-mean-for-africa/>
- Negnevitsky, M. (2011). *Artificial Intelligence: A Guide to Intelligent Systems*. Boston: Addison Wesley.
- Nenadic, I. (2018, March 8). *Data-driven Online Political Microtargeting: Hunting for Voters, Shooting Democracy?* Retrieved October 21, 2019, from Center for Media Pluralism and Media Freedom: <http://cmpf.eui.eu/data-driven-online-political-microtargeting-hunting-for-voters-shooting-democracy/>
- Newton, C. (2019, June 28). *Facebook's Supreme Court for Content Moderation is Coming into Focus*, Retrieved October 21, 2019, from The Verge: <https://www.theverge.com/interface/2019/6/28/18761357/facebook-independent-over-sight-board-report-zuckerberg>
- Nilsson, K. (2019, March 8). *Why AI Needs More Women.* Retrieved October 21, 2019, from Forbes: <https://www.forbes.com/sites/kimnilsson/2019/03/08/why-ai-needs-more-women/#6fc45cab7f90>
- Nissenbaum, H. (2004). *Privacy as Contextual Integrity*. Washington Law Review, 79(1), 101-139.
- NumberBoost. (2019). *NumberBoost*. Retrieved October 18, 2019, from NumberBoost: <https://www.numberboost.com/>
- Nwafor. (2019, April 15). *Nigeria is not ripe for Artificial Intelligence.* Retrieved October 21, 2019, from Vanguard: <https://www.vanguardngr.com/2019/04/nigeria-is-not-ripe-for-artificial-intelligence/>
- O'Dwyer, R. (2018, May 18). *Algorithms are making the same mistakes assessing credit scores that humans did a century ago.* Retrieved October 21, 2019, from QZ: <https://qz.com/i27678i/algorithms-are-making-the-same-mistakes-assessing-credit-scores-that-humans-did-a-century-ago/>
- OECD. (2001). *Understanding the Digital Divide*. Paris: OECD.

- Oghia, M. J. (2018, November). *Human Rights in AI: A Journalism & Media Perspective*. Paris, France. Retrieved September 2, 2019, from <https://www.youtube.com/watch?v=77LNQqgs3tU>
- Oleksy, W., Just, E., & Zapedowska-Kling, K. (2012). *Gender Issues in Information and Communication Technologies*. *Journal of Information, Communication and Ethics in Society*, 10(2), 107-120.
- O'Malley, J. (2018, January 12). *Captcha if you can: how you've been training AI for years without realising it*. Retrieved from Techradar: <https://www.techradar.com/news/captcha-if-you-can-how-youve-been-training-ai-for-years-without-realising-it>
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin.
- Orr, D. (2016, June 10). *At Last, A Cure for Feminism: Sex Robots*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/commentis-free/2016/jun/10/feminism-sex-robots-women-technology-objectify>
- Owino, J. (2019, April 18). *Agriculture Takes Lead in Adopting AI in Kenya*. Retrieved October 21, 2019, from Capital Business: <https://www.capitalfm.co.ke/bu-siness/2019/04/agriculture-takes-lead-in-adopting-ai-in-kenya/>
- Packin, N. G., & Lev Aretz, Y. (2018). *Learning Algorithms and Discrimination*. In W. Barfield, & U. Pagallo, *Research Handbook of Artificial Intelligence and Law* (pp. 88-113). Cheltenham: Edward Edgar Publishing.
- Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., CeLiK, Z. B., & Swami, A. (2017). *Practical Black-Box Attacks against Machine Learning*. ACM Asia Conference on Computer and Communications Security (pp. 1-14). Abu Dhabi: Cornell University: arXiv.
- Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.
- Parnell, T., & Dunner, C. (2018, March 20). *IBM Sets Tera-scale Machine Learning Benchmark Record with POWER9 and NVIDIA GPUs: Available Soon in PowerAI*. Retrieved October 21, 2019, from IBM Research Blog: <https://www.ibm.com/blogs/research/2018/03/machine-learning-benchmark/>
- Parsheera, S. (2018). *A Gendered Perspective on Artificial Intelligence, ITU Kaleidoscope 2018 – Machine Learning for a 5G Future* (pp. 1-7). Santa Fe: ITU.
- Patricia, S., & Steinberg, D. L. (1987). *Made to Order: The Myth of Reproductive and Genetic Progress*. New York: Pergamon Press.
- Paul, K. (2019, April 17). *Disastrous lack of diversity in AI industry perpetuates bias, study finds*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study>

- Phakathi, B. (2019, April 9). *Cyril Ramaphosa to chair fourth industrial revolution commission*. Retrieved October 21, 2019, from Business Day: <https://www.businesslive.co.za/bd/national/2019-04-09-cyril-ramaphosa-to-chair-fourth-industrial-revolution-commission/>
- Posetti, J. (2017). *Protecting journalism sources in the digital age*. Paris: UNESCO.
- Public Private Partnership Commission. (2017). *Digital Malawi Project (DMP) Resettlement Policy Framework*. Blantyre: February.
- Quach, K. (2019, March 5). *The Infamous AI Gaydar Study Was Repeated – and, No, Code Can't Tell If You're Straight or Not Just From Your Face*, Retrieved October 21, 2019, from The Register: [https://www.theregister.co.uk/2019/03/05/ai\\_gaydar/](https://www.theregister.co.uk/2019/03/05/ai_gaydar/)
- Raso, F. A., Hilligoss, H., Krishnamurthy, V, Bavitz, C, & Kim, L (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks*. Cambridge: The Berkman Klein Center for Internet & Society.
- Rense, S. (2018, February 12). *What Are 'Deepfakes,' and Why Are Pornhub and Reddit Banning Them?* Retrieved October 21, 2019, from Esquire: <https://www.esquire.com/lifestyle/sex/ai7043863/what-are-deepfakes-celebrity-porn/>
- Reporters Without Borders. (2018). *Online Harassment of Journalists*. Paris: Reporters Without Borders.
- Richardson, K. (2016). *Sex Robot Matters: Slavery, the prostituted and the Rights of Machines!* IEEE Technology and Society, 35(2), 46-53.
- Rieger, S. (2018, July 26). *At Least Two Malls Are Using Facial Recognition Technology to Track Shoppers' Ages and Genders Without Telling*. Retrieved October 21, 2019, from CBC: <https://www.cbc.ca/news/canada/calgary/calga-ry-malls-1.4760964>
- Robitzski, D. (2018, May 3). *AI Researchers Are Boycotting A New Journal Because It's Not Open Access*. Retrieved October 21, 2019, from Futurism: <https://futurism.com/artificial-intelligence-journal-boycot-open-access>
- robotcampaign. (2019, February 6). *Why We Must Discuss The Normalisation of Sex Dolls and Sex Robots in Our Society!* Retrieved October 18, 2019, from Campaign against Sex Robots: <https://campaignagainstsexrobots.org/>
- Rouvroy A. (2014). *Data Without (Any)Body? Algorithmic Governmentality as Hyper-dissident and The Role of Law as Technical Organ*. Conference on General Organology (pp. 1-2). Canterbury: University of Kent.
- Rouvroy, A. (2016). *Of Data and Men Fundamental Rights and Freedoms in A World of Big Data*. Strasbourg: Council of Europe.
- Rouvroy, A., & Stiegler, B. (2016). *The Digital Regime of Truth: From The Algorithmic Governmentality to A New Rule of Law*. La Deleuziana: Online Journal of Philosophy, 3(3), 6-29.

- Sachs, J. D. (2018). R&D, *Structural Transformation, and the Distribution*. In A. Agrawal, J. Gans, & A. Goldfarb, *The Economics of Artificial Intelligence* (pp. 329-348). Chicago: University of Chicago Press.
- Salmon, F., & Stokes, J. (2010, December 27). *Algorithms Take Control of Wall Street*. Retrieved October 21, 2019, from Wired: <https://www.wired.com/2010/12/ff-ai-flashtrading/>
- Sample, I. (2017a, November 1). *We can't compete: why universities are losing their best AI scientists*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/science/2017/nov/01/cant-compete-universities-losing-best-ai-scientists>
- Sample, I. (2017b, November 2). *Big tech firms' AI hiring frenzy leads to brain drain at UK universities*. Retrieved 16 October, 2019, from The Guardian: <https://www.theguardian.com/science/2017/nov/02/big-tech-firms-google-ai-hiring-frenzy-brain-drain-uk-universities>
- Sandelson, J. (2018, March 19). *The politics of AI and scientific research on sexuality*. Retrieved October 21, 2019, from LSE Engenderings: <https://blogs.lse.ac.uk/gender/2018/03/19/the-politics-of-ai-and-scientific-research-on-sexuality/>
- Santa Clara Principles. (2018, May 7). *The Santa Clara Principles*, Content Moderation at Scale Conference. Washington, D.C. Retrieved from The Santa Clara Principles on Transparency and Accountability in Content Moderation : <https://santaclara-principles.org/>
- Scanlon, T. (1972). *A Theory of Freedom of Expression*, *Philosophy and Public Affairs*, 1(2), 204-226.
- Schwartz, O. (2018, November 12). *You thought fake news was bad? Deep fakes are where truth goes to die*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>
- Scott, M., & Isaac, M. (2016, September 9). *Facebook Restores Iconic Vietnam War Photo It Censored for Nudity*. Retrieved October 21, 2019, from The New York Times: <https://www.nytimes.com/2016/09/09/technology/facebook-vietnam-war-photo-nudity.html>
- Scripps, A. (2017). *Artificial Intelligence as A Digital Privacy Protector*. *Harvard Journal of Law & Technology*, 31(1), 217-234.
- Shah, S. (2017, February 1). *Facebook? Censors Explicitly Sexual' Nude Statue of Neptune*. Retrieved October 21, 2019, from DigitalTrends: <https://www.digitaltrends.com/social-media/facebook-nude-statue/>
- Sharpe, A., & Raj, S. (2017, September 15). *Using AI to determine queer sexuality is misconceived and dangerous*. Retrieved October 21, 2019, from The Conversation: <http://theconversation.com/using-ai-to-determine-queer-sexuality-is-misconceived-and-dangerous-83931>

- Shead, S. (2017 December). *The UK government is giving AI startups access to serious computation power at a new 'Machine Intelligence Garage'*, Retrieved October 21, 2019, from Business Insider: <https://www.businessinsider.fr/us/uk-government-gives-ai-startups-access-to-computation-power-2017-12>
- Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J., CBauer, Z. (2018). *The AI Index 2018 Annual Report*, Stanford: AI Index Steering Committee, Human-Centered AI Initiative, Stanford University.
- Sibal, P. (2016, November 11). *Why Trump's Election Is A Failure Of Liberalism*. Retrieved October 21, 2019, from Huffpost: [https://www.huffingtonpost.in/prateek-sibal/why-trumps-election-is-a-failure-of-liberalism\\_a\\_21603465/](https://www.huffingtonpost.in/prateek-sibal/why-trumps-election-is-a-failure-of-liberalism_a_21603465/)
- Simonite, T. (2017, August 21). *Machines Taught by Photos Learn A Sexist View of Women*. Retrieved October 21, 2019, from Wired: <https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>
- Simonite, T. (2018a, January 11). *When It Comes to Gorillas, Google Photos Remains Blind*. Retrieved October 16, 2019, from Wired: <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>
- Simonite, T. (2018b, August 17). *AI Is the Future—But Where Are the Women?* Retrieved October 21, 2019, from WIRED: <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>
- Snow, J. (2018a, February 14). *'We're in a diversity crisis': cofounder of Black in AI on what's poisoning algorithms in our lives*. Retrieved October 21, 2019, from MIT Technology Review: <https://www.technologyreview.com/s/6io92/were-in-a-diversity-crisis-black-in-ai-founder-on-whats-poisoning-the-algorithms-in-our/>
- Snow, J. (2018b, February 26). *Bias already exists in search engine results, and it's only going to get worse*. Retrieved October 18, 2019, from MIT Technology Review: <https://www.technologyreview.com/s/6io275/meet-the-woman-who-searches-out-search-engines-bias-against-women-and-minorities/>
- Solon, O. (2017 October 4). *More than 70% of US fears robots taking over our lives, survey finds*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/technology/2017/oct/04/robots-artificial-intelligence-machine-us-survey>
- Solon, O., & Levin, S. (2016, December 16). *How Google's search algorithm spreads false information with a rightwing bias*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/technology/2016/dec/16/google-autocomplete-rightwing-bias-algorithm-political-propaganda>
- Stanford, S. (2018, August 29). *Artificial Intelligence: Salaries Heading Skyward*. Retrieved October 21, 2019, from Medium: <https://medium.com/towards-artificial-intelligence/artificial-intelligence-salaries-heading-skyward-e4ib2a7bba7d>
- Stathoulopoulos, K., & Mateos-Garcia, J. (2019). *Gender Diversity in AI Research*. London: Nesta.

- Steel, E., & Angwin, J. (2010, August 4). *On the Web's Cutting Edge Anonymity in Name Only*. Retrieved October 17 2019, from The Wall Street Journal: <https://www.wsj.com/articles/SB10001424052748703294904575385532109190198>
- Stokel-Walker, C. (2019, August 24). *Facebook's Ad Data May Put Millions of Gay People At Risk*. Retrieved October 21, 2019, from NewScientist: <https://www.newscientist.com/article/2214309-facebooks-ad-data-may-put-millions-of-gay-people-at-risk/>
- Stray, J. (2018, December). *More Algorithmic Accountability Reporting, and a lot of it Will be Meh*. Retrieved October 21, 2019, from Nieman Lab: <https://www.niemanlab.org/2018/12/more-algorithmic-accountability-reporting-and-a-lot-of-it-will-be-meh/>
- Tala. (2019). *About*. Retrieved October 18, 2019, from Tala: <https://tala.co/about/>
- Tech2News. (2019, January 2). *Researchers Create AI Worker so Real it even Cheats when Given Tough Tasks*. Retrieved October 18, 2019, from First Post: <https://www.firstpost.com/tech/science/researchers-create-ai-worker-so-real-it-even-cheats-when-given-tough-tasks-5822551.html>
- TensorFlow Lagos. (2019). *TensorFlow Lagos*. Retrieved October 18, 2019, from TensorFlow Lagos: <https://twitter.com/TensorflowLagos>
- The Economist. (2017, December 7). *Google leads in the race to dominate artificial intelligence*. Retrieved October 16, 2019, from The Economist: <https://www.economist.com/business/2017/12/07/google-leads-in-the-race-to-dominate-artificial-intelligence>
- The Guardian. (2019, May 6). *Nigerian develops AI platform that can translate over 2000 African languages*. Retrieved October 17 2019, from The Guardian: <https://www.theguardian.com/news/2019/may/06/nigerian-develops-ai-platform-that-can-translate-over-2000-african-languages/>
- The New York Stem Cell Foundation. (2019). *Institutional Report Card for Gender Equality*. Retrieved October 18, 2019, from The New York Stem Cell Foundation: [https://nyscf.org/wp-content/uploads/2017/06/2017\\_IWISE-Report-Card-i.pdf](https://nyscf.org/wp-content/uploads/2017/06/2017_IWISE-Report-Card-i.pdf)
- The Observatory of Economic Complexity. (2019). *Semiconductor Devices*. Retrieved October 16, 2019, from The Observatory of Economic Complexity: <https://oec.world/en/profile/hs92/854i/>
- The Presidency of the Republic of Ghana Communications Bureau. (2019, January 21). *Latest News*. Retrieved October 18, 2018, from The Presidency of the Republic of Ghana: <https://www.presidency.gov.gh/index.php/briefing-room/news-style-2/1059-president-akufo-addo-outlines-gov-ts-pillars-of-growth-for-science-and-technology>
- The Royal Society. (2018). *The Impact of Artificial Intelligence on Work*. London: The Royal Society.

- The University of Sheffield. (2018, December 20). *Threats to journalism explored during inaugural UNESCO lecture*. Retrieved October 18, 2019, from The University of Sheffield News: <https://www.sheffield.ac.uk/news/nr/unesco-chair-lecture-journalism-safety-impunity-1.822760>
- Thomas, R. (2015, July 27). *If You Think Women in Tech Is Just A Pipeline Problem, You Haven't Been Paying Attention*. Retrieved October 21, 2019, from Medium: <https://medium.com/tech-diversity-files/if-you-think-women-in-tech-is-just-a-pipeline-problem-you-haven-t-been-paying-attention-cb7a2073b996>
- Thomas, R. (2016, October 4). *The Real Reason Women Quit Tech (and How to Address It)*. Retrieved October 21, 2019, from Medium: <https://medium.com/tech-diversity-files/the-real-reason-women-quit-tech-and-how-to-address-it-6dfb606g29fd>
- Tonetti, C. J. (2018). *Nonrivalry and the Economics of Data*. Society for Economic Dynamics (pp. 477-489). Mexico City: Society for Economic Dynamics.
- Tschan, I., & Bekkoenova, A. (2018, November 22). *What does Artificial Intelligence mean for the future of Human Rights?* Retrieved October 21, 2019, from UNDP Europe and Central Asia: <https://www.eurasia.undp.org/content/rbec/en/home/blog/2018/what-does-artificial-intelligence-mean-for-the-future-of-human-r.html>
- Twitter Help Center. (2019, March). *Terrorism and violent extremism policy*. Retrieved September 4, 2019, from Twitter Help Center: <https://help.twitter.com/en/rules-and-policies/violent-groups>
- UN OHCHR (2018a, 24 May). *UN experts call on India to protect journalist RanaAyyub from online hate campaign*. Retrieved September 4, 2019, from United Nations Human Rights Office of the High Commissioner: [https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23126&L\\_angID=E](https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23126&L_angID=E)
- UN OHCHR. (2018b, September 3). *UN human rights expert says Facebook's 'terrorism' definition is too broad*. Retrieved September 4, 2019, from United Nations Human Rights Office of the High Commissioner: [https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23494&L\\_angID=E](https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23494&L_angID=E)
- UN Secretary-Generals High Level Panel on Digital Cooperation. (2019). *The Age of Digital Interdependence*. New York: The UN Secretary-General's High Level Panel on Digital Cooperation.
- UNESCO. (2005). *Towards Knowledge Societies*. Paris: UNESCO Publishing.
- UNESCO. (2012). *Policy Guidelines for the Development and Promotion of Open Access*. Paris: UNESCO.
- UNESCO. (2014). *Priority Africa at UNESCO: An Operational Strategy for Its Implementation 2014-2021*. Paris: UNESCO.
- UNESCO. (2015a). *Keystones to Foster Inclusive Knowledge Societies*. Paris: UNESCO Publishing.

- UNESCO. (2015b). *Knowledge societies: The way forward to build a better world*. Retrieved October 18, 2019, from UNESCO: [https://en.unesco.org/70years/knowledge-societies\\_way\\_forward\\_better\\_world](https://en.unesco.org/70years/knowledge-societies_way_forward_better_world)
- UNESCO. (2015c). *UNESCO Science Report: Towards 2030*. Paris: UNESCO.
- UNESCO. (2016, May 17). *Report Shows Homophobic and Transphobic Violence in Education to be A Global Problem*. Retrieved October 18, 2019, from UNESCO: <https://en.unesco.org/news/report-shows-homophobic-and-transphobic-violence-education-be-global-problem>
- UNESCO. (2017a). *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*, Paris: UNESCO. UNESCO. (2017b). *Protecting Journalism Sources in the Digital Age*. Paris: UNESCO.
- UNESCO. (2018a). *39 C/5 Approved Program and Budget 2018-2019*. Retrieved October 18, 2019, from UNESCO: [https://unesdoc.unesco.org/ark/48223/pf0000261648](https://unesdoc.unesco.org/ark:/48223/pf0000261648)
- UNESCO. (2018b). *Forum on Artificial Intelligence in Africa*. Retrieved October 18, 2019, from UNESCO: <https://en.unesco.org/artificial-intelligence/africa-forum>
- UNESCO. (2018c, July-September). *Lexicon for Artificial Intelligence*. Retrieved October 18, 2019, from UNESCO: <https://en.unesco.org/courier/2018-3/lexicon-artificial-intelligence>
- UNESCO. (2018d, November 16). *UNESCO hosts a workshop on Artificial Intelligence for Human Rights and SDGs at 2018 Internet Governance Forum*. Retrieved October 18, 2019, from UNESCO News: <https://en.unesco.org/news/unesco-hosts-workshop-artificial-intelligence-human-rights-and-sdgs-2018-internet-governance>
- UNESCO. (2019a). *UNESCO's Internet Universality Indicators: A Framework for Assessing Internet Development*. Paris: UNESCO.
- UNESCO. (2019b). *World Press Freedom Day*. Retrieved October 18, 2019, from UNESCO: <https://en.unesco.org/commemorations/worldpressfreedomday>
- UNESCO. (2019c, May 19). 206 EX/42. *Decisions Adopted by The Executive Board at Its 206<sup>th</sup> Session*. Paris, France.
- UNESCO. (2019d, October 9-23). 207 EX/SR.6. *Decisions Adopted by The Executive Board at Its 207<sup>th</sup> Session*. Paris, France.
- UNESCO. (2019e). *UNESCO's Internet Universality ROAM-X Indicators*. Retrieved October 18, 2019, from UNESCO: <https://en.unesco.org/themes/internet-universality-indicators>
- UNESCO. (2019f, September 28). *International Day for Universal Access to Information*. Retrieved October 18, 2019, from UNESCO: <https://en.unesco.org/commemorations/accesstoinformationday>

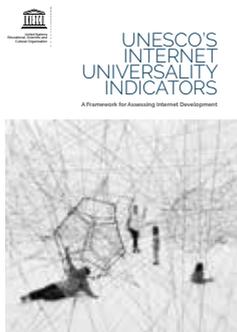
- UNESCO. (2019g, September 9). *UNESCO Engages Technology and Policy Experts for Human Centered AI in Africa*. Retrieved October 18, 2019, from UNESCO: [https:// en.unesco.org/news/unesco-engages-technology-and-policy-experts-human-centered-ai-africa](https://en.unesco.org/news/unesco-engages-technology-and-policy-experts-human-centered-ai-africa)
- UNESCO. (2019h). *In Focus Series of World Trend in Freedom of Expression and Media Development*. Retrieved November 15, 2019, from UNESCO: [httpsv/en.unesco.org/worLd-media-trends](https://en.unesco.org/world-media-trends)
- UNESCO Const, art. I, § 2, cl a. (1945, November 16). London, United Kingdom.
- UNESCO Institute of Statistics. (2019). *Information and communication technologies (ICT)*. Retrieved October 18, 2019, from UNESCO Institute of Statistics: [http:// uis.unesco.org/en/gLossary-term/information-and-communication-techno-Lo-gies-ict](http://uis.unesco.org/en/glossary-term/information-and-communication-technologies-ict)
- UNESCO; EQUALS Skills Coalition. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*, Paris: EQUALS.
- UNFE. (2013). *Definitions*. Retrieved November 4, 2019, from United Nations: Free and Equal: [httpsv/www.unfe.org/definitions/](https://www.unfe.org/definitions/)
- UNGA. (2015). *Information and communications technologies for development -A/70/L33*. New York: United Nations.
- UNGA A/73/348. (2018, August 29). *Promotion and Protection of The Right to Freedom of Opinion*. New York, United States of America.
- UNGA A/HRC/22/i7/Add.4. (2013, January 11). *Annual report of the United Nations High Commissioner for Human Rights*. New York, United States of America.
- UNGA A/HRC/38/35. (2018, April 6). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. Retrieved August 30, 2019, from United Nations: [https://documents-dds-ny.un.org/doc/ UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement](https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement)
- UNGA A/RES/68/167 (2019, December 18). *The right to privacy in the digital age. United Nations General Assembly* <https://undocs.org/A/RES/68/167>
- UNGA A/RES/70/1 para. 16 & 17. (2015, September 25). *Transforming our world: the 2030 Agenda for Sustainable Development*. New York, United States.
- UNGA A/RES/70/125 para.3. (2015, December 16). *Resolution adopted by the General Assembly on 16 December 2015*. New York, United States of America.
- UNGA Resolution 217 (1948, December 10). *Universal Declaration of Human Rights*. Paris, France.
- UNGA Resolution 2200A (XXI). (1966, December 19). *International Covenant on Civil and Political Rights*. New York, United States of America.
- UNHRC. (2011, July 21). *General Comment No. 34*. Geneva, Switzerland.
- UNHRC A/HRC/17/31. (2011). *Guiding Principles on Business and Human Rights*. Geneva: United Nations Publishing Service.

- UNHRC A/HRC/26/29. (2014, April 14). *Report of The Special Rapporteur on The Rights to Freedom of Peaceful Assembly and of Association*, Maina Kiai. New York, United States of America.
- UNHRC A/HRC/39/29. (2018, August 3). *The Right to Privacy in The Digital Age*. New York, United States of America.
- UNHRC A/HRC/40/63. (2019, February 27). New York, United States of America.
- UNHRC HRI/GEN/I/Rev.g (Vol. I). (1988, April 8). *General Comment No. 16 The Right to Respect of Privacy Family, Home and Correspondence, and Protection of Honour and Reputation*. New York, United States.
- United for Gender Parity. (2018). *Making Progress Toward Achieving Gender Parity*. Retrieved October 18, 2019, from United for Gender Parity: <https://www.un.org/gender/>
- United Nations Population Division. (2019). *World Population Prospects 2019*. Retrieved October 18, 2019, from United Nations Population: <https://population.un.org/wpp/>
- UNSG. (2017, September 15). *Interview: UN should be flagbearer when it comes to gender parity, stresses top official*. Retrieved October 21, 2019, from United Nations: <https://news.un.org/en/story/2017/09/564982-interview-un-should-be-flagbearer-when-it-comes-gender-parity-stresses-top>
- UNSG. (2018, July 12). *Secretary-General's High-level Panel on Digital Cooperation*. Retrieved October 21, 2019, from United Nations: <https://www.un.org/en/digital-cooperation-panel/>
- Value Colleges. (2019). *Top 50 Best Value Bachelor's in Artificial Intelligence and Machine Learning*. Retrieved October 18, 2019, from Value Colleges: <https://www.valuecolleges.com/ranking/best-value-machine-learning-bachelors/>
- Van der Spuy A. (2017). *What if we all governed the Internet? Advancing multistakeholder participation in Internet governance*. Paris: UNESCO.
- Vandenhoe, W. (2005). *Non-Discrimination and Equality in the View of the UN Human Rights Treaty Bodies*. Oxford, UK: Intersentia.
- Varian, H. (2018, June). *Artificial Intelligence, Economics, and Industrial Organisation*. In A. Agrawal, J. Gans, & A. Goldfarb, *The Economics of Artificial Intelligence: An Agenda* (pp. 399-419). Chicago: Chicago University Press.
- Varian, H. (2019, June). *Artificial Intelligence, Economics, and Industrial Organisation*. In A. Agrawal, J. Gans, & A. Goldfarb, *The Economics of Artificial Intelligence: An Agenda* (pp. 399-419). Chicago: Chicago University Press.
- Varley, C. (2018, April 6). *Are Sex Robots Just Turning Women into Literal Objects* Retrieved October 21, 2019, from BBC: <https://www.bbc.co.uk/bbcthree/article/8b-be0749-62ee-40fg-a8ac-a2d75lc474f6>

- Vayana, E., & Tasioulas, J. (2016). *The Dynamics of Big Data and Human Rights: The Case of Scientific Research*. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 28,1-14.
- Vincent, J. (2017 December 5). *Tencent says there are only 300,000 AI engineers worldwide, but millions are needed*. Retrieved October 21, 2019, from The Verge: <https://www.theverge.com/2017/12/5/16737224/global-ai-talent-shortfall-tencent-report>
- Vincent, J. (2017, August 22). *Transgender YouTubers Had Their Videos Grabbed to Train Facial Recognition Software*. Retrieved October 21, 2019, from The Verge: <https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset>
- Wajcman, J. (1991). *Feminism Confronts Technology*. Cambridge: Polity Press.
- Wajcman, J. (2010). *Feminist Theories of Technology*. *Cambridge Journal of Economics*, 34(1), 143-152.
- Wallis, J. (2018, September 27). *Is Artificial Intelligence Sexist?* Retrieved October 21, 2019, from The Globe and Mail: <https://www.theglobeandmail.com/business/careers/leadership/article-is-artificial-intelligence-sexist/>
- Wang, Y., & Kosinski, M. (2018). *Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation from Facial Images*. *Journal of Personality and Social Psychology*, 114(2), 246-257.
- Waters, S. (2018). *The Effects of Mass Surveillance on Journalists' Relations with Confidential Sources: A Constant Comparative Study*. *Digital Journalism*, 6(10), 1294-1313.
- Wiggers, K. (2018, December 28). *A researcher trained AI to generate Africa masks*. Retrieved October 21, 2019, from Venture Beat: <https://venturebeat.com/2018/12/28/a-researcher-trained-ai-to-generate-africa-masks/>
- Wilson, C. (2011, October 25). *Stanford's machine-learning class tackles the lazy hiker principle*. Retrieved October 21, 2019, from Slate: <https://slate.com/technology/2011/10/stanfords-machine-learning-class-tackles-the-lazy-hiker-principle.html>
- Wilson, M. (2019, March 19). *The World's First Genderless AI Voice Is Here. Listen Now*. Retrieved October 21, 2019, from Fast Company: <https://www.fastcompany.com/90321378/the-worlds-first-genderless-ai-voice-is-here-listen-now>
- Wolff, H. E. (2018, February 7). *How the continent's languages can unlock the potential of young Africans*, Retrieved October 21, 2019, from The Conversation: <http://theconversation.com/how-the-continent-languages-can-unlock-the-potential-of-young-africans-90322>
- Wollerton, M., & Crist, R. (2018, September 20). *Amazon Echo Dot, Basics Microwave, Echo Sub: Everything Amazon just announced*. Retrieved October 21, 2019, from CNET <https://www.cnet.com/news/new-amazon-echo-dot-microwave-sub-alexa-hardware-event-september-20-2018/>

- Women in AI. (2019). *About Women in AI*, Retrieved October 18, 2019, from Women in AI: <https://www.womeninai.co/>
- Women's Forum for The Economy & Society. (2019). *2019 Presentation*, Paris: Women's Forum for The Economy & Society.
- Working Group on Artificial Intelligence. (2018). *Mauritius Artificial Intelligence Strategy* Port Louis: Government of Mauritius.
- World Bank. (2013). *Inclusion Matters: The Foundation for Shared Prosperity*. Washington D.C.: The World Bank.
- WSIS-05/TUNIS/DOC/6(Rev. i)-E art. 80 & 87 (2005, November 18). *Tunis Agenda for The Information Society*. Tunis, Tunisia.
- Young, H. (2015, May 25). *The digital language divide*. Retrieved October 21, 2019, from The Guardian: <https://www.theguardian.com/education/ng-interactive/2015/may/28/language-barrier-internet-experience>
- Yulianto, B., & Shidarta, S. (2015). *Philosophy of Information Technology: Sex Robot and Its Ethical Issues*. International Journal of Social Ecology and Sustainable Development, 6(4), 67-76.
- Zenvus. (2019). *Home*. Retrieved October 18, 2019, from Zenvus: <https://www.zenvus.com/>
- Zerrour, L. (2019, March 18). *50 Millions DH pour Promouvoir La Recherche en Intelligence Artificielle*. Retrieved October 21, 2019, from Aujourd'hui Le Maroc: <http://aujourd'hui.ma/societe/50-millions-dh-pour-promouvoir-la-recherche-en-in-telligence-artificielle>
- Zuckerberg, M. (2018, November 15). *A Blueprint for Content Governance and Enforcement*. Retrieved October 21, 2019, from Facebook: <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>

# INDICADORES DE LA UNESCO SOBRE LA UNIVERSALIDAD DE INTERNET



Apuntar a una Internet que promueva las normas de los derechos humanos, apoye las sociedades del conocimiento inclusivas y fomente el desarrollo sostenible, esa es la base del concepto de universalidad de Internet, que ha guiado el posicionamiento de la UNESCO en cuestiones vinculadas a Internet desde 2015. Los Indicadores de la UNESCO sobre la universalidad de Internet constituyen una herramienta integral para evaluar el estado del desarrollo de Internet en un determinado país, permitiendo que surjan nuevas percepciones y mejoras de política basadas en la evidencia. La publicación presenta los 303 indicadores en 6 categorías, fuentes y medios de verificación, así como una guía de implementación para las evaluaciones nacionales voluntarias.

## SERIE DE LA UNESCO SOBRE LA LIBERTAD EN INTERNET



Desde 2011, la UNESCO ha publicado 11 ediciones como parte de la serie flagship sobre la libertad en Internet. Esta serie explora las cambiantes cuestiones legales y de política relacionadas con Internet, al tiempo que ofrece a los Estados Miembros y a otros interesados recomendaciones de política con el objetivo de promover un ambiente propicio para la libertad de expresión en la red. Estas publicaciones incluyen:

- *What if we all governed the Internet? Advancing multistakeholder participation in Internet governance (2017)*
- *Survey on Privacy in Media and Information Literacy with Youth Perspectives (2017)*
- *Protecting Journalism Sources in the Digital Age (2017)*
- *Human rights and encryption (2016)*
- *Privacy, free expression and transparency: Redefining their new boundaries in the digital age (2016)*
- *Principles for governing the Internet: A comparative analysis (2015)*
- *Countering online hate speech (2015)*
- *Building digital safety for journalism: A survey of selected issues (2015)*
- *Fostering freedom online: The role of Internet intermediaries (2014)*
- *Global survey on Internet privacy and freedom of expression (2013)*
- *Freedom of connection, freedom of expression: The changing legal and regulatory ecology shaping the Internet (2011)*

Todas las publicaciones pueden descargarse del siguiente sitio web:  
<http://en.unesco.org/unesco-series-on-internet-freedom>

# El aporte de la inteligencia artificial y las TIC avanzadas a las sociedades del conocimiento

Este informe reconoce la inteligencia artificial (IA) como una oportunidad para alcanzar los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas, gracias a su contribución al desarrollo de sociedades del conocimiento inclusivas.

Basándose en el marco DAAM de universalidad de Internet de la UNESCO, acordado por los Estados Miembros de la UNESCO en 2015, este estudio analiza

- cómo la IA y las tecnologías de la información y las comunicaciones (TIC) avanzadas impactarán en los Derechos Humanos en términos de libertades de expresión, privacidad, medios de comunicación, periodismo y no discriminación;
- cómo la Apertura debe informar los desafíos tecnológicos y de seguridad relacionados con la IA;
- cómo el Acceso a la IA depende de algoritmos, hardware, recursos humanos y datos, y
- cómo un enfoque de múltiples actores respecto de la gobernanza de la IA puede abordar los desafíos y las oportunidades para beneficio de la humanidad.

El estudio también ofrece un conjunto de opciones para la acción que pueden ayudar a informar el desarrollo de nuevos marcos de política, y un nuevo examen de las políticas en vigor, así como de otras acciones para todos los interesados: los Estados Miembros, el sector privado, la comunidad técnica, la sociedad civil y las organizaciones intergubernamentales, incluida la UNESCO.



Organización  
de las Naciones Unidas  
para la Educación,  
la Ciencia y la Cultura

Sector de la  
Comunicación  
e Información

