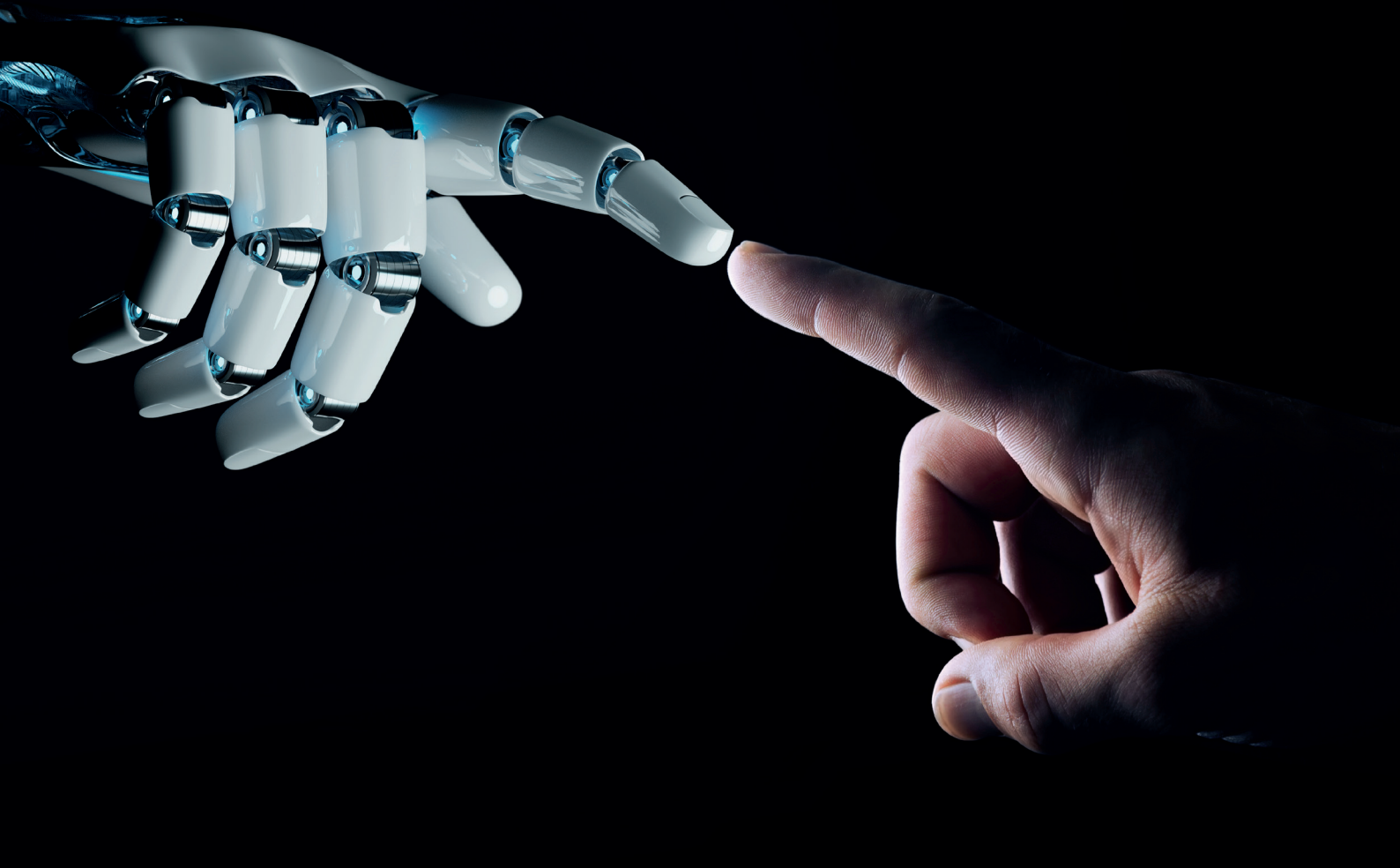




USAID
FROM THE AMERICAN PEOPLE



GUÍA DE INTELIGENCIA ARTIFICIAL (IA) ÉTICA



Esta Guía es un recurso para personas formuladoras de políticas en sectores locales que buscan comprender los retos, oportunidades y riesgos de la adopción de la IA.

Incluye una definición de IA, preocupaciones éticas relacionadas a la IA y cómo éstas pueden abordarse.

También incluye ejemplos para ilustrar los retos éticos de la IA, enfoques para abordar estos problemas, junto con preguntas de reflexión. Específicamente, esta Guía utiliza el sector de la educación como ejemplo para explorar los elementos de la IA ética, pero las lecciones también se pueden generalizar a otros sectores. Finalmente, la Guía incluye un Glosario y un Anexo con detalles adicionales de los estudios de casos destacado en el documento.

EXENCIÓN DE RESPONSABILIDAD Este informe es posible gracias al generoso apoyo del pueblo estadounidense a través de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID). Los contenidos son responsabilidad de DAI y no reflejan necesariamente los puntos de vista de USAID o del Gobierno de los Estados Unidos. Esta publicación/informe/guía fue producida bajo el Proyecto Fronteras Digitales de DAI (Acuerdo de Cooperación AID-OAA-A-17-00033) a pedido de USAID

CONTENTS

¿Qué es la IA?	4
¿Por qué utilizamos IA?	4
¿Cómo funciona?	4
¿Por qué la IA necesita datos?	6
¿Qué es la ética de la IA?	6
¿Por qué usamos IA en el sector educativo?	6
Por qué es importante la ética de la IA	7
¿Cuáles son los problemas morales relacionados con la IA?	8
¿Quién es responsable de los resultados de IA?	8
¿Por qué aparecen ALGUNOS problemas morales?	9
¿Cuáles son los riesgos?	10
¿Quiénes deben interesarse y conocer sobre la ética de la IA?.....	10
¿Los sistemas de IA están sesgados globalmente?	10
Cómo aplicar la ética de la IA	11
Glosario	16
Anexo: casos éticos de IA en detalle	18
Soluciones éticas de IA.....	20

¿QUÉ ES LA IA?

La inteligencia artificial (o IA) es un concepto que se refiere a algoritmos informáticos que resuelven problemas utilizando técnicas asociadas con la inteligencia humana: razonamiento lógico, representación del conocimiento, procesamiento del lenguaje y reconocimiento de patrones. La IA se usa a menudo para crear diferentes aplicaciones, desde simples *chatbots* hasta complejos programas de predicción de terremotos o delitos, entre muchas otras posibilidades.

¿POR QUÉ UTILIZAMOS IA?

La IA ofrece una interesante extensión de muchas de las capacidades humanas, tales como observación, procesamiento y toma de decisiones. Las salidas y resultados de los sistemas de IA ofrecen eficiencia y eficacia a los humanos que de otra manera no serían posibles. La capacidad de cómputo y los sistemas que utilizan IA pueden exceder capacidades cognitivas humanas, realizando un aprendizaje continuo, sin supervisión humana y basado en patrones que son humanamente imposibles de discernir (por ejemplo, la capacidad de identificar a las personas en función de su modo de andar sin siquiera verles la cara). Utilizando un impulso dinámico, la IA también puede crear incentivos instantáneos para el cumplimiento normativo (por ejemplo, una selección guiada de beneficios diseñada para promover un comportamiento económico específico).

¿CÓMO FUNCIONA?

En un sistema de IA, algoritmos predefinidos siguen una serie de instrucciones para transformar conjuntos de datos en resultados que pueden utilizarse para tomar decisiones, ya sea por el propio sistema informático o por un ser humano. Muchos algoritmos de IA aprenden directamente de los datos, mediante la identificación de patrones y relaciones, sin instrucciones basadas en reglas fijadas por personas como en el caso de las estadísticas tradicionales.¹ La “caja negra” de los sistemas de IA se refiere a las entradas y operaciones del sistema que no son visibles para el usuario final u otras partes y, a veces, incluso a las personas científicas de datos que construyen sistemas de IA. Los algoritmos que pueden actuar de forma independiente sin control o supervisión de humanos se consideran autónomos.

¿CUÁL ES LA DIFERENCIA ENTRE IA Y APRENDIZAJE AUTOMÁTICO?

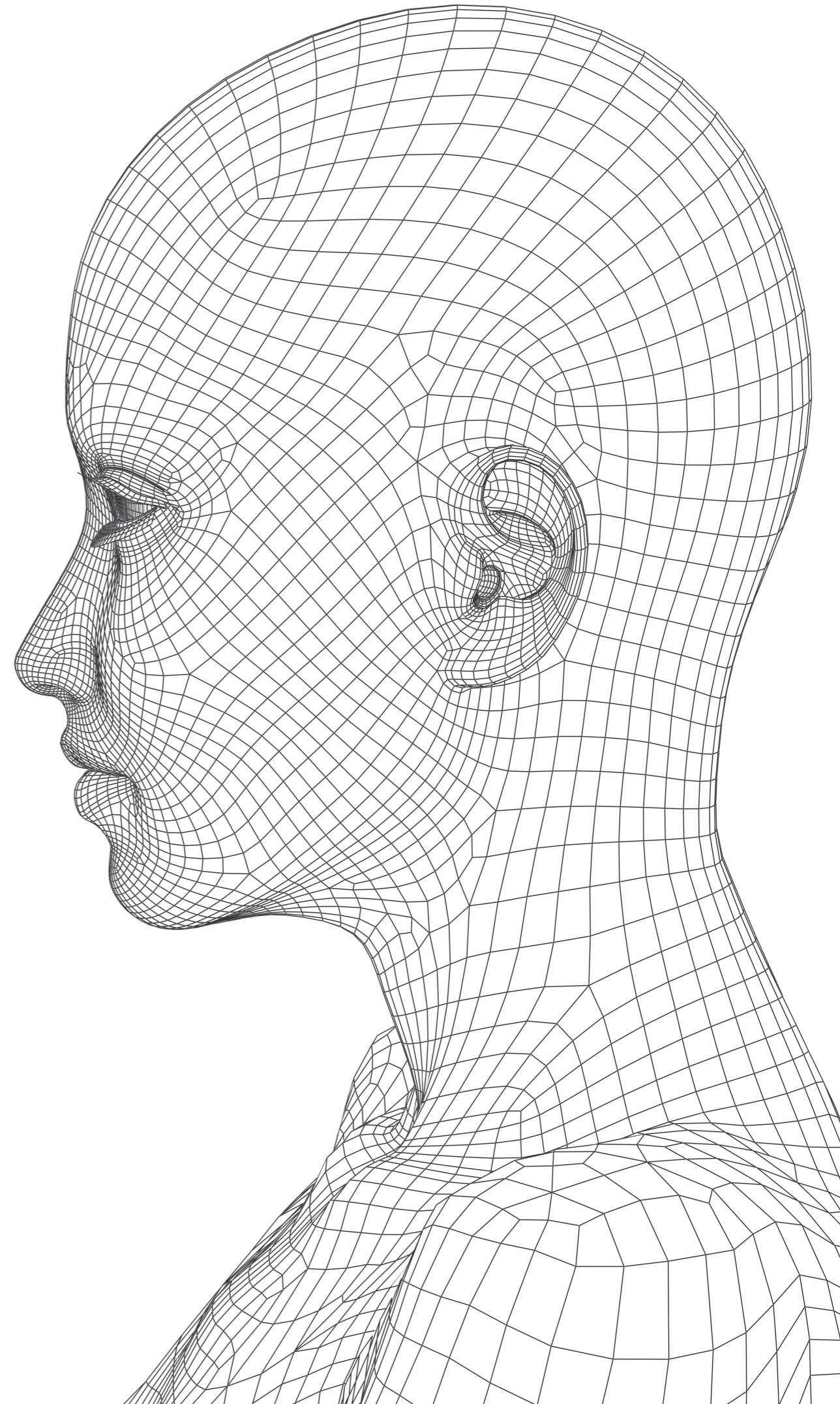
El aprendizaje automático (*Machine Learning*, o ML por sus siglas en inglés) habilita los sistemas de IA. Si bien esta técnica se enfoca en el aprendizaje y la predicción, las aplicaciones de IA a menudo se basan en las predicciones de ML para crear, planificar o tener un impacto en el mundo real. Las decisiones automatizadas pueden implementarse directamente (como en la robótica) o sugerirse a una persona tomadora de decisiones (como las recomendaciones de productos en las compras en línea).²

Para reflexionar

La IA depende en gran medida de los datos que recibe para la toma de decisiones. Si el conjunto de datos de entrada está sesgado o es corrupto, también lo serán las recomendaciones del sistema de IA.

¹ [Prasanna Lal Das \(2022\). Algorithms in Government: A Magic Formula or a Divisive Force?](#)

² [USAID Digital Strategy \(2020-2024\).](#)



¿POR QUÉ LA IA NECESITA DATOS?

Grandes conjuntos de datos, como los generados por un sistema educativo estatal o nacional, habilitan la IA. El concepto de IA tiene sus raíces en la década de 1950, y hoy está en boga debido a una mayor cantidad de datos y capacidad de cómputo. Por ejemplo, los conjuntos de datos sobre movimientos de población, la propagación de epidemias y el cambio climático, entre muchos otros, pueden mejorar significativamente las capacidades de la IA para monitorear tendencias emergentes y proveer evidencia para avanzar en los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas.

¿QUÉ ES LA ÉTICA DE LA IA?

La ética es la ciencia del comportamiento adecuado. Aristóteles argumentó que la ética es el estudio de las relaciones humanas en su forma más perfecta. Afirmó que es la base para crear un modelo óptimo de interrelaciones humanas y justas, asegurando una comunicación óptima entre las personas y un punto de referencia para crear una estructura de conciencia moral. La práctica de la ética de la IA es la consideración de problemas morales relacionados con la interacción de la tecnología, los humanos y la sociedad. Así como la IA evoluciona rápidamente, también lo hacen las consideraciones éticas de la IA; sin embargo, el objetivo general de la ética de la IA es buscar crear un modelo óptimo de interrelaciones entre humanos y tecnología.

¿POR QUÉ USAMOS IA EN EL SECTOR EDUCATIVO?

Los sistemas de IA modifican la naturaleza de los sistemas educativos a través de la capacidad de analizar conjuntos de datos grandes y complejos, automatizar procesos, comprender los cambios en el rendimiento de los estudiantes, crear nuevas formas de interacción entre estudiantes y profesores, personalizar métodos de aprendizaje y más. Simultáneamente, la posibilidad de una implementación descuidada de la IA crea graves riesgos éticos con consecuencias negativas a largo plazo, genera prejuicios, puede inhibir la motivación e incluso desencadenar malestar social.

Ver más definiciones de conceptos relacionados a la ética de la IA en el Glosario al final de la Guía.

Para reflexionar

El acceso a datos suficientes, imparciales y de buena calidad es un factor fundamental para lograr una IA ética y confiable.



Para reflexionar

No existe una definición común de la ética de la IA.

Para reflexionar

La ética de la IA actual se trata más de las preguntas correctas que de las respuestas correctas.

POR QUÉ ES IMPORTANTE LA ÉTICA DE LA IA

Mientras que la IA puede traer múltiples beneficios para las personas, hay algunas consideraciones éticas que no pueden ser ignoradas. Dado que estos sistemas se utilizan cada vez más en todos los sectores, los sistemas de IA pueden tener efectos significativos en la obtención de crédito, en el empleo, la educación, la competencia y más. Sin embargo, sin la ética incrustada en los algoritmos, es casi imposible garantizar que no causen más daño que bien. Con una adopción más amplia de esta tecnología en los últimos años, se ha propagado desconfianza en la información pública y en los aparatos de gobierno, considerándole responsable de perpetuar la discriminación en la prestación de servicios y el perfilar desfavorablemente a segmentos de la población, planteando numerosas preocupaciones morales.

La responsabilidad de las personas científicas de datos o los equipos de desarrollo de software y otros participantes en el ciclo de vida de la IA³ raramente se extiende a la ética de la IA. Éstos con frecuencia prestan más atención a qué tan bien un sistema o producto está realizando su función prevista que a las posibles implicaciones sociales y éticas. Por eso es importante involucrar a diferentes grupos de partes interesadas, como la sociedad civil, para garantizar un diseño e implementación éticos de los sistemas de IA.

CASO DE ESTUDIO: SOLUCIONES IMPERFECTAS DE IA EN LA EDUCACIÓN

País: Reino Unido **Año:** 2020

¿Qué sucedió? El algoritmo de inscripción de estudiantes favoreció a los estudiantes de escuelas privadas y áreas prósperas, dejando a los estudiantes de alto rendimiento de las escuelas públicas gratuitas afectados de manera desproporcionada.

¿Por qué es importante para los propósitos de esta guía? Muchos formuladores de políticas se enfocan solo en los efectos positivos que la IA aporta al sector de la educación, cuando primero se debería priorizar en la identificación y mitigación de riesgos.

¿CUÁLES SON LOS PROBLEMAS MORALES RELACIONADOS CON LA IA?

Los algoritmos de IA pueden ser opacos, complejos y sujetos a errores, sesgos, elaboración de perfiles, discriminación y otras prácticas injustas. Esto sucede, en parte, porque los algoritmos son creados por personas y las personas no son objetivas. Otra razón por la que esto podría suceder es porque existen sesgos y prejuicios históricos en los conjuntos de datos de los que aprenden los sistemas de IA, y si las personas no abordan los sesgos de datos, los sistemas de IA pueden replicarlos. Los problemas éticos aparecen en muchos aspectos de la adopción de la IA. Por ejemplo, en escenarios en donde los vehículos autónomos deben de tomar decisiones ante un choque inevitable;⁴ en casos donde los sesgos raciales, de género y edad en el reconocimiento facial; lo mismo con prejuicios hacia personas en situación de discapacidad; o bien en prácticas discriminatorias en los sectores bancario y de seguros, basadas en conjuntos de datos sesgados. Hay casos en los que las personas culpan a los algoritmos por injusticias que impactan negativamente sus vidas, como que sus hijos se vean privados de la admisión a una universidad o que jueces nieguen la libertad bajo fianza que depende de sistemas automatizados.⁵

¿QUIÉN ES RESPONSABLE DE LOS RESULTADOS DE IA?

Las personas interactúan con los algoritmos como curadoras de bases de datos, creadoras de algoritmos y consumidoras de sus resultados, así como modelos a seguir. Las personas pueden proporcionar un punto de control para el algoritmo, ya sea poniendo a prueba o validando las decisiones tomadas por éstos. Pero la responsabilidad de los errores de la IA es una gran preocupación moral. Hay debates en curso sobre quién es responsable de cualquier error o consecuencia negativa de los sistemas de IA: ¿la persona que desarrolló el algoritmo de IA, la propietaria del dispositivo inteligente, la operadora de los datos, quienes proporcionaron los datos de entrenamiento o alguien más? Incluso hay debates sobre si los algoritmos de IA pueden ser considerados como inventores o no. La falta de certidumbre en este frente puede seguir aumentando la desconfianza pública, generando riesgos y generando consecuencias duraderas para la innovación y la adopción de tecnología.

4 [Dilema del tranvía](#): "Un tranvía fuera de control se precipita por las vías y matará a cinco personas inocentes en el camino. Puede accionar un interruptor que dirigirá el carro a una vía diferente. Pero otro hombre está parado en esa segunda vía, y apretar el interruptor lo llevará a la muerte".

5 [Racism in, Racism out. A primer on algorithmic racism](#). Lilah Burke (2020). [The Death and Life of an Admissions Algorithm. Pretrial Algorithms \(Risk Assessment\)](#).

CASO DE ESTUDIO: SISTEMAS DE RECONOCIMIENTO FACIAL DEFECTUOSOS

País: Estados Unidos de América **Año:** 2018

¿Qué sucedió? El software de reconocimiento facial de Google tenía un sesgo contra personas afroamericanas. Los algoritmos de reconocimiento de imágenes en Google Photos clasificaban a estas personas como "gorilas".

¿Por qué es importante para los propósitos de esta guía? Actualmente, numerosas entidades del gobierno local se plantean el uso de sistemas de reconocimiento facial, sin tener en cuenta los riesgos y posibles efectos de su uso, como ilustra este caso.

Ver detalles en el anexo.

Para reflexionar

¿Cómo resuelven las máquinas el dilema del carro?

CASO DE ESTUDIO: IMPORTANCIA DE LOS DATOS PARA LA SALUD DE LAS MUJERES

¿Qué sucedió?: A medida que la IA se usa más ampliamente en el sistema de salud, la falta de datos de calidad suficientes sobre la fisiología de las mujeres y las intervenciones médicas y los resultados reproducirán los diagnósticos erróneos actuales.

¿Por qué es importante? La exclusión común de las mujeres como sujetos de prueba en muchas investigaciones médicas da como resultado una falta de datos relevantes sobre la salud de las mujeres que tiene implicaciones críticas. Esto a menudo no es considerado por los implementadores.

Ver detalles en el anexo.

¿POR QUÉ APARECEN ALGUNOS PROBLEMAS MORALES?

Las aplicaciones de IA como *chatbots*, sistemas de evaluación criminal, sistemas de reconocimiento facial, entre otros, se vuelven sexistas o racistas debido a los sesgos inherentes a los conjuntos de datos y, a menudo, a los datos de fuente abierta. El sistema de IA emite conclusiones a partir de los datos con los que es entrenado. Cuando los sistemas de aprendizaje profundo se entrenan en datos abiertos, nadie puede controlar exactamente lo que aprenden. Los datos históricos de entrenamiento a menudo están sesgados, es posible que no incluyan datos sobre mujeres, personas con discapacidad, pueblos indígenas o cualquier otro grupo que históricamente se encuentran en el lado desfavorable de la brecha digital. En esos casos, los algoritmos pueden considerar estas poblaciones como valores atípicos y crear modelos que conducen a errores y problemas éticos de manera que y esas poblaciones seguirán siendo invisibles en la era moderna.

Para reflexionar

Si la IA comete un error médico y genera un diagnóstico erróneo que conduce a la muerte, ¿quién es la persona responsable?

CASO DE ESTUDIO: PRÁCTICA DISCRIMINATORIA EN LA CONTRATACIÓN DE MUJERES

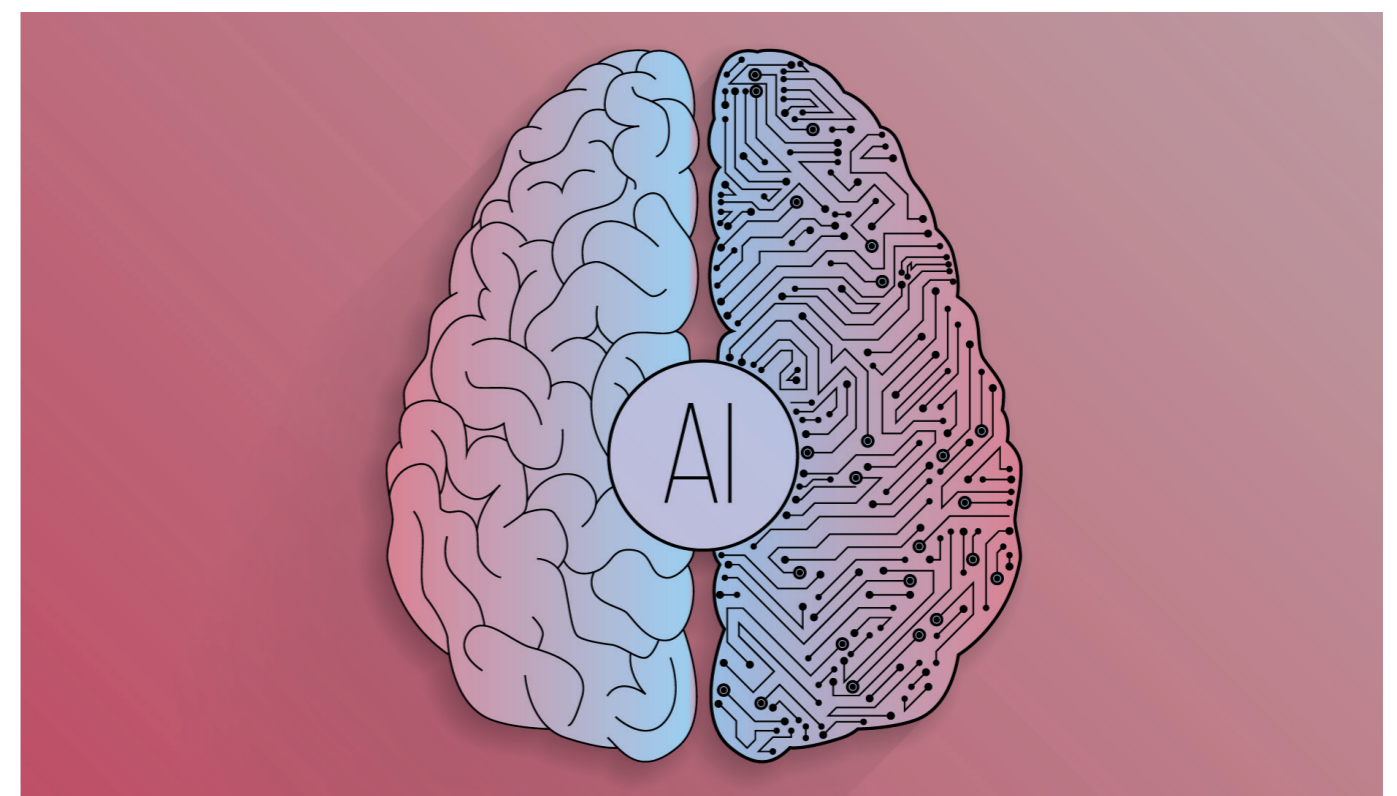
País: Estados Unidos de América

Año: 2018

¿Qué pasó?: Amazon abandonó su programa de contratación con IA debido a su sesgo contra las mujeres, porque se entrenó con los currículos de los candidatos que en su mayoría eran hombres.

¿Por qué es importante para efectos de esta guía? Equipos de Recursos Humanos tanto en los sectores privado como público están apostando al aprovechamiento de IA para eficientar los procesos de contratación, sin tener en cuenta ejemplos de casos como éste y los efectos negativos de una contratación sesgada.

Ver detalles en el anexo.



¿CUÁLES SON LOS RIESGOS?

Los problemas éticos pueden tener consecuencias distintas con diferentes niveles de severidad. Estos incluyen todo, desde el aumento de la desigualdad hasta los procesos de litigio extendidos y la resistencia social. La creación de perfiles y los sesgos de los algoritmos contra una raza o género en particular, o una categoría específica de personas, pueden afectar el funcionamiento de cualquier sistema, ya sea en la educación, la atención médica, las finanzas o incluso la democracia. Los valores humanos fundamentales como la privacidad personal, la protección de datos, la equidad y la autonomía se ponen en peligro. La IA también se puede usar de manera maliciosa y causar daños en casi cualquier campo para falsificar datos, robar contraseñas, interferir con el trabajo de otro software y máquinas, lo que socava aún más la confianza en la tecnología.

¿QUIÉNES DEBEN INTERESARSE Y CONOCER SOBRE LA ÉTICA DE LA IA?

Todos. La ciudadanía, las empresas, los gobiernos y el mundo académico son todas personas usuarias o adoptantes de aplicaciones y tecnologías de IA. Todos enfrentan retos éticos que impactan sus vidas de manera diferente. Y todos tienen una perspectiva para agregar a la conversación ética más amplia de IA para nuestra sociedad.

CASO DE ESTUDIO: IA CONTRA LAS PERSONAS CON DISCAPACIDAD

País: Caso global

¿Qué sucedió?: Las personas con discapacidad pueden interpretarse como valores atípicos en una aplicación de IA que puede imitar la discriminación directa e indirecta a la que se enfrentan en la sociedad.

¿Por qué es importante para efectos de esta guía? Al considerar el desarrollo y adopción de un sistema de IA a nivel local, es importante identificar si se cuenta con datos representativos que incluyan a personas en situación de discapacidad para evitar que puedan ser excluidos de los resultados o recomendaciones de dicho sistema.

Ver detalles en el anexo.

¿LOS SISTEMAS DE IA ESTÁN SESGADOS GLOBALMENTE?

El desarrollo de sistemas de IA, la recopilación de datos y la creación de estándares están ocurriendo en el Norte Global, donde se encuentran la mayoría de las instituciones de investigación de IA y grandes empresas tecnológicas. Un gran número de países del Sur Global aún experimentan pobreza de datos y están en proceso de consolidación de una infraestructura digital confiable y robusta. Por lo tanto, existen preocupaciones acerca de si las ideas occidentales de equidad deben considerarse universales⁶ y si se aplican de la misma manera en los países en desarrollo que en las economías avanzadas. Numerosas personas cuestionan la primacía de las tradiciones éticas occidentales en la mayoría de los sistemas de IA y se preguntan si la incorporación de creencias éticas basadas en sistemas alternativos inspirados, por ejemplo, en el budismo o Ubuntu, podría cambiar algunas suposiciones sobre la ética de la IA. Sin embargo, algunas naciones toman ciertas medidas para cambiar esta tendencia, un ejemplo son los Principios Maoríes para la Soberanía de Datos.⁷

CÓMO APLICAR LA ÉTICA DE LA IA

Los gobiernos de todo el mundo han comenzado a desarrollar estrategias tanto para aprovechar las oportunidades como para mitigar los riesgos relacionados con el uso de la IA. Un método que han empleado es el desarrollo de pautas éticas para el uso de IA y otros sistemas automatizados. Si bien muchas de estas guías son producidas por instituciones u organismos gubernamentales que carecen de atribuciones legales para su aplicación, sientan la base de reglas, comienzan a establecer estándares y crean conciencia entre los distintos actores gubernamentales. Además, las organizaciones del sector privado, las instituciones académicas y las organizaciones no gubernamentales⁸ también han creado sus propios principios éticos o pautas para un público más amplio. A continuación se comparten una serie de buenas prácticas en el avance hacia una IA ética.

I. CREAR CONCIENCIA E INICIAR EL DIÁLOGO

La ética de la IA es igualmente importante para el gobierno, las empresas y las personas; debe discutirse abierta y ampliamente. Sin embargo, el tema está fuera del radar para muchos. Es por eso que un buen punto de partida sería lanzar actividades de concientización y traer a la mesa a las distintas partes interesadas. Es importante garantizar que exista la participación y el debate más amplios posibles sobre cuestiones de ética de la IA, incluidos los organismos nacionales, subnacionales y municipales, las empresas (tanto grandes empresas como las pequeñas y medianas), las asociaciones profesionales y la ciudadanía.

CASO DE ESTUDIO: ALEMANIA, COMISIÓN DE ÉTICA

País: Alemania **Año:** 2018

¿Qué sucedió?: La Comisión de Ética de Datos se estableció para producir puntos de referencia éticos, pautas y recomendaciones para el desarrollo y uso de IA para el gobierno.

¿Por qué es relevante para efectos de esta guía? Esta buena práctica puede ser contextualizada por gobiernos locales que busquen contar con grupos de trabajo especializado en la materia.

Ver detalles en el anexo.



⁶ Decolonial AI Manifesto

⁷ Maori Principles for Data Sovereignty

⁸ Véase el booklet para emprendimientos y empresas [Ética de la IA para América Latina](#) de C Minds y Meta. Disponible en

2. ADOPTAR PRINCIPIOS RECTORES

Muchos países, ciudades y organizaciones han desarrollado principios y pautas, códigos de ética y herramientas de autoevaluación sobre IA para personas funcionarias públicas y organizaciones empresariales⁹. Al centro de cada documento hay una lista de principios rectores que no están normados por la ley pero ayudan a dar forma a la aplicación ética de la IA. En general, los siguientes cinco principios predominan en las distintas propuestas

Transparencia	No maleficencia y confianza	Equidad y justicia
La IA tiene un uso abierto y comprensible, auditable y documentado, de cualquier dato con medidas de seguridad claras y actualizadas.	La IA no daña o inflige el menor daño posible para alcanzar un resultado. Se garantiza la preparación continua para interferir en los sistemas de IA.	Los sistemas y soluciones de IA se subordinan a reglas y leyes definidas por humanos, lo que garantiza los derechos humanos por encima de todo.
Autonomía y responsabilidad	Explicabilidad	
El sistema de IA revela quién es responsable de los trabajos y soluciones de IA. Al mismo tiempo, se refiere a que los humanos ceden voluntariamente parte del poder de toma de decisiones a la IA	Los sistemas de IA garantizan la necesidad de comprender y exigir responsabilidades en los procesos de toma de decisiones de la IA.	

Hay otros principios adoptados por diferentes organizaciones y que se encuentran en documentos reconocidos internacionalmente, por ejemplo, en la Declaración de Montreal para una IA responsable¹⁰ o los principios de la Organización para la Cooperación y el Desarrollo Económicos (OCDE)¹¹, que vale la pena conocer en detalle.

CASO DE ESTUDIO: CANADÁ ESTABLECE PRINCIPIOS RECTORES

País: Canadá **Año:** 2018

¿Qué sucedió?: La Junta del Tesoro desarrolló un conjunto de principios rectores para ayudar a los funcionarios gubernamentales a explorar la IA de una manera que esté “gobernada por valores, ética y leyes claras”.

¿Por qué es relevante para efectos de esta Guía? Se pueden tomar en cuenta aspectos y aprendizajes de la experiencia canadiense en el diseño de principios rectores alineados a la realidad local de gobiernos estatales y municipales.

Ver detalles en el Anexo.

⁹ Véase, por ejemplo, la [Moratoria de las Naciones Unidas sobre el uso de IA que amenace los derechos humanos](#), o la herramienta de autoevaluación [fAIR LAC a la mano](#).

¹⁰ [The Montréal Declaration for responsible AI development \(2018\)](#)

¹¹ [Principios de IA de la OCDE](#)

3. DESARROLLAR UN MARCO LEGAL PARA LA ÉTICA DE LA IA

Casi todas las iniciativas desarrolladas en la actualidad han resultado en pautas y marcos que son recomendados y no obligatorios. El desarrollo de estándares éticos para el procesamiento de datos utilizando IA puede ser un paso importante hacia el establecimiento de una base legal para la IA ética. El estándar puede adoptar la forma de recomendaciones metodológicas para desarrollar códigos éticos para las organizaciones que participan en el desarrollo y adopción de tecnologías de IA.

CASO DE ESTUDIO: DECLARACIÓN DE DERECHOS DE IA

País: Estados Unidos de América

Año: 2022

¿Qué sucedió? Se estableció la Declaración de Derechos de IA y se introdujeron las protecciones que las personas deben de tener ante su uso.

¿Por qué es relevante para efectos de esta guía? Se presentan 5 principios que deben guiar el diseño, uso e implementación de sistemas automatizados para proteger a la ciudadanía. Estos principios pueden ser considerados como una buena práctica a considerar en el proceso de identificación de principios aplicables en contextos locales.

Ver detalles en el anexo.

4. REDACTAR LINEAMIENTOS ESPECÍFICOS DEL SECTOR PARA EL USO DE LA IA

La ética de la IA varía cuando se trata de sectores específicos. Los principios comunes de equidad y justicia pueden no ser suficientes cuando se trata de atención médica y diagnóstico médico, ventas de comercio electrónico o educación. Es fundamental acordar un conjunto de principios y directrices relevantes para diferentes sectores. Un acuerdo sobre este tema puede ser beneficioso para la aplicación de la IA de forma ubicua.

CASO DE ESTUDIO: LIBRO DE APRENDIZAJE

País: Reino Unido

Año: 2020

¿Qué sucedió?: El gobierno desarrolló un Libro de aprendizaje para autoridades públicas para comprender mejor cómo funciona la IA y cómo se puede aplicar en el sector público.

¿Por qué es relevante para efectos de esta guía? El Reino Unido es un país líder en el desarrollo de lineamientos para el uso responsable y ético de la IA. Este ejercicio representa una buena práctica que agrupa ejes rectores en la materia.

Ver detalles en el anexo.

5. FOMENTAR LA TRANSPARENCIA Y EL DEBATE SOBRE NUEVOS USOS DE DATOS CON IMPLICACIONES ÉTICAS A MEDIDA QUE SURJAN

La economía de datos es un campo altamente dinámico. La adopción de leyes, normas y estándares a menudo no sigue el ritmo de la introducción y el uso de tecnologías. Los debates permanentes y prácticos sobre las implicaciones éticas emergentes son importantes si se desea una transformación digital sostenible y basada en datos. Una estrecha colaboración con la sociedad civil podría coadyuvar al monitoreo de los cambios y garantizar la transparencia.

CASO DE ESTUDIO: EXPLICACIONES ABIERTAS DE IA

País: Singapur **Año:** 2020

¿Qué sucedió? La startup de IA UCARE.AI reveló los parámetros exactos utilizados en el desarrollo del modelo de IA a su clientela en el sector salud, declarando el uso de IA en sus procesos de análisis y predicción.

¿Por qué es relevante para efectos de esta guía? Este ejemplo sienta un precedente de caso de negocio exitoso donde hay un balance entre transparencia en el uso de IA y privacidad de los datos de personas usuarias.

Ver detalles en el anexo.

6. COMPROMETERSE CON EL DESARROLLO DE PRINCIPIOS INTERNACIONALES DE ÉTICA DE IA

Países como el Reino Unido, Canadá, Alemania, Japón, Argentina, Emiratos Árabes Unidos y organizaciones internacionales como la Comisión Europea, la OCDE, GSMA y muchas otras contribuyen activamente a la agenda de ética de IA al emitir códigos de ética, pautas, etc., u organizar consorcios como el *Global Partnership for AI* (GPAI) para trabajar juntos en vías y soluciones basadas en valores para la IA.¹² Están en el límite de darse cuenta de cómo la adopción ética de la IA puede garantizar avances tecnológicos y crecimiento económico inclusivo y sostenible. Unir fuerzas con socios internacionales es importante para facilitar la adopción de IA confiable.

CASO DE ESTUDIO: PRINCIPIOS DE IA DE LA OCDE

País: Global **Año:** 2019

¿Qué sucedió? Los Principios de IA de la OCDE cuentan entre sus adherentes a 38 países miembro (incluidos México y otros 3 países de la región) y los socios no miembros. Estas directrices formaron la base de los Principios de IA del G20.

¿Por qué es relevante para efectos de esta guía? Es importante que el gobierno local que está explorando el potencial uso de IA tenga en consideración si su país se ha adherido a los principios de la OCDE para tomarlos como punto de partida de cualquier desarrollo de políticas públicas.

Ver detalles en el anexo.

7. UTILIZAR LAS INICIATIVAS DEL SECTOR PRIVADO PARA EL BIEN PÚBLICO

El sector privado también está implementando herramientas de IA disponibles públicamente que ayudan a mitigar los riesgos de IA. Los ejemplos incluyen la herramienta AI Fairness 360 de IBM, People + AI Research (PAIR) de Google, el kit de herramientas de auditoría y sesgo Aequitas de la Universidad Carnegie Mellon, Fairlearn, Datasheets for Datasets de Microsoft y Model Card de Tensorflow. Por ejemplo, la herramienta IBM AI Fairness es un completo conjunto de herramientas de métricas de código abierto para verificar sesgos no deseados en conjuntos de datos y modelos de aprendizaje automático, y algoritmos de última generación para mitigar dicho sesgo y que muchos profesionales aplican.¹³ USAID también apoya estos esfuerzos. En 2021, presentó la Gestión de Proyectos de Aprendizaje Automático en el Desarrollo Internacional, un conjunto de herramientas prácticas para aquellos en gobiernos o sectores públicos que trabajarán con científicos de datos para desarrollar herramientas de IA.¹⁴



¹² [The Global Partnership on Artificial Intelligence \(GPAI\)](#)

¹³ [AI Fairness 360](#)

¹⁴ [AID Managing Machine Learning Projects in International Development: A Practical Guide](#)

GLOSARIO

Las definiciones a continuación se derivan de múltiples estudios de USAID¹⁵, a menos que se especifique lo contrario.

Adopción: Cambios que ocurren cuando las personas o instituciones comienzan a usar una nueva tecnología y la incorporan en sus rutinas o procesos existentes. Por ejemplo, las personas que usan una cuenta de dinero móvil para recibir remesas y pagar facturas se considerarían “adoptantes”, mientras que las que hacen un retiro único para vaciar una cuenta de transferencia de efectivo no lo harían.

Algoritmo: Es un procedimiento paso a paso para convertir cualquier entrada dada en salidas útiles. Un algoritmo informático sigue una serie de instrucciones para transformar las entradas (datos) en salidas que pueden utilizarse para tomar decisiones, ya sea por el sistema informático o por un ser humano.

Internet de las cosas (IoT): Se refiere a dispositivos y máquinas conectados que recopilan datos, los conectan con análisis inteligentes y adaptan su comportamiento/respuestas en función de la información en la red de comunicación. Los teléfonos inteligentes son dispositivos IoT.

Ciberseguridad: La prevención de daños, protección y restauración de computadoras, sistemas de comunicaciones electrónicas, servicios de comunicaciones electrónicas, comunicaciones por cable y comunicaciones electrónicas, incluida la información contenida en ellos, para garantizar su disponibilidad, integridad, autenticación, confidencialidad y no repudio.

Higiene cibernética: Las prácticas y los pasos que toman las personas usuarias de computadoras y otros dispositivos para mantener la salud del sistema y mejorar la seguridad en línea. Estas prácticas a menudo son parte de una rutina para garantizar la seguridad de la identidad y otros detalles que podrían ser robados o corrompidos¹⁶.

Alfabetización digital: La capacidad de acceder, administrar, comprender, integrar, comunicar, evaluar y crear información de manera segura y adecuada a través de dispositivos digitales y tecnologías en red para participar en la vida económica y social¹⁷. Esto puede incluir competencias que se denominan alfabetización informática, alfabetización en tecnología de la información y las comunicaciones, alfabetización en información y alfabetización mediática.

Aprendizaje automático: Es el proceso estadístico de derivar una regla o patrón de una gran cantidad de datos para predecir datos futuros.

Datos abiertos: Se refiere a los datos disponibles gratuitamente y almacenados deliberadamente en un formato de datos fácil de leer, particularmente por otras computadoras, y por lo tanto reutilizables.

Privacidad de datos: El derecho de un individuo o grupo a mantener el control y la confidencialidad de la información sobre ellos mismos, especialmente cuando esa intrusión resulta de la recopilación y el uso indebido o ilegal de datos sobre la o las personas en cuestión.

Protección de datos: La práctica de garantizar la protección de los datos contra el acceso, uso, divulgación, interrupción, modificación o destrucción no autorizados, para brindar confidencialidad, integridad y disponibilidad.

Economía digital: El uso de la infraestructura digital y de Internet por parte de individuos, empresas y gobiernos para interactuar entre sí, participar en actividades económicas y acceder a bienes y servicios digitales y no digitales. Una diversa gama de tecnologías y plataformas facilitan la actividad en la economía digital; sin embargo, gran parte de la actividad depende en cierta medida de Internet, los teléfonos móviles, los datos digitales y los pagos digitales.

Infraestructura digital: Los componentes fundamentales que permiten las tecnologías y los servicios digitales. Los ejemplos de infraestructura digital incluyen cables de fibra óptica, torres de telefonía móvil, satélites, centros de datos, plataformas de software y dispositivos de usuario final.

Plataforma: Un grupo de tecnologías utilizadas como base sobre la cual se pueden construir otras tecnologías o ejecutar aplicaciones y servicios. Por ejemplo, Internet es una plataforma que habilita aplicaciones y servicios web.

¹⁵ [USAID DECA Toolkit](#) | [USAID Digital Strategy](#) | [USAID Managing Machine Learning Projects in International Development: A Practical Guide](#)

¹⁶ [What is Cyber Hygiene? A Definition of Cyber Hygiene, Benefits, Best Practices, and More](#)

¹⁷ [A Global Framework of Reference on Digital Literacy Skills for Indicator](#)

ANEXO: CASOS ÉTICOS DE IA EN DETALLE

Ejemplos de casos relacionados a la ética de la IA, tanto buenas prácticas como impactos negativos.

LA IA DEGRADA A LOS ESTUDIANTES

En el Reino Unido, en un contexto de pandemia, el algoritmo de matriculación de estudiantes favoreció a los estudiantes de escuelas privadas y áreas prósperas, lo que afectó de manera desproporcionada a los alumnos de alto rendimiento de las escuelas públicas. A muchos estudiantes se les han revocado sus plazas universitarias como resultado de este sistema: el 40% de los estudiantes recibieron notas inferiores a las previstas. Cuando salieron a la calle y a los juzgados a reclamar, la situación obligó al gobierno a retractarse de las notas. Las revisiones posteriores sugirieron que los algoritmos podrían haber estado sesgados (reforzando los prejuicios en los datos históricos, además de favorecer a las escuelas más pequeñas). Los críticos también discreparon con las limitadas herramientas de participación y responsabilidad que el gobierno proporcionó a los estudiantes y padres. [Fuente.](#)

ENFRENTAR PROBLEMAS DE PRIVACIDAD Y VIGILANCIA MIENTRAS SE INTEGRA LA IA EN LA EDUCACIÓN

La inteligencia artificial puede ayudar a los estudiantes a obtener comentarios útiles más rápido y, entre otras cosas, reducir la carga de los profesores. Los sistemas de inteligencia artificial pueden rastrear cómo el usuario interactúa con las cosas; la experiencia resultante proporciona una experiencia personalizada. En educación, esto puede incluir sistemas que identifiquen fortalezas y debilidades y patrones en el desempeño de los estudiantes. Si bien los maestros hacen esto hasta cierto punto en su enseñanza, monitorear y rastrear las conversaciones en línea y las acciones de los estudiantes también puede limitar la participación de los estudiantes y hacerlos sentir inseguros cuando asumen la responsabilidad de sus ideas. [Fuente.](#)

SISTEMAS DE RECONOCIMIENTO FACIAL DEFECTUOSOS

Google es uno de los líderes en IA; sin embargo, su software de reconocimiento facial mostró un sesgo

contra personas afroamericanas. En varios casos declarados, los algoritmos de reconocimiento de imágenes en Google Photos clasificaban a las personas afroamericanas como “gorilas”. La empresa ha elegido limitar el servicio para evitar correr el riesgo de otra clasificación errónea. [Fuente.](#)

LA FALTA DE DATOS DESAGREGADOS POR GÉNERO CONDUCE A RESULTADOS FATALES

Los maniqués de prueba de accidentes automovilísticos se introdujeron por primera vez en la década de 1950 y durante décadas se basaron en el percentil 50 de hombres. El maniquí más utilizado mide 1,77 m de alto y pesa 76 kg (significativamente más alto y más pesado que una mujer promedio). Esta es una de las razones por las que las mujeres tienen un 17% más de probabilidades de morir y un 47% más de sufrir lesiones en accidentes que los hombres. En otro ejemplo, un estudio de 12 de los monitores de actividad física más comunes encontró que subestimaron los pasos durante las tareas domésticas hasta en un 74 por ciento y las calorías consumidas durante estas tareas hasta en un 34 por ciento. Las mujeres están significativamente excluidas de la investigación médica porque los hallazgos no están desagregados por género. La falta de datos desagregados significa que la mitad de la población está mal representada o no tiene igual acceso a servicios o tecnologías públicas o privadas. [Fuente.](#)

PRÁCTICA DISCRIMINATORIA EN LA CONTRATACIÓN DE MUJERES

Amazon abandonó su programa de contratación de IA debido a su sesgo contra las mujeres. El algoritmo comenzó a entrenarse con los currículos de los candidatos a puestos de trabajo de los últimos diez años. Debido a que la mayoría de los solicitantes eran hombres, desarrolló un sesgo para preferir a los hombres y penalizó las características asociadas con las mujeres. El programa no logró eliminar el sesgo de género porque estaba profundamente integrado en los conjuntos de datos. [Fuente.](#)

LA IA POLICIAL ACTÚA CONTRA LAS PERSONAS DE RAZA NEGRA

Un estudio de las herramientas de IA que utilizan las autoridades estadounidenses para determinar la probabilidad de que un delincuente reincida descubrió que los algoritmos producían resultados diferentes para las personas de raza negra y blanca en las mismas condiciones. Además, la controversia en torno al uso de la IA por parte de las fuerzas del orden, una vigilancia policial significativamente predictiva, ha generado fuertes críticas debido a su efecto discriminatorio. [Fuente.](#)

BOT DE IA RACISTA

En 2016, Microsoft lanzó un bot de IA en Twitter que podía interactuar y aprender de las personas usuarias de la plataforma de redes sociales. Sin embargo, se volvió racista y sexista a las pocas horas de conocer los datos abiertos en Twitter. Microsoft lo cerró menos de un día después de su lanzamiento. [Fuente.](#)

SERVICIOS DE TRADUCCIÓN HABILITADOS POR IA SESGADOS HACIA LOS HOMBRES

La herramienta de Google Translate cambia sistemáticamente el género de las traducciones. Los estereotipos se cuelan en las traducciones porque Google optimiza las traducciones para el idioma inglés. Como mostró algún experimento, en muchos casos, la herramienta cambia el género de la palabra. Por ejemplo, la frase “vier Historikerinnen und Historiker” (cuatro historiadores masculinos y femeninos) se traduce como “cuatro historiadores” (cuatro historiadores masculinos) en español, con resultados similares en italiano, francés y polaco. Las historiadoras son eliminadas del texto. [Fuente.](#)

IA Y SALUD DE LA MUJER

La exclusión común de las mujeres como sujetos de prueba en investigaciones médicas da como resultado una falta de datos relevantes sobre la salud de las mujeres. La enfermedad cardíaca, por ejemplo, se ha considerado tradicionalmente como una enfermedad predominantemente masculina, con pautas clínicas “basadas en evidencia” basadas en la fisiología masculina. Esto ha llevado a enfermedades cardíacas masivas mal diagnosticadas o subdiagnosticadas en mujeres. A medida que la IA se usa más ampliamente en el sistema de salud, la falta de datos de calidad suficientes sobre la fisiología de las mujeres y las intervenciones médicas y los resultados reproducirán los diagnósticos erróneos actuales. [Fuente.](#)

IA CONTRA LAS PERSONAS CON DISCAPACIDAD

Hay ejemplos de discriminación involuntaria contra las personas con discapacidad por parte de las aplicaciones de IA. Las personas con discapacidad pueden interpretarse como valores atípicos, o como una aplicación de IA que puede imitar la discriminación directa e indirecta a la que se enfrentan en la sociedad. Por ejemplo, los sistemas de inteligencia artificial programados con datos de empleados anteriores pueden interpretar una discapacidad revelada como una característica negativa si los solicitantes anteriores con discapacidades fueron frecuentemente excluidos en las primeras etapas. [Fuente.](#)

EMBARAZO ADOLESCENTE “PREDICHO” POR IA

En 2018, el Ministerio de la Primera Infancia de la provincia de Salta en Argentina y Microsoft presentaron un sistema algorítmico para predecir el embarazo adolescente. La llamaron Plataforma Tecnológica de Intervención Social. El objetivo era pronosticar qué mujeres de áreas de bajos ingresos quedarían embarazadas durante los siguientes cinco años utilizando el algoritmo. Nunca se hicieron explícitas las implicaciones de ser declarada “predestinada” a la maternidad, o cómo el saberlo ayudaría a prevenir la concepción adolescente. El sistema se basó en datos—incluyendo edad, etnia, país de origen, discapacidad y si la casa del sujeto tenía agua caliente en el baño—de 200.000 residentes en la ciudad de Salta, incluidas 12.000 mujeres y niñas entre las edades de 10 y 19. La Plataforma Tecnológica para la Intervención Social no fue objeto de una revisión formal, y sus efectos en las niñas y mujeres no han sido examinados debido a la ausencia total de una legislación nacional de IA. Hizo falta transparencia y rendición de cuentas en la publicación de información sobre los resultados; se descubrió que la base de datos del sistema sólo contenía información sobre grupos raciales y socioeconómicos, no sobre el acceso a la educación sexual o la anticoncepción, que son ampliamente reconocidos por las organizaciones de salud pública como los métodos más efectivos para reducir la tasa de embarazo adolescente. [Fuente.](#)

SOLUCIONES ÉTICAS DE IA

COMISIÓN ALEMANA DE ÉTICA DE DATOS

En Alemania, se estableció una Comisión de ética de datos para producir puntos de referencia éticos, pautas y recomendaciones para el desarrollo y uso de IA para el gobierno. El resultado de su trabajo fue una publicación de lineamientos éticos con un conjunto de [Recomendaciones Específicas para la Acción](#), con la intención de “proteger al individuo, preservar la cohesión social y salvaguardar y promover la prosperidad en la era de la información”. En particular, en Alemania, muchas organizaciones privadas establecen sus propias pautas generales. Por ejemplo, Deutsche Telekom estableció [Directrices para la inteligencia artificial](#) que describen cómo se debe utilizar la IA en Deutsche Telekom y cómo se deben desarrollar los productos basados en la IA.

PRINCIPIOS RECTORES DE CANADÁ

Canadá es ejemplar en el despliegue de herramientas para personas funcionarias públicas que les ayudan a explorar la IA de manera que se “rija por valores, ética y leyes claros”. El enfoque canadiense es integral: proporciona [principios rectores](#) para garantizar el uso ético de la IA, una lista de empresas que buscan vender soluciones de IA al gobierno y una [herramienta de autoevaluación](#). Esta última ayuda a los organismos gubernamentales a evaluar los riesgos de implementar un sistema de toma de decisiones automatizado. Se presenta en forma de un cuestionario de 80 puntos relacionado con procesos comerciales, datos, diseño de sistemas, algoritmos y decisiones de diseño de sistemas. Los resultados proporcionados por la evaluación informan al organismo sobre el impacto potencial de la IA propuesta y brindan información sobre los requisitos aplicables.

18 [National Artificial Intelligence Initiative Act of 2020](#)

19 [Algorithmic Discrimination Protections](#)

20 [S.2551 - AI Training Act](#)

LEYES DE IA EN LOS ESTADOS UNIDOS DE AMÉRICA

La Iniciativa Nacional de Inteligencia Artificial¹⁸ se adoptó en 2020. Sin embargo, esta iniciativa no ha definido el sesgo ni mencionado el sesgo de género a partir de esta revisión. Una nueva [Declaración de derechos de la IA](#) ha introducido cinco protecciones que las personas deben tener en la era de la IA. Específicamente, la Protección contra la discriminación algorítmica declara: “No debe sufrir discriminación por parte de los algoritmos, y los sistemas deben usarse y diseñarse de manera equitativa”.¹⁹ En octubre de 2022, la [Capacitación en Inteligencia Artificial para la Adquisición de Fuerza Laboral \(también conocida como Ley de Capacitación en IA\)](#)²⁰ fue adoptada por el poder ejecutivo. El propósito del programa es garantizar que la fuerza laboral pública tenga conocimiento de las capacidades y riesgos asociados con la IA. Varias agencias gubernamentales federales y estatales y el sector privado han lanzado iniciativas para prevenir el sesgo algorítmico en los sectores público y privado. Una iniciativa de la industria liderada por Data & Trust Alliance ha desarrollado el [Algorithmic Bias Safeguards for Workforce](#), un cuestionario estructurado que las empresas pueden usar para adquirir software para evaluar a los trabajadores

EXPLICACIONES ABIERTAS DE IA DE SINGAPUR

UCARE.AI, una empresa emergente con sede en Singapur, ofrece un producto de predicción de costos basado en IA en su plataforma que trabaja con hospitales para entregar estimaciones precisas de las facturas hospitalarias a los pacientes. Para generar una mayor confianza en el uso de la IA, UCARE.AI tuvo en cuenta ser transparente en su uso de la IA con las partes involucradas. UCARE.AI no solo reveló los parámetros exactos utilizados en el desarrollo del modelo de IA para su clientela, sino que también brindó explicaciones detalladas sobre todos los algoritmos que tuvieron algún impacto en las operaciones, los ingresos o la base de dicha clientela. Al darse cuenta de que la precisión de la proyección de facturas es apreciada por los hospitales y pacientes, UCARE.AI tomó la decisión consciente de declarar el uso de IA en su análisis y predicción de los montos de las facturas a personas administradoras de datos y sus pacientes. [Fuente](#).

LIBRO DE APRENDIZAJE PARA AUTORIDADES PÚBLICAS EN REINO UNIDO

El gobierno del Reino Unido desarrolló un [Libro de aprendizaje para autoridades públicas](#) para comprender mejor cómo funciona la IA y cómo se puede aplicar en el sector público. Un último documento ilustrativo son las [Directrices para la adquisición de IA](#) que proporcionan un conjunto de principios sobre cómo comprar tecnología de IA, así como información sobre cómo abordar los desafíos que pueden surgir durante la adquisición.

ÉTICA DE LA IA: UN TRABAJO CONJUNTO

Reconociendo que los temas relevantes para la IA trascienden las fronteras, los países también están adoptando cada vez más enfoques regionales para la IA, incluidos los esfuerzos coordinados en la Unión Europea y la Unión Africana, entre los estados nórdicos-bálticos y las naciones árabes, y dentro de el G7 y el G20. La OCDE también ha fortalecido sus esfuerzos relacionados con la IA en los últimos años, encabezados por el Observatorio de Políticas de IA de la OCDE. De hecho, los Principios de IA de la OCDE adoptados en 2019 son los primeros estándares intergubernamentales sobre IA. La OCDE creó sus directrices que proporcionan una lista de principios generales y recomendaciones de políticas con el objetivo de guiar a los gobiernos, organizaciones e individuos en el diseño y ejecución de sistemas de IA de una manera que priorice los mejores intereses de las personas y garantice que los diseñadores y operadores rindan cuentas por su correcto funcionamiento. [Fuente](#).



Creative Commons Atribución 4.0 Internacional (CC BY 4.0)

Este es un informe de acceso abierto distribuido bajo los términos de la licencia Creative Commons Atribución 4.0 Internacional (CC BY), que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se acredite a los autores originales y la fuente, y se indiquen cualquier modificación o adaptación.



USAID
FROM THE AMERICAN PEOPLE

U.S. Agency for International Development

1300 Pennsylvania Avenue, NW

Washington, DC 20523

Tel: (202) 712-0000

www.usaid.gov

