

Hacia la igualdad sustantiva en el ámbito de la inteligencia artificial (IA):

políticas transformadoras de IA para la
igualdad de género y la diversidad

Noviembre de 2024



GPAI

THE GLOBAL PARTNERSHIP
ON ARTIFICIAL INTELLIGENCE

Este informe se planificó antes de la integración de la Global Partnership on Artificial Intelligence (GPAI) y la Organización para la Cooperación y el Desarrollo Económico (OCDE) a mediados de 2024. Por consiguiente, el informe no se sometió a la aprobación de los miembros de la GPAI y la OCDE y no debe considerarse que refleja sus posiciones.

Agradecimientos

Este informe fue elaborado en el marco del proyecto “Hacia la Diversidad Real y la Igualdad de Género en el Ámbito de la Inteligencia Artificial: Prácticas Prometedoras y Recomendaciones basadas en Evidencia”, bajo la dirección de las personas coordinadoras del proyecto y la orientación del Grupo Consultor de Personas Expertas, con el apoyo del Grupo de Trabajo de la AGIA sobre Uso Responsable de la IA. El Grupo de Trabajo de la AGIA sobre Uso Responsable de la IA acordó desclasificar este informe y ponerlo a disposición del público.

Codirección:

Paola Ricaurte Quijano*,

Tecnológico de Monterrey y *Berkman Klein Center for Internet & Society, Harvard University*

Benjamin Prud’homme†,

Mila - *Quebec Artificial Intelligence Institute*

El informe fue preparado por **Isadora Hellegren Létourneau‡**, de la organización Mila, con las contribuciones de las personas codirectoras del proyecto **Paola Ricaurte Quijano*** —del Tecnológico de Monterrey y de *Berkman Klein Center for Internet & Society, Harvard University*— y **Benjamin Prud’homme†** —de la organización Mila—, así como con las de las siguientes personas que contribuyeron al proyecto en calidad de consultoras: **Anita Gurumurthy‡**, **Nandini Chami‡**, **Malavika Rajkumar‡** y **Merrin Muhammed Ashraf‡** —de *IT for Change*— y **Wanda Muñoz‡** —de la Red de Investigación Feminista en Inteligencia Artificial—.

Agradecemos a las personas integrantes del Grupo Consultor de Personas Expertas su dedicación y el apoyo brindado a lo largo del proyecto: **Lucia Velasco***, de *School of Transnational Governance, European University Institute*; **Ricardo Baeza-Yates***, de *Institute for Experiential AI, Northeastern University*; **Hawa Bi Khan†**, Consultoría de Atención a la Salud en *Digital Health, AI and Sustainability*; **Karen de Brouwer Vásquez†**, de la Secretaría de Relaciones Exteriores, México; **Caroline M. Coward†**, de *Jet Propulsion Laboratory/Caltech*; **Kudakwashe Dandajena***, de *African Institute for Mathematical Sciences (AIMS)* y de *University of the Western Cape*; **Gillian Dowie†**, de IDRC; **Laurent Elder†**, de IDRC; **Golnoosh Farnadi†**, de *McGill University* y de Mila; **Dafna Feinzholtz***, de la UNESCO; **Jessica Fjeld†**, de *De|Center* y de *Berkman Klein Center for Internet & Society, Harvard University*; **Karine Gentelet†**, de *Université du Québec en Outaouais (UQO) & International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA)*; **Alison Gillwald‡**, de *Research ICT Africa*; **Gloria Guerrero‡**, de ILDA; **Toshiya Jitsuzumi***, de *Chuo University*; **Nicole Kaniki†**, de *Senomi Solutions Inc.*; **Ching-Yi Liu***, de *National Taiwan University*; **Nicole Osayande†**, de Mila; **Michael O’Sullivan***, de *University of Auckland*; **Inese Podgaiska***, de *Association of Nordic Engineers*; **Catherine Régis***, de *Université de Montréal* y Mila; **Michael Running Wolf†**, de *Indigenous AI* y Mila; **Juliana Sakai***, de *Transparência Brasil*; **Patricia Shaw†**, de *Beyond Reach Consulting Limited*; **Prateek Sibal†**, de la UNESCO; **Andrew Spoprie†**, de *iNZight Analytics Ltd*; **Anupama Srikonda†**, de *Indian School of Business*; **Ruhiya Kristine Steward†**, de IDRC; **Elissa Strome†**, de *Canadian Institute for Advanced Research (CIFAR)*; **Jaco du Toit***, de la UNESCO; **Eliane Ubalijoro†**, de CIFOR-ICRAF; **Jamila Venturi†**, de *Derechos Digitales*; y **Jeff Ward***, de *Animikii Indigenous Technology*.

También deseamos expresar nuestro profundo agradecimiento a los miembros del Grupo Consultor de Personas Expertas por las importantes perspectivas, la experiencia y los conocimientos técnicos que aportaron para dar forma a las etapas iniciales del informe. Muchas gracias, **Anaelia Altagracia Ovalle**[‡], de *University of California, Los Angeles*; **Jake Okechukwu Effoduh**[‡], de *Lincoln Alexander School of Law of Toronto Metropolitan University*; **Maui Hudson (Whakatōhea)**[‡], de *Te Kotahi Research Institute, University of Waikato*; **Petra Molnar**[‡], de *Berkman Klein Center for Internet & Society, Harvard University, Refugee Law Lab, York University*; **Rosario Patricia Galarza Meza**[‡], de *International Disability Alliance (IDA)*; y **Umut Pajaro Velasquez**[‡], consultor e investigador independiente sobre la IA y la gobernanza de internet. También agradecemos a **Carolina Botero Cabrera**[‡], de *Karisma Foundation*, y a **Mathieu Marcotte**[‡], de CEIMIA, por compartir sus conocimientos especializados.

Quisiéramos extender nuestro agradecimiento a **Ana Gabriela Ayala Núñez**[‡], del Tecnológico de Monterrey; a **Ivanna Martínez Polo**[‡], del Tecnológico de Monterrey; a **Leslie Evelin Salgado Arzuaga**[‡], de *University of Calgary*; a **Shazade Jameson**[‡], de *Tilburg University*; a **Shamira Ahmed**[‡], de *Data Economy Policy Hub, Tilburg University*; a **Florian Lebre**[‡], de *Université Laval*; a **Gargi Sharma**[‡], de *CLIMA Fund*; a **Selene Yang**[‡], de la mujeres en Inteligencia Artificial y de Geochicas, y a **Razieh Shirzadkhani**[‡], de Mila, por sus trascendentes contribuciones de consultoría en varias etapas del proyecto.

Agradecemos a nuestras organizaciones aliadas en la ejecución del proyecto —Derechos Digitales, *Research ICT Africa* y *Data Pop Alliance*— por su dedicación para habilitar, organizar y llevar a cabo numerosas consultas con múltiples y diversas partes interesadas dispersas geográfica y culturalmente. Quisiéramos enfatizar nuestra gratitud hacia todas las personas que participaron en las consultas regionales y con grupos específicos, tanto anónima como públicamente (véase el apéndice).

Extendemos nuestro reconocimiento a las personas representantes de las organizaciones de la sociedad civil por ayudar a reconocer y presentar las voces de las comunidades marginadas. Muchas gracias, **Otto Mazariegos**[‡], de la Asociación Nacional de Ciegos de Guatemala y de la Red Latinoamericana de Organizaciones de Personas con Discapacidad y sus Familias (RIADIS); **Abia Akram**[‡], de *STEPS Pakistan*; **Andrea Gunraj**[‡], de *Canadian Womens' Foundation*; **Dragana Kaurin**[‡], de *Berkman Klein Center for Internet & Society, Harvard University*; **Pierrine Leukes**[‡], de *Research ICT Africa*; y **Nancy Gros-Louis McHugh**[‡], de *First Nations of Quebec and Labrador Health and Social Services Commission*.

Los resultados de este proyecto no habrían sido posibles sin los esfuerzos de quienes lo propusieron inicialmente: **Wanda Muñoz**[‡], de la Red de Investigación Feminista en Inteligencia Artificial; **Paola Ricaurte Quijano**[‡], del Tecnológico de Monterrey y de *Berkman Klein Center for Internet & Society, Harvard University*; **Catherine Régis**[‡], de *Université de Montréal* y Mila; **Inese Podgaiska**[‡], de *Association of Nordic Engineers*; **Juliana Sakai**[‡], de *Transparência Brasil*; **Toshiya Jitsuzumi**[‡], de *Chuo University*; **Ricardo Baeza-Yates**[‡], de *Institute for Experiential AI, Northeastern University*; **Celine Caira**[‡] y **Luis Aranda**[‡], de la OCDE; **Prateek Sibal**[‡] y **Vanessa Dreier**[‡], de la UNESCO; y **Benjamin Prud'homme**[‡], de Mila.

Quisiéramos agradecer a **Ekua Quansah**[‡], de *Quansah Consulting*, su trabajo en la revisión del lenguaje inclusivo [en la versión en inglés], a **Zofia Laubitz**[‡] por su trabajo de corrección de pruebas, a **Leonardo Studio Design**[‡] por el diseño gráfico del informe, a **Alvaro Muñoz**[‡] por su trabajo de traducción del inglés al español, y a **Ariana Seferiades**[‡], de Mila, por la revisión en español.

La AGIA desea reconocer los incansables esfuerzos de colegas que trabajan en Mila, en CEIMIA (*International Centre of Expertise in Montréal on Artificial Intelligence* o *Centre d'expertise internationale de Montréal en intelligence artificielle*) y en el Grupo de Trabajo de la AGIA sobre Uso Responsable de la IA. Agradecemos especialmente su dedicado trabajo a **Niobe Haitas**[‡], Directora de Proyectos de Mila, durante la etapa inicial del proyecto. Agradecemos profundamente a **Laëtitia Vu**, **Camille Seguin**, y a **Stephanie King**, de CEIMIA, y a **Anna Jahn**[‡], de Mila, su compromiso y su apoyo a lo largo del proyecto. Para concluir, valoramos enormemente la dedicación de quienes copresidieron el Grupo de Trabajo: **Amir Banifatemi**[†], de *AI Commons*, y **Francesca Rossi**[†], de *IBM Research*, así como a **Catherine Régis**^{*}, de *Université de Montréal* y de Mila, y **Raja Chatila**^{*}, de *Sorbonne University*, quienes ocuparon esos cargos anteriormente.

* Persona experta

** Persona observadora

† Persona especialista invitada

‡ Bajo contrato con el Centro de Apoyo Especialista para contribuir a sus proyectos

Este trabajo cuenta con licencia conforme a las condiciones que establece la Licencia de Atribución de Creative Commons 4.0 que autoriza su uso sin restricciones, siempre que se dé crédito a las personas autoras y a la fuente original. La licencia se encuentra disponible en <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>

Cómo citar este documento

AGIA 2024. *Hacia la igualdad sustantiva en el ámbito de la inteligencia artificial (IA): políticas transformadoras de IA para la igualdad de género y la diversidad*, Informe, Noviembre de 2024, Alianza Global sobre la Inteligencia Artificial.

Proyecto realizado con el apoyo de



Mila



ceimia

Contenido

Prólogo	7
Resumen ejecutivo	8
Recomendaciones clave para formular políticas transformadoras de IA	10
Introducción	13
PARTE I. Consultas, desafíos cruciales y vías posibles	18
Consultas Regionales y con Grupos Específicos	19
Consultas Regionales y con Grupos Específicos: Resultados	21
Desafíos cruciales para garantizar la igualdad de género y la diversidad en el ámbito de la IA	21
Amalgama entre acceso e inclusión	22
Exclusión del Conocimiento e Invisibilización	23
Distribución inequitativa de recursos	25
Vías para afrontar los desafíos en igualdad de género y diversidad en ecosistemas de IA	26
Desarrollar la capacidad del sector público y de la población en general	27
Incentivar el diseño inclusivo y democratizar las innovaciones	28
Garantizar la rendición de cuentas de las partes programadoras, proveedoras y responsables del despliegue	29
Garantizar la participación significativa de los grupos marginados	30
Breve resumen de los desafíos y de las vías	31
PARTE II. Marco conceptual	32
Marco conceptual	32
Un enfoque transformador de la igualdad sustantiva	33
Un enfoque socio-técnico de los sistemas de IA	36
El ciclo de vida de la IA	37
Las etapas del ciclo de vida de los sistemas de IA	37
Oportunidades para intervenir	38
El marco de políticas transformadoras en el ámbito de la IA	40
Dimensiones clave	40

PARTE III. Prácticas prometedoras y recomendaciones prometedoras	43
Prácticas prometedoras hacia una igualdad sustantiva en el ámbito de la IA	44
Recursos para la capacitación y la educación pública	45
El diálogo global de la UNESCO	45
<i>Data Justice Policy Brief</i>	46
<i>AI & Equality Human Rights Toolbox</i>	47
<i>Indigenous Pathfinders in AI</i>	48
Prácticas de diseño tecnológico inclusivo y de innovaciones democráticas	49
Metodología <i>Design from the Margins (DFM)</i>	49
Red de Investigación Feminista en Inteligencia Artificial (f<A+i>r)	50
Iniciativa <i>Fixing the bAIs</i>	51
<i>Indigenous Jobs Map</i>	52
Medidas aplicables a la rendición de cuentas	53
<i>Global Index on Responsible AI</i>	53
<i>The Migration and Technology Monitor</i>	55
<i>The Algorithmic Justice League</i>	56
Iniciativas de inclusión y participación significativas	57
<i>Māori Data Governance Model</i>	57
El taller de RIADIS	58
<i>The Gender and Responsible Artificial Intelligence Network (GRAIN)</i>	59
Recomendaciones clave para formular políticas transformadoras de inteligencia artificial	60
Diseños inclusivos e innovaciones democráticas	61
Participación significativa en la gobernanza de la IA	64
Transparencia y rendición de cuentas para prevenir daños	66
Acceso efectivo a la justicia	68
Notas de pie de página	71
Referencias	73
Apéndice: Lista de participantes en consultas regionales	80

Prólogo

La IA brinda grandes oportunidades para el bien social, pero esta abundancia de posibilidades puede y debe aprovecharse de forma equitativa. En este informe nos proponemos ofrecer una vía hacia la igualdad sustantiva en los ecosistemas de IA contemplando acciones que posibiliten la participación de todas las voces. Nuestro objetivo es transformar los paradigmas sobre quiénes pueden participar en el ámbito de la IA, dónde y cuándo a fin de democratizar su desarrollo, su despliegue y su gobernanza.

Las pruebas son claras: los sistemas de IA no son neutrales. Reproducen los modelos mundiales, los valores culturales, el conocimiento y los lenguajes de los contextos en los que se conciben y así replican o amplifican desigualdades sistémicas basadas en género, raza, origen étnico, habilidades, clase social y educación, entre otros factores. Las políticas deben evitar que los sistemas de IA tengan impactos diferenciados en comunidades diversas y que amplíen las brechas existentes entre diferentes países y también dentro de cada uno de ellos. Sin embargo, no existe una perspectiva única para comprender cómo se reproducen estas desigualdades ni un planteamiento único para contemplar sus causas y ofrecer soluciones a lo largo del ciclo de vida de la IA. Se necesitan esfuerzos en todos los niveles y aspectos para articular respuestas que conduzcan a tomar acciones transformadoras.

En este informe reconocemos las causas fundamentales de la desigualdad global y nos enfocamos en los daños actuales que la IA está causando en la vida de millones de personas. Adoptamos un enfoque sistémico, participativo y socio-técnico centrado en los derechos humanos y la justicia social para impulsar un cambio transformador. Las desigualdades en los ecosistemas de IA incluyen los sesgos en los datos o en los modelos algorítmicos, pero van más allá de estos. Por ello matizamos y enfatizamos la inclusión significativa en los procesos de desarrollo y en la toma de decisiones relativas a los sistemas de IA. También reconocemos que los beneficios del desarrollo tecnológico, la riqueza, el poder, el conocimiento y la infraestructura están concentrados actualmente en las manos de pocas personas, mientras la mayoría global padece sus costos en forma de explotación de recursos en sus territorios, trabajo en condiciones precarias y extractivismo de datos.

Es importante señalar que nos enfocamos en identificar formas concretas para avanzar de manera diferente. Por lo tanto, junto con valiosas iniciativas impulsadas por diversas partes interesadas, este informe ofrece una hoja de ruta para personas responsables de tomar decisiones, con recomendaciones y ejemplos de prácticas prometedoras que pueden servir como referencia para lograr un cambio transformador. Además, brinda orientaciones para alinear el diseño, el desarrollo y el despliegue de la IA teniendo en cuenta los principios del bien común y el bienestar de las poblaciones locales y globales.

En 2023, el Grupo de Trabajo de la AGIA sobre Uso Responsable de la IA nos confió la codirección del proyecto *Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity* (Hacia la igualdad sustantiva en el ámbito de la inteligencia artificial (IA): políticas transformadoras de IA para la igualdad de género y la diversidad), un honor por el que quisiéramos expresar nuestro agradecimiento. Con este informe y otras actividades del proyecto esperamos contribuir a la construcción de ecosistemas de IA verdaderamente inclusivos, equitativos y justos.

Paola Ricaurte Quijano,

Profesora Titular, Tecnológico de Monterrey, y Profesora Asociada en el Berkman Klein Center for Internet & Society de la Universidad de Harvard

Benjamin Prud'homme,

Vicepresidente de Políticas, Seguridad y Asuntos Globales, Equipo de Liderazgo, Mila



Resumen

El rápido avance de la inteligencia artificial (IA) está transformando las sociedades e impulsando el crecimiento económico con un gran potencial para mejorar la vida y el desarrollo socio-económico en el plano mundial. Sin embargo, existe el riesgo de que esta agrave las desigualdades existentes al reflejar y magnificar los sesgos sociales, especialmente los que afectan a grupos históricamente marginados. Desafíos como la discriminación, las injusticias, los sesgos y los estereotipos dañinos persisten a lo largo del ciclo de vida de la IA y afectan muchos aspectos de la vida humana. Se necesitan urgentemente marcos regulatorios sólidos para mitigar estas disparidades, prevenir daños y trabajar hacia la igualdad sustantiva y la diversidad en los ecosistemas de IA.

El informe *Hacia la igualdad sustantiva en el ámbito de la inteligencia artificial (IA): políticas transformadoras de IA para la igualdad de género y la diversidad* tiene como objetivo fortalecer la capacidad de los Estados y otras partes interesadas para fomentar ecosistemas de IA inclusivos, equitativos y justos. Examina prácticas prometedoras, proporciona perspectivas para políticas públicas y ofrece recomendaciones prácticas para mejorar la igualdad de género y la diversidad en la IA y la formulación de políticas en este ámbito. La *Guía de Medidas para Implementar las Recomendaciones sobre Políticas Transformadoras de IA* ofrece orientaciones adicionales para su aplicación.

Recomendaciones clave:

1. **Diseños inclusivos e innovaciones democráticas:** Integrar acciones afirmativas y medidas para la inclusión institucional, y fomentar el diseño de tecnología inclusiva.
2. **Participación significativa en la gobernanza de la IA:** Fomentar y garantizar la participación significativa de los grupos marginados en la gobernanza de la IA para mejorar las políticas sobre IA.
3. **Transparencia y rendición de cuentas para prevenir daños:** Establecer salvaguardias *ex ante* y mecanismos de rendición de cuentas entre todas las partes interesadas de la IA con el fin de prevenir daños y garantizar la equidad.
4. **Acceso efectivo a la justicia:** Implementar acciones que garanticen que los grupos marginados tengan acceso a recursos legales contra la discriminación y los sesgos derivados de la IA.

Las perspectivas y las recomendaciones se basan en consultas regionales y con grupos específicos, así como en un marco conceptual basado en un enfoque de derechos humanos en la IA. Mediante estas consultas, el informe examina el estado actual de la igualdad de género y la diversidad en el sector de la IA, resalta las disparidades persistentes y presenta vías y prácticas prometedoras hacia la igualdad sustantiva (real) en la IA. El marco parte de la premisa de que la desigualdad es estructural y de que se puede abordar y remediar mediante medidas apropiadas y transformadoras.

Las políticas transformadoras sobre IA afrontan las causas fundamentales de la desigualdad para lograr una igualdad sustantiva en el ámbito de la IA y más allá. La igualdad sustantiva que se logre en el ámbito de la IA por medio de un cambio transformador significará un avance en los derechos humanos e impulsará el desarrollo económico y social. Con políticas transformadoras sobre IA podemos mejorar la calidad, la facilidad de uso y la efectividad de los sistemas de IA, contribuyendo así a crear un futuro más equitativo, sostenible y próspero para todas las personas.



Mujeres africanas intercambian opiniones sobre un comando en una computadora. Fotografía: Iwaria Inc. en Unsplash.

Resumen de recomendaciones clave para formular políticas transformadoras de inteligencia artificial

Este resumen de recomendaciones clave presenta en forma breve las medidas concretas que las personas responsables de políticas públicas pueden adoptar para integrar los principios de igualdad de género y diversidad de manera efectiva en los marcos de las políticas, el derecho, la reglamentación y las prácticas relacionadas con la IA. Las recomendaciones para formular políticas transformadoras de IA se agrupan en las cuatro categorías siguientes:

- Diseños inclusivos e innovaciones democráticas
- Participación significativa en la gobernanza de la IA
- Transparencia y rendición de cuentas para prevenir daños
- Acceso efectivo a la justicia

Diseños inclusivos e innovaciones democráticas

1. Involucrar a grupos históricamente marginados en roles técnicos y no técnicos en todo lo concerniente al ecosistema de IA.

Implementar acciones afirmativas en todo el ecosistema de IA para involucrar a las mujeres y a otros grupos históricamente marginados en roles técnicos y no técnicos en todo lo concerniente al ecosistema de IA, para incrementar así la diversidad de perspectivas. Asignar recursos para identificar y eliminar las barreras que enfrenta la representación de la diversidad. Esto incluye garantizar una educación accesible e inclusiva más allá de los ecosistemas de IA.

2. Fortalecer las capacidades hacia una inclusión institucional

Invertir —dentro de las instituciones y agrupaciones públicas y privadas— en el desarrollo de capacidades y la sensibilización sobre las experiencias y los derechos de los grupos históricamente marginados. Mantener un diálogo constante con las personas representantes de los grupos marginados para comprender y eliminar las barreras específicas a las que se enfrentan.

3. Permitir el procesamiento de categorías especiales de datos

Permitir el procesamiento de categorías especiales de datos en determinadas circunstancias excepcionales, con fundamento en un interés público significativo, para lograr la igualdad y la no discriminación. A fin de prevenir resultados discriminatorios, las partes proveedoras de sistemas de IA deben verificar la ausencia de sesgos sistémicos y garantizar que los conjuntos de datos sean representativos de la diversidad. Esto debe realizarse sin contravenir los derechos de protección de datos personales.

4. Financiar enfoques de investigación y diseño de tecnología transformadora en las innovaciones de la IA.

Financiar investigaciones y otorgar subvenciones y reconocimiento público para incentivar la aplicación de enfoques de diseño tecnológico inclusivo y transformador en la IA, tales como los que se basan en principios de diseño tecnología feminista. Estos enfoques abordan las brechas existentes entre la justicia de carácter técnico y la justicia política. Apoyar innovaciones en sistemas de IA que se alineen con estos principios promueve aplicaciones, prácticas y procesos más equitativos y justos.

Participación significativa en la gobernanza de la IA

5. Promover una participación pública y comunitaria efectiva

Emplear diversas metodologías de participación pública a nivel nacional e internacional. Incluir a los grupos marginados en las discusiones nacionales sobre la gobernanza de la IA y amplificar la mayoría global en los foros internacionales sobre el mismo tema. Habilitar la participación de personas representantes de grupos marginados asignando presupuestos para gastos de participación que garanticen que los procesos de información y consulta sean accesibles, gratuitos y claros.

6. Invertir en el desarrollo de capacidades entre grupos marginados

Financiar y apoyar programas educativos, estructuras de redes y otros recursos que se propongan desarrollar las habilidades y la confianza entre los grupos marginados para que participen de manera significativa o lideren activamente los procesos que atienden sus necesidades. Trabajar con las comunidades marginadas y con las organizaciones que las representen para que organicen sus propias sesiones de sensibilización y sus consultas sobre temas relacionados con la IA.

7. Legislar sobre los derechos a la participación pública *ex ante*

Basar los procesos de toma de decisiones en IA en los derechos de participación pública *ex ante*, como los establecidos en la Convención de Aarhus de la Comisión Económica de las Naciones Unidas para Europa (CEPE). La aplicación de estos principios en los procesos de toma de decisiones en IA permite que las partes afectadas, así como las organizaciones de la sociedad civil y el público en general, puedan impugnar las consecuencias de las decisiones algorítmicas mediante el razonamiento y la deliberación en el ámbito público.

8. Proteger los derechos colectivos sobre datos e IA

Revisar los marcos de derechos que se ven afectados por los sistemas y procesos de IA, como los marcos de derechos de propiedad intelectual y de datos, para 1) salvaguardar la soberanía sobre los datos y los conocimientos de las comunidades indígenas y los grupos marginados, incluidas las minorías lingüísticas, religiosas y étnicas; y, 2) garantizar el derecho a beneficiarse del progreso científico.



Delegada con hiyab y traje formal hablando al micrófono en el podio durante una conferencia política.



Transparencia y rendición de cuentas para prevenir daños

9. Establecer el derecho a la información en los sistemas de IA y mejorar la transparencia algorítmica

Establecer el derecho a la información en los sistemas de IA. Este derecho debe otorgar a las personas el acceso a detalles claros y accesibles sobre cuándo se emplea la IA, qué algoritmos se utilizan, qué datos se emplean como datos de entrada y qué criterios se aplican en los procesos de toma de decisiones. Exigir una mayor transparencia algorítmica permite que las personas afectadas negativamente por sistemas de IA puedan impugnar sus resultados. También fomenta la innovación tecnológica para hacer frente a limitaciones como la opacidad del comportamiento, y mejorar la interpretabilidad y la explicabilidad.

10. Habilitar y realizar evaluaciones obligatorias del impacto en los derechos humanos (EIDH)

Propiciar y conducir estas evaluaciones del impacto con políticas que orienten sobre la manera de implementarlas. Las evaluaciones deberán determinar si los riesgos de daño son aceptables en el marco de los derechos fundamentales e incluir obligaciones claras para eliminar o prevenir dichos riesgos. Las evaluaciones también deberán considerar y comparar posibles enfoques no tecnológicos para identificar las medidas menos intrusivas en los derechos humanos.

11. Formular medidas de rendición de cuentas aplicables a los sistemas y procesos algorítmicos del sector público

En el ámbito de las contrataciones públicas, desarrollar directrices específicas para la IA con el fin de proteger los derechos humanos y el debido proceso que hagan referencia a las complejidades y los riesgos que representan los sistemas y procesos algorítmicos de IA. Promover iniciativas de datos abiertos para construir bibliotecas abiertas de algoritmos utilizados en los sistemas del sector público. Garantizar que las personas responsables de políticas públicas reciban capacitación para que puedan llevar a cabo eficazmente la debida diligencia en la contratación de sistemas de IA.

Acceso efectivo a la justicia

12. Fortalecer la responsabilidad contextual para la no discriminación en los sistemas de IA

Fortalecer la responsabilidad contextual por la no discriminación en los sistemas de IA en proporción a otras medidas de rendición de cuentas, como el grado de transparencia, de interpretabilidad y de explicabilidad. Deberá revisarse la reglamentación sobre responsabilidad por los daños causados por productos defectuosos y sobre responsabilidad subjetiva (por culpa) para reflejar con precisión las complejidades de los sistemas de IA y la toma de decisiones basada en datos. La rendición de cuentas efectiva en el desarrollo y despliegue de la IA tiene en cuenta características específicas como la opacidad, la explicabilidad, el comportamiento autónomo, la adaptación continua y la predictibilidad limitada. Trazar una vía hacia la responsabilidad en materia de IA para garantizar una rendición de cuentas apropiada entre las partes públicas y privadas proveedoras y encargadas de su despliegue.

13. Instar a las instituciones promotoras de la igualdad a iniciar procedimientos

Instar a las instituciones promotoras de la igualdad —incluidas las instituciones nacionales de derechos humanos y otras organizaciones de interés público— a movilizarse en aras del interés público. Deben también autorizar a estas instituciones a presentar demandas ante las autoridades supervisoras incluso sin necesidad de identificar a las partes denunciantes. Aliviar la carga de la prueba y dotar a estos organismos de la autoridad legal y la capacitación necesaria para abordar de manera efectiva la discriminación y los daños que causan los sistemas de IA y los procesos relacionados.



14. Aliviar la carga de la prueba para las personas demandantes

Examinar y revisar las reglas a fin de aliviar la carga de la prueba impuesta a las partes demandantes (World Commission on the Ethics of Scientific Knowledge and Technology, 2005). Las normas existentes sobre responsabilidad por los daños causados por productos defectuosos suelen exigir a las partes perjudicadas que demuestren la relación causal entre los defectos del producto y los daños específicos que ocasione. Considerar la posibilidad de ajustar estas reglas para que las partes demandantes puedan probar sus casos y exigir una indemnización con mayor facilidad.

Introducción

El rápido avance de la inteligencia artificial (IA) está transformando las industrias e impulsando el crecimiento económico con un gran potencial para mejorar la vida y el desarrollo socio-económico en el plano mundial. Sin embargo, también puede agravar desigualdades ya existentes. Los sistemas de IA pueden reflejar y amplificar los sesgos sociales, así como reproducir desigualdades sociales, que afectan en particular a los grupos históricamente marginados. A medida que aumenta el uso de sistemas y procesos de IA en áreas críticas como la justicia penal, la educación, atención a la salud y el empleo, se vuelve imperativo abordar las crecientes desigualdades y la falta de diversidad en los ecosistemas de IA y en la formulación de las políticas correspondientes.

Reconocer la igualdad de género y la diversidad como una prioridad en los procesos de toma de decisiones en materia de IA es fundamental para alcanzar los Objetivos de Desarrollo Sostenible (ODS). Las crecientes desigualdades y la falta de diversidad en la IA representan un riesgo significativo para alcanzar los ODS al pasar por alto las necesidades específicas, los conocimientos locales especializados y los contextos culturales de los grupos históricamente marginados y excluidos. Al pasar por alto las perspectivas de la mayoría global en los debates mundiales sobre las políticas se corre el riesgo de agravar aún más las desigualdades existentes a nivel global (UNCTAD, 2021). Existe una necesidad urgente de marcos regulatorios sólidos para mitigar estas disparidades, para prevenir daños y para garantizar que el desarrollo de los sistemas y procesos de IA sea beneficioso para todas las personas (Chauhan y Kshetri, 2022; Dankwa-Mullan et al, 2021; Dignum, 2023; Joyce et al, 2021; UNESCO, 2022; Vinuesa et al, 2020).

El informe *Hacia la igualdad sustantiva en el ámbito de la inteligencia artificial (IA): políticas transformadoras de IA para la igualdad de género y la diversidad* proporciona a los ecosistemas de IA —y en particular a los Estados como garantes de los derechos— recomendaciones prácticas sobre políticas y perspectivas derivadas de prácticas promotoras sobre cómo integrar de manera efectiva los enfoques de igualdad de género y diversidad a lo largo del ciclo de vida de la IA y cómo formular las políticas en este ámbito.

Sistemas de Inteligencia Artificial (IA)

Un sistema de IA es un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de los datos de entrada que recibe, cómo generar información de salida como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos reales o virtuales. Una vez desplegados, los distintos sistemas de IA presentan diversos niveles de autonomía y varían en su capacidad de adaptación.

(OECD, 2024b)

El informe surge de una iniciativa del Grupo de Trabajo de la AGIA sobre Uso Responsable de la IA y analiza las experiencias y conocimientos locales especializados de cinco regiones en todo el mundo. En las consultas regionales y las de grupos específicos participaron aproximadamente 200 personas que representaron a más de 50 países y a una diversa gama de comunidades e identidades. Esto incluyó a personas representantes del ámbito académico, la sociedad civil, la industria y el gobierno. Al priorizar las consultas se buscó escuchar y comprender una multiplicidad de voces, perspectivas y experiencias en sus contextos específicos. Las perspectivas y las recomendaciones presentadas en este informe se basan en los resultados de las consultas, así como en una extensa revisión bibliográfica y documental. Estos resultados señalan los desafíos críticos identificados que se afrontan para alcanzar la igualdad de género y la diversidad en los ecosistemas y procesos de IA, desafíos que ilustran las profundas desigualdades arraigadas en los ecosistemas de IA.

“El punto de partida debe ser el análisis de los daños que la población esté experimentando y que pueda experimentar. Esto exige escuchar a quienes han tenido afectaciones, así como a quienes han dedicado muchos años a identificar y a responder a estos daños. Las mujeres, los grupos minoritarios y las personas marginadas, en particular, son afectadas desproporcionadamente por los sesgos en la IA. Debemos hacer esfuerzos serios para incluir a estos grupos en toda discusión sobre la gobernanza de la IA.”

Volker Türk,
Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2023

Los desafíos críticos ponen de relieve cómo las nuevas formas de discriminación y daños individuales y comunitarios que surgen a la par con los sistemas y procesos de IA afectan de forma desproporcionada a grupos históricamente marginados y excluidos. Los sistemas de IA pueden, por ejemplo, afectar desproporcionadamente a personas que enfrentan barreras basadas en el sexo, la identidad de género, la expresión de género y/o la orientación sexual, incluidas las mujeres (Costanza-Chock, 2018; Donnelly y Stapleton, 2022; Hamidi *et al*, 2018). Los sistemas y procesos de IA han facilitado aún más la violencia de género facilitada por la tecnología (TFGBV, por sus siglas en inglés) como, por ejemplo: la distribución de imágenes sin consentimiento, el acoso y el acecho en línea. El anonimato y la escala en que operan las plataformas impulsadas por IA agravan estos abusos. La extracción de características de género en funciones de biometría también plantea problemas éticos y de privacidad, en particular para las comunidades LGBTIQ+ (Ovalle *et al*, 2023a). Muchos sistemas de IA, como las tecnologías de escaneo corporal que evalúan la desviación y el riesgo, por ejemplo, se basan a menudo en supuestos cisnormativos y heteronormativos, lo que conduce a la discriminación de las personas transgénero (Costanza-Chock, 2018).

Grupos históricamente marginados

Entre los grupos de personas históricamente marginadas y excluidas, se incluyen, entre otras, a las mujeres, los pueblos indígenas, las personas racializadas, las personas con discapacidad, las personas que enfrentan barreras por su orientación sexual, su sexo, su identidad y/o expresión de género y sus características sexuales (SOGIESC, por sus siglas en inglés)¹, las personas en movimiento² y las personas de bajos ingresos o provenientes de entornos de escasos recursos. No existe un consenso sobre la definición o el uso de los términos “marginación” o “grupos históricamente marginados”. La marginación generalmente se refiere a una forma de exclusión, por medio de elementos económicos, políticos, culturales o sociales, y es “*una modalidad de desventaja aguda y persistente, enraizada en desigualdades sociales subyacentes*” (UNESCO, 2009). El concepto “históricamente marginado” resalta el hecho de que la marginación de ciertos grupos se ha expandido durante varias décadas o incluso siglos, y que los sistemas que perpetúan dicha marginación y sus efectos son intergeneracionales.

(Adaptado de UNESCO, 2009)

Otro desafío importante es la falta de diversidad e inclusión en todo el ecosistema de la IA (Cheong *et al*, 2021; Suresh *et al*, 2022). La exclusión en los roles técnicos y no técnicos limita la diversidad de perspectivas, lo que conduce a sistemas de IA que perpetúan la exclusión de conocimientos y que refuerzan estereotipos dañinos y discriminación (por ejemplo, Badaloni y Rodà, 2022; Buolamwini y Gebru, 2018; Hull, 2023; Ricaurte, 2019; UNESCO, 2024a). Los sistemas sesgados de toma de decisiones en cuanto a beneficios sociales, por ejemplo, pueden asignar menos recursos a personas con discapacidad y limitar las oportunidades de empleo para los grupos marginados.

La interseccionalidad de las identidades de personas pertenecientes a grupos históricamente marginados puede resultar en desventajas acumuladas, ya que los efectos negativos de los sistemas de IA se entrecruzan con desigualdades sociales más amplias (Buolamwini y Gebru, 2018; Forster, 2022; Kong, 2022; Ricaurte, 2022; Ricaurte y Zasso, 2023; UNCTAD, 2021; UNHRC, 2020). Por ejemplo, los sistemas IA han identificado erróneamente muletas como armas, exponiendo así a personas con discapacidad a un riesgo de daño mayor del que ya se enfrenta en situaciones riesgosas (UNHRC, 2021). De manera similar, el despliegue de tecnologías de vigilancia y del análisis predictivo de datos en contextos migratorios se enfoca a menudo en comunidades marginadas, lo que agrava tensiones ya existentes y contribuye a la comisión de violaciones de los derechos humanos (Dumbrava, 2021; McGregor y Molnar, 2023). Estos ejemplos ponen de relieve las desigualdades profundamente arraigadas y los desafíos críticos para lograr la igualdad de género y la diversidad en los ecosistemas y procesos de IA que requieren esfuerzos políticos concertados.

Discriminación algorítmica

La discriminación algorítmica se produce cuando, sin justificación alguna, los sistemas de IA tratan o afectan a las personas de manera diferente por motivos de raza, color, origen étnico, sexo, religión, edad, discapacidad o de otras clasificaciones protegidas (White House, 2022). Este complejo problema surge de la interacción entre el despliegue tecnológico, las prácticas sociales y los objetivos políticos, y va más allá de los sesgos en los datos e incluye todos los casos en los que la IA perpetúa la desigualdad por medio de interacciones sociales.

(Adaptado del documento *Global Index on Responsible AI*, 2024)



Una mujer presta atención a la presentación de una colega. Fotografía: Christina @ wocintechchat.com en Unsplash.

En vista de estos desafíos, este informe ofrece un enfoque integral para abordar la desigualdad de género y la ausencia de diversidad e inclusión en los sistemas IA por medio de recomendaciones clave para formular políticas transformadoras de inteligencia artificial. Dichas recomendaciones clave impulsan y se cimentan en la implementación de los principios de IA de la OCDE (OECD, 2024a), la Recomendación sobre la Ética de la IA de la UNESCO (UNESCO, 2022), los Objetivos de Desarrollo Sostenible (ODS) (United Nations, 2024), el Pacto Digital Global (United Nations, 2024b) y otros trabajos que abogan por la alineación de los sistemas de IA con los marcos de derechos humanos para avanzar en la igualdad de género y la diversidad en los ecosistemas de IA y en la formulación de las políticas relacionadas (World Economic Forum, 2024)³. La gobernanza internacional de los sistemas y procesos de IA es un factor clave para alcanzar los ODS. El marco de políticas transformadoras en el ámbito de la IA presentado aquí se funda en la premisa de que la desigualdad es estructural y, por lo tanto, es posible abordarla y corregirla mediante medidas adecuadas y transformadoras.

Las recomendaciones clave para formular políticas transformadoras de inteligencia artificial encaminadas a lograr la igualdad sustantiva están ancladas en un enfoque basado en los derechos humanos para los sistemas y procesos de IA. Este enfoque integra los derechos humanos en el centro de los ecosistemas de IA eliminando limitaciones y mejorando las capacidades para garantizar el disfrute de los derechos humanos en la totalidad de los ecosistemas de IA y en la formulación de las políticas correspondientes. Con las recomendaciones se busca lograr la igualdad sustantiva abordando las barreras estructurales que impiden que los grupos históricamente marginados y excluidos se beneficien de los avances en el desarrollo de la IA y se pretende instar a las mujeres y a otros grupos marginados a la acción para lograr el goce pleno de sus derechos.

Igualdad sustantiva

El concepto de “igualdad sustantiva” se refiere a la igualdad *de facto* (igualdad de hecho o igualdad real) o a la igualdad de los resultados (Fredman, 2016; UN Women, 2019). Esto incluye garantizar que todas las personas tengan igualdad de oportunidades y de acceso a ellas, así como un entorno favorable para lograr resultados iguales. La igualdad sustantiva planteada en este informe incorpora una comprensión de la equidad⁴, la igualdad de género⁵ y la diversidad que aboga por la igualdad de oportunidades y de resultados mediante medidas que reconozcan las diferencias, fomenten la expresión y la participación, y eliminen la discriminación, la marginación y la distribución desigual del poder y de los recursos. La igualdad sustantiva es tanto un objetivo en sí misma como una agenda transformadora y como un medio para alcanzar otros objetivos de desarrollo sostenible, incluyendo los de salud, crecimiento económico y educación.

(Buckup, 2009; OECD, 2015; Ostry *et al*, 2018).

Lograr la igualdad sustantiva mediante un cambio transformador en el ámbito de la IA no solo significará un avance en materia de derechos humanos, sino que también impulsará el desarrollo económico y social. Incorporar principios de igualdad de género y diversidad guiados por el marco de los derechos humanos es esencial para construir ecosistemas de IA propicios para la innovación, y ecosistemas que sean inclusivos y que estén alineados con las necesidades de la sociedad, siendo estos compromisos formales de la OCDE y de todos los Estados Miembros de la UNESCO. Con la implementación de políticas inclusivas en el ámbito de la IA podemos mejorar la calidad, la facilidad de uso y la efectividad de los sistemas y procesos de IA, contribuyendo así a lograr un futuro más equitativo, sostenible y próspero para todas las personas y comunidades.



Manifestación de un grupo de activistas LGBTQI+ en contra del Proyecto de Ley de Discriminación Religiosa en Sídney, Australia. Fotografía: Nikolas Gannon en Unsplash.

Resumen del informe

Parte I. Consultas, desafíos cruciales y vías posibles

El informe comienza con una descripción de las consultas regionales y de las consultas con grupos específicos realizadas en colaboración con las organizaciones socias ejecutoras del proyecto, con el fin de comprender mejor las necesidades no satisfechas, las barreras sistémicas y las soluciones estratégicas propuestas. A esta sección le sigue una visión general de los resultados de las consultas en la que se identifican los desafíos cruciales y las posibles vías para lograr la igualdad de género y la diversidad en los ecosistemas de IA y la formulación de las políticas relacionadas. Las consultas fundamentan el contenido de este informe.

Parte II. Marco conceptual

Después de los desafíos cruciales y las vías posibles, el informe presenta el marco conceptual que sustenta el análisis de prácticas prometedoras y el desarrollo de recomendaciones clave para las personas responsables de políticas públicas.

Este marco conceptual se divide en tres partes: en primer término, explica el concepto de igualdad sustantiva y el de cambio transformador como un enfoque. En segundo lugar, adopta un enfoque que contempla tanto los aspectos sociales como los técnicos de los sistemas de IA y de sus ciclos de vida, y muestra las oportunidades para intervenir en cada fase de la cadena de producción de la IA. En tercer lugar, presenta el marco de políticas transformadoras en el ámbito de la IA con sus tres dimensiones clave: 1) Reparar desventajas sistémicas (el derecho a la inclusión); 2) Rectificar el déficit democrático (el derecho a la participación); y 3) Revertir errores de reconocimiento (el derecho a la dignidad).



Un joven en sudadera de capucha sobresale entre una muchedumbre. Fotografía: Jake Weirick en Unsplash.

Parte III. Prácticas prometedoras y recomendaciones

La siguiente sección presenta prácticas prometedoras que incluyen la promoción de los derechos y la participación significativa o la inclusión de grupos marginados, o tienen dicho propósito, o son lideradas por dichos grupos. Este conjunto de prácticas es ilustrativo y su objetivo es ofrecer a las personas responsables de políticas públicas y a otras partes interesadas ejemplos concretos de acciones que pueden reproducirse o adaptarse a sus contextos.

Finalmente, el informe proporciona un panorama detallado de las recomendaciones clave para formular políticas transformadoras en el ámbito de la inteligencia artificial. Las recomendaciones para formular políticas transformadoras de IA se agrupan en las cuatro categorías siguientes que reflejan las dimensiones clave del marco:

- Diseños inclusivos e innovaciones democráticas
- Participación significativa en la gobernanza de la IA
- Transparencia y rendición de cuentas para prevenir daños
- Acceso efectivo a la justicia

Con estas categorías, las principales recomendaciones para formular una política transformadora en el ámbito de la IA adoptan un enfoque holístico hacia la igualdad sustantiva.



PARTE I.

Consultas, desafíos cruciales y vías posibles



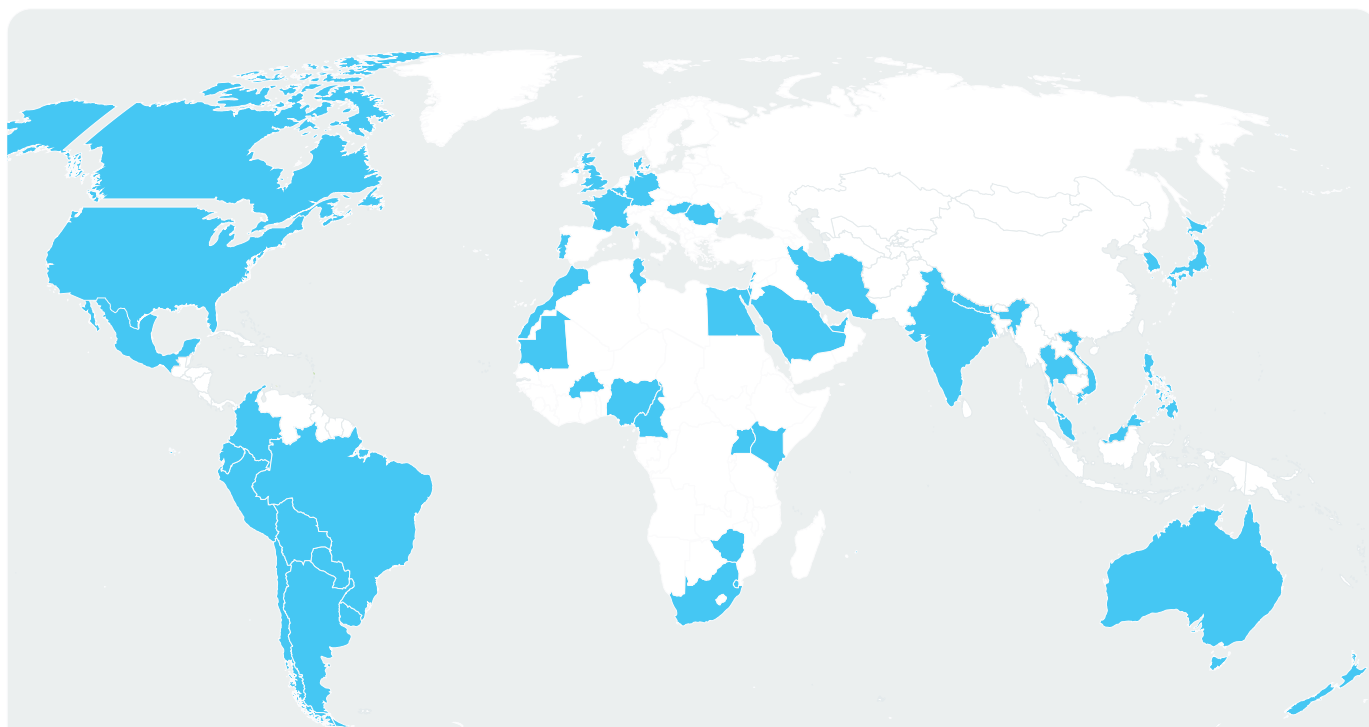
Tres mujeres de diferentes generaciones sentadas una tras otra contra una pared azul, vestidas con coloridos atuendos tradicionales. Fotografía: Srimathi Jayaprakash en Unsplash.



Consultas regionales y con grupos específicos

El proyecto involucró a especialistas locales de cinco regiones del mundo, con aproximadamente 200 participantes que representan a más de 50 países, y una diversidad de comunidades e identidades. En el proyecto participaron personas representantes del ámbito académico y de la sociedad civil, la industria y el gobierno, quienes compartieron sus perspectivas en consultas regionales o en grupos específicos. Esta sección describe esas consultas, junto con algunas consideraciones y limitaciones.

El equipo del proyecto colaboró con organizaciones de investigación regionales y con organizaciones de la sociedad civil (organizaciones socias ejecutoras) para llevar a cabo las consultas regionales. Entre las organizaciones socias ejecutoras del proyecto se incluyeron Derechos Digitales, Research ICT Africa y Data Pop Alliance; también contribuyeron varias personas investigadoras asociadas en diferentes regiones del mundo. Las organizaciones ejecutoras identificaron, analizaron y categorizaron a las partes interesadas y a las organizaciones clave de cada región, optimizando así la organización de las consultas regionales. Enfatizando un alcance global, internacional, interseccional y con múltiples partes interesadas, las organizaciones colaboradoras se aseguraron de trabajar de manera inclusiva con una amplia gama de participantes.



Mapa mundial que representa las regiones geográficas donde tuvieron lugar las consultas. Imagen: Leonardo Studio Design. ⁶

Las consultas regionales se llevaron a cabo en cinco regiones:

- África Subsahariana
- Oriente Medio y África del Norte (MENA, por sus siglas en inglés)
- América Latina
- América del Norte y Europa
- Asia y el Pacífico



MÁS DE 50 PAÍSES | **MÁS DE 200 PERSONAS PARTICIPANTES**
(180 EN CONSULTAS REGIONALES Y MÁS DE 20 EN CONSULTAS DE GRUPOS ESPECÍFICOS)

Además de las consultas regionales, el proyecto se propuso canalizar las perspectivas de grupos marginados y personas que se identifican como miembros de ellos, incluyendo —entre otros— a mujeres, pueblos indígenas, personas racializadas, personas con discapacidad, personas que enfrentan barreras en función de diversas características relacionadas con la orientación sexual, la identidad de género, la expresión de género y las características sexuales (SOGIESC, por sus siglas en inglés), personas en movimiento, personas de bajos ingresos o provenientes de entornos de escasos recursos, y personas hablantes y/o revitalizadoras de lenguas indígenas y minorizadas, entre otros, que se encuentran en diversas regiones y contextos representando a diferentes grupos clave.

Las consultas adoptaron diversas formas, tales como entrevistas individuales, mesas redondas y contribuciones por escrito, con el objetivo general de captar una amplia gama de perspectivas y voces. Las personas participantes representaron una diversidad de comunidades e identidades, incluidos diferentes grupos clave. Además, se contó en el proyecto con la contribución de un Grupo Consultor de Personas Expertas (GCPE) y de un Grupo Asesor del Proyecto (GAP). El GCPE desempeñó sus labores de especialista para el equipo del proyecto en las etapas formativas. El grupo incluyó a personas con discapacidad, personas con diversas características SOGIESC, pueblos indígenas, grupos racializados y personas representantes de personas en movimiento provenientes, en su totalidad, de diferentes regiones del mundo. Los miembros del GCPE aportaron valiosas perspectivas interseccionales y sobre la diversidad. Para validar los pasos metodológicos del proyecto se formó un GAP para consultas periódicas. Este grupo contó entre sus miembros con personas expertas de AGIA y con especialistas externos. El equipo del proyecto analizó las perspectivas y discusiones generadas durante las entrevistas de forma temática como datos cualitativos con el fin de identificar temas novedosos, convergencias y divergencias de opiniones, consenso entre participantes, prácticas prometedoras, lecciones aprendidas e intercambio de conocimientos.

Consideraciones y limitaciones

Sensibilidades culturales y legales

Las cuestiones que pueden considerarse controvertidas, tabú o ilegales en áreas específicas —como los relacionados con la orientación sexual y la expresión de género— afectaron la cantidad de participantes dispuestos o capaces de participar en entrevistas. Las actitudes locales, el grado de aceptación y las leyes pueden limitar la aceptación y protección para estas personas.

Barreras por diferencias de lenguaje

El proyecto se llevó a cabo principalmente en inglés, con la opción de participar en español y en francés. Esto generó una desigualdad en las oportunidades para participar.

Acceso y límite de tiempo

Las consultas fueron virtuales, lo que requería acceso a un dispositivo con conexión a Internet, por lo que quedaron excluidas personas que ya se consideraban marginadas digitalmente. Las preguntas técnicas en los cuestionarios limitaron la accesibilidad dependiendo del nivel de conocimientos previos sobre IA. La brevedad del tiempo asignado también supuso una limitación tanto para la participación como para la calidad de los resultados obtenidos.

Representación desigual

La participación de representantes de la sociedad civil y del ámbito académico fue más extensa en comparación con la de representantes del gobierno o del sector industrial. Además, la mayoría de quienes participaron contaban con un alto nivel educativo, lo que no representa adecuadamente las voces de personas con menos educación formal o pertenecientes a grupos marginados. La representación de participantes y grupos de género no binario fue menos prominente que la de otras identidades.

Limitaciones de representación por región y por país

No se incluyeron en el informe algunos países y regiones debido a restricciones de tiempo, accesibilidad y recursos. En particular, los países del Caribe no estuvieron representados en el proceso consultivo.

Consultas regionales y con grupos específicos: Resultados

En esta sección se exponen las principales conclusiones derivadas de la identificación y el análisis empírico en profundidad de las opiniones, perspectivas, experiencias y mejores prácticas de las múltiples partes interesadas en relación con la IA, la igualdad de género y la diversidad, en su interacción con barreras sistémicas y soluciones estratégicas para alcanzar la igualdad de género y la diversidad en los ecosistemas de IA. En la síntesis de las conclusiones se presta especial atención a representar voces y perspectivas diversas de múltiples grupos y regiones. El siguiente resumen no representa las opiniones matizadas de todas las personas participantes, pero ofrece un panorama general de las experiencias declaradas sobre los desafíos críticos y las vías posibles para abordar la igualdad de género y la diversidad en todo el ciclo de vida de la IA y en la formulación de las políticas en este ámbito. Los resultados están categorizados como sigue: En la primera parte de esta sección se exponen y describen los tres principales desafíos para implementar la igualdad de género y la diversidad en los ecosistemas de IA, según quienes participaron. En la segunda parte se presentan cuatro vías para abordar la igualdad de género y la diversidad en la IA. Finalmente, se ofrece un resumen breve de los desafíos y las vías propuestas.

Desafíos cruciales para garantizar la igualdad de género y la diversidad en el ámbito de la IA

Las consultas regionales y las de grupos específicos revelaron tres desafíos principales para implementar la igualdad de género y la diversidad en los ecosistemas de IA.

■ Amalgama entre acceso e inclusión

Un tema recurrente en las consultas regionales, especialmente en el África Subsahariana, fue la necesidad de diferenciar entre el mero acceso a productos y servicios de IA y la inclusión significativa en los ecosistemas de IA. Quienes participaron afirmaron claramente lo que comprende una inclusión significativa: una transformación sistémica en el diseño y desarrollo de la IA que ponga al timón a los grupos históricamente marginados.

■ Exclusión del conocimiento e invisibilización

Los modelos de IA no representan la diversidad y perpetúan la exclusión del conocimiento y otros tipos de invisibilización, lo que resulta en injusticia epistémica. Las consultas realizadas en América Latina, África Subsahariana y Asia, así como con comunidades indígenas en América del Norte, mostraron la incompatibilidad entre los modos dominantes de desarrollo de IA y sus ontologías de conocimientos.

■ Distribución inequitativa de recursos

La distribución de recursos en el ecosistema económico internacional de IA es inequitativa. En todas las consultas regionales se identificó como un desafío crítico la exclusión sistémica de los beneficios del desarrollo en la economía de IA entre los grupos históricamente marginados, especialmente entre la mayoría global. La explotación laboral y el extractivismo de datos continúan perpetuando esta desigualdad, mientras que la cadena de valor de la IA afianza la injusticia ambiental.



Grupo de mujeres en Sierra Leona. Fotografía: Annie Spratt en Unsplash.



Amalgama entre acceso e inclusión

Inclusión

El concepto “inclusión” se refiere a asegurar la participación plena y significativa de los grupos marginados en todos los sectores, en todos los niveles, con efectos positivos y mensurables en su calidad de vida y en el acceso irrestricto a sus derechos humanos en igualdad de condiciones con los demás grupos. La inclusión requiere a) medidas para instar a la acción y apoyar a los grupos marginados, y b) medidas estructurales, políticas e institucionales para corregir las injusticias sistémicas en todos los niveles. Una iniciativa inclusiva comprende diversidad necesariamente y crea un entorno de respeto mutuo, sensación de pertenencia, confianza, apoyo y participación.

(DESA-UN, 2009)



“La inclusión en el ámbito de la inteligencia artificial no es simplemente cuestión de añadir o dar acceso. De hecho, se trata de un acceso que produce un florecimiento o de un acceso que genera beneficios. De modo que si el acceso no produce beneficios, ese acceso es excluyente”.

*Angella Ndaka,
Centre for African Epistemic Justice, Kenya*

“[La igualdad de género y la diversidad], en gran medida, [se reduce] a reclutar más mujeres. Pero el entorno al que las estamos integrando no es cordial”

*Luisa Olaya,
GIZ (FAIR Forward), Alemania*



“Debemos ser cuidadosos al crear soluciones genéricas de género y diversidad porque cada país o región tiene sus propias reglas de diferenciación social y sus prioridades sobre inclusión”

*Anónimo,
Asia*



Un tema recurrente en las consultas regionales, especialmente en el África Subsahariana, fue la necesidad de diferenciar entre el mero acceso a productos y servicios de IA y la inclusión significativa en los ecosistemas de IA. Enfocarse únicamente en asegurar el acceso desalienta la inclusión significativa y reduce la programación que transforma la perspectiva de género a iniciativas aisladas que no son integrales, sostenibles o transformadoras. Las personas participantes en las consultas regionales fueron rotundas sobre lo que, por otro lado, comprende una inclusión significativa: la transformación sistémica en la gobernanza, el diseño y el desarrollo de la IA que ponga al timón a los grupos históricamente marginados. También consideran que las medidas específicamente contextuales para mejorar la autodeterminación de los grupos marginados en los ecosistemas de IA son esenciales para abordar las diferencias sociales y la injusticia estructural, y para trascender enfoques alejados de la igualdad de género y la diversidad en la gobernanza de la IA.



Exclusión del Conocimiento e Invisibilización



“La falta de diversidad en la representación puede conducir a que los equipos de desarrollo de IA ignoren las necesidades y perspectivas de las mujeres, la gente de color y otros grupos marginados”

*Anónimo,
Asia*

“El tema más urgente es la representación de personas no binarias, trans e indígenas, como desarrolladoras y como líderes en tecnología.”

*Marcos Cornelio Sánchez Ramírez,
PIT Policy Lab, México*



“Alinear nuestras tecnologías con nuestra historia y nuestra cultura es realmente importante, y la soberanía y la reciprocidad son valores clave en el ámbito de la IA.”

*Florian Lebet,
Comunidades indígenas, Canadá*

Los modelos de IA no representan la diversidad y perpetúan la exclusión del conocimiento y otros tipos de invisibilización, lo que redundaría en injusticia epistémica: la exclusión del conocimiento, las culturas y los lenguajes. En las consultas regionales realizadas en América Latina, África Subsahariana y Asia, así como en los diálogos con comunidades indígenas en América del Norte, se enfatizó la incompatibilidad entre los modos dominantes de desarrollo de IA y sus ontologías de conocimientos. Quienes participaron hicieron hincapié en su preocupación sobre el refuerzo de las jerarquías interseccionales de diferencias sociales en los ecosistemas de IA. Los sistemas de IA incorporan los valores y las prácticas culturales de los países y espacios donde se conciben y se desarrollan. Esto crea incongruencias cuando estas tecnologías se aplican en sociedades y culturas globales diversas. Los conjuntos de datos que sustentan estos sistemas a menudo carecen de diversidad representativa, lo que conduce a la invisibilización y a la falta de representación de grupos históricamente marginados. Esto también puede conducir a la amplificación de los prejuicios sociales existentes. Además, los sesgos de las instituciones desarrolladoras de IA pueden imponer eliminaciones de tipo cultural o incluso violar los derechos humanos. El diálogo con las comunidades indígenas reveló la enorme preocupación sobre la incompatibilidad entre los modos dominantes de desarrollo de IA y sus ontologías del conocimiento. En las consultas regionales realizadas en Asia se señaló la escasez de recursos como un obstáculo importante para abordar estas injusticias epistémicas en los sistemas de IA.

Injusticia epistémica

El concepto “injusticia epistémica” se refiere al trato injusto de personas o grupos en su capacidad de ser quienes tienen conocimientos o contribuyen a su obtención (Fricker, 2007; Hull, 2023). Este concepto destaca el modo en que los prejuicios y sesgos pueden socavar o hacer caso omiso del conocimiento, las experiencias y las perspectivas de personas o grupos marginados.

- La injusticia hermenéutica ocurre cuando existe una brecha en la comprensión colectiva que impide a las personas marginadas encontrar el sentido de sus experiencias. Esto sucede porque sus experiencias sociales no son bien comprendidas o reconocidas por la sociedad en general, a menudo debido a la falta de un lenguaje o de conceptos en común.
- La injusticia testimonial se presenta cuando la credibilidad de una persona hablante se juzga de manera injusta debido a prejuicios. Por ejemplo: una persona de un grupo marginado puede ver que no se dé crédito a sus palabras o puede verse ignorada simplemente por su identidad, independientemente de la validez de la información que proporcione.
- La injusticia contributiva surge como resultado de la injusticia testimonial, cuando las personas marginadas son excluidas de participar en la generación y validación del conocimiento. Sus perspectivas y contribuciones no son valoradas ni reconocidas, lo que limita su efecto en los procesos de toma de decisiones y en la comprensión más amplia de la sociedad.



Grupo de activistas con megáfono protestando en las calles; concepto de huelga y manifestación. Fotografía: Halfpoint en iStock.



Distribución inequitativa de recursos



“Cuando hablamos del ciclo de vida de la IA, también estamos hablando de la extracción de minerales en las minas de cobalto en el Congo. ¿Hasta qué punto estamos considerando las implicaciones en materia de género en el desarrollo de IA en ese contexto?”

*Kelly Stone,
AI Observatory, Sudáfrica*

“Las bonitas palabras sobre las empresas emergentes y la liberación económica ocultan la realidad de que [...] experimentamos otra forma de colonización [...]”

*Maha Jouini,
African Center for Artificial Intelligence
and Digital Technology, Mauritania*



Dos de los obstáculos importantes para el acceso a los beneficios resultantes de la economía de la IA identificados en todas las consultas regionales fueron a) la distribución desigual de recursos en el ecosistema económico internacional de la IA, y b) la exclusión sistémica de mujeres y de otros grupos históricamente marginados, fenómenos que afectan particularmente a la mayoría global. La explotación laboral y el extractivismo de datos continúan perpetuando esta desigualdad, mientras que la cadena de valor de la IA afianza la injusticia ambiental.

Una preocupación apremiante que surgió en la consulta regional de Asia fue la explotación laboral generalizada de quienes, en la mayoría global, son las partes (personas) titulares de los datos. Las labores pesadas en las cadenas de valor de la IA —como el etiquetado de datos, la anotación de datos e imágenes, la transcripción y la traducción, así como la moderación de contenido— son realizadas desproporcionadamente por personal contratista empleado y localizado en el terreno de la mayoría global. Las consultas en América Latina también subrayaron que algunas formas de trabajo en la industria de la IA son valoradas menos que otras, y que estas labores pesadas se reservan comúnmente para personas marginadas que forman parte de la mayoría global. Esto tiende a invisibilizar estas contribuciones —y las de las mujeres y otros grupos marginados desde una perspectiva más amplia— dentro del ecosistema global en el ámbito de la IA.

También se identificó como una preocupación el extractivismo de datos, donde las personas en la mayoría global, así como los grupos y personas marginadas en el mundo minoritario, se ven reducidas a fungir como meras partes proveedoras de datos sin procesar y como consumidoras finales de productos y servicios de IA. La consulta regional en América Latina destacó que la norma de consentimiento libre e informado, comúnmente adoptada en los marcos legislativos de protección de datos personales, puede no ser adecuada para enfrentar el extractivismo de datos. Este enfoque contratista no da cuenta de las distorsiones que surgen debido al desequilibrio de poder entre las partes proveedoras de sistemas de IA y las partes titulares de los datos. Debido a que las corporaciones digitales transnacionales controlan las infraestructuras digitales fundamentales que sustentan gran parte de la vida económica y social, las personas se ven a menudo cooptadas en condiciones de servicio injustas, especialmente en modelos de negocio basados en la vigilancia de datos.

La consulta regional en el Oriente Medio y en África del Norte destacó algunas preocupaciones sobre el modo en que las cadenas de valor de la IA perpetúan dependencias económicas neocoloniales. La capacidad de los países de la mayoría global para desarrollar IA en sus propios términos, para su propio beneficio y de acuerdo con sus propias prioridades, está gravemente limitada. Las personas participantes opinaron que la distribución actual de los recursos en el ámbito de la IA podría intensificar la desigualdad global a largo plazo. Por último, como se subrayó en las consultas regionales en África Subsahariana y como también lo hicieron las personas participantes indígenas, la cadena de valor de la IA fortalece la injusticia ambiental mediante la extracción continua de recursos y otras actividades que afectan desproporcionadamente a las comunidades locales de la mayoría global.

Vías para afrontar los desafíos en igualdad de género y diversidad en ecosistemas de IA

Las consultas regionales identificaron cuatro vías principales para enfrentar los desafíos que presenta la igualdad sustantiva:

- **Desarrollar la capacidad del sector público y de la población en general**
Fortalecer el conocimiento, la comprensión general y la capacidad del sector público y de la población en general para involucrarse con los ecosistemas y el desarrollo de IA.
- **Alentar el diseño inclusivo en el sector privado**
Alentar al sector privado mediante diversas medidas para integrar procesos de diseño inclusivo a lo largo de todo el ciclo de vida de la IA
- **Garantizar la rendición de cuentas de las partes programadoras, proveedoras y operadoras**
Instituir marcos regulatorios legalmente vinculantes para garantizar la rendición de cuentas por los daños y violaciones de derechos cometidos por entidades públicas y del sector privado que diseñen, posean, desplieguen y operen sistemas de IA
- **Garantizar la participación significativa de los grupos marginados**
Amplificar las voces y expandir los espacios para la participación significativa en el desarrollo de IA y en los procesos de toma de decisiones relacionados con la IA.



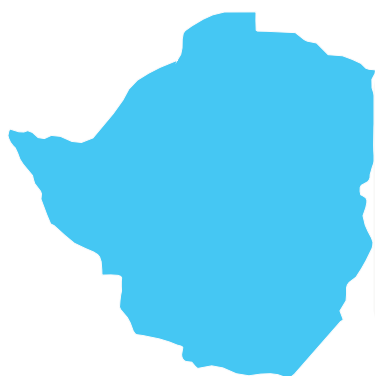
Cinco personas de diversidad manifiesta conversan sentadas a la mesa. Fotografía: Christina Morillo en Unsplash.



Desarrollar la capacidad del sector público y de la población en general

“Los sistemas de IA o algorítmicos se ven como algo destinado a un pequeño grupo de personas expertas. Apenas se tiene conciencia de la manera en que nos afectan indiscriminadamente y en la que se integran en nuestra vida cotidiana”.

*Dr Preeti Raghunath,
University of Sheffield, Malasia*



“El mejor recurso que podemos proporcionar a las personas responsables de políticas públicas es la conciencia. Instruirlas y hacerles entender los objetivos de la IA y lo que esta puede aportar a la sociedad”.

*Kudakwashe Dangajena,
African Union, Zimbabwe*

Las consultas regionales subrayaron que fortalecer la capacidad de las personas responsables de políticas públicas y de la población en general para participar en los procesos de toma de decisiones relacionados con la IA es determinante para abordar las desigualdades existentes. El fortalecimiento de la capacidad del funcionariado público y la educación cívica desarrollarán el conocimiento, la comprensión y la capacidad de todas las partes para involucrarse en los ecosistemas de IA.

Se identificó el desarrollo de la capacidad del sector público como un aspecto clave para asegurar la incorporación efectiva de medidas que garanticen la igualdad sustantiva en la gobernanza de la IA. Según varias de las personas que participaron en las consultas, los programas de educación cívica deben centrarse en promover la comprensión pública de los debates determinantes sobre la ética en el ámbito de la IA y sobre los principios normativos de una IA basada en los derechos humanos, y deben disipar a la vez el tecno-pesimismo. Esto implica crear mecanismos más inclusivos y participativos para que las personas participen en la configuración de las trayectorias que siga la IA.



Un hombre y una mujer observan una pantalla. Fotografía: Desola Lanre-Ologun en Unsplash.



Incentivar el diseño inclusivo y democratizar las innovaciones



“La IA debe ser un código abierto. En las comunidades deberíamos poder abrir, examinar, auditar y ver lo que esté haciendo el código”.

*Gerardo López Gómez,
Derechos Digitales, Ecuador*

“Uno puede inspirarse en un concepto, pero este debe estar arraigado, contextualizado y —en última instancia— diseñado, desarrollado, entrenado, desplegado, aprendido y refinado en el país que se pretenda beneficiar. No creo que haya atajos para eso”.

*Kelly Stone,
AI Observatory, Sudáfrica*



“La integración de la diversidad y la igualdad de género requiere la adopción de una estrategia interseccional —impulsada por la educación, centrada en la colaboración y fundamentada en políticas públicas— que dé prioridad máxima a la creación de una IA ética, inclusiva y equitativa”.

*Anónimo,
Asia*

En las consultas se identificó la necesidad de alentar la integración de procesos de diseño inclusivo a lo largo de todo el ciclo de vida de la IA. En las consultas regionales en Asia, África Subsahariana y América Latina se destacó la importancia del diseño y desarrollo tecnológico inclusivos para lograr resultados efectivos en igualdad de género y diversidad en el ecosistema de IA. Esto requiere alentar a todas las partes interesadas a priorizar el desarrollo de la IA de forma ética, inclusiva y equitativa. También implica proporcionar incentivos a las instituciones desarrolladoras de IA para que adapten sus modelos a las necesidades y condiciones locales, contextualizándolos para garantizar su adecuación a los contextos sociales, culturales y ambientales pertinentes donde se desplieguen. La diversidad lingüística es un aspecto de esta contextualización y surgió como una dimensión primordial en la consulta regional en Asia.

Las prácticas de diseño inclusivo pueden fomentarse por medio de políticas públicas que democratizen la innovación. Estas políticas deben proporcionar a diversas comunidades acceso a recursos, capacitación e infraestructura, a fin de habilitarlas para desarrollar tecnologías adaptadas a sus necesidades, en sus propios términos.



Garantizar la rendición de cuentas de las partes programadoras, proveedoras y operadoras



“No creo que las empresas tecnológicas puedan autorregularse y ser responsables en la rendición de cuentas. Es imprescindible que rindan cuentas ante nuestras instituciones democráticas”.

*Anónimo,
Tailandia*



“En la contratación y adquisición de tecnologías, los Estados deben establecer normas mínimas de transparencia, auditabilidad y participación social”.

*José Alfredo Hau Caamal,
Mozilla Foundation e Ação Educativa, Brasil*

Todas las personas participantes consideraron como prioridad esencial exigir rendición de cuentas a las partes programadoras, proveedoras y operadoras de sistemas de IA. Hubo consenso entre las consultas regionales sobre la necesidad de crear marcos regulatorios jurídicamente vinculantes para garantizar la rendición de cuentas. Debe responsabilizarse a las entidades públicas y del sector privado que diseñen, posean, desplieguen y operen sistemas de IA por los daños y violaciones de derechos que provoquen.

Aunque se consideraron importantes, las auditorías externas y las evaluaciones de los efectos sobre los derechos humanos para incorporar la diversidad de género y la igualdad en el sector de IA no son una solución definitiva. Se mencionaron los riesgos que conlleva dejar a cargo del sector privado la vigilancia de estos procesos, y también se mencionaron las lecciones aprendidas al examinar los sectores ambiental y financiero en materia de corrupción y captura de datos. Las partes interesadas estatales y de la sociedad civil deben estar preparadas y tener la capacidad para realizar auditorías y evaluaciones que prioricen el interés público.

Como enfatizaron quienes participaron en la consulta regional en América Latina, la transparencia y la auditabilidad del código fuente son fundamentales para garantizar la rendición de cuentas. En las consultas regionales realizadas en el Oriente Medio, África del Norte y América Latina se subrayó la necesidad de examinar —como una prioridad decisiva— el creciente riesgo de un nexo entre la vigilancia pública y la privada con IA. Quienes participaron en la consulta realizada en el Oriente Medio y en África del Norte destacaron el rol de los Estados para establecer tanto requisitos en cuanto a la debida diligencia en materia de derechos humanos, como mecanismos de rendición de cuentas y reparación de daños en el desarrollo, uso, examen y contratación de tecnologías digitales.



Garantizar la participación significativa de los grupos marginados



“Un factor que dificulta las cosas es la desigualdad prevalente en América Latina sobre quién puede y quién no puede hablar de tecnología, ya que en gran medida todavía se presume que las únicas personas que pueden criticar a la IA son las dedicadas a la ingeniería”.

*Marcos Cornelio Sánchez Ramírez,
PIT Policy Lab, México*

“Las conversaciones sobre la gobernanza de la IA son convocadas por personas que están en el poder y, por lo tanto, permanecen en esos círculos retroalimentándose entre sí y creando oportunidades entre ellos, lo que limita su alcance en lugar de ampliar el interés público”.

*Dr Preeti Raghunath,
University of Sheffield, Malasia*

En las consultas se subrayó la necesidad de democratizar el desarrollo y los procesos de toma de decisiones en torno a la IA mediante una participación significativa. Las personas participantes apoyaron la transición de un enfoque limitado y realizado por personas expertas, a un proceso más participativo que abarque a toda la sociedad, con la participación extensa de la ciudadanía. Este cambio es esencial para abordar posibles injusticias y sesgos, así como la opacidad de las tecnologías de IA.

Las personas participantes de Asia, África Subsahariana y América Latina destacaron los beneficios que aporta codiseñar sistemas de IA con la participación del público usuario final y promover colaboraciones interdisciplinarias. Involucrar al público en estos debates es fundamental para desarrollar sistemas de IA sensibles a los contextos del mundo real y a las perspectivas en diversidad, con un enfoque particular en los derechos de quienes corren mayor riesgo de enfrentar perjuicios.

Garantizar mecanismos multipartitas efectivos con participación inclusiva y significativa es fundamental, según quienes participaron en todas las consultas. Todas las partes —incluyendo los gobiernos, el sector privado, la sociedad civil, las comunidades técnicas y académicas, así como la población usuaria— deben participar con sus respectivos roles conforme a los principios de cooperación digital global establecidos en la Agenda de Túnez en la Cumbre Mundial sobre la Sociedad de la Información (CMSI). Actualmente, la sociedad civil y los grupos marginados están subrepresentados. Se necesitan medidas específicas, como la asignación de fondos, la planificación presupuestaria y el establecimiento de alianzas, para que estos grupos puedan participar de manera efectiva en la formulación de políticas. Además, en la consulta en Asia se destacó la necesidad de prevenir la captura del tema por parte de las élites, donde grupos dominantes desvían hacia sus propios intereses los fondos públicos destinados a la población en general. Los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos (UNHRC, 2011) se consideraron esenciales para garantizar la implementación de procesos de debida diligencia en todo el ciclo de vida de la IA y para garantizar una participación significativa —incluida la de personas con discapacidad— y la realización de evaluaciones de los efectos.

Breve resumen de los desafíos y de las vías

Las consultas regionales y las de grupos específicos revelaron tres desafíos principales para implementar la igualdad de género y la diversidad en los ecosistemas de IA: 1) la amalgama entre acceso e inclusión; 2) la exclusión del conocimiento y la invisibilización; y 3) la distribución desigual de los recursos.

Las consultas regionales identificaron cuatro vías principales para abordar los desafíos que presenta la igualdad sustantiva: 1) desarrollar la capacidad del sector público y de la población en general; 2) alentar el diseño inclusivo en el sector privado; 3) garantizar la rendición de cuentas por parte de las partes programadoras, proveedoras y operadoras; y 4) garantizar la participación significativa de los grupos marginados.



Una mujer en silla de ruedas se desplaza por el corredor de un edificio de ladrillo. Fotografía: Marcus Aurelius en Unsplash.



PARTE II.

Marco conceptual



Un primer plano de una mano vívidamente pintada de colores mixtos sobre un fondo azul. Fotografía: Alexander Grey en Unsplash.

Un enfoque transformador de la igualdad sustantiva

Con fundamento en enfoques de género transformadores, este informe promueve un enfoque transformador de las políticas en el ámbito de la IA para lograr la *igualdad sustantiva*.

Igualdad sustantiva

El concepto de “igualdad sustantiva” se refiere a la igualdad *de facto* (igualdad de hecho o igualdad real) o a la igualdad de los resultados (Fredman, 2016; UN Women, 2019).

Esto incluye garantizar que todas las personas tengan igualdad de oportunidades y de acceso a ellas, así como un entorno favorable para lograr resultados iguales. La igualdad sustantiva planteada en este informe incorpora una comprensión de la equidad⁷, la igualdad de género⁸ y la diversidad que aboga por la igualdad de oportunidades y de resultados mediante medidas que reconozcan las diferencias, fomenten la expresión y la participación, y eliminen la discriminación, la marginación y la distribución desigual del poder y de los recursos.

La igualdad sustantiva se fundamenta en el marco de los derechos humanos y se refleja en varios derechos que se fortalecen entre sí. La igualdad de género, la participación significativa y la inclusión tienen el objetivo de garantizar que puedan disfrutar plenamente de sus derechos todas las personas, independientemente de su sexo o características SOGIESC, o de su edad, discapacidad, origen étnico, religión, posición económica o cualquier otra condición social. La igualdad de género se reconoce como un principio central de los derechos humanos y figura entre los Objetivos de Desarrollo Sostenible (ODS). Este concepto garantiza que las personas, sin importar su género, tengan igualdad de oportunidades y estén libres de toda discriminación⁹. La participación significativa y la inclusión son derechos humanos establecidos en convenciones internacionales que garantizan que las personas desempeñen un rol genuino y efectivo en los procesos de toma de decisiones que afecten su vida. Lograr la igualdad sustantiva es un elemento primordial para llevar a la realidad los derechos de igualdad de género, la participación significativa y la inclusión para todas las personas.



Siete mujeres indias tejiendo sentadas en círculo.
Fotografía: AROYBARMAN en iStock.



Río que serpentea por un bosque de manglares en Krabi, Tailandia. Fotografía: Jitti Narksompong en iStock.



Igualdad sustantiva en el derecho internacional sobre los derechos humanos

Existen nueve instrumentos internacionales fundamentales de derechos humanos (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, Office of the High Commissioner for Human Rights, 2024b). A continuación se presentan algunos que reflejan los principios de la igualdad sustantiva:

La Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW, por sus siglas en inglés)

En materia de igualdad de género, la CEDAW pasa de un enfoque de igualdad de jure en la ley y en las políticas a un enfoque de la igualdad de facto al abordar de manera efectiva las barreras sistémicas que impiden la participación igualitaria de mujeres y niñas, en toda su diversidad, en todos los ámbitos de la vida (IWRRAW, sin fecha; Shapiro, sin fecha). El Comité de la CEDAW reconoce que la discriminación de género está inextricablemente ligada a otros factores como raza, origen étnico, religión, salud, posición, edad, clase, casta, orientación sexual e identidad de género (Campbell, 2015).

El Pacto Internacional de Derechos Económicos, Sociales y Culturales (PIDESC)

El PIDESC exige la no discriminación sustantiva en el disfrute de los derechos sociales y económicos, sin discriminación alguna por razones de raza, color, sexo, lengua, religión, opiniones políticas o de otro tipo, origen nacional o social, propiedades, nacimiento ni de ninguna otra condición (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 1966).

La Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial (ICERD, por sus siglas en inglés)

La ICERD impone a los Estados Partes la obligación de adoptar medidas específicas de políticas que garanticen a los grupos marginados debido a razones de raza, color, ascendencia u origen nacional o étnico, así como a las personas que pertenezcan a estos grupos, el disfrute pleno e igualitario de los derechos humanos y de las libertades fundamentales (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 1965).

La Convención sobre los Derechos de las Personas con Discapacidad (CDPD)

La CDPD reorienta el enfoque en materia de discapacidad pasando del modelo médico a un modelo fundamentado en los derechos humanos que incluye los cambios sistémicos en los aspectos jurídicos, sociales y de formulación de políticas necesarios para formar sociedades inclusivas. La Convención establece medidas específicas para que los Estados —como garantes del cumplimiento de los derechos— eliminen toda barrera y posibiliten a las personas con discapacidad el acceso y pleno disfrute de sus derechos (Goldschmidt, 2017).

La igualdad sustantiva se basa en el principio de que el derecho humano a la igualdad debe responder a las necesidades de las personas que estén en desventaja o que sean menospreciadas, excluidas o ignoradas. Es responsabilidad de todos los Estados —como garantes de los derechos— la de proteger, fomentar y garantizar el cumplimiento de los derechos humanos. Los enfoques fundamentados en los derechos humanos para la formulación de políticas, proyectos, programas y sistemas generalmente incluyen las siguientes actividades clave:

- Reconocimiento de los derechos humanos y de los instrumentos relacionados que resulten afectados por la iniciativa o que la afecten.
- Empoderamiento de quienes son titulares de derechos y su acceso a la justicia para que puedan reclamar sus derechos de manera individual y colectiva.
- Desarrollo de la capacidad de las instituciones y el personal garantes de los derechos para proteger y promover los derechos humanos.
- Participación significativa de los grupos marginados.

Entre las obligaciones específicas de los Estados se incluyen la eliminación de las barreras sistémicas contra la igualdad de género y la eliminación de las diversas formas de discriminación que contribuyen a perpetuar la marginación. También se incluye el fomento activo de prácticas inclusivas para crear un entorno en el que todas las personas puedan participar plenamente y beneficiarse de los recursos y oportunidades que ofrece la sociedad.¹⁰

Desde la perspectiva de la igualdad sustantiva, las políticas deben abordar las causas fundamentales para corregir las desigualdades de género, eliminar las barreras estructurales —incluidos los desequilibrios de poder y el afianzamiento geopolítico— e instar a las poblaciones marginadas a la acción. Esto requiere atender factores sociales interseccionales como edad, identidad de género, orientación sexual, discapacidad, origen étnico, religión, condición migratoria o económica, entre otros, para garantizar que todas las personas, independientemente de su identidad, estén incluidas en todos los aspectos de la sociedad y tengan el poder de influir en los resultados. Dicho de otra manera, las políticas deben ser transformadoras.



Un grupo diverso de personas, principalmente mujeres, y personas de First Nations (Estados Unidos), todas con pintura facial, protestando.

La igualdad real se puede lograr mediante el uso de medidas específicas para el cambio transformador (UN Women, 2015). Las políticas, programas y presupuestos efectivamente inclusivos para la igualdad de género y la diversidad son los recursos que permiten la expansión estratégica y significativa de la igualdad sustantiva y les garantizan a los grupos marginados su acceso a los derechos humanos. El concepto de “igualdad sustantiva” proporciona un fundamento sólido para impulsar un cambio transformador que aborde tanto los desafíos que presentan la desigualdad y la exclusión a lo largo de todo el ciclo de vida de la IA, como también la formulación de las políticas en este ámbito. La siguiente sección describe el enfoque socio-técnico hacia los sistemas de IA y el ciclo de vida de la IA necesario para la aplicación satisfactoria de políticas sobre IA transformadoras.

Cambio transformador

Buscar un cambio transformador o adoptar un enfoque transformador de género implican abordar las causas fundamentales de la desigualdad para hacer que las instituciones y relaciones sociales sean más inclusivas y equitativas eliminando barreras estructurales o jurídicas e instando a los grupos marginados a la acción (UNRISD, 2017). Las acciones para lograrlo incluyen la introducción de cambios en la legislación y en las políticas, la adaptación y mejora de sistemas y servicios, la redistribución de recursos y la modificación de normas, creencias o prejuicios (incluidos los estereotipos), comportamientos y prácticas perjudiciales o negativas.

(UNICEF, 2021)

Un enfoque socio-técnico de los sistemas de IA

Trabajar por la igualdad de género y la diversidad en el ámbito de la IA requiere considerar todo el ciclo de vida de la IA desde una perspectiva socio-técnica. Un enfoque socio-técnico de los sistemas de IA y de todo el ciclo de vida de la IA reconoce que el rendimiento y los efectos de los sistemas de IA se derivan de la interacción entre el diseño técnico, la infraestructura y las dinámicas e incentivos sociales. Se suele describir los sistemas de IA como sistemas automatizados o como un conjunto de técnicas informáticas que pueden tener un efecto en sus entornos. Las definiciones de los sistemas de IA apegadas a tales términos los describen como “sistemas automatizados que pueden influir en sus entornos” (OCDE, 2024b). Estas descripciones reconocen el rol fundamental que los sistemas de IA pueden tener o llegarán a tener para influir sobre el medio ambiente, las prácticas sociales, las normas, los derechos y las estructuras (UNESCO, 2024c). Los enfoques basados en modelos para formular políticas sobre la IA suelen adoptar esta descripción de los sistemas de IA y del ciclo de vida de la IA.

Sistemas de inteligencia artificial (IA)

Un sistema de IA es un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de los datos de entrada que recibe, cómo generar información de salida como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos reales o virtuales. Una vez desplegados, los distintos sistemas de IA presentan diversos niveles de autonomía y varían en su capacidad de adaptación

(OECD, 2024b)

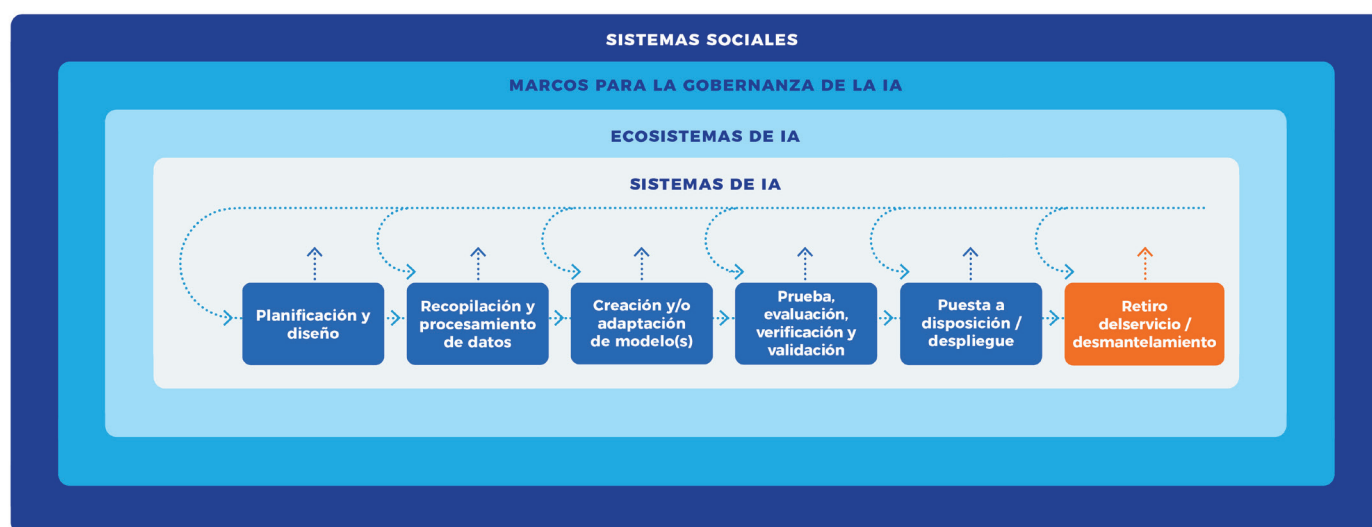
Las políticas sobre la IA se centran a menudo en enfoques basados en modelos para alcanzar los objetivos de diversidad, equidad e inclusión. Estos enfoques —como la eliminación técnica de sesgos— tienen como objetivo corregir las inexactitudes en los datos causadas por sesgos sociales y mejorar los sistemas de IA para lograr en estos una representación justa, así como un trato equitativo y no discriminatorio (Cachat-Rosset y Klarsfeld, 2023). Por ejemplo, eliminar sesgos en los datos que subestiman a las mujeres puede corregir sesgos en la puntuación de las evaluaciones de desempeño, mientras que seleccionar mejores variables proxy puede reducir la discriminación racial en algoritmos de riesgo de reincidencia o en modelos cuyo objetivo es predecir la reincidencia (el “rasgo objetivo”). Usar “registros de arrestos” como variable proxy, por ejemplo, para predecir la reincidencia refleja la elaboración sesgada de perfiles raciales ya presente entre las fuerzas policiales (Benjamin, 2019). La desigualdad de género y la discriminación social se pueden abordar con la incorporación de ajustes matemáticos, con diseños meticulosos, con la recopilación de datos que reflejen la diversidad y que sean representativos, así como con pruebas y supervisión colaborativas (Hellman, 2023; Jora *et al*, 2022; Solaiman *et al*, disponible próximamente). Sin embargo, la precisión de los datos no puede garantizar por sí sola la equidad cuando existen sesgos sistémicos en el rasgo que se fije como objetivo (Joyce *et al*, 2021; Kong, 2022; Ovalle *et al*, 2023a). Mientras que los enfoques centrados en modelos corrigen inexactitudes en los datos, un enfoque socio-técnico aborda las causas fundamentales de la desigualdad integrada en las dimensiones socio-técnicas del ciclo de vida de la IA (Iason, 2022). Este enfoque destaca el efecto mutuo entre los elementos sociales y técnicos.



Gente cruzando una calle. Fotografía: bee32 en iStock.

El ciclo de vida de la IA

Una interpretación socio-técnica de los sistemas de IA considera a la sociedad y a la tecnología como un sistema coherente —de hecho, una multiplicidad de sistemas— y un proceso recurrente. Esta interpretación incluye las normas, prácticas, materiales e infraestructuras en los que —y mediante los cuales— operan los sistemas tecnológicos. También abarca narrativas, imaginarios colectivos, ecosistemas, intereses geopolíticos y económicos, y la IA como tema de estudio (Ricaurte y Zasso, 2021). Todos estos elementos sociotécnicos se interconectan durante todo el ciclo de vida de los sistemas de IA (Chen y Metcalf, 2024). El ciclo de vida del sistema de IA descrito en este informe se basa en conceptualizaciones existentes y las integra con sistemas socio-técnicos adicionales clave en la formulación de políticas sobre igualdad de género y diversidad¹¹.



(Imagen adaptada de OECD, 2024a; Ricaurte, 2024)

Las etapas del ciclo de vida de los sistemas de IA

Planeación y diseño

Esta etapa incluye el planteamiento de la problemática. Las partes involucradas definen las metas, objetivos y alcance del sistema de IA, así como los problemas clave u oportunidades que el sistema de IA debe abordar.

Recopilación de datos y (pre)procesamiento

En esta etapa las partes investigadoras y desarrolladoras identifican las fuentes de datos necesarias para entrenar y validar los modelos de IA. Entre estas pueden incluirse bases de datos, programas para Interfaz de Programación de Aplicaciones (API, por sus siglas en inglés), sensores e inserción manual de datos. Es importante determinar las prácticas de recopilación de datos más pertinentes para la operación a fin de garantizar que se incluyan todos los elementos y procesos necesarios. A este paso le sigue el preprocesamiento de datos. Este comprende limpiar los datos, revisar los valores faltantes y transformarlos en un formato adecuado para modelar el sistema de IA. Otros procesos implican la normalización (ajustando los datos a un intervalo estándar), la configuración de las características (ajustar la escala de los datos) y el diseño de características (crear nuevas características a partir de los datos existentes).

Creación de modelos

Esta etapa consiste en modelar los datos adquiridos utilizando modelos de IA según los requisitos del problema. Por ejemplo, usar modelos basados en visión para datos de imágenes y modelos basados en lenguaje para datos textuales.

Pruebas, verificación y validación

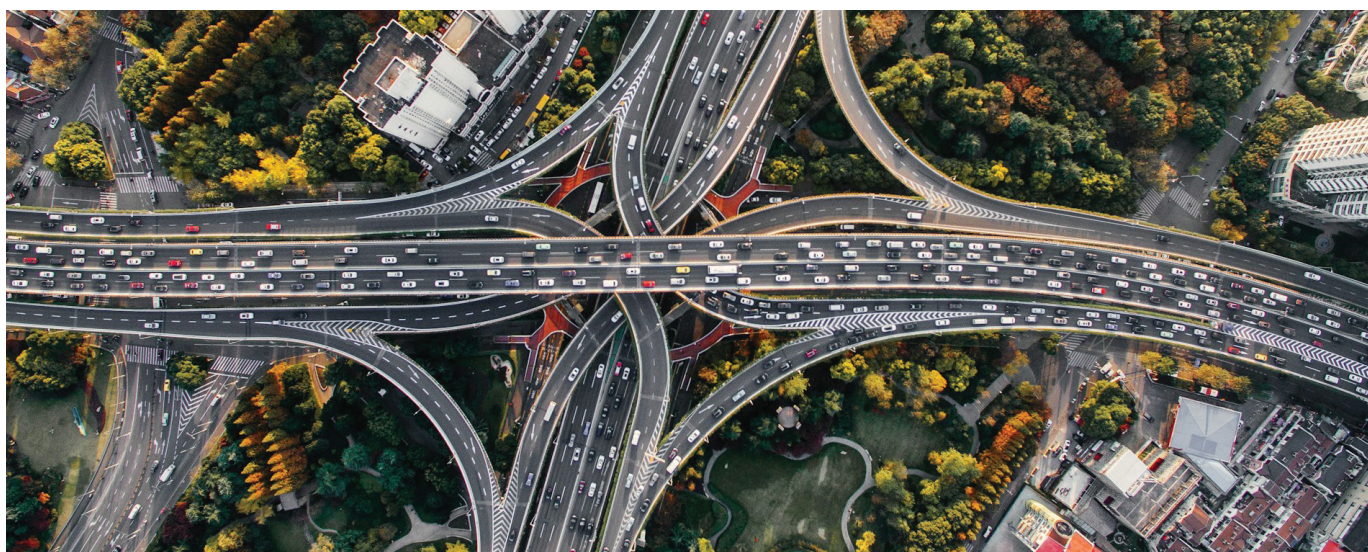
La realización de pruebas, la verificación y la validación son actividades clave durante el ciclo de vida de la IA. Esto comprende usar parámetros de referencia para identificar errores, sesgos o limitaciones en los modelos de IA. Otras actividades comportan validar los modelos con técnicas como pruebas A/B (un método de comparación) para garantizar la precisión y la fiabilidad.

Despliegue, operación y vigilancia

Las personas responsables del despliegue de un sistema de IA fijan un cronograma para tal fin antes de ponerlo a disposición del público usuario. En esta etapa es fundamental colaborar con equipos expertos en tecnología de la información y administradores de sistemas para integrar el de IA con los sistemas, bases de datos e interfaces del público usuario existentes. La aplicación de mecanismos de monitoreo y registro para rastrear el funcionamiento del sistema de IA e identificar todo problema que se presente será útil en la etapa siguiente. Durante la etapa de operación y monitoreo, las personas a cargo del despliegue mantienen y monitorean continuamente el rendimiento del sistema de IA.

Retiro del servicio o desmantelamiento

La decisión de retirar del servicio o de desmantelar un sistema de IA puede producirse en cualquier momento durante su ciclo de vida.



Perspectiva aérea de una intrincada red de autopistas, con una maraña de carreteras, puentes y numerosos vehículos. Fotografía: Denys Nevozhai en Unsplash.

Oportunidades de intervención

Contemplar el ciclo de vida del sistema de IA desde una perspectiva socio-técnica ayuda a identificar oportunidades de intervención para crear sistemas de IA más inclusivos y equitativos. La desigualdad, la inequidad y los diferentes tipos de daños o efectos negativos ocurren en todas las etapas que componen los ciclos de vida de los sistemas de IA. Muchos de estos desafíos provienen de injusticias epistémicas que pueden ocurrir cuando las perspectivas y las experiencias de las personas o comunidades son ignoradas o socavadas en etapas clave en el ciclo de vida del sistema de IA debido a una limitada representación y reconocimiento de los diversos tipos de conocimientos y experiencia especializada que intervienen (Cheong et al, 2021; Nihei, 2022). Comprender y manejar los elementos sociales y técnicos en juego en cualquier contexto relacionado con los sistemas de IA es, por lo tanto, un elemento clave para intervenir con eficacia.

Transparencia, interpretabilidad y explicabilidad en los sistemas y procesos de IA

La transparencia implica que las instituciones desarrolladoras de sistemas documenten el proceso de selección de los conjuntos de datos, variables y modelos utilizados en el desarrollo, el entrenamiento, la validación y las pruebas de la IA. Deberán detallar las medidas que garanticen la calidad de los datos y de la información de salida, e indicar el nivel de confianza en dicha información del sistema, con la intervención humana necesaria en caso de obtener información de salida poco fiable.

La transparencia también incluye documentar los métodos para la detección de sesgos potenciales.

Deberá ser obligatorio efectuar una verificación y una validación independientes en los sistemas que afecten significativamente la vida y el bienestar.

El concepto interpretabilidad se refiere a proporcionar información clara sobre los procedimientos que aplican los algoritmos. Esto ayuda a las partes interesadas a comprender el funcionamiento interno de los sistemas de IA.

La explicabilidad comprende detallar las decisiones específicas que tomen los algoritmos. Es un factor crucial en las políticas públicas y en otros contextos en los que las decisiones algorítmicas puedan beneficiar de forma injusta a determinados grupos. Garantizar que los grupos marginados tengan derecho a recibir explicaciones evita que se perpetúe la desigualdad. Distinguir entre explicaciones genuinas y justificaciones a posteriori es esencial para preservar la responsabilidad y la confianza en los sistemas de IA.

(Adaptado de ACM, 2022)

Los sistemas socio-técnicos que interactúan con el ciclo de vida de los sistemas de IA comprenden los ecosistemas de IA, los marcos de gobernanza de IA y los sistemas sociales vinculados al contexto económico, así como los aspectos de financiamiento, relaciones sociales, normas, valores, prácticas, instituciones, infraestructuras y entornos. Los ecosistemas de IA comprenden a todas las partes actoras o interesadas que intervienen en el ciclo de vida del sistema de IA, o que se ven afectadas por él. Garantizar la participación de las principales partes interesadas teniendo en cuenta su posición es fundamental para que esta sea significativa. Integrar a las partes interesadas e interdisciplinarias pertinentes en los procesos de toma de decisiones ayuda, por ejemplo, a evitar que los modelos hereden o amplifiquen sesgos dañinos y prácticas discriminatorias. Los marcos de gobernanza de la IA incluyen marcos, políticas, legislación y reglamentos a nivel nacional, internacional o mundial que configuran el uso, el desarrollo, la experiencia y los efectos de los sistemas de IA. Identificar y desarrollar marcos, estrategias y mecanismos aplicables a todos los niveles y subsanar las deficiencias para garantizar la igualdad de género y la diversidad son cometidos importantes para garantizar la transparencia, la rendición de cuentas y el cumplimiento de las reglas en los sistemas y procesos de IA. Los sistemas sociales, como el contexto económico y el financiamiento, influyen mucho en las posibilidades de cambio. Para lograr una inclusión significativa es fundamental identificar y abordar las barreras erigidas contra la igualdad de género y la diversidad garantizando un acceso equitativo a los recursos financieros, el conocimiento y las redes. Del mismo modo, es importante transformar las normas y prácticas institucionales y sociales mediante la toma de conciencia, la educación, las políticas institucionales y otras herramientas para garantizar estrategias de IA eficaces. La siguiente sección presenta el marco de políticas transformadoras en el ámbito de la IA para lograr la igualdad sustantiva y presenta una base holística para las intervenciones relacionadas con políticas sobre IA.

El marco de políticas transformadoras en el ámbito de la IA

El marco de políticas transformadoras en el ámbito de la IA adopta un enfoque transformador para lograr la igualdad sustantiva en todo el ciclo de vida de la IA y en la formulación de las políticas relacionadas.¹² Las políticas transformadoras en el ámbito de la IA integran los derechos humanos en el núcleo de los ecosistemas de IA, eliminando limitaciones y mejorando las capacidades para garantizar el disfrute de los derechos humanos en la totalidad del entorno tecnológico.

El marco pretende instar a las mujeres y a otros grupos marginados a la acción para lograr el goce pleno de sus derechos. Esto incluye eliminar tanto la discriminación directa como la indirecta y apoyar acciones afirmativas para promover la igualdad sustantiva (Gurumurthy y Deepti, 2023). El marco propone medidas para garantizar la implementación de todo cambio necesario para eliminar las barreras institucionales que perpetúan la exclusión de las mujeres y de otros grupos marginados en las etapas de desarrollo, toma de decisiones y gobernanza de la IA. Reconocer la exclusión estructural y la distribución desigual del poder y los recursos permite intervenir en los modelos de toma de decisiones impulsados por IA y en otros aspectos del diseño, desarrollo y despliegue de la IA. Esto aumenta las posibilidades de lograr resultados equitativos y produce resultados más justos.

La siguiente sección describe las formas principales en las que un ecosistema de IA transformador puede promover la igualdad de género y la diversidad.

Dimensiones clave

El marco de políticas transformadoras en el ámbito de la IA hacia la igualdad de género y la diversidad se basa en tres dimensiones clave de la igualdad sustantiva. El marco adopta un enfoque sistémico para integrar los derechos humanos (Bartoletti y Xenidis, 2023) y profundiza en estas dimensiones para que puedan ser transformadoras las políticas y la legislación relacionadas con igualdad de género y diversidad en la innovación de los sistemas de IA. En esta sección se describen las siguientes tres dimensiones fundamentales:

- 1. Reparación de desventajas sistémicas:
el derecho a la inclusión**
- 2. Rectificación del déficit democrático:
el derecho a la participación**
- 3. Reversión de errores de reconocimiento:
el derecho a la dignidad**

1. Reparación de desventajas sistémicas: el derecho a la inclusión

Un marco de políticas transformadoras para la igualdad de género y la diversidad puede abordar la desventaja sistémica existente en los ecosistemas de IA promoviendo activamente el derecho a la inclusión. La desventaja sistémica puede entenderse como “privar a una persona de oportunidades genuinas para seguir los caminos que ella misma se proponga” mediante la asignación injusta de recursos, bienes y oportunidades en una sociedad (Fredman, 2016). Los ecosistemas de IA pueden perpetuar las desventajas sistémicas mediante la injusticia distributiva; por ejemplo: la asignación excluyente de oportunidades en los mercados en forma de discriminación racial en los algoritmos de anuncios de viviendas o la exclusión de servicios públicos en forma de discriminación en la automatización de sistemas de protección social (Stauffer, 2023). También contribuyen a perpetuar la injusticia redistributiva, por ejemplo, cuando para consolidar monopolios intelectuales los grandes consorcios se apropian de lo que podrían considerarse plataformas de datos compartidos (Rikap, 2022).

Un marco de políticas transformadoras puede fomentar activamente el derecho a la inclusión tomando las siguientes acciones:

- Eliminar los daños que produce la discriminación directa e indirecta en los ecosistemas de IA.
- Establecer —entre las partes programadoras y proveedoras de ecosistemas de IA— la obligación de promover procesos de diseño equitativos y acciones afirmativas para mujeres y otros grupos marginados (Bartoletti y Xenidis, 2023).
- Orientar la innovación pública y privada hacia la generación de beneficios para los grupos marginados. Esto incluye, por ejemplo, presupuestos destinados al apoyo de innovaciones de IA que mejoren la accesibilidad y la seguridad social de poblaciones racializadas.

- Proteger las plataformas de datos compartidos y la innovación en sistemas abiertos de IA contra la apropiación privada mediante el uso de marcos de propiedad intelectual adecuados y basados en el principio de ciencia abierta a fin de democratizar los beneficios del progreso científico (Krishna, 2020).

2. Rectificación del déficit democrático: el derecho a la participación

Un marco de políticas transformadoras para la igualdad de género y la diversidad debe abordar el déficit democrático en los ecosistemas de IA mediante el derecho a la participación y la creación de espacios para las voces de los grupos marginados. Esto permite a los garantes de derechos preservar la premisa básica de la democracia representativa: que cada persona sujeta a una política tenga voz en su formulación (Pande, 2002).

Los sesgos, la discriminación y la injusticia en los modelos de IA representan un déficit democrático en los ecosistemas de IA y son producto de la exclusión interseccional y de la desigualdad perpetuada en los espacios donde se crea la IA (West et al, 2019). Teniendo en cuenta que los espacios de innovación en sistemas de IA son en su mayoría propiedad de empresas transnacionales con sede en países de minorías globales, las iniciativas de diversidad y equidad tienden a reducirse a soluciones técnicas superficiales, como agregar más datos a la red neuronal o aumentar la visibilidad de otras identidades (Molnar, 2024; Niesen, 2016; Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2023). Además, se tiende a aplicar los sistemas de IA de mayor riesgo (identificación biométrica, reconocimiento de emociones y vigilancia automatizada) sobre grupos de personas ya marginadas por nuestros sistemas —personas en movimiento (personas migrantes, refugiadas y otras personas desplazadas) o personas en conflicto con la ley, por ejemplo— que son excluidas a menudo de las protecciones estatales garantizadas a la comunidad política nacional de ciudadanos (#ProtectNotSurveil, 2024).

Dos tipos de medidas políticas resultan primordiales:

- Instituir mecanismos sociales para realizar un escrutinio público y obtener retroalimentación en los procesos privados de fijación de normas sobre sistemas de IA.
- Garantizar los derechos de acceso a la información y a la justicia, así como el derecho a la participación pública en la toma de decisiones sobre leyes y políticas relacionadas con la IA (Smuha, 2021).

3. Reversión de errores de reconocimiento: el derecho a la dignidad

Los errores de reconocimiento —denigración, humillación y falta de valoración de las personas— constituyen una violación del principio de que toda la humanidad está constituida de seres iguales (Fredman, 2016). El estigma, los estereotipos, la humillación y la violencia “por motivos de género, raza, discapacidad, orientación sexual u otros factores sociales locales” (Fredman, 2016) constituyen un ataque a la identidad personal: la reivindicación que tiene toda persona de “ser un ente social, racional y capaz de actuar de manera autónoma” (Martineau et al, 2012).



Grupo de personas, principalmente mujeres, protestando por sus derechos en Puebla de Zaragoza, México. Fotografía: Domingo Alvarez E en Unsplash.

Por un lado, la interacción entre las máquinas y el ser humano en el desarrollo y el despliegue de sistemas de IA formaliza la discriminación; por otro, cosifica y amplifica las formas existentes de desigualdad social. Esto lo reconocen como opresión algorítmica algunas personas académicas feministas (Benjamin, 2019; Eubanks, 2017; Noble y Roberts, 2019; West, 2020). Dicha opresión algorítmica se manifiesta no solo en la negación injusta de oportunidades para los grupos marginados (lo que se comentó en la dimensión 1), sino también como injusticia representacional (como los generadores de imágenes de IA que sexualizan a mujeres racializadas o los algoritmos de medios sociales que amplifican el sexismo y la misoginia).

Un marco de políticas transformadoras para la igualdad de género y la diversidad debe revertir estos errores de reconocimiento y acreditar la lucha moral de los grupos marginados para que se reconozca su identidad personal, protegiendo y promoviendo el derecho a la dignidad. La dignidad es fundamental en toda sociedad que realmente adopte la diversidad. El derecho a la dignidad consagrado en la jurisprudencia de los derechos humanos exige que los Estados reconozcan la necesidad de una acción concertada —y que la faciliten— contra todo proceso descalificador que estigmatice a una persona o grupo por cualquier motivo, así como contra los mecanismos institucionales que perpetúen dicho proceso.

Esto implica una acción política de doble enfoque:

- Instituir la obligación positiva de las partes proveedoras y usuarias de sistemas de IA de prevenir la discriminación algorítmica (directa y por intermediación).
- Establecer salvaguardias radicales de transparencia y rendición de cuentas en los ecosistemas de IA con auditorías algorítmicas obligatorias ex ante, y asegurar que los secretos comerciales y las leyes sobre patentes no interfieran con las medidas de rendición de cuentas destinadas a eliminar los errores de reconocimiento en los sistemas de IA.



Excavación minera en una montaña. Fotografía: Vlad Chețan en Pexels.

PARTE III.

Prácticas prometedoras y recomendaciones



Mujer de pelo trenzado en un campo de hierba alta y seca, con los brazos cruzados por encima de la cabeza levantados hacia un cielo nublado y vestida con una prenda estampada de colores. Fotografía: Ian Kiragu en Unsplash.

Prácticas prometedoras hacia una igualdad sustantiva en el ámbito de la IA

En esta sección se presenta un conjunto de prácticas prometedoras de todo el mundo que demuestran cómo hacer progresar de manera efectiva en los ecosistemas de IA las tres dimensiones clave de la igualdad sustantiva descritas en el marco de políticas transformadoras en el ámbito de la IA —abordar la desventaja sistémica, rectificar el déficit democrático y revertir los errores de reconocimiento—. Los aprendizajes derivados de cada una se resumen como perspectivas para políticas públicas.

Las prácticas abarcan una variedad de acciones —creadas por diversas partes interesadas, tales como agencias multilaterales, organizaciones de la sociedad civil, organismos públicos en el plano nacional y organizaciones en redes— que fomentan de manera efectiva los objetivos de igualdad de género y diversidad en los ecosistemas de IA. Estas prácticas fueron identificadas durante las consultas regionales o hechas con grupos específicos. Las iniciativas tienen el objetivo de incluir los derechos y la participación significativa de varios grupos marginados —ya sea a nivel local, regional o global—, o tienen ese objetivo, o trabajan para lograrlo. Este conjunto es ilustrativo y no exhaustivo; está destinado a proporcionar a las personas responsables de políticas públicas y a otras partes interesadas ejemplos concretos de las acciones que pueden reproducirse, apoyarse o adaptarse a otros contextos.

Las prácticas prometedoras se clasifican como ejemplos de las siguientes categorías (entre otras):

- **Recursos para la capacitación y la educación pública**
Guía experta y recursos para las personas responsables de políticas públicas y para la población en general.
- **Prácticas de diseño tecnológico inclusivo y de innovaciones democráticas**
Modelos tecnológicos —desarrollados por organizaciones sin fines de lucro y agencias públicas— que integran de manera eficiente los objetivos de igualdad de género y diversidad en las etapas apropiadas del ciclo de vida de la IA.
- **Medidas aplicables a la rendición de cuentas**
Índices, iniciativas de monitoreo y medidas de sensibilización que promueven la rendición de cuentas de las partes proveedoras y responsables del despliegue de IA a fin de respetar, proteger y promover los derechos humanos de todas las personas.
- **Iniciativas de inclusión y participación significativas**
Iniciativas que promueven la participación activa y significativa, así como la cogobernanza en los ecosistemas de datos e IA.



Una mujer sostiene un papel que dice “SOMOS IGUALES”. Fotografía: Jacob Wackerhausen en iStock.

Recursos para la capacitación y la educación pública

Diálogo mundial de la UNESCO

Nombre de la iniciativa:

[Artificial Intelligence and Gender Equality: Key Findings of UNESCO's Global Dialogue](#)

Organización:

[United Nations Educational, Scientific and Cultural Organization \(UNESCO\)](#)

País/Región:

Global

Descripción

El informe *Global Dialogue* de la UNESCO se enfoca en integrar la igualdad de género en las políticas y prácticas relacionadas con los sistemas de IA. El informe examina e identifica las lagunas en la integración de la igualdad de género en los principios existentes de la IA. Ofrece recomendaciones y un plan de acción que orienta a cada grupo de partes interesadas sobre la forma de poner en práctica las recomendaciones. El plan de acción se concentra en cinco aspectos: 1) sensibilización; 2) garantizar que la igualdad de género siga siendo una prioridad; 3) creación de coaliciones; 4) capacitación y financiamiento; y 5) investigación, monitoreo y financiamiento. Este recurso se creó en diálogo con expertos mundiales en IA e igualdad de género de diversos grupos de personas interesadas y se financió mediante una asociación con la Agencia Alemana de Cooperación Internacional (GIZ, por sus siglas en inglés).

Cómo fomenta dicha iniciativa la igualdad sustantiva

La iniciativa promueve intervenir en pro de una igualdad de género transformadora. Al abogar por la inclusión de una perspectiva de igualdad de género en el desarrollo y la gobernanza de la IA se fija el objetivo de eliminar las barreras que perpetúan la desigualdad de género. Al recomendar formas de capacitar, habilitar e instar a las mujeres a participar, la iniciativa hace frente a las causas fundamentales de las desventajas sistémicas. También se propone rectificar el déficit democrático habilitando a los defensores de la igualdad de género para que identifiquen oportunidades y desarrollen aptitudes con fines de emancipación enfocándose en el aumento de la capacidad, la sensibilización y los conocimientos en materia de IA. El propósito es revertir los errores de reconocimiento al abordar los sesgos y promover una narrativa más inclusiva en el desarrollo y el despliegue de sistemas de IA.

Perspectivas para políticas públicas

■ Pasar de los “comportamientos personales” al cambio de sistema

Haciendo hincapié en los imperativos de la igualdad de género, el informe recomienda que la industria de la IA y las personas responsables de políticas públicas dejen de centrarse en las personas a título individual y que se enfoquen en cambiar el sistema. Esto les permitirá adoptar principios, marcos y recomendaciones que transformen la perspectiva de género.

■ Orientar sobre la forma de poner en práctica los principios, los marcos y las recomendaciones que transformen la perspectiva de género

El informe destaca la falta de directrices para la puesta en práctica de principios como la equidad y la transparencia. Las personas responsables de políticas públicas deben basarse en este y en otros recursos para reflejar los principios en acción. También deben formular una orientación y garantizar el desarrollo de ecosistemas eficaces de IA que transformen la perspectiva de género.

Data Justice Policy Brief

Nombre de la iniciativa:

[Data Justice Policy Brief](#)

País/Región:

Global

Organización:

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

Descripción

El objetivo del documento *Data Justice Policy Brief* [Resumen de políticas hacia la justicia en la gestión de datos] es incorporar en las políticas y en las prácticas de los sistemas de IA los principios de justicia en el manejo de datos. Este informe proporciona un marco integral para que las personas responsables de políticas públicas garanticen que el desarrollo y el despliegue de la IA sean justos, transparentes e inclusivos. El documento subraya la importancia de abordar los retos sociales y económicos relacionados con la proliferación de sistemas y procesos de IA. Ofrece, asimismo, recomendaciones prácticas para crear políticas que mitiguen los daños, promuevan la rendición de cuentas y garanticen que los beneficios de la IA se distribuyan de manera equitativa entre todos los sectores de la sociedad. El objetivo final es apoyar un entorno de IA más justo y ético.

Cómo fomenta dicha iniciativa la igualdad sustantiva

La iniciativa aborda la desventaja sistémica promoviendo políticas sobre IA justas y transparentes que se basen en los principios de inclusión, equidad, justicia económica y reparación de daños. Subsana de ese modo el déficit democrático al abogar por procesos inclusivos en la formulación de políticas que comprendan una diversidad de partes interesadas, mejorando así la representación y la rendición de cuentas. Además, al reconocer y garantizar que las comunidades marginadas desempeñen un papel por designio y al promover la transparencia y la rendición de cuentas en las prácticas y sistemas de datos e IA, la iniciativa salvaguarda el derecho a la dignidad, revirtiendo los errores de reconocimiento.

Perspectivas para políticas públicas

■ **Participación democrática de las comunidades afectadas, por designio**

La aplicación de justicia en materia de datos exige que las personas responsables de políticas públicas identifiquen a todas las partes interesadas que puedan verse afectadas por la recopilación y el uso de datos y por las actividades basadas en ellos. Las partes interesadas esenciales son las titulares de los datos —de forma individual y colectiva—, así como las generadoras primarias de los datos. Garantizar que su participación se integre democráticamente en el diseño, desarrollo y despliegue de sistemas de uso intensivo de datos, incluidos los sistemas y procesos de IA.

■ **Fomentar la transparencia en la gestión de datos y en las prácticas y sistemas de IA**

El informe recomienda que las personas que dirijan la recopilación y el uso de datos e innovaciones basadas en datos estén obligadas a hacer pública la información sobre los datos que recopilan y sobre la forma en que se utilizan, incluida la información sobre los datos de entrada y los algoritmos de IA. También recomienda que faciliten esta información directamente a las personas y comunidades afectadas.

AI & Equality Human Rights Toolbox

Nombre de la iniciativa:

[AI & Equality Human Rights Toolbox](#)

País/Región:

Global

Organización:

[Women at the Table](#) y [École Polytechnique Fédérale de Lausanne \(EPFL\)](#), en consulta con [OHCHR](#)

Descripción

AI & Equality Toolbox [Herramienta de Derechos Humanos en IA e Igualdad] es una plataforma educativa diseñada como catalizador para generar debates fundamentados y colaboración para crear IA con una perspectiva basada en los derechos humanos. La plataforma ofrece una metodología, talleres y cursos en línea con una serie de recursos seleccionados y organizados que se dirigen a personas tecnológas, practicantes de IA, personas expertas en ciencias sociales, activistas, personas responsables de políticas públicas y el público en general. La iniciativa se enfoca en los conceptos de contexto y propósito, así como en la colaboración multidisciplinaria y la participación comunitaria, para introducir la equidad y la inclusión en las innovaciones de la IA. Es una iniciativa de la organización *Women at the Table* en colaboración con la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH) y con la EPFL. Su curso en línea está disponible en el portal de aprendizaje del Centro de la Sorbona para la IA (SCAI, por sus siglas en inglés) en colaboración con la Universidad de la Sorbona.

Cómo fomenta dicha iniciativa la igualdad sustantiva

Al promover un enfoque en la IA basado en los derechos humanos, la plataforma aborda la desventaja sistémica garantizando la inclusión de los grupos marginados en el desarrollo de la IA, reduciendo los sesgos y promoviendo la equidad. Fomenta debates fundamentados y colaboración entre personas tecnológas y responsables de políticas públicas, así como con el público en general, garantizando la inclusión de las comunidades marginadas en la adopción de políticas relacionadas con la IA. Así corrige el déficit democrático. Finalmente, por medio de la educación y la colaboración interdisciplinaria, la plataforma desafía el mito de la neutralidad de la IA, y crea un entorno en el que las personas tecnológas y responsables de políticas públicas y las personas afectadas por los sistemas de IA comparten un vocabulario común que reconoce e integra una variedad de perspectivas sociales.

Perspectivas para políticas públicas

■ **La adopción de un enfoque basado en los derechos humanos**

Adoptar un enfoque basado en los derechos humanos a lo largo del ciclo de vida de la IA garantiza que los sistemas de IA se desarrollen, se desplieguen y se auditen con apego a los derechos humanos, fomentando debates fundamentados y un vocabulario compartido sobre el uso ético de los datos. Con los derechos humanos como punto de referencia se subraya la obligación que tienen las instituciones desarrolladoras de IA de cumplir sus obligaciones de transparencia y explicabilidad, y de garantizar el “derecho a saber” de las personas afectadas por los sistemas de IA.

■ **Una plataforma para la colaboración multidisciplinaria y la participación comunitaria**

La iniciativa *AI & Equality Human Rights Toolbox* tiene como objetivo crear una comunidad global de personas dedicadas a la investigación, a las ciencias sociales y al procesamiento de datos, así como a activistas en materia de IA que colaboren e intercambien ideas sobre la forma en que los sistemas de IA pueden servir al bien público y prevenir daños. La iniciativa fomenta y promueve la participación de la comunidad y el desarrollo de capacidades mediante una plataforma con recursos tales como cursos en línea y oportunidades para el debate. Al promover espacios de colaboración multidisciplinaria y de participación comunitaria basados en los derechos humanos, las personas responsables de políticas públicas pueden contribuir a aumentar los conocimientos, la capacidad y la sensibilización sobre la IA.



Indigenous Pathfinders in AI

Nombre de la iniciativa:

[Indigenous Pathfinders in AI](#)

País/Región:

Canadá

Organización:

[Mila - Quebec AI Institute](#) y [Indspire](#)

Descripción

Indigenous *Pathfinders* in AI [Personas Indígenas Pioneras en IA] es un programa que brinda una trayectoria profesional transformadora para instar al talento indígena a aprender, desarrollar y encabezar la evolución de la IA. Arraigado en las cosmovisiones comunitarias e indígenas, el programa tiende puentes entre las perspectivas indígenas y los sistemas y procesos de la IA. Dirigido por la organización Mila - Québec AI Institute en colaboración con Indspire, el programa instruye a las personas participantes sobre aspectos técnicos y no técnicos de la IA, incluyendo temas como la IA responsable y la gobernanza ética de los datos. Las personas participantes reciben estipendios y tienen cubiertos los gastos de viaje, alojamiento y alimentación, lo que les permite participar plenamente en el programa sin restricciones económicas. La iniciativa *Pathfinders* está creando nuevas oportunidades para que las personas indígenas innovadoras moldeen el futuro de la IA e impulsen soluciones eficaces y centradas en la comunidad.

Cómo fomenta dicha iniciativa la igualdad sustantiva

Las barreras sistémicas, incluidos los problemas financieros, el acceso limitado a las oportunidades de capacitación y la infraestructura digital inadecuada, restringen tanto el desarrollo de aptitudes como el conocimiento de las oportunidades profesionales que existen en el sector de la IA. La iniciativa *Indigenous Pathfinders* in AI hace presenta la igualdad sustantiva como una iniciativa específica que desmantela las barreras que se oponen a la innovación democrática y la participación significativa en la gobernanza de la IA. El programa pretende fomentar la inclusión anticipada y significativa en los ecosistemas de IA potenciando el talento y las perspectivas indígenas para configurar los sistemas y procesos de IA desde su concepción, y abordando las exclusiones epistémicas. Al abordar estas brechas de manera proactiva, el programa garantiza un futuro más equitativo y diverso. Este enfoque aborda las desventajas sistémicas arraigadas en las desigualdades históricas y fortalece la capacidad de participación significativa, corrigiendo el déficit democrático.

Perspectivas para políticas públicas

■ **Soluciones sobre la IA impulsadas por la comunidad**

Además de crear vías profesionales que promuevan los conocimientos sobre IA entre personas indígenas con talento, el programa *Pathfinders* involucra activamente a las voces indígenas en el desarrollo de soluciones de IA adaptadas a las necesidades propias de sus comunidades. *Pathfinders* establece una nueva norma para integrar las perspectivas indígenas en todo el proceso de desarrollo de la IA, desde su diseño hasta la gobernanza ética de los datos. Esta iniciativa introduce un enfoque transformador de la educación en IA, al tiempo que habilita a las comunidades indígenas para configurar el futuro de la IA de forma que refleje sus valores y prioridades.

■ **Un enfoque holístico que da prioridad al conocimiento indígena**

El entorno de aprendizaje del programa *Pathfinders* está moldeado por perspectivas indígenas, y su núcleo se basa en las cuatro R de la sabiduría indígena: responsabilidad, respeto, reciprocidad y relevancia. Holística por naturaleza, la organización *Pathfinders* pone énfasis en el bienestar de las personas participantes, brindando acceso a la iniciativa *Elders and knowledge keepers* [Personas adultas mayores y guardianes del saber], así como sesiones dedicadas a enseñanzas culturales. Al habilitar al talento indígena para abordar los desafíos que está en una posición ideal para entender, el programa fomenta el desarrollo de soluciones sobre IA que abordan cuestiones sociales importantes para sus comunidades y definidas por ellas.



Prácticas de diseño tecnológico inclusivo y de innovaciones democráticas

Metodología *Design from the Margins (DFM)*

Nombre de la iniciativa:

[Design From the Margins \(DFM\) Methodology](#)

Organización:

[The De|Center](#)

País/Región:

Global

Descripción

Design from the Margins (DFM) [Diseño desde la marginación] es una metodología desarrollada para crear tecnología equitativa centrándose en las experiencias que experimentan los grupos marginados o descentrados durante todo el ciclo de vida de la IA. La metodología DFM reside en The De|Center, un centro interdisciplinario dedicado a la investigación centrada en la comunidad, la rendición de cuentas empresarial y el diseño tecnológico para reducir los daños que ocasionan las tecnologías existentes y emergentes. La metodología incorpora la investigación centrada en la comunidad, la reducción de daños y un profundo compromiso con los grupos marginados (usuarios descentrados), las personas que corren más riesgo de sufrir los efectos dañinos de la tecnología. Las mejoras introducidas en plataformas como Grindr, Signal y WhatsApp son ejemplos del éxito de este enfoque. La iniciativa se financia con subvenciones y donaciones.

Cómo fomenta dicha iniciativa la igualdad sustantiva

DFM se centra en la reducción de daños y pretende transformar las herramientas tecnológicas para evitar que se utilicen para perpetuar la opresión sistémica. Así aborda las desventajas sistémicas. La metodología implica una profunda investigación centrada en la comunidad, labores de colaboración e intervenciones que integran los principios de seguridad, dignidad y derechos de los grupos marginados. DFM corrige el déficit democrático y ayuda a garantizar que los avances tecnológicos estén al servicio de todas las personas, especialmente de las más afectadas por la injusticia. La iniciativa reconoce que la tecnología produce mejores resultados para todas las personas cuando se diseña para las más afectadas por las injusticias sistémicas. El enfoque también revierte los errores de reconocimiento ya que las voces marginadas se amplifican y se concentran gracias a la metodología y al trabajo relacionado.

Perspectivas para políticas públicas

■ Creación o apoyo de un fondo de ayuda mutua

Por medio de un fondo de ayuda mutua, los grupos marginados pueden acceder a un apoyo financiero sin condiciones en forma de ayuda en efectivo en especie. Se anima a las empresas que se benefician de la metodología DFM a contribuir con fondos. Con la creación o el apoyo que brinden a fondos similares, las personas responsables de políticas públicas pueden habilitar a los grupos marginados para que puedan dirigir el desarrollo que buscan.

■ Convertir daños y carencias en cambios viables

La metodología DFM hace posibles las intervenciones tecnológicas al transformar los daños y las injusticias que se han documentado en cambios viables en la tecnología relacionada con la IA y en su práctica. Esto incluye también otras preocupaciones expresadas por los grupos marginados. Las personas responsables de políticas públicas también pueden apoyar prácticas similares que logren cambios concretos que sean producto de las experiencias que han vivido las personas más afectadas.



La Red de Investigación Feminista en Inteligencia Artificial (f<A+i>r)

Nombre de la iniciativa:

La red f<A+i>r

Organización:

<A+> Alliance for Inclusive Algorithms

País/Región:

Global

Descripción

La Feminist AI Research Network —conocida como f<A+i>r (Red de Investigación Feminista en IA)— se propone crear tecnologías inclusivas en el ámbito de la IA abordando los sesgos sistémicos, raciales e interseccionales. El Directorio Global, al que sólo se puede acceder por invitación, incluye a aproximadamente 100 personas representantes del ámbito académico, a activistas y a profesionales feministas de la IA de diferentes campos dedicadas a investigaciones multidisciplinarias e innovaciones feministas en IA para promover un cambio transformador. La red f<A+i>r fomenta el intercambio de conocimientos Sur-Sur-Norte-Sur y destaca las innovaciones feministas a nivel mundial. La red f<A+i>r fue creada en 2020 por la iniciativa <A+> Alliance for Inclusive Algorithms (<A+> Alianza para Algoritmos Inclusivos), con el apoyo del Centro Internacional de Investigaciones para el Desarrollo de Canadá (IDRC, por sus siglas en inglés), la Fundación Oak y las divisiones de Derechos Humanos y Derechos de las Mujeres del Departamento Federal de Asuntos Exteriores de Suiza (FDFA, por sus siglas en inglés).

Cómo fomenta dicha iniciativa la igualdad sustantiva

La red f<A+i>r fomenta el objetivo de abordar las desventajas sistémicas al combatir activamente los sesgos sistémicos, raciales e interseccionales en el ámbito de las tecnologías de IA. Se propone corregir los daños futuros mediante el uso de datos, algoritmos, modelos, políticas y sistemas nuevos que puedan ser investigados y sometidos a pruebas para lograr un cambio transformador. Además, corrige el déficit democrático al fomentar la inclusión y apoyar el desarrollo de capacidades de equipos feministas de innovación del conjunto de la mayoría global. La iniciativa repara la falta de reconocimiento produciendo modelos interdisciplinarios innovadores y efectivos que aprovechan las tecnologías emergentes para corregir los sesgos y barreras de la vida real que afectan los derechos, la representación y la igualdad de las mujeres y de otros grupos marginados.

Perspectivas para políticas públicas

■ La promoción de sistemas de IA transformativos

Al financiar y apoyar de otras formas el desarrollo y la puesta en práctica de sistemas y procesos de IA que integren activamente los principios tecno-feministas y corrijan los sesgos sistémicos tanto de género como los interseccionales (creando nuevos conjuntos de datos, algoritmos y modelos), las personas responsables de políticas públicas refuerzan la igualdad de género y la diversidad en la IA.

■ Colaboración e intercambio de conocimientos

La red de tecnólogas feministas de la mayoría global contribuye al desarrollo de soluciones de IA específicas para cada contexto que aborden las necesidades y desafíos regionales. Esto conduce a políticas de IA más inclusivas y equitativas a nivel global.



Iniciativa *Fixing the bAIs*

Nombre de la iniciativa:

[Fixing the bAIs: Using AI to Correct Gender Bias in AI](#)

Organización:

MullenLowe MENA for [Aurora 50](#)

País/Región:

Emiratos Árabes Unidos / MENA

Descripción

La campaña *Fixing the bAIs* pretende abordar y mitigar los sesgos de género en los sistemas de IA mediante el entrenamiento de los conjuntos de datos de esta y usando también IA. Las mujeres están enormemente subrepresentadas en los conjuntos de datos utilizados para entrenar la IA. La campaña corrige los sesgos de representación en los bancos comunes de datos de entrenamiento utilizando herramientas de IA tales como Midjourney, Dall-e y Stable Diffusion para la generación de imágenes, y valiéndose de comandos de texto que tienen en cuenta el género. Las imágenes están exentas de regalías y de derechos de autor a fin de fomentar su difusión, lo que servirá para conformar diversos conjuntos de datos que se utilizarán para entrenar la IA. También fomenta la colaboración y el diálogo entre todas las partes para promover las mejores prácticas y soluciones innovadoras para la reducción de sesgos en la IA.

Cómo fomenta dicha iniciativa la igualdad sustantiva

La campaña *Fixing the bAIs* aborda las desventajas sistémicas esforzándose por reducir la discriminación y promover un trato justo entre diferentes perfiles demográficos mediante la corrección de sesgos en sistemas de IA. Corrige el déficit democrático al mejorar la transparencia y la rendición de cuentas en el desarrollo de IA, garantizando así la participación de las comunidades marginadas. La campaña también revierte los errores de reconocimiento al diseñar intencionalmente sistemas de IA inclusivos en cuanto al género: proporcionando conjuntos de datos de entrenamiento con diversidad de género y de licencia abierta, desafiando los roles de género dominantes y garantizando que estén mejor representados los grupos marginados y que sus contribuciones sean reconocidas.

Perspectivas para políticas públicas

■ **La obligación de disponer de conjuntos de datos con diversidad y entrenamiento con perspectiva de género**

Las personas responsables de políticas públicas pueden reducir los sesgos en los sistemas de IA ordenando el uso de conjuntos de datos con diversidad y procesos de entrenamiento con enfoque de género. Al incorporar prácticas de uso de datos inclusivos que tengan en cuenta la diversidad y al garantizar que los conjuntos de datos de entrenamiento representen varios perfiles demográficos, los sistemas de IA pueden reflejar mejor y atender las necesidades de todos los grupos sociales.

■ **Requisitos de divulgación**

Las personas responsables de políticas públicas pueden garantizar la mejora continua de los sistemas de IA y la confianza pública en ellos adoptando reglamentos que obliguen a las organizaciones a divulgar sus esfuerzos y avances en la mitigación de los sesgos de dichos sistemas. Con auditorías periódicas y la publicación de informes sobre los sesgos en los sistemas de IA aunadas a medidas correctivas se promueven la transparencia y la rendición de cuentas en el desarrollo de la IA.



Indigenous Jobs Map

Nombre de la iniciativa:
[Indigenous Jobs Map](#)

País/Región:
Australia

Organización:
[Australia National Research Agency](#)
[\(Commonwealth Scientific and Industrial Research Organisation: CSIRO\)](#)

Descripción

La iniciativa *Indigenous Jobs Map* [Mapa de Empleos para Comunidades Indígenas] preparado por la organización CSIRO, proporciona una visión integral de las oportunidades de trabajo en las comunidades indígenas de Australia. Su objetivo es fomentar el empleo y el desarrollo económico en las comunidades indígenas mediante mapas y visualización de los datos laborales actuales. El Mapa de Empleos para Comunidades Indígenas es un análisis de datos creado con la intención de ayudar a la Agencia Nacional de Investigación de Australia a implementar de manera efectiva su Plan de Acción de Reconciliación para mejorar los resultados laborales de la fuerza de trabajo aborígen e isleña del Estrecho de Torres en el mercado laboral nacional. La iniciativa pretende mejorar la transparencia y el acceso a la información sobre empleos, permitiendo una mejor toma de decisiones y planificación tanto para quienes buscan empleo como para quienes lo proporcionan.

Cómo fomenta dicha iniciativa la igualdad sustantiva

El Mapa de Empleos para Comunidades Indígenas mejora el acceso a las oportunidades de trabajo para las comunidades indígenas ilustrando con mapas los datos laborales actuales, promoviendo el desarrollo económico y reduciendo las disparidades en el empleo. De ese modo aborda las desigualdades sistémicas. Al proporcionar información sobre empleos de manera transparente y accesible, la iniciativa garantiza que se incluyan las voces y necesidades indígenas en los procesos de toma de decisiones, fomentando de ese modo una mayor inclusión en la planificación y en las políticas. Así corrige el déficit democrático. El mapa resalta las contribuciones y el potencial de las comunidades indígenas, contrarrestando estereotipos y garantizando que sus habilidades y sus oportunidades sean reconocidas y valoradas, revirtiendo los errores de reconocimiento.

Perspectivas para políticas públicas

■ **Acción afirmativa centrada en datos para crear estrategias de empleo**

El Mapa de Empleos para Comunidades Indígenas utiliza datos detallados sobre puestos de trabajo para crear estrategias de empleo adaptadas para las comunidades indígenas, ayudando a las personas responsables de políticas públicas a abordar las disparidades laborales y a promover el crecimiento económico. Esta iniciativa destaca la importancia de un enfoque de igualdad sustantiva en la IA mostrando la forma en que los análisis de datos pueden ayudar a las personas responsables de políticas públicas a adoptar las acciones afirmativas pertinentes.

■ **Mejoras en los aspectos de transparencia y acceso**

La iniciativa ofrece una plataforma transparente y accesible para visualizar las oportunidades de empleo. Las personas responsables de políticas públicas pueden utilizar recursos similares para garantizar a las comunidades indígenas el fácil acceso a la información sobre puestos de trabajo con fundamentos sólidos para la toma de decisiones y la planificación. Con este diseño tecnológico inclusivo se ayuda a cerrar la brecha existente en la información y se insta a las personas indígenas a participar en el mercado laboral.



Medidas aplicables a la rendición de cuentas

The Global Index on Responsible AI

Nombre de la iniciativa:

[The Global Index on Responsible AI](#)

Organización:

[Global Center on AI Governance \(GCG\)](#)

País/Región:

Global

Descripción

El *Global Index on Responsible AI* [Índice Global de IA Responsable] es una iniciativa cuyo propósito es evaluar y promover prácticas responsables de IA a escala mundial. Proporciona una evaluación integral de las políticas y prácticas sobre IA de los países, centrándose en la inclusión, la rendición de cuentas y la ética. El índice constituye un recurso multidimensional y un punto de referencia —que comprende datos representativos a escala mundial— con el cual las personas responsables de políticas públicas y quienes se dedican a la investigación y el periodismo pueden seguir y medir el progreso de las medidas que los países estén adoptando para garantizar el disfrute y la incorporación de los derechos humanos en los sistemas de IA. El índice lo dirige el GCG con el apoyo del Centro Internacional de Investigaciones para el Desarrollo de Canadá, el Gobierno de Canadá y la organización USAID.

Cómo fomenta dicha iniciativa la igualdad sustantiva

El índice insta a los países a adoptar políticas que reduzcan los sesgos y la discriminación en los sistemas de IA, garantizando que los grupos marginados se beneficien de las tecnologías de IA y abordando la desventaja sistémica. También mide los compromisos y las prácticas adoptadas en cada país para garantizar que la IA, en lugar de perjudicar, promueva la igualdad de género y el empoderamiento y participación de las mujeres. El índice se elaboró en consulta con las partes afectadas a fin de rectificar el déficit democrático y se basa en los principios de accesibilidad y apertura, inclusión y participación, y en la ambición de reflejar de forma justa los contextos y realidades locales. Al revertir los errores de reconocimiento, el índice pone de relieve las mejores prácticas que garantizan una representación justa de todos los grupos sociales en el desarrollo y despliegue de la IA, combatiendo los estereotipos y fomentando la inclusión.

Una IA Responsable

El Índice Global de IA Responsable define la “Inteligencia Responsable” como el diseño, desarrollo, despliegue y gobernanza de la IA de manera que respete y proteja todos los derechos humanos y que salvaguarde los principios de ética en sistemas de IA en cada etapa del ciclo de vida y cadena de valor de la IA. Exige que todas las partes interesadas que intervengan en el ecosistema de IA asuman la responsabilidad por los efectos humanos, sociales y ambientales de sus decisiones. El diseño, despliegue y gobernanza responsables de la IA deben ser proporcionales al propósito de su uso y responder a las necesidades tecnológicas de las personas y las sociedades a las que pretende servir.

(Adaptado del documento *Global Index on Responsible AI*, 2024)

Perspectivas para políticas públicas

■ Puntos de referencia para la reglamentación

La iniciativa proporciona puntos de referencia que tienen en cuenta las responsabilidades de las partes interesadas a lo largo de todo el ciclo de vida y ecosistema de la IA. Las personas responsables de políticas públicas pueden utilizarlos para promulgar reglamentos que exijan transparencia en las operaciones de IA y mecanismos claros de rendición de cuentas para garantizar el uso de los sistemas de IA de una manera responsable que no perpetúe ningún sesgo.

■ Cooperación internacional mediante supervisión colaborativa

La iniciativa promueve la cooperación internacional mediante supervisión colaborativa. Las personas responsables de políticas públicas pueden asimismo fortalecer y apoyar el rol de las comunidades mundiales en la supervisión colaborativa y en la rendición de cuentas de las partes interesadas en la IA. La cooperación internacional en materia de IA responsable es un compromiso compartido entre países de todo el mundo.



Un primer plano de manos entrelazadas; una de ellas lleva un anillo. Fotografía: Nathan Dumlao en Unsplash.



The Migration and Technology Monitor

Nombre de la iniciativa:

[The Migration and Technology Monitor](#)

Organización:

[Refugee Law Lab](#)

País/Región:

Global

Descripción

El *Migration and Technology Monitor* o MTM [Monitor de Migración y Tecnología] es una agrupación de personas periodistas, cineastas, académicas y de comunidades internacionales que trabajan para mostrar cómo las tecnologías fronterizas en evolución afectan a las personas migrantes, basándose en las experiencias y perspectivas de estas. Es una comunidad construida por y para personas en movimiento. Hace un seguimiento de las tecnologías de vigilancia, de la automatización y del uso de la IA para detectar, rastrear y tomar decisiones sobre las personas en movimiento. El MTM apoya a las comunidades de personas migrantes auspiciando una confraternidad para relatar sus experiencias sobre la forma en que las ha afectado la vigilancia fronteriza. La iniciativa pretende difundir sus conocimientos y promover el trabajo participativo conectando análisis rigurosos con las políticas y la promoción de derechos para hacer partícipes al público, a los medios de comunicación y a las personas responsables de políticas públicas.

Cómo fomenta dicha iniciativa la igualdad sustantiva

El Monitor aborda las desventajas sistémicas al incluir a las personas en movimiento —tales como las personas refugiadas y las personas desplazadas— en el diseño, el desarrollo y el despliegue de sistemas de IA. La iniciativa mitiga los daños sistémicos poniendo en primer plano las experiencias vividas y destacando las prácticas discriminatorias que tienen lugar en las fronteras y en entornos humanitarios. Al hacer partícipes en las conversaciones políticas sobre tecnología a las personas migrantes —a menudo excluidas de los procesos democráticos— el Monitor aborda el déficit democrático y consagra el concepto “Nada sobre nosotras sin nosotras”, que indica que no deberá formularse ninguna política pública sin la participación significativa de las comunidades y las personas afectadas (originalmente “Nothing about us without us” en inglés, que se ha traducido convencionalmente como “Nada sobre nosotros sin nosotros” y que podría entenderse también como “Nada sobre nuestra situación sin nuestra intervención”)¹³. Además, la iniciativa revierte los errores de reconocimiento haciendo hincapié en las experiencias de las personas migrantes, garantizando una representación justa en los debates políticos y desafiando las narrativas excluyentes.

Perspectivas para políticas públicas

■ **Informes y supervisión transparentes**

Documentar las prácticas discriminatorias y los efectos dañinos que experimentan las personas en movimiento fomenta la rendición de cuentas. Las personas responsables de políticas públicas pueden promover la rendición de cuentas exigiendo informes públicos detallados y una supervisión independiente de las herramientas de IA utilizadas en fronteras y en contextos humanitarios para garantizar que respeten las normas de los derechos humanos.

■ **Supervisión comunitaria**

Apoyar la supervisión comunitaria de los sistemas de inteligencia artificial que afectan a las personas migrantes garantiza que las personas afectadas tengan un rol directo en la evaluación y la rendición de cuentas de las tecnologías que afectan su vida. Las personas responsables de políticas públicas pueden mejorar la rendición de cuentas y la capacidad de respuesta en el despliegue de la IA estableciendo mecanismos para efectuar auditorías dirigidas por la comunidad y ciclos de retroalimentación.



The Algorithmic Justice League

Nombre de la iniciativa:

[The Algorithmic Justice League](#)

Organización:

[The Algorithmic Justice League](#)

País/Región:

Estados Unidos / América del Norte

Descripción

La iniciativa *Algorithmic Justice League* (AJL) [Liga para la Justicia Algorítmica] combina arte, investigación y promoción de derechos para la sensibilización sobre los daños y los sesgos de la IA a fin de desarrollar la capacidad de mitigar sus efectos. Mediante campañas de sensibilización pública, programas educativos y desarrollo de capacidades, la iniciativa fortalece a las personas más afectadas por los sistemas de IA y habilita a las personas responsables de políticas públicas, a las dedicadas a la investigación y a las profesionales del sector para garantizar que los sistemas de IA se desarrollen y se desplieguen de forma que proteja los derechos humanos y promueva la justicia social. Una de sus actividades más importantes es la de enfrentar la injusticia interseccional del sexismo y el racismo en el ecosistema específico de las tecnologías de reconocimiento facial. La AJL obtiene financiamiento de la Ford Foundation y la McArthur Foundation, así como de contribuciones individuales.

Cómo fomenta dicha iniciativa la igualdad sustantiva

Al abogar por una IA justa y equitativa, la AJL se propone reducir la discriminación sistémica contra las comunidades marginadas. Aborda las desventajas sistémicas y revierte los errores de reconocimiento descubriendo y mitigando los prejuicios en las tecnologías de IA, garantizando que estos sistemas no perpetúen las desigualdades existentes. A base de auditorías y campañas para la promoción de derechos que visibilizan el racismo, el sexismo, el capacitismo y otras formas de discriminación, esta iniciativa pone de manifiesto las preocupaciones esenciales sobre las violaciones del derecho a la autonomía y a la dignidad. Promueve la transparencia y la rendición de cuentas en el desarrollo de la IA, fomentando la participación y el diálogo públicos. Esto ayuda a democratizar la elaboración de políticas de IA garantizando que se incluyan diversas voces en los debates sobre el rol de la tecnología en la sociedad.

Perspectivas para políticas públicas

■ **Sensibilización pública a través del arte y el relato de experiencias**

Además de participar en charlas y producir material educativo, la AJL genera sensibilización pública sobre los efectos de los sistemas y procesos de IA valiéndose del arte y el relato de experiencias, amplificando los relatos personales de la gente cuyas vidas han sido directamente afectadas por algoritmos injustos. Apoyando y facilitando iniciativas dedicadas a la sensibilización pública sobre los efectos dañinos de los sistemas y procesos de inteligencia artificial, las personas responsables de políticas públicas pueden contribuir a lograr procesos de gobernanza de la IA más efectivos e inclusivos.

■ **Promoción de derechos mediante la documentación colectiva**

La AJL invita, a toda persona que desee hacerlo, a enviar y compartir las experiencias que haya vivido con sistemas y procesos de IA tales como los programas de reconocimiento facial en los puntos de control de la Administración de Seguridad en el Transporte (TSA, por sus siglas en inglés). Estos relatos permiten comprender mejor las experiencias de los grupos marginados frente a cuestiones bien documentadas tales como la discriminación racial, las preocupaciones sobre la privacidad o las consecuencias de negarse a participar. Las personas responsables de políticas públicas pueden proponer o apoyar iniciativas que inviten a las personas a compartir sus experiencias que hayan tenido a fin de documentar los daños que hayan sufrido para comprenderlos mejor y para abrir las posibilidades de hacer justicia.



Iniciativas de inclusión y participación significativas

Māori Data Governance Model

Nombre de la iniciativa:

[Māori Data Governance Model](#)

País/Región:

Nueva Zelanda

Organización:

[Te Kāhui Raraunga Charitable Trust](#)

Descripción

La iniciativa *Māori Data Governance Model* [Modelo de Gobernanza de Datos Maoríes] de Te Kāhui Raraunga se centra en garantizar que los datos maoríes se rijan por principios y valores maoríes. Los datos maoríes representan una *taonga* (posesión preciada) que requiere modelos de protección y cuidado culturalmente fundamentados. El modelo aboga por la soberanía de los datos, haciendo hincapié en que sean las comunidades maoríes las que ejerzan el control, la propiedad y la aplicación de los datos. El modelo ofrece orientación para la gobernanza de los datos maoríes en todo el sistema, en consonancia con las responsabilidades del gobierno conforme al *te Tiriti o Waitangi* (el Tratado de Waitangi). La iniciativa es el resultado de un proceso de codiseño entre personas maoríes en posición de liderazgo, personas representantes de organizaciones maoríes con intereses en los datos, personas maoríes expertas en datos individuales y agencias de la Corona.

Cómo fomenta dicha iniciativa la igualdad sustantiva

La iniciativa promueve la soberanía de los datos, garantizando que sean las comunidades maoríes las que controlen sus datos y los beneficios derivados. Al alinear las prácticas en manejo de datos con los principios maoríes apoya el desarrollo y reduce las desventajas sistémicas arraigadas en desigualdades históricas. En el proceso de codiseño intervienen personas en posición de liderazgo, organizaciones y personas expertas, todas ellas maoríes, lo que garantiza que las voces maoríes formen parte integral de las decisiones sobre gobernanza de datos. Este enfoque inclusivo refuerza la participación democrática y se apega a las obligaciones del Tratado de Waitangi. Al centrar y amplificar las voces maoríes, la iniciativa garantiza que se reconozcan y se respeten las perspectivas y los derechos maoríes en la gobernanza de datos, corrigiendo la falta de representación histórica, así como la exclusión o injusticia epistémica.

Perspectivas para políticas públicas

- **Proceso de codiseño basado en tratados fundados en el concepto “Nada sobre nosotras sin nosotras” con el que se indica que no deberá formularse ninguna política pública sin la participación significativa de las comunidades y las personas afectadas** [originalmente “Nothing about us without us” en inglés, que se ha traducido convencionalmente como “Nada sobre nosotros sin nosotros” y que podría entenderse también como “Nada sobre nuestra situación sin nuestra intervención”].

La iniciativa hace hincapié en un proceso de codiseño basado en tratados, en el que son las comunidades maoríes las que definen los resultados y en quienes reside la autoridad para tomar las decisiones. Al reconocer el derecho de las comunidades maoríes a administrar sus plataformas de datos compartidos como parte integrante de su derecho a la autodeterminación y como principio estructuralmente habilitador, la iniciativa garantiza su participación e inclusión significativas en la gobernanza de datos.

- **Amplificación de las voces indígenas**

Habilitar a las comunidades indígenas para que tengan más voz en la administración de sus recursos de datos colectivos es esencial para prevenir injusticias por falta de representación y para evitar la apropiación cultural. Al abordar las deficiencias de censos anteriores y dar prioridad a las voces maoríes en el proceso de codiseño, la iniciativa garantiza que la gobernanza de los datos refleje las perspectivas y los derechos de las comunidades maoríes. Esta práctica mejora la representación y fomenta la precisión y el respeto en la administración de los datos maoríes.



The RIADIS Workshop

Nombre de la iniciativa:

[RIADIS Workshop to Raise Awareness on Artificial Intelligence \(AI\) and Assistive Technologies among Persons with Disabilities in Latin America](#) (en español)

País/Región:

América Latina

Organización:

[RIADIS](#) con financiamiento de la Christian Blind Mission

Descripción

La Red Latinoamericana de Organizaciones No Gubernamentales de Personas con Discapacidad y sus Familias (RIADIS) es una red regional dedicada a promover y defender los derechos humanos y la inclusión social de las personas con discapacidad. La red se centra en la promoción de derechos, la capacitación y la colaboración para influir en las políticas y crear conciencia sobre los derechos y las necesidades de las personas con discapacidad. El taller se organizó para presentar y compartir buenas prácticas, acciones e investigaciones en IA, así como tecnologías de apoyo para personas con discapacidad. En el taller se contó con herramientas como la de conversión de voz a texto y se debatió sobre los riesgos que representan los sistemas de IA para las personas con discapacidad en el mercado laboral si no se las consulta a lo largo del ciclo de vida de la IA¹⁴.

Cómo fomenta dicha iniciativa la igualdad sustantiva

La iniciativa RIADIS demuestra estrategias prácticas para centrar las necesidades y preocupaciones de las personas con discapacidad en la innovación de la IA. Al garantizar la inclusión de sus voces en todas las fases del ciclo de vida de la IA, la RIADIS ayuda a abordar la exclusión sistémica. Los talleres específicos proporcionan a las personas con discapacidad los conocimientos y recursos necesarios para participar en los debates y en la toma de decisiones, instándolas a abogar por la inclusividad en el diseño y a exigir rendición de cuentas a las instituciones desarrolladoras de IA, rectificando así el déficit democrático. Al concentrarse en las necesidades y las voces de las personas con discapacidad, la RIADIS se propone revertir los errores de reconocimiento que estas han experimentado en la historia. La iniciativa promueve su visibilidad y garantiza que se valoren y se reconozcan sus contribuciones.

Perspectivas para políticas públicas

■ Empoderamiento por medio de la educación

Los talleres específicos que empoderan e instan a las personas con discapacidad a participar en los procesos de toma de decisiones relacionados con la IA permiten a quienes participan comprender la problemática, abogar por sus necesidades y exigir diseños inclusivos a las instituciones desarrolladoras de IA. Las personas responsables de políticas públicas pueden reproducir este enfoque financiando y apoyando programas educativos que se propongan desarrollar las habilidades y la confianza entre los grupos marginados para participar de manera significativa en los debates sobre políticas y en los procesos de diseño tecnológico.

■ Adopción de marcos éticos específicos para la discapacidad

La adopción de marcos éticos específicos para la discapacidad en el diseño y el desarrollo de la IA garantiza que se tengan en cuenta las necesidades y los derechos específicos de las personas con discapacidad a lo largo de todo el ciclo de vida de la IA. Las personas responsables de políticas públicas deben exigir la integración de estos marcos en los procesos de desarrollo de la IA para promover las consideraciones éticas que aborden específicamente los temas que atañen a la discapacidad¹⁵.

The Gender and Responsible Artificial Intelligence Network (GRAIN)

[Red Género e Inteligencia Artificial Responsable]

Nombre de la iniciativa:

[The Gender and Responsible Artificial Intelligence Network \(GRAIN\)](#)

País/Región:

Senegal, Nigeria, Uganda

Organización:

[Initiative Prospective Agricole et Rurale \(IPAR\)](#), [Centre for the Study of the Economy in Africa \(CSEA\)](#) y [Sunbird AI](#)

Descripción

La Red en pro de una IA Responsable y con Inclusión de Género (GRAIN, por sus siglas en inglés) es una iniciativa que se centra en integrar la igualdad de género y las prácticas responsables en el ámbito de la IA en el África Subsahariana. Pone en contacto a organizaciones, universidades y centros de operaciones (*hubs*) para colaborar en materia de IA e inclusión de género. Su objetivo es aprender, compartir, abogar por sus derechos y proponer soluciones para vincular el desarrollo de la IA con la igualdad de género, apoyando el desarrollo inclusivo en la región. La red es un consorcio de grupos de reflexión y organizaciones no gubernamentales y forma parte del programa Inteligencia Artificial para el Desarrollo en África (conocido como AI4D Africa). La red opera con el apoyo financiero del Centro Internacional de Investigaciones para el Desarrollo de Canadá y la Agencia Sueca de Cooperación Internacional para el Desarrollo (Sida, por sus siglas en inglés).

Cómo fomenta dicha iniciativa la igualdad sustantiva

La red GRAIN reduce las desventajas sistémicas conectando organizaciones, universidades y centros de operaciones (*hubs*), garantizando que el desarrollo de la IA en el África Subsahariana incluya la diversidad entre sus perspectivas. La iniciativa aborda problemas, tales como la escasa representación de las mujeres y la falta de atención a las lenguas africanas y a las soluciones de desarrollo inclusivas con perspectiva de género. Promueve la colaboración y la toma de decisiones inclusivas y multidisciplinarias, haciendo partícipes a diversas partes interesadas en los debates sobre IA y género. Este enfoque ayuda a corregir el déficit democrático. La red aboga por sistemas de IA que reconozcan y respeten las identidades y las contribuciones de todas las personas, fomentando un entorno tecnológico más inclusivo y que tenga en cuenta el contexto, y revirtiendo los errores de reconocimiento que han experimentado las comunidades marginadas en su historia.

Perspectivas para políticas públicas

■ Desarrollo de políticas inclusivas sobre IA

La red GRAIN subraya la necesidad de políticas de IA que aborden la discriminación interseccional y la representación insuficiente de ciertos grupos para garantizar que en el desarrollo de la IA se incluya la perspectiva de la diversidad, especialmente en relación con las mujeres del África Subsahariana, y la atención a las lenguas africanas. Las personas responsables de políticas públicas deberían adoptar marcos que impongan la obligación de tener en cuenta la interseccionalidad y el contexto local en el desarrollo de la IA para reducir las desventajas sistémicas.

■ Colaboración multidisciplinaria y participativa

La iniciativa GRAIN enfatiza la colaboración inclusiva y multidisciplinaria con la participación de varias partes interesadas en los debates sobre IA y género. Las personas responsables de políticas públicas deben fomentar la participación de las partes afectadas en la formulación de políticas sobre IA para crear sistemas de IA más representativos que tengan en cuenta los contextos pertinentes.

Recomendaciones clave para formular políticas transformadoras de inteligencia artificial

El informe *Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity* (Hacia la igualdad sustantiva en el ámbito de la inteligencia artificial (IA): políticas transformadoras de IA para la igualdad de género y la diversidad) proporciona un marco integral y recomendaciones clave para lograr la igualdad sustantiva intrínseca en los ecosistemas de IA y en la formulación de políticas relacionadas. Las siguientes recomendaciones describen las medidas concretas que las personas responsables de políticas públicas pueden adoptar para integrar los principios de igualdad de género y diversidad de manera efectiva en los marcos de las políticas, el derecho, la reglamentación y las prácticas relacionadas con la IA.

Las recomendaciones sobre políticas transformadoras de la IA se basan en los resultados de las consultas realizadas tanto a nivel regional como con grupos específicos, las perspectivas para políticas públicas derivadas de prácticas prometedoras, y reflejan las dimensiones clave del marco de políticas transformadoras en el ámbito de la IA. Los desafíos identificados en las consultas —tales como la amalgama entre acceso e inclusión significativa, la exclusión del conocimiento y la distribución inequitativa de los recursos— ponen de manifiesto las profundas desigualdades existentes en el ecosistema de la IA. Estos desafíos exigen esfuerzos concertados en la formulación de políticas que 1) aborden de manera eficaz las barreras estructurales que impiden que los grupos históricamente marginados y excluidos se beneficien con los avances en el desarrollo de la IA, y 2) se propongan habilitar a estos grupos marginados para que gocen plenamente de sus derechos. Las recomendaciones para formular políticas transformadoras de IA se agrupan en las cuatro categorías siguientes:

- **Diseños inclusivos e innovaciones democráticas**
- **Participación significativa en la gobernanza de la IA**
- **Transparencia y rendición de cuentas para prevenir daños**
- **Acceso efectivo a la justicia**

Con estas categorías, las principales recomendaciones para formular una política transformadora en el ámbito de la IA adoptan un enfoque holístico hacia la igualdad sustantiva. En primer lugar, el **diseño inclusivo y la innovación democrática** son fundamentales para abordar las desventajas sistémicas a las que se enfrentan las mujeres y otros grupos marginados y salvaguardar el derecho a la inclusión. Las personas responsables de políticas públicas pueden contribuir activamente a ello mediante la acción afirmativa: invirtiendo en el desarrollo de capacidades para la inclusión institucional, permitiendo el procesamiento de categorías especiales de datos, así como financiando enfoques de investigación y diseño de tecnologías de carácter transformador. En segundo lugar, garantizar una **participación significativa en la**

gobernanza de la IA corrige el déficit democrático y salvaguarda el derecho a la participación. Esto incluye promover el compromiso público efectivo y la participación comunitaria, invertir en el desarrollo de capacidades entre los grupos marginados, legislar sobre los derechos de participación pública y salvaguardar los derechos colectivos relacionados con los datos y la IA.

En tercer lugar, la **transparencia y la rendición de cuentas son esenciales para prevenir daños**, corregir el déficit democrático y revertir los errores de reconocimiento. Estas categorías permiten a las personas e instituciones defensoras de la igualdad sustantiva y a otras partes interesadas examinar los sistemas y procesos de IA, detectar sesgos y responsabilizar a las partes públicas y privadas proveedoras y responsables del despliegue de los efectos perjudiciales o discriminatorios. Las recomendaciones incluyen garantizar el derecho a la información, mejorar la transparencia algorítmica y establecer una rendición de cuentas clara entre todas las partes interesadas involucradas en los ecosistemas de IA mediante evaluaciones de impacto en los derechos humanos y directrices para la contratación pública.

Por último, **garantizar el acceso efectivo** a la justicia revierte los errores de reconocimiento y salvaguarda el derecho a la dignidad. Al reforzar la responsabilidad contextual, facultar a las instituciones promotoras de la igualdad y aliviar la carga de la prueba, las personas responsables de políticas públicas pueden mejorar el acceso a la justicia de las personas y grupos que sufren discriminación y daños en relación con los sistemas y procesos de IA. Juntas, las categorías abordan todas las dimensiones del marco de políticas transformadoras en el ámbito de la IA para lograr la igualdad real de manera efectiva.

Las recomendaciones deberán contemplarse y aplicarse según el contexto. Para aplicarlas, las personas responsables de políticas públicas deben tener en cuenta las consideraciones contextuales —geográficas, sociales, culturales, históricas, económicas, jurídicas y políticas— que moldean los ecosistemas de IA en diferentes regiones y adaptar dichas recomendaciones como corresponda. Esto también incluye considerar las variaciones en los marcos jurídicos y normativos, las capacidades técnicas, la disponibilidad de recursos y la cooperación de las partes interesadas, así como identificar y abordar los obstáculos para su aplicación. En la *Guía de Medidas para Implementar las Recomendaciones sobre Políticas Transformadoras de IA* encontrará orientaciones adicionales para su aplicación. Al comprometernos a adoptar políticas transformadoras de IA para lograr la igualdad de género y la diversidad podemos avanzar hacia una igualdad sustantiva en los ecosistemas de IA y en la formulación de las políticas relacionadas.

Diseños inclusivos e innovaciones democráticas

El diseño inclusivo y la innovación democrática son fundamentales para abordar las desventajas sistémicas a las que se enfrentan las mujeres y otros grupos marginados y salvaguardar el derecho a la inclusión. La falta de diversidad y representación, de sensibilización y comprensión de experiencias y perspectivas diversas, y de los procesos que permiten una inclusión significativa en todos los ecosistemas de IA representan un desafío crítico para llevar a la práctica una igualdad real. Las personas responsables de políticas públicas pueden contribuir activamente en los procesos de diseño inclusivo e innovación democrática mediante la acción afirmativa invirtiendo en el desarrollo de capacidades para la inclusión institucional, permitiendo el procesamiento de categorías especiales de datos, así como financiando enfoques de investigación y diseño de tecnologías de carácter transformador.

1. Involucrar a grupos históricamente marginados en roles técnicos y no técnicos en todo lo concerniente al ecosistema de IA

Implementar acciones afirmativas en todo el ecosistema de IA para involucrar a las mujeres y a otros grupos históricamente marginados en roles técnicos y no técnicos en todo lo concerniente al ecosistema de IA, para incrementar así la diversidad de perspectivas. Asignar recursos para identificar y eliminar las barreras que enfrenta la representación de la diversidad. Esto incluye garantizar una educación accesible e inclusiva más allá de los ecosistemas de IA. Los equipos que incluyen diversidad y los enfoques interdisciplinarios aportan perspectivas más amplias al desarrollo de la IA, haciendo que los sistemas sean más inclusivos, equitativos y transformadores (Buhl, 2023). Para garantizar aún más el cumplimiento de los estándares en igualdad de género y diversidad, se debe exigir que los procesos de diseño de IA obtengan la aprobación de los consejos consultivos sobre ética y los programas de certificación.



Mujer nativa maorí comentando asuntos de negocios con otra mujer ante una computadora tipo tableta en un lugar de trabajo en Nueva Zelandia. Fotografía: comers74 en iStock.

2. Fortalecer las capacidades hacia una inclusión institucional

Invertir —dentro de las instituciones y organizaciones públicas y privadas— en el desarrollo de capacidades y la sensibilización sobre las experiencias y los derechos de los grupos históricamente marginados, como las personas que enfrentan barreras por motivos de sexo, orientación sexual, identidad o expresión de género, personas racializadas, personas con discapacidad o pueblos indígenas (Bartoletti y Xenidis, 2023). Mantener un diálogo constante con las personas representantes de los grupos marginados para comprender y eliminar las barreras específicas a las que se enfrentan. El desarrollo de capacidades y la sensibilización para la inclusión institucional fortalecerán la capacidad de las instituciones para ser más inclusivas con los grupos marginados, tanto en su papel de empleadoras como en su función de partes proveedoras de servicios. Formular y revisar periódicamente las directrices para la infraestructura de comunicaciones y para las comunicaciones escritas y verbales a fin de garantizar que no representen barreras de accesibilidad, que no discriminen por el uso de determinado lenguaje o determinadas expresiones, o para evitar que no representen la diversidad de la sociedad. Esto incluye la realización de evaluaciones de accesibilidad y el establecimiento de listas para verificar la accesibilidad y la inclusión para todos los foros y las actividades en línea. Estas medidas contribuyen a una inclusión significativa al promover prácticas inclusivas y una mayor comprensión dentro de las instituciones sobre la forma de admitir experiencias y perspectivas diversas. Cada persona dedicada a la investigación o que forme parte del personal debe comprender lo que deba cambiar específicamente en sus prácticas, protocolos y directrices para contribuir a alcanzar el objetivo de la institución de ser más inclusiva y respetuosa de la diversidad.

3. Permitir el procesamiento de categorías especiales de datos

Permitir el procesamiento de categorías especiales de datos en determinadas circunstancias excepcionales, con fundamento en un interés público significativo, para lograr la igualdad y la no discriminación. Esto debe realizarse sin contravenir los derechos de protección de datos personales, de forma similar a lo dispuesto en el considerando 44c del Reglamento de Inteligencia Artificial de la Unión Europea. A fin de prevenir obtener información discriminatoria de salida, las partes proveedoras de sistemas de IA deben verificar la ausencia de sesgos sistémicos y garantizar que los conjuntos de datos sean representativos de la diversidad. Esto puede requerir el procesamiento de datos personales de carácter delicado para evaluar la forma en que atributos protegidos como la raza y el género podrían corresponderse con variables indirectas en el modelo y perpetuar la discriminación (Deck *et al*, 2024; Dwork *et al*, 2012). Este procesamiento puede restringirse dependiendo de la legislación sobre protección de datos. Evaluar el interés público sustancial para lograr la igualdad y la no discriminación respetando los derechos de protección de datos personales es esencial para proteger y promover los derechos humanos. Los marcos sólidos de protección de datos están distribuidos de forma dispar en el mundo, pero constituyen una necesidad para prevenir la discriminación y la explotación en el ámbito de la IA. Apoyar la investigación para ampliar las perspectivas colectivas sobre las plataformas de datos compartidos y los enfoques de interés público para conformar la aplicación de esta recomendación (Cofone, 2023).

4. Financiar enfoques de investigación y diseño de tecnología transformadora en las innovaciones de la IA.

Financiar investigaciones y otorgar subvenciones y reconocimiento público para incentivar la aplicación de enfoques de diseño tecnológico inclusivo y transformador en la IA, tales como los que se basan en principios de diseño tecnología feminista. Estos enfoques abordan las brechas existentes entre la justicia de carácter técnico y la justicia política. Apoyar innovaciones en sistemas de IA que se alineen con estos principios promueve aplicaciones, prácticas y procesos más equitativos y justos.



Una mujer australiana y dos mujeres más jóvenes miran la bandera aborigen. Townsville, Australia, durante la semana NAIDOC. Fotografía: Ann Smith en iStock.

Fomentar innovaciones en IA con base en diseños tecnológicos feministas comporta:

Integrar conocimientos consagrados que reflejen la diversidad de identidades

Respetar la autonomía y la dignidad de todas las personas, especialmente entre las comunidades marginadas, como un principio fundamental de diseño en el desarrollo de la IA.

- Fomentar, entre las comunidades y personas históricamente excluidas, la aplicación del principio “Nada sobre nosotras sin nosotras” —que indica que no deberá formularse ninguna política pública sin la participación significativa de dichas comunidades y personas afectadas— como norma de referencia para lograr la plena participación y la igualdad de oportunidades para sí mismas, por sí mismas y consigo mismas. Reconocer esto como parte del principio de “no hacer daño”.
- Se recomienda codiseñar sistemas de IA con personas que enfrentan barreras por motivos de sexo, orientación sexual, identidad de género y/o expresión de género para garantizar innovaciones que tengan en cuenta el género.

Adoptar parámetros de equidad interseccional en los modelos de IA

Implementar parámetros de equidad interseccional para igualar el rendimiento de los sistemas de IA entre los distintos subgrupos interseccionales. Estos parámetros deberán cumplir lo siguiente:

- Considerar múltiples atributos protegidos.
- Proteger todos los valores de intersección de estos atributos como, por ejemplo, a las mujeres racializadas.
- Seguir protegiendo los valores de atributos individuales como, por ejemplo, a las mujeres.
- Garantizar la protección de los grupos marginados más afectados por la discriminación en la sociedad. Rectificar las diferencias sistemáticas debidas a la opresión estructural, en lugar de codificarlas (Buolamwini, 2023; Buolamwini y Gebru, 2018; Ovalle *et al*, 2023b).

Desarrollar y promover las lenguas y culturas indígenas y minorizadas en el desarrollo de la IA.

- Promover tanto las iniciativas públicas a gran escala como los proyectos más pequeños impulsados por la comunidad para fomentar la participación activa de esta en la creación de conjuntos de datos de IA y esquemas algorítmicos, así como en la formulación de usos concretos. Establecer incentivos e infraestructuras a nivel nacional y subnacional en las lenguas locales¹⁶.



Fila de fichas negras que forman la palabra FEMINISTA sobre una superficie de madera color marrón. Fotografía: Jen Theodore en Unsplash.

Participación significativa en la gobernanza de la IA

Garantizar una participación significativa en la gobernanza de la IA corrige el déficit democrático y salvaguarda el derecho a la participación, lo que genera mejores políticas para todas las personas. A menudo se confunde el acceso con la inclusión, pero el acceso por sí solo no garantiza ni una participación significativa ni la inclusión. Las políticas, la legislación y las normas pueden prevenir y mitigar los daños con mayor eficacia cuando las personas responsables de políticas públicas escuchan y colaboran con las personas más afectadas por los sistemas y procesos de IA, y con quienes tienen experiencia en identificar y abordar sus efectos nocivos. Entre las recomendaciones se incluye promover el compromiso público efectivo y la participación comunitaria, invertir en el desarrollo de capacidades entre los grupos marginados, legislar sobre los derechos de participación pública y salvaguardar los derechos colectivos relacionados con los datos y la IA.

5. Promover una participación pública y comunitaria efectiva

Fomentar la equidad y la inclusión en el diseño, desarrollo, despliegue y gobernanza de los sistemas de IA empleando diversas metodologías de participación pública en los planos nacionales e internacionales. Esto implica incluir a los grupos marginados en las discusiones nacionales sobre la gobernanza de la IA y amplificar la mayoría global en los foros internacionales sobre el mismo tema. Utilizar encuestas deliberativas, jurados comunitarios, asambleas de ciudadanos, conferencias de consenso, minipúblicos deliberativos, deliberación en línea, realización de presupuestos participativos, prácticas iterativas y participativas de investigación y diseño, asambleas comunitarias y mesas redondas comunitarias de referencia (OECD, 2021)¹⁷. Dialogar con las comunidades marginadas y las organizaciones representativas de grupos marginados para comprender y eliminar las barreras específicas a las que se enfrentan (United Nations, 2019, 2021). Asignar partidas presupuestarias para cubrir los costos de participación de las personas representantes de grupos marginados —tales como la remuneración por su tiempo y sus conocimientos especializados— y los costos de los ajustes razonables adicionales a los que se recurra para garantizar la accesibilidad (por ejemplo: contratación de intérpretes de lengua de señas o intérpretes de lenguas indígenas). Estos esfuerzos también deben garantizar que los procesos de información y consulta sean accesibles, gratuitos y comprensibles para los grupos marginados, utilizando el lenguaje y los canales y formatos de comunicación adecuados.

6. Invertir en el desarrollo de capacidades entre grupos marginados

Financiar y apoyar programas educativos, estructuras de redes y otros recursos que se propongan desarrollar las habilidades y la confianza entre los grupos marginados para que participen de manera significativa o dirijan activamente los procesos que atienden sus necesidades. Trabajar con comunidades marginadas y las organizaciones que las representen para que organicen sus propias sesiones de sensibilización y sus consultas sobre temas relacionados con la IA, mejorando así su comprensión y su participación. Invertir en sesiones de sensibilización pública sobre IA, derechos individuales, instituciones pertinentes y mecanismos para buscar justicia y reparación de los daños.

7. Legislar sobre los derechos a la participación pública *ex ante*

Basar los procesos de toma de decisiones en IA en los derechos de participación pública *ex ante*, como los establecidos en la Convención de Aarhus de la Comisión Económica de las Naciones Unidas para Europa (CEPE). La Convención de Aarhus garantiza los derechos de la sociedad en materia de acceso a la información, la participación del público en la toma de decisiones y el acceso a la justicia en asuntos ambientales (CEPE, 1998). La aplicación de estos principios en los procesos de toma de decisiones en IA permite que las partes afectadas, así como las organizaciones de la sociedad civil y el público en general, puedan impugnar las consecuencias de las decisiones algorítmicas mediante el razonamiento y la deliberación en el ámbito público. (UNECE, sin fecha). Esta recomendación afirma el papel crucial del público y permite un mayor escrutinio y rendición de cuentas en los procesos de toma de decisiones relacionados con la IA.

8. Proteger los derechos colectivos sobre datos e IA

Revisar los marcos de derechos que se ven afectados por los sistemas y procesos de IA, como los marcos de derechos de propiedad intelectual y de datos, para 1) salvaguardar la soberanía sobre los datos y los conocimientos de las comunidades indígenas y los grupos marginados, incluidas las minorías lingüísticas, religiosas y étnicas; y, 2) garantizar el derecho a beneficiarse del progreso científico. Esta recomendación contribuye a lograr un reparto más equitativo de los beneficios. Garantizar sólidas salvaguardias institucionales para proteger los conjuntos de datos del sector social, especialmente cuando exista el riesgo de apropiación de funciones esenciales para el desarrollo mediante modelos de IA (por ejemplo, en la salud, la educación y el bienestar). Garantizar que los marcos de gobernanza de datos e IA protejan la innovación pública con reformas en las leyes de propiedad intelectual que fomenten ecosistemas no comerciales de IA que contribuyan al bienestar social y a los derechos colectivos. El derecho de todas las personas y comunidades a beneficiarse del progreso científico debe incluir los ecosistemas de la IA. Esto adquiere especial importancia en vista de la marginación de las culturas locales y de los sistemas de conocimientos en los modelos dominantes de datos e IA. Los regímenes de acceso y uso que rigen los conjuntos de datos públicos deben promover los objetivos de democracia, desarrollo y derechos humanos sin erosionar el bienestar en la sociedad y en el entorno (Gurumurthy y Deepti, 2023).

En los párrafos siguientes se presentan vías posibles hacia la protección de los datos colectivos y de los derechos relacionados con la IA:

■ Implementar el acceso condicional a los datos públicos

El acceso a los datos de dominio público y de gobierno abierto debe ser condicional, con limitaciones de acuerdo con la finalidad y cláusulas claras sobre la extinción del acceso. Establecer salvaguardias institucionales sólidas para los conjuntos de datos de los sectores sociales —tales como los de salud, educación y bienestar— a fin de garantizar que los modelos de IA respeten los principios del servicio público y protejan a las comunidades marginadas. (Tomasso Fia, 2021).

■ Imponer limitaciones de uso justo

Imponer imitaciones de uso justo a la forma en que los modelos de IA aprenden con los datos de entrenamiento y la forma en que los utilizan, a fin de evitar la especulación abusiva en el mercado mediante la reutilización.

■ Proponer licencias colectivas

Desarrollar propuestas de licencias colectivas que equilibren los derechos morales de las personas creadoras con los sistemas de valores que consideran el patrimonio común intelectual como patrimonio público.

■ Garantizar la reciprocidad en los fondos comunes de datos

Establecer garantías de reciprocidad en los espacios comunes de datos, garantizando que las instituciones desarrolladoras de modelos privados que utilicen capas de datos públicos estén obligadas a compartir los datos con los que cuenten y a enriquecer los recursos comunes.



Personas de comunidades nativas estadounidenses en el Chumash Day Powwow and Inter-tribal Gathering. Fotografía: Hanna Tor en iStock.

Transparencia y rendición de cuentas para prevenir daños

La transparencia y la rendición de cuentas son esenciales para prevenir daños, corregir el déficit democrático y revertir errores de reconocimiento. Estos conceptos permiten a las partes interesadas examinar los sistemas y procesos de IA, detectar sesgos y responsabilizar a las partes públicas y privadas proveedoras y responsables del despliegue de los efectos perjudiciales o discriminatorios. Capacitan a los grupos marginados para desafiar las prácticas discriminatorias y fomentan la confianza pública garantizando que los procesos de toma de decisiones sean abiertos y comprensibles. Las siguientes recomendaciones garantizan el derecho a la información, mejoran la transparencia algorítmica y establecen una rendición de cuentas clara entre todas las partes interesadas involucradas en los ecosistemas de IA mediante evaluaciones de impacto en los derechos humanos y directrices para la contratación pública.

9. Establecer el derecho a la información en los sistemas de IA y mejorar la transparencia algorítmica

Establecer el derecho a la información en los sistemas de IA. Este derecho debe otorgar a las personas el acceso a detalles claros y accesibles sobre cuándo se emplea la IA, qué algoritmos se utilizan, qué datos se emplean como datos de entrada y qué criterios se aplican en los procesos de toma de decisiones. Esta recomendación refuerza los requisitos de transparencia algorítmica y permite a las personas afectadas negativamente por los sistemas de IA impugnar sus resultados (Grochowski *et al*, 2021; Kossow *et al*, 2021). La información proporcionada debe ser comprensible para el público en general. Para prevenir la explotación de la propiedad intelectual o de los secretos comerciales como medio para evadir la rendición de cuentas, las obligaciones de transparencia deberán prevalecer sobre tales reclamaciones (Fink y Finck, 2022). Exigir una mayor transparencia algorítmica también fomenta las innovaciones tecnológicas para hacer frente a las limitaciones como la opacidad del comportamiento, y mejorar la interpretabilidad y la explicabilidad (Bommasani *et al*, 2024; Winfield *et al*, 2021).

Entre las posibles vías adicionales para mejorar el derecho a la información relacionada con la IA y la transparencia algorítmica se incluyen las siguientes:

- Examinar y considerar la revisión de la legislación sobre secretos comerciales para establecer límites claros y excepciones de interés público para encontrar un equilibrio entre los derechos de propiedad con la transparencia (Kilic, 2024).
- Considerar la posibilidad de establecer una nueva norma de adecuación *post facto* para la transparencia algorítmica. Una norma de este tipo requeriría que las inferencias de los sistemas y procesos de IA en el funcionamiento público ofrecieran información clara y accesible sobre los datos, el modelo y los criterios utilizados en los procesos de toma de decisiones. Esta norma garantiza que las decisiones estén respaldadas por justificaciones registradas, mejorando la explicabilidad y permitiendo auditorías públicas para mitigar los efectos perjudiciales (Mathews y Sinha, 2020).
- Considerar una combinación de explicación del modelo (es decir, el acoplamiento de un sistema de IA opaco con un modelo interpretable y transparente que capte plenamente la lógica del sistema opaco), inspección del modelo (es decir, una representación que permita comprender algunas propiedades específicas de un modelo opaco o sus predicciones) y explicación de los resultados (es decir, una explicación de los resultados de una IA opaca en un caso concreto).
- Considerar la posibilidad de prohibir los modelos de caja negra, especialmente en los casos de alto riesgo y en los procesos públicos de toma de decisiones, y apoyar el uso de modelos interpretables (Rudin y Radin, 2019).

10. Habilitar y realizar evaluaciones obligatorias del impacto en los derechos humanos (EIDH)

Propiciar y conducir estas evaluaciones del impacto con políticas que orienten sobre la manera de implementarlas. Las partes públicas y privadas proveedoras y responsables del despliegue tienen la obligación de respetar los derechos humanos y los principios internacionales, incluida la aplicación de la diligencia debida en materia de derechos humanos y evaluaciones de impacto a lo largo de todo el ciclo de vida de la IA (Pacto Mundial Digital, 2024). Las evaluaciones deberán determinar si los riesgos de daño son aceptables en el marco de los derechos fundamentales e incluir obligaciones claras para eliminar o prevenir dichos riesgos (Jakubowska *et al*, 2024). Las evaluaciones también deberán considerar y comparar posibles enfoques no tecnológicos para identificar las medidas menos intrusivas en los derechos humanos. Si la evaluación revela que existen motivos graves de preocupación, las instituciones responsables del despliegue de los sistemas de IA deberán abstenerse de usarlos.

Entre las posibles vías hacia la obligatoriedad de las EIDH para los sistemas de IA figuran las siguientes:

- Considerar la posibilidad de exigir evaluaciones de impacto *ex ante* exhaustivas que se centren explícitamente en el género y la diversidad.
- Considerar la posibilidad de ordenar que los resultados de dichas evaluaciones de los efectos se presenten a un organismo regulatorio o de supervisión para su estudio y aprobación antes de que pueda desplegarse el sistema de IA en cuestión. Los procesos de estudio y revisión deben apegarse a las normas pertinentes de representación para abordar eficazmente las consideraciones de igualdad sustantiva.
- Considerar la posibilidad de prohibir determinados sistemas y procesos de IA si no se justifican los efectos potenciales o reales conforme a la legislación internacional sobre derechos humanos (McGregor y Molnar, 2023).
- Considerar la posibilidad de establecer multas o sanciones contra las organizaciones que no realicen las evaluaciones de sus efectos exigidas.
- Considerar la posibilidad de establecer la presunción jurídica de que el despliegue de un sistema de IA sin una evaluación exhaustiva de sus efectos constituye una prueba *prima facie* de discriminación.
- Considerar la posibilidad de establecer mecanismos de revisión y reevaluación periódica de los riesgos.

11. Formular medidas de rendición de cuentas aplicables a los sistemas y procesos algorítmicos del sector público

En el ámbito de las contrataciones públicas, desarrollar directrices específicas para la IA con el fin de proteger los derechos humanos y el debido proceso que hagan referencia a las complejidades y los riesgos que representan los sistemas y procesos algorítmicos de IA (Hickok, 2022). Promover iniciativas de datos abiertos para construir bibliotecas abiertas de algoritmos utilizados en los sistemas del sector público. Garantizar que las personas responsables de políticas públicas reciban capacitación para que puedan llevar a cabo eficazmente la debida diligencia en la contratación de sistemas de IA. Considerar la posibilidad de establecer mecanismos de moratoria o la prohibición de los Sistemas de Información Automatizados (SIA) y los sistemas y procesos de IA que puedan afectar los derechos humanos.



Un mujer india de la generación del milenio explica a su colega un proyecto de colaboración; sentadas al escritorio comparten ideas, hablan y colaboran. Fotografía: fizkes en iStock.

Acceso efectivo a la justicia

Garantizar el acceso efectivo a la justicia es fundamental para lograr la igualdad sustantiva. Esto exige la rendición de cuentas, revierte los errores de reconocimiento y salvaguarda el derecho a la dignidad. Al reforzar la responsabilidad contextual, facultar a las instituciones promotoras de la igualdad y aliviar la carga de la prueba, las personas responsables de políticas públicas pueden mejorar el acceso a la justicia de las personas y grupos que sufren discriminación y daños en relación con los sistemas y procesos de IA.

12. Fortalecer la responsabilidad contextual para la no discriminación en los sistemas de IA

Fortalecer la responsabilidad contextual por la no discriminación en los sistemas de IA en proporción a otras medidas de rendición de cuentas tales como el grado de transparencia, de interpretabilidad y de explicabilidad. Deberá revisarse la reglamentación sobre responsabilidad por los daños causados por productos defectuosos y sobre responsabilidad subjetiva (por culpa) para reflejar con precisión las complejidades de los sistemas de IA y la toma de decisiones basada en datos. La rendición de cuentas efectiva en el desarrollo y despliegue de la IA tiene en cuenta características específicas como la opacidad, la explicabilidad, el comportamiento autónomo, la adaptación continua y la predictibilidad limitada. Trazar una vía hacia la responsabilidad en materia de IA para garantizar una rendición de cuentas apropiada entre las partes públicas y privadas proveedoras y encargadas de su despliegue.

Entre las vías posibles se incluyen:

- Introducir condiciones de uso y fomentar la conducción de pruebas experimentales (*sandboxes*) de IA en ámbitos inexplorados para que el sector público establezca la normativa. Esto establecerá límites claros para un uso seguro y reducirá los daños, y también posibilitará la innovación y las pruebas iterativas.
- Considerar un marco de responsabilidad mixta para las partes proveedoras y responsables del despliegue de sistemas de IA que equilibre la responsabilidad subjetiva (por culpa) y la responsabilidad objetiva en relación con otras medidas de rendición de cuentas.
- Evaluación de la normativa de responsabilidad centrada en el software (EPRS, 2024).
- Exigir a las partes públicas y privadas proveedoras y responsables del despliegue que eliminen todo daño descubierto e de conflicto discriminatorio derivado de los procesos asociados con desarrollo y el despliegue de los sistemas en determinados usos.
- Indemnizar por los daños y perjuicios identificados materiales e inmateriales sufridos como resultado de discriminación algorítmica.

13. Instar a las instituciones promotoras de la igualdad a iniciar procedimientos

Instar a las instituciones promotoras de la igualdad —incluidas las instituciones nacionales de derechos humanos y otras organizaciones de interés público— a movilizarse en aras del interés público. Autorizar a estas instituciones a presentar demandas ante las autoridades supervisoras incluso sin necesidad de identificar a las partes denunciantes (Equinet, 2023)¹⁸. Aliviar la carga de la prueba y dotar a estos organismos de la autoridad legal y la capacitación necesaria para abordar de manera efectiva la discriminación y los daños que causan los sistemas de IA y los procesos relacionados.

14. Aliviar la carga de la prueba para las personas demandantes

Examinar y revisar las reglas a fin de aliviar la carga de la prueba impuesta a las partes demandantes (World Commission on the Ethics of Scientific Knowledge and Technology, 2005). Las normas existentes sobre responsabilidad por los daños causados por productos defectuosos suelen exigir a las partes perjudicadas que demuestren la relación causal entre los defectos del producto y los daños específicos que ocasione. Dada la complejidad y la escasa predictibilidad de los sistemas de IA, este requisito representa un desafío. Considerar la posibilidad de ajustar estas reglas para que las partes demandantes puedan probar sus casos y exigir una indemnización con mayor facilidad.

Aligeramiento de la carga de la prueba en los casos de discriminación y de daños y perjuicios en el ámbito de la IA: Conclusiones de la propuesta de Directiva de la UE sobre Responsabilidad por Daños y Perjuicios en el ámbito de la IA

■ **Introducción de una “presunción de causalidad”**

En la propuesta de Directiva de la UE sobre responsabilidad en materia de IA se introduce una presunción de causalidad que presupone un nexo causal entre el incumplimiento de un deber de diligencia derivado del derecho de la UE o nacional (la culpa) y la información de salida producida por el sistema de IA o la no producción de una información de salida del sistema de IA que haya provocado los daños en cuestión.

Este principio equilibra los intereses de las partes demandantes y demandadas en los litigios sobre responsabilidad por los daños causados por productos defectuosos sin aumentar los riesgos de responsabilidad que podrían obstaculizar la innovación y la adopción de productos y servicios basados en la IA. Corresponde a las instituciones proveedoras, operadoras y usuarias de sistemas de IA demostrar la validez y la imparcialidad de sus sistemas reduciendo, a las partes perjudicadas, la carga de probar la conexión entre las operaciones del sistema de IA y los resultados perjudiciales (European Commission, 2022; European Parliament, 2023).

■ **Concesión a los tribunales nacionales de la facultad para ordenar la divulgación de pruebas en relación con sistemas de IA de alto riesgo sospechosos de causar daños**

Las personas perjudicadas por sistemas de IA enfrentan dificultades para cumplir los requisitos relacionados con la reclamación de responsabilidad debido a la complejidad de estos sistemas. Conceder a los tribunales nacionales la facultad para ordenar la divulgación de pruebas en relación con sistemas de IA de alto riesgo sospechosos de causar daños puede mejorar la rendición de cuentas (World Commission on the Ethics of Scientific Knowledge and Technology, 2005). Esto se apega al principio de precaución, a condición de que los tribunales sigan los principios de necesidad y proporcionalidad y de que eviten las solicitudes genéricas.



Siluetas de una persona con el puño en alto, con un vibrante atardecer en sepia graduado de fondo. Fotografía: Miguel Bruna en Unsplash.

Conclusiones y exploración de vías futuras

Las recomendaciones clave para lograr una política de IA transformadora que se presentan en este informe ofrecen una vía integral para lograr la igualdad sustantiva intrínseca en los ecosistemas de IA y en la formulación de políticas relacionadas. Categorizadas para reflejar las dimensiones clave del marco de políticas transformadoras en el ámbito de la IA, las recomendaciones se centran en cuatro categorías: diseño inclusivo e innovación democrática, participación significativa en la gobernanza de la IA, transparencia y rendición de cuentas para la prevención de daños y acceso efectivo a la justicia. En conjunto, estos pilares crean un fundamento sólido para los ecosistemas de IA y para la formulación de las políticas correspondientes para progresar hacia la igualdad sustantiva.

La aplicación efectiva de estas recomendaciones requerirá inversiones intencionadas, asignaciones presupuestarias, desarrollo de capacidades, colaboración de las partes interesadas e innovaciones normativas para sortear los obstáculos y las complejidades que comporta el rápido desarrollo de la IA. Y lo que es más importante, requerirá un compromiso permanente para enfrentar las barreras tecnológicas, jurídicas, financieras o de otro tipo, y para elevar y empoderar activamente a las voces típicamente excluidas para lograr un cambio transformador en beneficio de todas las personas.

La exploración de vías futuras, la inversión continua en investigación interdisciplinaria, intersectorial y multisectorial, y la colaboración contribuirán a lograr sistemas y procesos de IA más inclusivos, equitativos y transformadores. Avanzar en nuestra comprensión del modo en que los sistemas y procesos de IA repercuten en otros marcos de derechos, ensanchar las perspectivas colectivas sobre, por ejemplo, los derechos a la privacidad, el derecho a la libertad de expresión, la educación y la salud, o los derechos medioambientales y laborales, así como continuar el estudio de las plataformas de datos compartidos y los enfoques de interés público también conformarán las futuras aplicaciones de las políticas transformadoras. Las pruebas iterativas continuas y la innovación en entornos normativos seguros permitirán incrementar el aprendizaje y el desarrollo de marcos para la formulación de políticas eficaces. Además, será necesaria la colaboración nacional e internacional entre sectores y regiones para adaptar las políticas a los diversos contextos locales, garantizando que los sistemas y la gobernanza de la IA no solo sean equitativos en el plano global, sino también sensibles a las variaciones regionales en los entornos jurídicos, culturales y económicos. Invertir en la investigación para seguir evaluando el impacto de las políticas, las leyes y los reglamentos en el entorno en rápida evolución de la IA permitirá a las personas responsables de políticas públicas seguir mejorando y refinando las políticas y las prácticas transformadoras.



Imagen detallada en blanco y negro de manos entrelazadas, destacando una pulsera en una y un anillo en la otra.
Fotografía: Aarón Blanco Tejedor en Unsplash.

En última instancia, si se aplican con el compromiso de tener en cuenta enfoques sensibles al contexto y una innovación continua, estas recomendaciones nos conducirán a un futuro en el que los beneficios de los sistemas y procesos de la IA se compartirán equitativamente, garantizando un mundo más igualitario, diverso e inclusivo para todas las personas.

Notas de pie de página

- ¹ En este informe empleamos el término SOGIESC como término inclusivo de toda orientación sexual, identidad de género, expresión de género y características sexuales, incluidos los rasgos intersexuales. Algunas personas también utilizan SOGI (orientación sexual, identidad de género) o SOGIE (orientación sexual, identidad de género y expresión de género). El informe reconoce que el acrónimo se refiere a toda la humanidad con sus orientaciones sexuales e identidades de género, incluidas las personas cisgénero y heterosexuales. Sin embargo, el término LGBTQI+ hace hincapié específicamente en personas concretas con identidades marginadas (por ejemplo, transexuales, no binarias, lesbianas, etc.). Más información en Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (2024a) y en Smith (2023).
- ² Entre las personas en movimiento se incluye a las refugiadas, las solicitantes de asilo, las migrantes y las personas desplazadas internas (IDP, por sus siglas en inglés).
- ³ Véase, por ejemplo, la *Women for Ethical AI Action Request* (Solicitud de Acción de Mujeres por una IA Ética), en la que se pide a todos los países que se sumen a las actividades de la UNESCO para apoyar la misión de impulsar acciones sistemáticas hacia una IA que incluya la perspectiva de género (UNESCO, 2024b).
- ⁴ La equidad reconoce que cada persona vive circunstancias diferentes, y asigna las oportunidades y recursos exactos correspondientes para alcanzar un resultado igualitario. Véase International Women's Day (2023).
- ⁵ El informe adopta un enfoque no binario de la igualdad de género y hace referencia a las barreras a las que se enfrentan no solo las mujeres, sino también las personas con diversidad en su orientación sexual, identidad y expresión de género y características sexuales (SOGIESC).
- ⁶ Representación proveniente de diez países de África (Burkina Faso, Camerún, Kenya, Mauricio, Nigeria, Sudáfrica, Suazilandia, Túnez, Uganda y Zimbabue); diez países de América Latina (Argentina, Bolivia, Brasil, Chile, Colombia, Ecuador, México, Paraguay, Perú y Uruguay); diez países de América del Norte y Europa (Alemania, Bélgica, Canadá, Dinamarca, Estados Unidos, Francia, Hungría, Portugal, Reino Unido y Rumanía); nueve países de la región Oriente Medio y Norte de África (Egipto, Irán, Israel, Líbano, Mauritania, Marruecos, Arabia Saudita, Túnez y Emiratos Árabes Unidos); y once países de la región Asia-Pacífico (Australia, India, Japón, Malasia, Nepal, Nueva Zelandia, Filipinas, Singapur, Corea del Sur, Tailandia y Vietnam).
- ⁷ La equidad reconoce que cada persona vive circunstancias diferentes, y asigna las oportunidades y recursos exactos correspondientes para alcanzar un resultado igualitario.
- ⁸ El informe adopta un enfoque no binario de la igualdad de género y hace referencia a las barreras a las que se enfrentan no solo las mujeres, sino también las mujeres y personas con diversidad SOGIESC.
- ⁹ Este principio está consagrado en diversos instrumentos internacionales de derechos humanos, tales como la Declaración Universal de Derechos Humanos y la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW) (Office of the High Commissioner for Human Rights, 1979).
- ¹⁰ Encontrará ejemplos en CEDAW (UN Women, 2009), Art. 2(3) y 7, y en Office of the High Commissioner for Human Rights, 2006.
- ¹¹ Las descripciones de esta sección se basan en la visualización que presenta la OCDE del ciclo de vida de los sistemas de IA y en la descripción que hace la UNESCO del ciclo de vida de la IA y de sus distintas etapas, desde la investigación y su concepción hasta su desmontaje y terminación (UNESCO, 2024c, p. 10). Sin embargo, hace más hincapié en la recopilación y preprocesamiento de datos y en la elaboración de modelos estadísticos (Ovalle et al, 2023b). Las dimensiones clave señaladas en el documento *OECD Framework for the Classification of AI Systems* [Marco de la OCDE para la Clasificación de los Sistemas de IA] sirven de base para conformar la descripción de los sistemas sociales (OECD, 2022).



- ¹² El marco se basa en los de IA ya existentes y va un paso más adelante para afirmar el deber del Estado de eliminar activamente la discriminación y de abordar la injusticia estructural.
- ¹³ El lema “Nada sobre nosotras sin nosotras” se basa en este principio de participación y ha sido utilizado como “Nada sobre nosotros sin nosotros” por las Organizaciones de Personas con Discapacidad a lo largo de los años como parte del movimiento mundial para lograr la plena participación y la igualdad de oportunidades para, por y con las personas con discapacidad. <https://www.un.org/esa/socdev/enable/iddp2004.htm>.
- ¹⁴ Conforme a las recomendaciones emitidas por el Relator Especial sobre los Derechos de las Personas con Discapacidad.
- ¹⁵ Conforme a las recomendaciones emitidas por el Relator Especial sobre los Derechos de las Personas con Discapacidad.
- ¹⁶ Véase, por ejemplo, la iniciativa de España de desarrollar un modelo fundacional de IA entrenada en español y en lenguas cooficiales —en colaboración con instituciones, tales como el Barcelona Supercomputing Center y la Academia Española de la Lengua—, iniciativa que ejemplifica la transparencia y la accesibilidad (Computerworld Spain, 2024).
- ¹⁷ El Principio 1.3 de la OCDE sobre Transparencia y Explicabilidad subraya la necesidad de abordar los problemas de transparencia promoviendo la argumentación pública y estableciendo entidades consagradas, si es necesario, a mejorar la conciencia y la comprensión de los sistemas de IA. Esto fomenta la aceptación de las tecnologías de IA y la confianza en ellas.
- ¹⁸ Facultad jurídica de tomar conocimiento de un asunto sin recibir una denuncia formal.



Referencias

ACM (2022), “Statement on Principles for Responsible Algorithmic Systems”, *ACM Technology Policy Office*, Washington, DC, <https://www.acm.org/binaries/content/assets/public-policy/final-joint-ai-statement-update.pdf>.

Badaloni, S. y A. Rodà (2022), “Gender knowledge and Artificial Intelligence”, documento presentado en el 1st Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming, BEWARE-22, conjuntamente con AIxIA 2022, University of Udine, Udine, Italia.

Bartoletti, I. y R. Xenidis (2023), “Study on the Impact of Artificial Intelligence Systems, their Potential for Promoting Equality, Including Gender Equality, and the Risks They May Cause in Relation to Non-discrimination”, Gender Equality Commission (GEC) and the Steering Committee on Anti-Discrimination, Diversity and Inclusion (CDADI), Consejo de Europa, Estrasburgo, <https://rm.coe.int/prems-112923-gbr-2530-etude-sur-l-impact-de-ai-web-a5-1-2788-3289-7544/1680ac7936>.

Benjamin, R. (ed.) (2019), *Captivating technology: Race, carceral technoscience, and liberatory imagination in everyday life*, Duke University Press, Durham, NC.

Bommasani, R. *et al* (2024), The Foundation Model Transparency Index v1.1, <https://crfm.stanford.edu/fmti/paper.pdf>.

Buckup, S. (2009), *The price of exclusion: The economic consequences of excluding people with disabilities from the world of work*, Employment Working Paper No. 43, ILO, https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40ed_emp/%40ifp_skills/documents/publication/wcms_119305.pdf

Buhl, N. (2023), Human-in-the-Loop Machine Learning (HITL) Explained, Encord Blog, <https://encord.com/blog/human-in-the-loop-ai/>.

Buolamwini, J. (2023), *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*, Random House, Nueva York.

Buolamwini, J. y T. Gebru (2018), “Gender shades: Intersectional accuracy disparities in commercial gender classification”, en *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Association for Computing Machinery, Nueva York, <https://proceedings.mlr.press/v81/buolamwini18a.html>.

Cachat-Rosset, G. y A. Klarsfeld (2023), “Diversity, equity, and inclusion in Artificial Intelligence: An evaluation of guidelines”, *Applied Artificial Intelligence*, vol. 37(1), artículo 2176618, <https://doi.org/10.1080/08839514.2023.2176618>.

Campbell, M. (2015), “CEDAW and women’s intersecting identities: A pioneering new approach to intersectional discrimination”, *Revista Direito GV*, vol. 11(2), págs. 479–504, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2726534.

Chauhan, P. y N. Kshetri (2022), “The role of data and artificial intelligence in driving diversity, equity, and inclusion”, *Computer*, vol. 55, págs. 88–93, <https://ieeexplore.ieee.org/abstract/document/9755238>.

Chen, B.J. y J. Metcalf (2024), *Explainer: A Sociotechnical Approach to AI Policy*, Data & Society Policy Brief, https://datasociety.net/wp-content/uploads/2024/05/DS_Sociotechnical-Approach_to_AI_Policy.pdf.

Cheong, M.C.Y. K. Leins y S. Coghlan (2021), “Computer science communities: Who is speaking, and who is listening to the women? Using an ethics of care to promote diverse voices”, en *FAccT ‘21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Nueva York, <https://api.semanticscholar.org/CorpusID:231639365>.



Cofone, I. (2023), “The Privacy Fallacy: Harm and Power in the Information Economy”, *Cambridge University Press*, Cambridge, Reino Unido, <https://www.cambridge.org/core/books/privacy-fallacy/547578F2A1AE0C40963105CE-066B412E>.

Computerworld Spain (2024), “Spain will create foundational AI model in local languages”, *Computerworld*, <https://www.computerworld.com/article/1612218/spain-will-create-foundational-ai-model-in-local-languages.html>

Costanza-Chock, S. (2018), “Design justice, A.I., and escape from the matrix of domination”, *Journal of Design and Science*, <https://doi.org/10.21428/96c8d426>.

Dankwa-Mullan, I. et al (2021), “A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle”, *Journal of Health Care for the Poor and Underserved*, vol. 32, págs. 300–317, <https://muse.jhu.edu/article/789672>.

Deck, L. et al (2024), *Implications of the AI Act for Non-Discrimination Law and Algorithmic Fairness*, <https://arxiv.org/html/2403.20089v1>.

Dignum, V. (2023). “Responsible artificial intelligence: Recommendations and lessons learned”, en D. O. Eke, K. Wakunuma, y S. Akintoye (eds.), *Responsible AI in Africa: Challenges and Opportunities*, Springer International Publishing, Cham, Switzerland, https://doi.org/10.1007/978-3-031-08215-3_9.

Donnelly, N. and L. Stapleton (2022), The social impact of data processing: The case of gender mapped to sex, *IF-AC-PapersOnLine*, vol. 55(39), págs. 117–122, <https://doi.org/10.1016/j.ifacol.2022.12.021>.

Dumbrava, C. (2021), “Artificial intelligence at EU borders: Overview of applications and key issues, In-Depth Analysis”, *Think Tank, Comunidad Europea*, [https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA\(2021\)690706](https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA(2021)690706).

Dwork, C. et al (2012), “Fairness through awareness”, en *ITCS ‘12: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, <https://doi.org/10.1145/2090236.2090255>.

Equinet (2023), *Ensuring European AI that Protects and Promotes Equality for All*, <https://equineteurope.org/wp-content/uploads/2023/07/Recommendations-amendments.pdf>.

Eubanks, V. (2017), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, Nueva York.

European Commission (2022), *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, Comisión Europea, Bruselas, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0496>.

European Parliament (2023), *Artificial intelligence liability directive*, Briefing, EU Legislation in Progress, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI\(2023\)739342_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf).

Fink, M. y M. Finck (2022), “Reasoned A(I)dmistration: explanation requirements in EU law and the automation of public administration”, *European Law Review*, vol. 47(3), págs. 376–392.

Forster, M. (2022), *Refugee Protection in the Artificial Intelligence Era. A Test Case for Rights*, Chatham House, Londres.

Fredman, S. (2016), “Substantive equality revisited”, *International Journal of Constitutional Law*, vol. 14(3), págs. 712–738, <https://doi.org/10.1093/icon/mow043>.

Fricker, M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford Academic, Oxford.

Goldschmidt, J.E. (2017), “New perspectives on equality: Towards transformative justice through the Disability Convention?”, *Nordic Journal of Human Rights*, vol. 35(1), págs. 1–14, <https://doi.org/10.1080/18918131.2017.1286131>.



Global Center on AI Governance (2024), “Methodology: Our Multidimensional Framework”, *Global Index on Responsible AI*, <https://www.global-index.ai/methodology>

Grobelnik, M. *et al* (2024), “What is AI? Can you make a clear distinction between AI and non-AI systems?”, *The AI Work*, OECD.AI Policy Observatory,, <https://oecd.ai/en/work/definition>.

Grochowski, M. *et al* (2021), “Algorithmic transparency and explainability for EU consumer protection: Unwrapping the regulatory premises”, *Critical Analysis of Law*, vol. 8(1), págs. 43–63, <https://doi.org/10.33137/cal.v8i1.36279>.

Gurumurthy, A. y B. Deepti (2023), “Reframing AI governance through a political economy lens”, *IT for Change*, https://itforchange.net/sites/default/files/add/IT%20for%20Change_Reframing%20AI%20Governance%20through%20a%20Political%20Economy%20Lens.pdf.

Hacker, P. (2024), “Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence: Complementary Impact Assessment”, *European Parliamentary Research Service*, [https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU\(2024\)762861_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf).

Hellman, D. (2023), “Big data and compounding injustice”, *Journal of Moral Philosophy*, Vol. 21(1-2), pp. 62–83, <https://doi.org/10.1163/17455243-20234373>.

Hickok, M. (2024), “Public procurement of artificial intelligence systems: New risks and future proofing”, *AI & Society*, vol. 39, págs. 1213–1227, <https://doi.org/10.1007/s00146-022-01572-2>.

Hull, G. (2023), “Dirty data labeled dirt cheap: Epistemic injustice in machine learning systems”, *Ethics and Information Technology*, vol. 25(3), artículo 38, <https://doi.org/10.1007/s10676-023-09712-y>.

Iason, G. (2022), “Toward a theory of justice for artificial intelligence”, *Daedalus*, vol. 151(2), págs. 218–231, https://doi.org/10.1162/daed_a_01911

International Women’s Day (2023), *Equality versus Equity: What’s the difference as we #EmbraceEquity for IWD 2023 and beyond?*, <https://www.internationalwomensday.com/Missions/18707/Equality-versus-Equity-What-s-the-difference-as-we-EmbraceEquity-for-IWD-2023-and-beyond>.

IWRAW (sin fecha), *Substantive equality*, <https://cedaw.iwraw-ap.org/cedaw/cedaw-principles/cedaw-principles-over-view/substantive-equality/>.

Jakubowska, E. *et al* (2024), “EU’s AI Act fails to set gold standard for human rights”, *Análisis conjunto*, <https://www.amnesty.eu/wp-content/uploads/2024/04/EUs-AI-Act-fails-to-set-gold-standard-for-human-rights.pdf>.

Jora, R.B. *et al* (2022), “Role of Artificial Intelligence (AI) in meeting Diversity, Equality and Inclusion (DEI) goals”, en *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1.

Joyce, K. *et al* (2021), “Toward a sociology of artificial intelligence: A call for research on inequalities and structural change”, *Socius*, vol. 7, artículo 2378023121999581, <https://doi.org/10.1177/2378023121999581>.

Kilic, B. (2024), “Into Uncharted Waters: Trade Secrets Law in the AI Era”, CIGI Papers No. 295, *Centre for International Governance Innovation*, <https://www.cigionline.org/static/documents/no.295.pdf>.

Kong, Y. (2022), “Are ‘intersectionally fair’ AI algorithms really fair to women of color? A philosophical analysis”, en *FACCT ‘22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Nueva York, <https://doi.org/10.1145/3531146.3533114>.

Kossow, N., S. Windwehr y M. Jenkins (2021), “Algorithmic transparency and accountability”, *Transparency International Anti-Corruption Helpdesk Answer*, *Transparency International*, https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf.



- Krishna, V.K. (2020), "Open science and its enemies: Challenges for a sustainable science–society social contract", *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 6(3), artículo 61, <https://doi.org/10.3390/joit-mc6030061>.
- Martineau, W., N. Meer y S. Thompson (2012), "Theory and practice in the politics of recognition and misrecognition", *Res Publica*, vol. 18, págs. 1–9, <https://doi.org/10.1007/s11158-012-9181-7>.
- Mathews, H.V. y A. Sinha (2020), "Use of Algorithmic Techniques for Law Enforcement", *Economic and Political Weekly*, vol. 55(23), <https://www.epw.in/journal/2020/23/special-articles/use-algorithmic-techniques-law-enforcement.html>.
- McGregor, M. y P. Molnar (2023), *Digital Border Governance: A Human Rights Based Approach*, University of Essex y Alto Comisionado de las Naciones Unidas para los Derechos Humanos, <https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf>.
- Molnar, P. (2024), "The deadly digital frontiers at the border", *Time*, 21 de mayo, <https://time.com/6979557/unregulated-border-technology-migration-essay/>.
- Nihei, M. (2022), "Epistemic injustice as a philosophical conception for considering fairness and diversity in human-centered AI principles", *Interdisciplinary Information Sciences*, vol. 28(1), págs. 35–43, <https://doi.org/10.4036/iis.2022.A.01>.
- Noble, S. y S. Roberts (2019), "Technological elites, the meritocracy, and postracial myths in Silicon Valley", en *RACISM POSTRACE*, R. Mukherjee, S. Banet-Weiser y H. Gray (eds.), *Racism Postrace*, Duke University Press, Durham, NC. Reporte #: 6. <https://escholarship.org/uc/item/7z3629nh>.
- OECD (2021), *Open government and citizen participation*, <https://www.oecd.org/en/topics/open-government-and-citizen-participation.html>.
- OECD (2022), "The OECD Framework for the Classification of AI Systems", *OECD.AI Policy Observatory*, https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/02/oecd-framework-for-the-classification-of-ai-systems_336a8b57/cb6d9eca-en.pdf.
- OECD (2024a), "OECD AI Principles overview", *OECD.AI Policy Observatory*, <https://oecd.ai/en/ai-principles>.
- OECD (2024b), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, París, <https://doi.org/10.1787/623da898-en>.
- Office of the High Commissioner for Human Rights (1965), *International Convention on the Elimination of All Forms of Racial Discrimination*, <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-elimination-all-forms-racial>.
- Office of the High Commissioner for Human Rights (1966), *International Covenant on Economic, Social and Cultural Rights*, <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights>.
- Office of the High Commissioner for Human Rights (1979), *Convention on the Elimination of All Forms of Discrimination against Women*, Nueva York, 18 de diciembre de 1979, <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women>.
- Office of the High Commissioner for Human Rights (2023), *Digital Border Governance: A Human Rights Based Approach*, <https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf>.
- Office of the High Commissioner for Human Rights (2024a), *LGBTI people*, https://www.ohchr.org/en/topic/lgbti-people?gad_source=1.



Office of the High Commissioner for Human Rights (2024b), *The Core International Human Rights Instruments and their monitoring bodies*, <https://www.ohchr.org/en/core-international-human-rights-instruments-and-their-monitoring-bodies>.

Ostry, J.D. et al (2018), “Economic gains from gender inclusion: New mechanisms, new evidence”, IMF Staff Discussion Note, IMF, <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2018/10/09/Economic-Gains-From-Gender-Inclusion-New-Mechanisms-New-Evidence-45543>

Ovalle, A. D. Liang y A. Boyd (2023a), “Should they? Mobile biometrics and technopolicy meet queer community considerations”, en *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, artículo 32, págs. 1-10 <https://dl.acm.org/doi/abs/10.1145/3617694.3623255>.

Ovalle, A. et al (2023b), “Factoring the matrix of domination: A critical review and reimagination of intersectionality in AI fairness”, en *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, <https://doi.org/10.1145/3600211.36047>.

Pande, R (2003). “Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India.”, *American Economic Review*, vol. 93 (4), <https://www.aeaweb.org/articles?id=10.1257/000282803769206232>.

#ProtectNotSurveil coalition (2024), “Joint statement – A dangerous precedent: How the EU AI Act fails migrants and people on the move”, *accessnow* <https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/>.

Ricaurte, P. (2024), “A sociotechnical approach to the AI Lifecycle”, *Feminist AI Research Network*, disponible próximamente.

Ricaurte, P. (2019), “Data epistemologies, the coloniality of power, and resistance”, *Television & New Media*, vol. 20(4), págs. 350–365, <https://doi.org/10.1177/1527476419831640>.

Ricaurte, P. (2022). *Ethics for the majority world: AI and the question of violence at scale*. Media, Culture & Society, 44(4), 726-745. <https://doi.org/10.1177/01634437221099612>.

Ricaurte, P. y M. Zasso (2023), “AI, ethics, and coloniality: A feminist critique”, en M. Cebral-Loureda, E.G. Rincón-Flores y G. Sanchez-Ante (eds.), *What AI Can Do*, Chapman and Hall/CRC, Nueva York.

Rikap, C. (2022), “The expansionary strategies of intellectual monopolies: Google and the digitalization of healthcare”, *Economy and Society*, vol. 52(1), págs. 110–136, <https://doi.org/10.1080/03085147.2022.2131271>.

Rudin, C. y J. Radin (2019), “Why are we using black box models in AI when we don’t need to? A lesson from an explainable AI competition”, *Harvard Data Science Review*, vol. 1, <https://doi.org/10.1162/99608f92.5a8a3a3d>.

Shapiro, J. (sin fecha), *CEDAW as a tool for promoting substantive gender equality*, <http://www.cedaw.org.tw/en/upload/media/Capacity%20Building/1-1The%20Mechanism%20of%20CEDAW%20Committee.pdf>.

Smith, R.A. (2023), “From LGBTQIA+ to SOGIESC: Reframing sexuality, gender, and human rights”, *Open Global Rights*, Nueva York.

Smuha, N.A. (2021), “Beyond the individual: governing AI’s societal harm”, *Internet Policy Review*, vol. 10(3), <https://doi.org/10.14763/2021.3.1574>.

Solaiman, I. et al (disponible próximamente), “Evaluating the social impact of generative AI systems in systems and society”, en P. Hacker, A. Engel, S. Hammer y B. Mittelstadt (eds.), *Oxford Handbook on the Foundations and Regulation of Generative AI*, Oxford University Press, Oxford.



Stauffer, B. (2023), *Automated Neglect: How The World Bank's Push to Allocate Cash Assistance Using Algorithms Threatens Rights*, Human Rights Watch, <https://www.hrw.org/report/2023/06/13/automated-neglect/how-world-banks-push-allocate-cash-assistance-using-algorithms>.

Suresh, H. *et al* (2022), "Towards intersectional feminist and participatory ML: A case study in supporting femicide counterdata collection", en *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Nueva York, <https://api.semanticscholar.org/CorpusID:249872586>.

Türk, V. (2023), *Artificial intelligence must be grounded in human rights, says High Commissioner*, Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, Ginebra.

UNCTAD (2021), *Technology and Innovation Report 2021*, <https://unctad.org/page/technology-and-innovation-report-2021>.

UNECE (sin fecha), *Introduction*, <https://unece.org/environment-policy/public-participation/aarhus-convention/introduction>.

UNECE (1998), *UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters*, <https://unece.org/environment-policy/public-participation/aarhus-convention/text>.

UNESCO (2009), *Concept paper on marginalization*, documento presentado en la Décima reunión del Grupo de Trabajo sobre la Educación para Todos (EPT), París.

UNESCO (2022), *Recommendation on the Ethics of Artificial Intelligence*, UNESCO, <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

UNESCO (2024a), *Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes*, <https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes>.

UNESCO (2024b), *Statement W4EAI, Women for Ethical AI Action Request*, SHS/2024/PI/H/1, <https://unesdoc.unesco.org/ark:/48223/pf0000388754/PDF/388754eng.pdf.multi>.

UNESCO (2024c), *Ethics of Artificial Intelligence: The Recommendation*, <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

UNESCO (2024d), *Ethical Impact Assessment*, Global AI Ethics and Governance Observatory, <https://www.unesco.org/ethics-ai/en/eia>.

UNESCO, IDB y OECD (2022), *The Effects of AI on the Working Lives of Women*, UNESCO, París.

UNHRC (2011), *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, A/HRC/17/31, https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

UNHRC (2020), *Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*, A/HRC/44/57, <http://undocs.org/en/A/HRC/44/57>.

UNHRC (2021), *Report of the Special Rapporteur on the Rights of Persons with Disabilities*, A/HRC/49/52, <http://docs.org/en/A/HRC/49/52>.

UNICEF (2021), *UNICEF Gender Policy and Action Plan 2022-2025: Gender-Transformative Programming*, <https://www.unicef.org/lac/en/media/43146/file>.

United Nations (2006), *Convention on the Rights of Persons with Disabilities and Optional Protocol*, <https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>.



United Nations (2019), *United Nations Disability Inclusion Strategy*, Naciones Unidas, Nueva York, https://www.un.org/en/content/disabilitystrategy/assets/documentation/UN_Disability_Inclusion_Strategy_english.pdf.

United Nations (2021), *Report of the Secretary-General's Task Force on Addressing Racism and Promoting Dignity for All in the United Nations Secretariat*, Naciones Unidas, Nueva York, https://hr.un.org/sites/hr.un.org/files/sap_final_report_0.pdf.

United Nations (2024), *The 17 Goals*, <https://sdgs.un.org/goals>.

United Nations (2024b), *Global Digital Compact*, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf.

UNRISD (2017), *Transformative Policies for Sustainable Development: What Does It Take?*, <https://www.unrisd.org/en/library/publications/transformative-policies-for-sustainable-development-what-does-it-take>.

UN Women (2009), *Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)*, <https://www.un.org/womenwatch/daw/cedaw/text/econvention.htm#intro>.

UN Women (2015), "Chapter 1: Substantive equality for women: The challenge for public policy", en *Progress of the World's Women 2015–2016*, UN Women, <https://progress.unwomen.org/en/2015/pdf/ch1.pdf>.

UN Women (2019), *CEDAW-based Legal Review: A Brief Guide*, UN Women in Ukraine, Kiev, https://eca.unwomen.org/sites/default/files/Field%20Office%20ECA/Attachments/Publications/2019/CEDAW%20Legal_ENG_compressed%20%281%29.pdf.

Vinuesa, R. et al (2020), "The role of artificial intelligence in achieving the Sustainable Development Goals", *Nature Communications*, vol. 11(1), artículo 1, <https://doi.org/10.1038/s41467-019-14108-y>.

West, S.M. (2020), "Redistribution and recognition: A feminist critique of algorithmic fairness", *Catalyst: Feminism, Theory, Technoscience*, vol. 6(2), <https://doi.org/10.28968/cfft.v6i2.33043>.

West, S.M., M. Whitaker y K. Crawford (2019), "Discriminating Systems: Gender, Race, and Power in AI - Report", *AI Now Institute*, <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2>.

White House (2022), *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, White House Office of Science and Technology Policy, <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>.

Winfield, A.F.T. et al (2021), "IEEE P7001: A Proposed Standard on Transparency", *Frontiers in Robotics and AI*, vol. 8, artículo 665729, <https://doi.org/10.3389/frobt.2021.665729>.

World Commission on the Ethics of Scientific Knowledge and Technology (2005), *The Precautionary Principle*, UNESCO, Paris, <https://unesdoc.unesco.org/ark:/48223/pf0000139578>.

World Economic Forum (2024), *AI Value Alignment: Guiding Artificial Intelligence Towards Shared Human Goals*, White Paper, World Economic Forum, https://www3.weforum.org/docs/WEF_AI_Value_Alignment_2024.pdf.



Apéndice: Lista de Participantes en Consultas Regionales

Esta es una lista de todas las personas y organizaciones participantes en las consultas regionales que han dado su consentimiento para ser nombradas en este informe.

Data Pop Alliance condujo las consultas regionales correspondientes a las regiones siguientes.

Asia

1. **Apekchya Rana Khatri**,
UNFPA Nepal
2. **Gaurav Sharma**,
GIZ India
3. **Dr. Preeti Raghunath**,
University of Sheffield
4. **Philippe Doneys**,
Asian Institute of Technology
5. **Hazel T. Biana**,
Research Fellow, Social Development
Research Center, De La Salle University
6. **Manila Hoang**,
Vietnam Foreign Ministry
7. **Eleonore Fournier-Tombs**,
United Nations University
8. **Omran Najjar**,
Humanitarian OSM StreetMap Team

Middle East and North Africa (MENA)

1. **Mohammad Reza Bateni**,
Amirkabir University of Technology
2. **Hayfa Khalfaoui**
3. **Maha Jouini**,
African Center for AI and Digital Technology
4. **Sukaina Al-Nasrawi**, United Nations Economic
and Social Commission for Western Asia
(UNESCWA)
5. **Fouad Mrad**

6. **Grace S. Thomson**,
Center for AI and Digital Policy (CAIDP)
7. **Ibrahim Helmy Emara**,
Tanta University
8. **Dr. Houda Chihi**,
Tunisia Telecom
9. **Limor Ziv**,
Humane AI
10. **Prof. Hazem Abdelazim**,
ESLSCA University
11. **Arwa ElGhadban**,
Technical University of Munich
12. **Ihab Shouly**

North America and Europe (NE)

1. **Mariana Roza Paz**,
Datasphere Initiative
2. **Yavar Baradaran-Khosrshohahi**
3. **Till Koebe**,
Saarland Informatics Campus
4. **H. Gijs van den Dool**,
Investigador independiente
5. **Luisa Olaya**,
GIZ (Fair Forward)
6. **Ruth Schmidt**,
GIZ (Fair Forward)



Derechos Digitales condujo la consulta regional correspondiente a la región siguiente.

América Latina y el Caribe (LAC, por sus siglas en inglés)

1. Maria Encalada Córdova
2. Tarcizio Roberto da Silva
3. Luz Elena González
4. Edson Prestes
5. Marianela Milanes
6. Silvana Bahia
7. Gisele da Silva Craveiro
8. Eduardo Carrillo
9. Lutiana Valadares Fernandes Barbosa
10. Zinnya del Villar Islas
11. Cristina Pombo
12. Viviane Ceolin Dallasta Del Grossi
13. Danielle Costa Carrara Couto.
14. Luis Fernando Arias
15. Silvana Fumega
16. Valentina Rozo Ángel
17. Paz Peña
18. Daniela Camila de Araújo
19. Fernanda K. Martins
20. Virginia Pardo
21. Hugo Alvaro Miranda Colque
22. Virginia Brussa
23. Paola Ricaurte Quijano
24. Fernanda Campagnucci
25. Gustavo Macedo
26. Maximiliano Maneiro Vaz
27. Andres Lombana Bermudez
28. Ivana Feldfeber
29. Augusto Luciano Mathurin
30. Iván Terceros
31. Eliana Quiroz
32. María Lorena Flórez Rojas
33. Natali Viezna Figueroa Figueroa
34. Juan de Brigard
35. Juan Pablo Marín Díaz
36. Marcela Mattiuzzo
37. Thiane de Nazaré

Paola Ricaurte Quijano condujo la consulta regional correspondiente a la región siguiente.

América Latina y el Caribe (LAC, por sus siglas en inglés)

1. Gerardo López Gómez
2. José Alfredo Hau Caamal
3. Verónica Aguilar Martínez
4. Jhonnatan Rangel
5. Marcos Cornelio Sánchez Ramírez
6. Miriam Hernández
7. Kupijy Vargas
8. Manuela Yah Panti
9. Yamili Chan Canul
10. Andrés Ta Chikinib
11. Mariana Díaz

ICT Research Africa condujo la consulta regional correspondiente a la región siguiente.

África Subsahariana (SSA, por sus siglas en inglés)

1. Sandra Aceng
2. Adedeji Adeniran
3. Adebisi Adewusi
4. Shariffa Amolo
5. Houda ChiChi
6. Kudakwashe Dandajena
7. Laetitia Delaunay-Badolo
8. Khetsiwe Dlamini
9. Sinclair Ebimowei
10. Alison Gillwald
12. Rigobert Kenmogne

13. Jonas Kgomo
14. Slindile Khumalo
15. Cheruiyot Kiprono
16. Anja Kovacs
17. Michael Magoronga
18. Lorraine Matanda
19. Tholang Mathapo
20. Bhekani Mbuli
21. Katileho Mokoena
22. Margaret Muga
23. Judith Murungi
24. Ernest Mwebaze
25. Angella Ndaka
26. Nelson Nkari
27. Adaeze P. Nwoba
28. Dorcas Nyamwaya
29. Durnebi Okwasah
29. Titilola Helen Olojede
30. Eniola Olowu
31. Lavina Ramkisoon
32. Kelly Stone
33. Oluwatoyin Taiwo
34. Cynthia Tapera
35. Moses Thiga
36. Israel Olatunji Tijani
37. Simone Toussi
38. Josephine Zingani
39. Mpho Moyo
40. Pierrine Leukes