



Hacia un

Índice Global de IA responsable



Promoviendo una agenda global sobre IA responsable

Contexto

Con el rápido **auge de las tecnologías impulsadas por la IA** en los últimos años, **se han logrado avances en el desarrollo de principios, directrices y regulaciones** para regir su desarrollo y uso. Es fundamental que nuestro punto de partida para todos estos esfuerzos sea el respeto de los derechos humanos internacionales, la protección y promoción de la democracia y la inclusión.

Si bien la IA tiene el potencial de usarse para el bien, como avanzar en la realización de los objetivos de desarrollo, cada vez hay más pruebas de que la IA puede usarse para socavar la democracia y afectar negativamente nuestros derechos y libertades.

Para proteger los derechos humanos y promover los valores democráticos, **debemos avanzar en la implementación** de los principios de la ética de la IA en la práctica.



Hacia un Índice Global de IA responsable

META

Fomentar el desarrollo y uso de IA que promueva y respete los derechos humanos, fortalezca la democracia y la inclusión, y responda a las necesidades socioeconómicas locales.

Hacia un Índice Global de IA responsable

Midiendo la evolución del compromiso y el progreso en la implementación de principios y prácticas de IA responsable

- ¿Existen las condiciones para apoyar y promover el uso responsable de la IA?
- ¿Cómo luce el progreso? ¿Podemos monitorearlo a lo largo del tiempo?



Objetivos del Índice en IA Responsable

Establecer puntos de referencia sistematizados basados en los derechos humanos para una IA responsable

Evaluar las condiciones a nivel país para apoyar y promover la IA responsable

Brindar evidencia a los tomadores de decisiones de todo el mundo para impulsar el uso responsable de la IA y promover la inclusión digital

Realizar los principios a través de la implementación, la evaluación y la rendición de cuentas



Construyendo sobre un movimiento global

**El Índice Global sobre IA
Responsable se basará y
complementará las
actividades de actores
clave**

UNESCO Recommendation on the Ethics of AI

UNESCO AI and Gender Equality Report

OECD Principles on Artificial Intelligence (AI Policy Observatory Working Groups) GPAI Working Group on Responsible AI

The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems

WEF agenda for scaling Responsible AI

Freedom Online Coalition Taskforce on AI and Human Rights

United Nations High-Level Agenda Digital Cooperation

G20 AI Human-Centred Principles

Indigenous Protocol and Artificial Intelligence Working Group

Decolonising AI Manifesto

UNESCO Report of COMEST on Robotics Ethics



Algorithmic and Automated decision-making Accountability and Impact Assessments

Data for Development

Responsible AI / ML

Ethics Assessments and Readiness

Inclusive Design and Diversity

Transparency; Explainability; &
Interpretability

Human Rights

Gender

Fairness / Bias

Privacy, Data Ethics and Protection

Security

Environmental Impact and Sustainability

Digital Literacy and Participation

AI Indexes

Global Index on Responsible AI

ACLU-Algorithmic Equity Toolkit | Doteveryone / Ada Lovelace Consequence Scanning | Canadian Algorithmic Impact Assessment | IEEE Ethics Certification Program for Autonomous and Intelligent Systems | Open Government Partnership: Algorithmic Accountability for the Public Sector | Amsterdam Algorithm Register | Artificial Intelligence Incident Database | The Assessment List on Trustworthy AI | Ada Lovelace Institute | The Information Accountability Foundation: AI Impact Assessment | European Parliament: Governance Framework for Algorithmic Accountability and Transparency | Canadian Algorithmic Impact Assessment | AI Now Institute | Alan Turing Institute: Understanding AI Ethics and Safety | Microsoft Responsible AI Toolbox | PWC Responsible AI Framework & Toolkit | TensorFlow: Responsible AI Toolkit | AI4People | Personal Data Protection Commission Singapore: Model AI Governance Framework | Microsoft Interpret ML | Responsible AI Institute Certification | United Nations Interregional Crime and Justice Research Institute (UNICRI) | World Economic Forum: AI Procurement in a Box | Institute for Ethical AI & Machine Learning: AI-RFX Procurement Framework | Artificial Intelligence (AI) Ethics for Nonprofits Toolkit | AI Ethics Impact Group | AI Ethics Lab Toolbox | Machine Intelligence Garage / Digital Catapult: Ethics Framework | Data Ethics: Data Ethics Readiness Test | Design Ethically Framework and Toolkits | Inter-American Development Bank | Ethical AI Advisory | Ethics Kit – Design Better | Omidyar Network Ethical Operating System Toolkit | Ethics & Algorithms Toolkit | The Information Accountability Foundation: Ethical Data Impact Assessments and Oversight Models | World Economic Forum Empowering AI Leadership | AISP Toolkit: Centering Racial Equity Throughout Data Integration | Amazon Conditional Demographic Disparity | Microsoft Inclusive Design | USC Diversity Toolkit | Alan Turing: Project ExplAIIn | European Parliament: Governance Framework for Algorithmic Accountability and Transparency | Human-Centered Artificial Intelligence (HCAI) Framework | Google Model Cards | IBM AI Explainability 360 | IEEE Ethics Certification Program for Autonomous and Intelligent Systems | Lime (Github) | Know Your Data | Data Ethics: Digital Self-Defence | Human Rights, Ethical and Social Impact Assessment | Humanitarian Data Science and Ethics Group: Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector | SIENNA (Stakeholder-Informed Ethics for New Technologies with High Socio-economic and Human Rights Impact) | The Danish Institute for Human Rights | Women at the Table: AI & Equality Human Rights | World Wide Web: Open Questions About Gender Inclusion | Aequitas | Audit AI | Australian Human Rights Commission | Black Box Auditing and Certifying and Removing Disparate Impact | Fairlearn | Google PAIR | IndiaAI: Mitigating Bias in AI | LinkedIn Fairness Toolkit | Microsoft AI Fairness Checklist | TensorFlow | Digital Civil Society Lab: Digital Impact Toolkit | Data Ethics: Digital Self-Defence | Know Your Data | GIZ / Data Security Council of India | United Nations Global Pulse: Building Ethics into Privacy Frameworks for Big Data and AI | Google Tensorflow: Privacy in Machine Learning | Open Data Institute: Data Ethics Canvas | Presido (Github): Anonymizer for Text and Images | United Kingdom: Data Ethics Framework | Adversarial ML Threat Matrix | Adversarial Roustness Toolbox (ART) | Artificial Intelligence Incident Database | Counterfeit | Code Carbon | USAID: Managing ML Projects in International Development | Carbon Emissions and Large Neural Network Training | Experiment-Impact-Tracker | ML CO₂ Impact | Nethope Digital Skills Framework | Microsoft Judgment Call / Community Jury | Berkman Klein Centre: Digital Citizenship+ (Plus) Resource Platform | Stanford AI Index | Government AI Readiness Index | Global Data Barometer | The Global AI Index | AI Global Surveillance Index | ICT Development Index | Global Innovation Index | Global Competitiveness Index | Digital Rights Corporate Accountability Index | Artificial Intelligence and Democratic Values Index



Áreas de mejora

Predominantemente enfocados y desarrollados a partir del Norte Global

Las experiencias de los países en desarrollo no se reflejan adecuadamente

Los derechos humanos no son centrales en muchas herramientas

No se tienen en cuenta las capacidades de los países emergentes para desarrollar y utilizar la IA de manera responsable

Datos primarios limitados, especialmente del Sur Global

No hay un instrumento comparativo global representativo para medir el progreso en la implementación de IA responsable

Elementos centrales del GI

Establecer principios básicos basados en los derechos humanos para medir la IA responsable

Evaluación holística de la medida en que los principios se aplican a través de las dimensiones centrales

Recopilación de datos primarios y secundarios en torno a indicadores clave

Red global de investigadores independientes

Resultados contextualizados: informados a nivel local y regional con foco en el Sur Global

Más de 100 países para informar el análisis comparativo

Características

Accesibilidad y Apertura

El Índice Global debe ser accesible para otras partes interesadas e investigadores que deseen comprender o reutilizar los datos. Una metodología demasiado compleja socavará la transparencia del proyecto y su utilidad para otros colectivos.

Inclusión y Participación

El Índice Global se creará conjuntamente con una amplia gama de actores relevantes para garantizar su valor en el terreno, con énfasis específico en incluir los aportes y las perspectivas de los grupos marginados.

Reflejar adecuadamente los contextos y las realidades locales

El Índice Global debe reflejar fielmente los contextos y las realidades locales, y medir el nivel de IA responsable de un país en relación con los recursos disponibles.

Un enfoque basado en los derechos humanos

Responsabilidad de los estados

Responsabilidad corporativa

Responsabilidad de la sociedad civil

para asegurar las condiciones del país para el uso responsable de la IA

Promover el desarrollo responsable de sistemas de IA respetuosos de los derechos. Supervisar el impacto de la IA en los derechos humanos y proporcionar respuestas para las violaciones de esos derechos

para promover y respetar los derechos humanos en el desarrollo y uso de la IA

Desarrollar y utilizar sistemas de IA de manera que promuevan los derechos humanos y fortalezcan la democracia. Respetar los derechos humanos en todas las cadenas de suministro

para promover y monitorear los derechos humanos

Abogar por el respeto de los derechos humanos en el uso y desarrollo de la IA. Monitorear el uso estatal y corporativo de la IA y promover la conciencia de los derechos.



Incorporando el Enfoque Basado en los DDHH

- Diseño de contenido basado en derechos:
 - ◆ Derechos civiles y políticos
 - ◆ Derechos socioeconómicos
 - ◆ Derechos laborales
 - ◆ Derechos de las generaciones futuras y derechos ambientales
 - ◆ Derechos comunitarios y culturales
- Diseño de indicadores: indicadores de progreso y de inputs además de indicadores de impacto y resultados
- Modelo de evaluación: realización progresiva, sujeta a los recursos disponibles

[‘Designing a rights-based Global Index on Responsible AI’](#)



Derechos humanos y áreas clave de medición

Categorías de derechos

**Igualdad de género y no
discriminación**

Derechos socioeconómicos

**Libertades individuales y
valores democráticos**

Áreas clave

- Esfuerzos contra la discriminación, el acoso y la violencia
 - Brecha digital entre y dentro de las naciones
 - Promoción de mujeres y colectivos subrepresentados en IA y STEM
 - Reducir y mitigar sesgos
-
- Transparencia y participación en las decisiones sobre el uso público de la IA
 - Inclusión en infraestructura y acceso de IA
 - Concienciación y alfabetización
 - Uso de la IA para promover la atención médica y el bienestar
-
- Seguridad y protección de la IA
 - Acceso a la información
 - Toma de decisiones automatizada, autonomía y supervisión humana
 - Privacidad y protección de datos
 - Protección de denunciantes
 - Uso pacífico y proporcionado de la IA
 - Desinformación, Libertad de expresión y discurso de odio



Derechos humanos y áreas clave de medición

Categorías de derechos

Empleo

Sustentabilidad y derechos de generaciones futuras

Sector privado y derechos humanos

Derechos comunitarios y culturales

Áreas clave

- Protección contra la pérdida de empleo y desarrollo de nuevas habilidades
- Normativas de empleo en plataformas

- Derechos de los niños
- Sostenibilidad del medio ambiente

- Responsabilidad Corporativa
- Normativas antimonopolio

- Protección del patrimonio lingüístico y cultural
- Control comunitario sobre datos/uso de IA local

Construyendo la red para la implementación

Hubs regionales

Estamos en el proceso de identificar hubs regionales en regiones de todo el mundo para:

- reclutar y capacitar investigadores del país
- apoyar en consultas regionales
- apoyar en pruebas piloto de instrumentos en la región
- supervisar las respuestas de los investigadores nacionales
- dar seguimiento a las respuestas de la encuesta del gobierno
- realizar análisis regional para el informe del Índice Global

Investigadores nacionales

Hemos estado participando en un ejercicio preliminar para identificar investigadores de países y redes de investigadores, particularmente en el sur global, para comprender áreas potenciales para el desarrollo de capacidades en las áreas temáticas de la IA responsable.



SIGAMOS EN CONTACTO



r^{aii}@d4d.net



@D4Dnetwork
@ResponsibleAiIndex



Próximamente

¡Dejanos tu contacto completando el [formulario!](#)





¡Muchas gracias!