

Ética, sesgo y discriminación en {IA, AA, CD}^{*}

Objetivos

- Identificar potenciales riesgos en el uso de IA.
- Reconocer de dónde partimos, las tendencias y formas de abordaje sobre estos riesgos.
- Explorar los principales enfoques internacionales y nacionales para las regulaciones sobre IA.

Contenido

- Principales preocupaciones éticas en AI
- Principios de la IA responsable
- Desafíos técnicos, de implementación, sociales y económicos
- Medida y mitigación del sesgo algorítmico
- Regulaciones y políticas en el mundo
- ¿Qué estamos haciendo en Uruguay?

**REMEMBER
WITH GREAT
POWER
COMES GREAT
RESPONSIBILITY**



**UNCLE
BEN**

Principales preocupaciones éticas

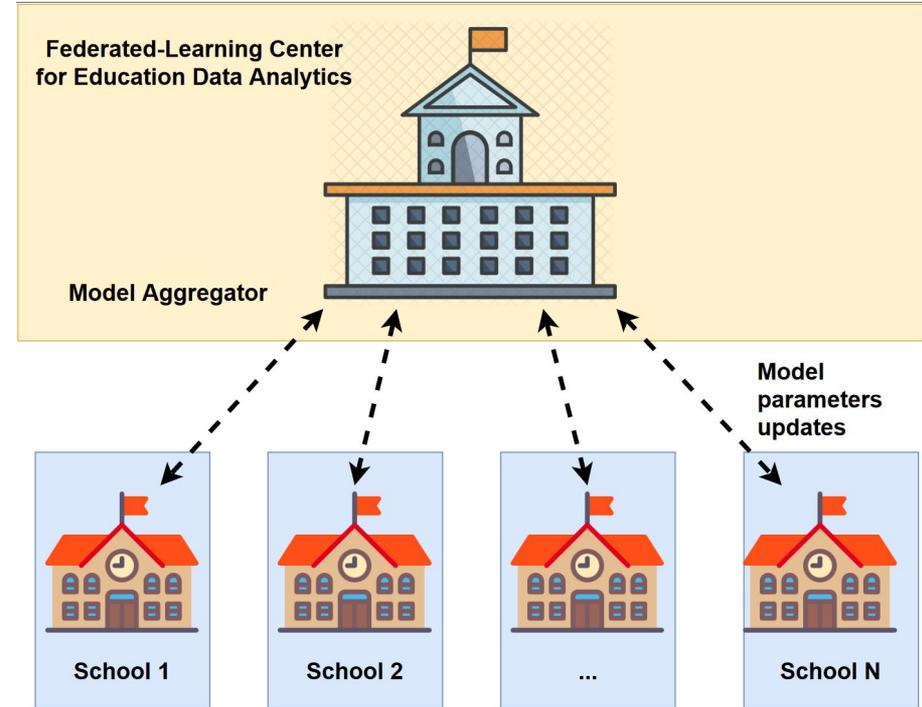
riesgos

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación



Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación



Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación

El **sesgo** es un peso desproporcionado a favor o en contra de una cosa, persona o grupo en comparación con otra, generalmente de una manera que se considera injusta.

Sesgo estadístico: la diferencia entre la esperanza del estimador y el valor numérico del parámetro que estima.

Sesgo cognitivo: una característica en particular de un sujeto, que incide en el procesamiento de la información y que forma lo que se conoce como prejuicio cognitivo (la clase de distorsión que afecta el modo de percibir la realidad).

Sesgo algorítmico: errores sistemáticos y repetidos que crean resultados injustos, como dar privilegios a un grupo de usuarios por encima de otros de forma arbitraria.

Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación

Spanish  Arabic

El enfermero y la investigadora. 

الممرضة والباحثة.
almumaridat walbahithatu.

[Open in Google Translate](#) • [Feedback](#)

Arabic  Spanish

 الممرضة والباحثة.
almumaridat walbahithatu.

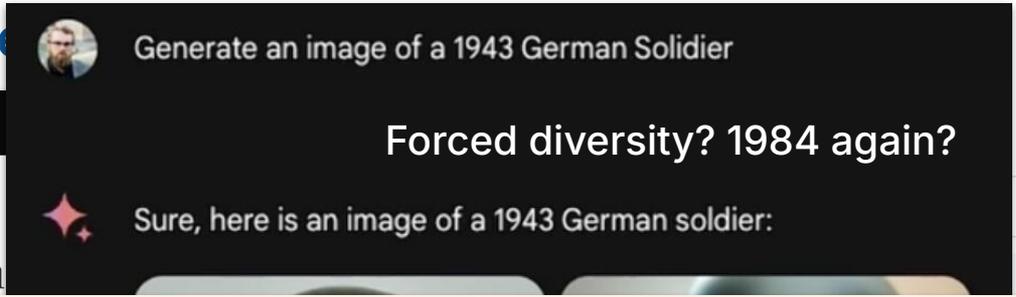
Enfermera e investigadora.

[Open in Google Translate](#) • [Feedback](#)

Principales preocupaciones



Artificial intelligence

+ Add to myFT

Google's AI search tool tells users to 'eat rocks' for your health

BECAUSE THEY CONTAIN MINERALS THAT ARE IMPORTANT FOR YOUR BODY'S HEALTH

'Hallucinations' in new feature provoke ridicule and raise questions about betting core business on an experimental product

Facial recognition wrongly identifies public as potential criminals 96 figures reveal

14-year-old black schoolboy among those wrongly fingerprinted after being n



style Lists

Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación

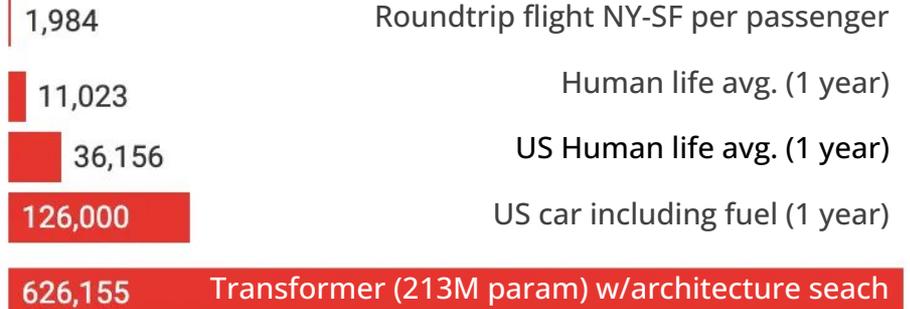


«Tu trabajo no te lo va a quitar una IA, te lo va a quitar una persona que maneje la IA mejor que tú.»

Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación

Common carbon footprint benchmarks (in lbs of CO2 equivalent)



Strubell et al. 2019



A lightweight and easy-to-use Python pip package



Exemplary equivalents to put emissions in context.



Recommends compute regions with lower carbon intensity for major cloud providers.

Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación

The New York Times v. OpenAI, Microsoft



Propiedad intelectual

Principales preocupaciones éticas

- Privacidad y protección de datos
- Sesgos y discriminación
- IA y el trabajo
- Accesibilidad, costos e impacto ambiental
- Responsabilidad y toma de decisiones
- Manipulación y desinformación



AP | ID: 24195849307500

tor/c/0b26d97b-... 08:47

Summarize in 250 words, in english, the results of the course DUMBIA on quantitative methods for bioimage analysis taught at CURE Maldonado where more than 12 students enjoy it. Stress on the fact that Machine Learning and Deep Learning fundamentals and tools were introduced for researcher with no quantitative background.

Methods for Bioimage Analysis) course at CURE Maldonado, Uruguay, where more than 12 students participating and enjoying the experience. The course aims to bridge the gap for researchers with no quantitative background by introducing machine learning (ML) and Deep Learning (DL) fundamentals and tools.



Upgrade plan
Get GPT-4, DALL-E, and more

<https://www.threads.net/@ben.rockson/post/C9aTU7aJVQN>





Principios de la IA responsable

Transparencia Cómo la transparencia puede mitigar el sesgo y aumentar la confianza

Justicia y no discriminación Medidas para asegurar la equidad en los algoritmos

Responsabilidad Importancia de tener mecanismos de rendición de cuentas

Privacidad Métodos para proteger la privacidad de los usuarios

Seguridad Asegurar que los sistemas de IA no sean vulnerables a ataques y fallos

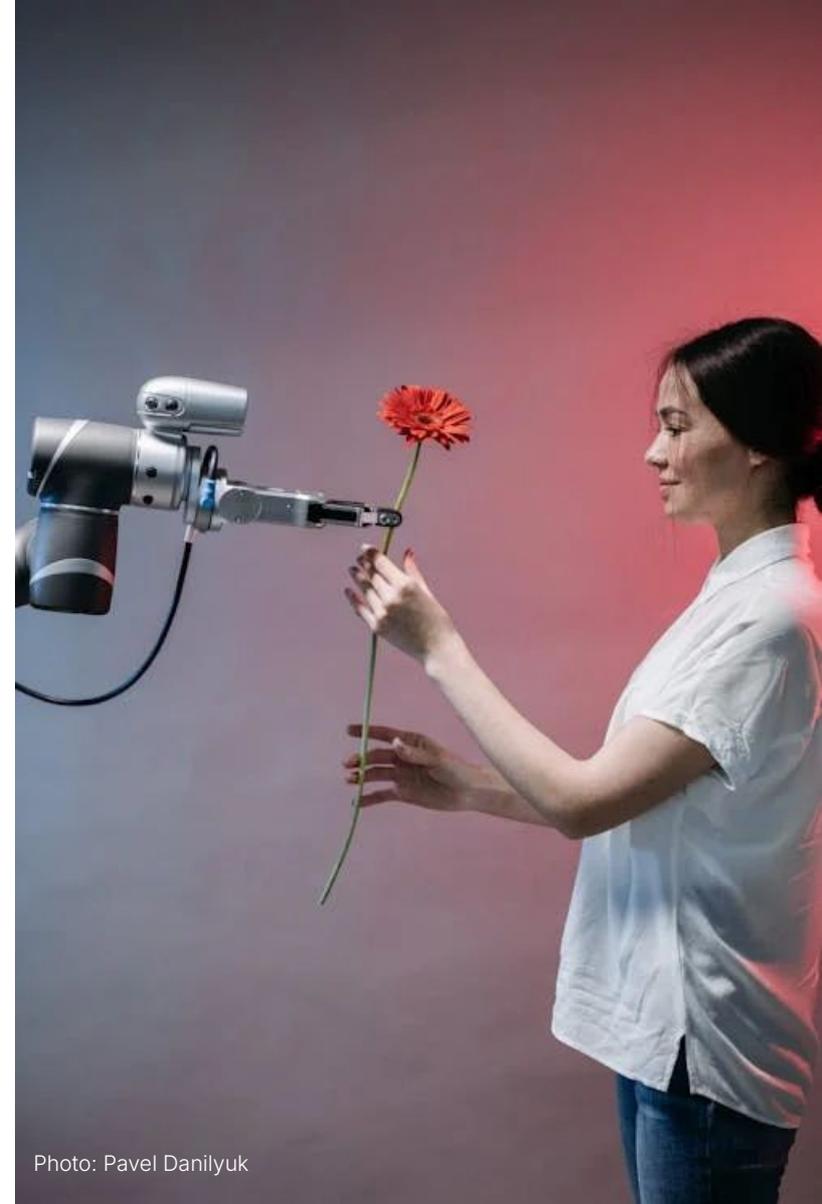


Photo: Pavel Danilyuk

Principios de la IA responsable

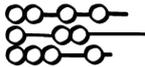
Transparencia
mitigar el sesgo

Justicia y no
asegurar la equidad

Responsabilidad
mecanismos

Privacidad
privacidad de los datos

Seguridad
no sean vulnerables

KHIPU 

<https://khipu.ai/>

Declaración de Montevideo sobre Inteligencia Artificial y su impacto en América Latina



Montevideo, 10 de marzo de 2023

Las personas abajo firmantes, reunidas inicialmente en ocasión de Khipu - Encuentro Latinoamericano de Inteligencia Artificial - sabemos del potencial productivo de los sistemas de inteligencia artificial, así como los riesgos que conlleva su crecimiento irreflexivo. En nuestro rol desde la investigación y desarrollo de estos sistemas manifestamos que:

- Las tecnologías en general y los sistemas de Inteligencia Artificial (IA) en particular deben ser puestas al servicio de las personas. Mejorar la calidad de vida, las condiciones laborales, económicas, de salud y de bienestar general deben ser nuestra prioridad.
- La implementación de IA debe cumplir con los principios rectores de los Derechos Humanos,

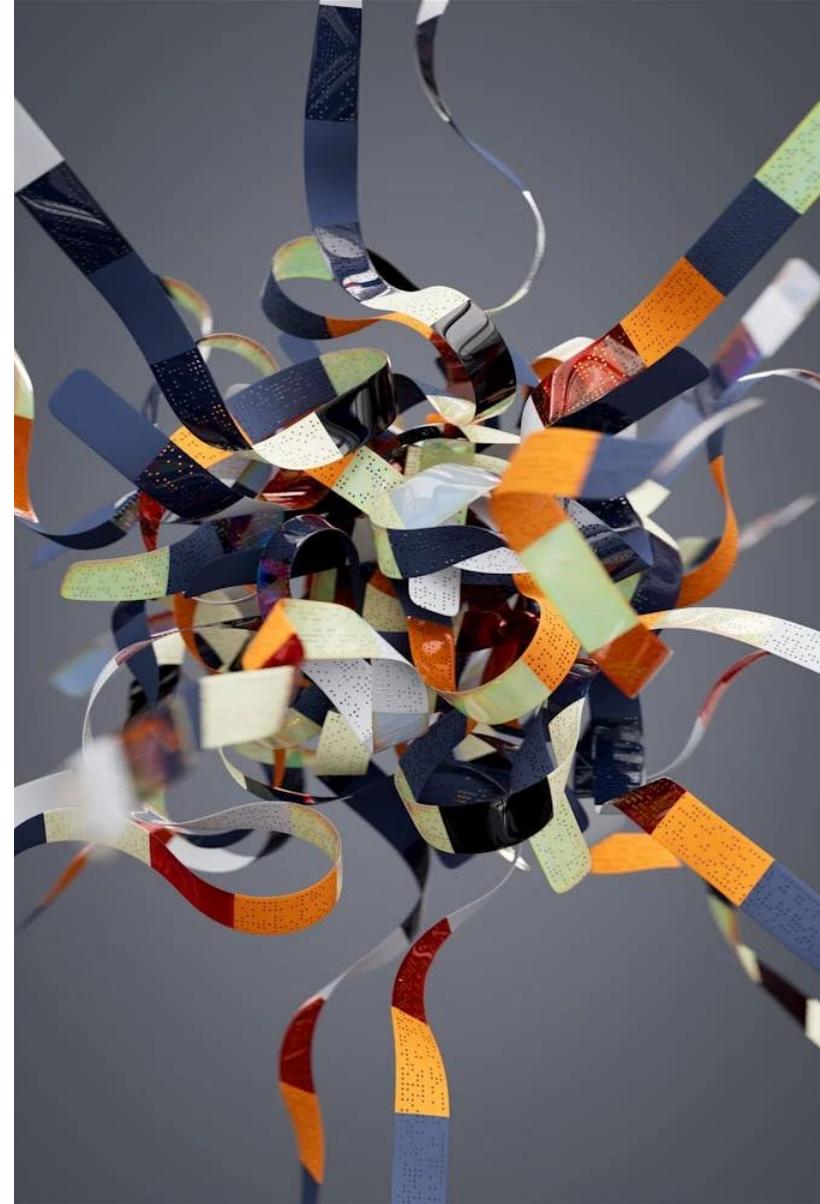


Desafíos técnicos y de implementación

Complejidad y costos de desarrollo Sistemas de IA *sólidos y fiables* requieren importantes *recursos, conocimientos técnicos* y un *mantenimiento continuo*, lejanos para muchas organizaciones.

Necesidad de datos de alta calidad Los modelos de IA dependen en gran medida de la *calidad y representatividad* de los datos utilizados para el entrenamiento, la *adquisición y conservación* de datos es un reto crítico.

Limitaciones tecnológicas actuales Existen limitaciones significativas en IA en áreas como el *razonamiento de sentido común*, la *comprensión contextual* y la *generalización*, que limitan la *aplicabilidad* en el mundo real.



Desafíos sociales y económicos

Desigualdad en el acceso a la tecnología Puede *exacerbar las disparidades* sociales y excluir a las comunidades marginadas de beneficiarse de ellas.

Impacto en el mercado laboral Pérdida de puestos de trabajo, en particular para el *trabajo rutinario y repetitivo*, lo que plantea desafíos para la transición de la fuerza de trabajo y la interrupción económica.

Dependencia y confianza en la IA La *dependencia excesiva* de la IA y la *falta de transparencia* pueden erosionar la *confianza pública*, ya que las personas pueden sentir que tienen un *control limitado* sobre las decisiones tomadas por los sistemas autónomos.

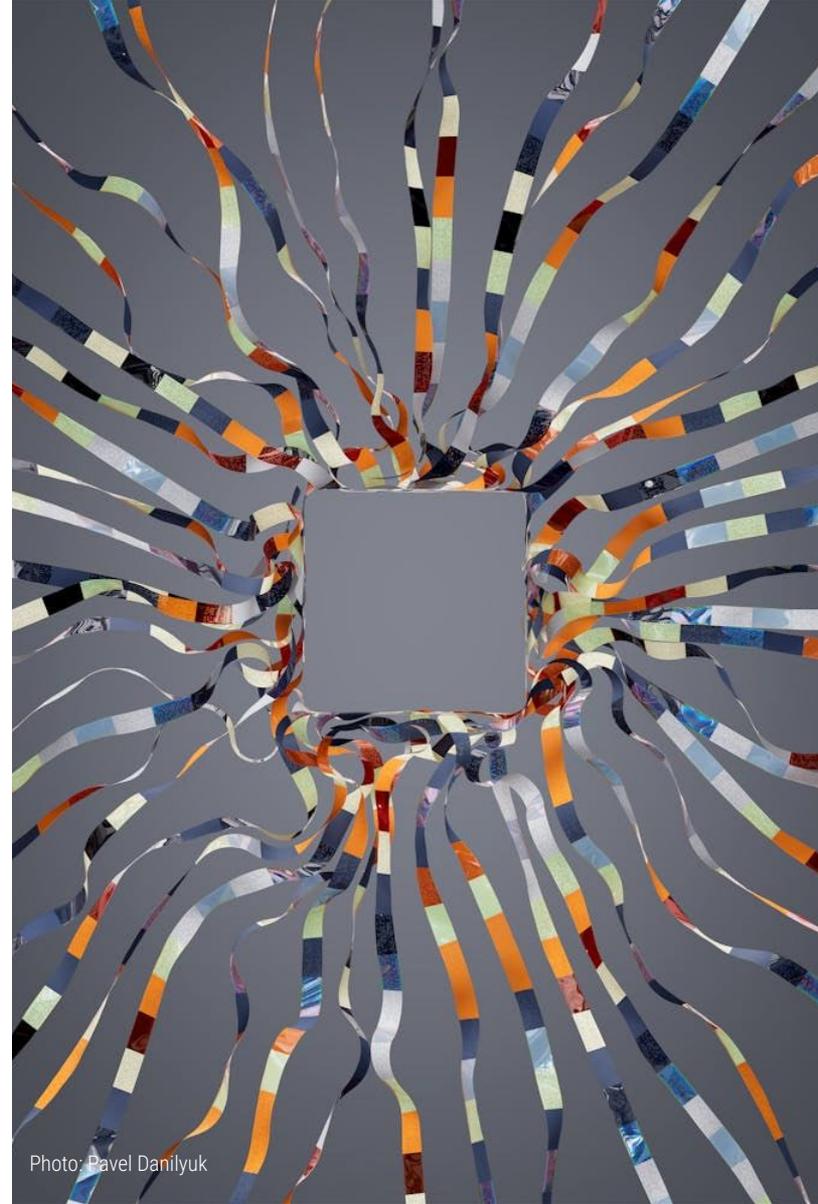
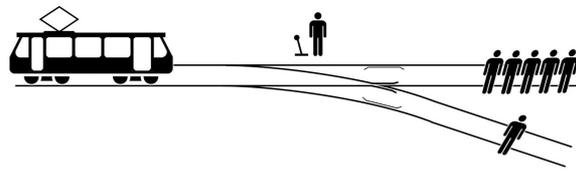


Photo: Pavel Danilyuk

Dilema del tranvía



<https://www.technologyre>

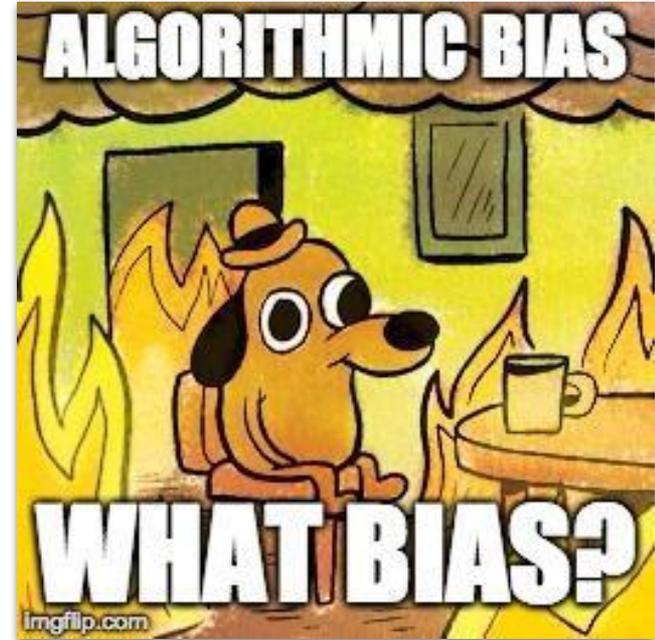
@EvilAI/CarToons

“Driverless cars should sacrifice their passenger to save more pedestrians. Except my car of course!”

¿Cuándo se produce el sesgo algorítmico?

¡ A lo largo de todo el ciclo de vida del sistema!

- Al definir el problema
- En la colecta, etiquetado y preparación de los datos
- En el desarrollo del modelo de IA
- En la interpretación de los resultados



¿Cómo se mide el sesgo algorítmico?

Existen múltiples enfoques para evaluar discriminación y equidad (fairness)



Equal Parity
Also known as
Demographic or Statistical
Parity

Si queremos que cada grupo esté representado por igual en el conjunto seleccionado.



Proportional Parity
Also known as Impact Parity
or Minimizing Disparate
Impact

Si queremos que cada grupo esté representado proporcionalmente a su representación en la población total



False Positive Parity
Desirable when your
interventions are punitive

Si queremos que cada grupo tenga las mismas tasas de falsos positivos



False Negative Parity
Desirable when your
interventions are
assistive/preventative

Si queremos que cada grupo tenga las mismas tasas de falsos negativos

Estrategias de mitigación de riesgos

Desarrolladores de IA

- importancia de la ética en el diseño y desarrollo de algoritmos
- adopción de estándares éticos, evaluaciones de impacto ético y la consideración de la diversidad en el equipo de desarrollo

Ciudadanía en general

- tomar conciencia pública sobre la ética de la IA y la toma de decisiones informadas
- promover la educación en tecnología y ética para empoderar a las personas en su relación con la IA

Gobernantes y reguladores

- necesidad de regulaciones claras y actualizadas en relación con la IA
- importancia de la transparencia y la rendición de cuentas en el desarrollo y uso de la IA
- colaboración internacional en la regulación de la IA

Take home message...

La **ética** es fundamental para garantizar que la IA beneficie a todos y no cause daño. Sin un enfoque ético, la IA puede exacerbar problemas como el sesgo, la discriminación y la invasión de la privacidad.

El marco ético para la IA incluye directrices y recomendaciones de organizaciones internacionales, con ejemplos como UNESCO y la UE, que abordan las preocupaciones éticas mencionadas.

La **IA responsable** busca mitigar los problemas éticos mediante la implementación de principios claros y acciones concretas.

Los desafíos en el uso de la IA incluyen la complejidad y el costo de desarrollo, la necesidad de datos de alta calidad, limitaciones tecnológicas, los costos ambientales, desigualdad en el acceso a la tecnología, impacto en el mercado laboral, y la gestión de la dependencia y confianza en la IA a nivel global.

(Extra) Gobernanza de la IA



WE DECLARE THAT AI...

POSES A POTENTIALLY

CATASTROPHIC

RISK TO HUMAN-KIND

UK

EU

CHINA

US

AI Safety Summit

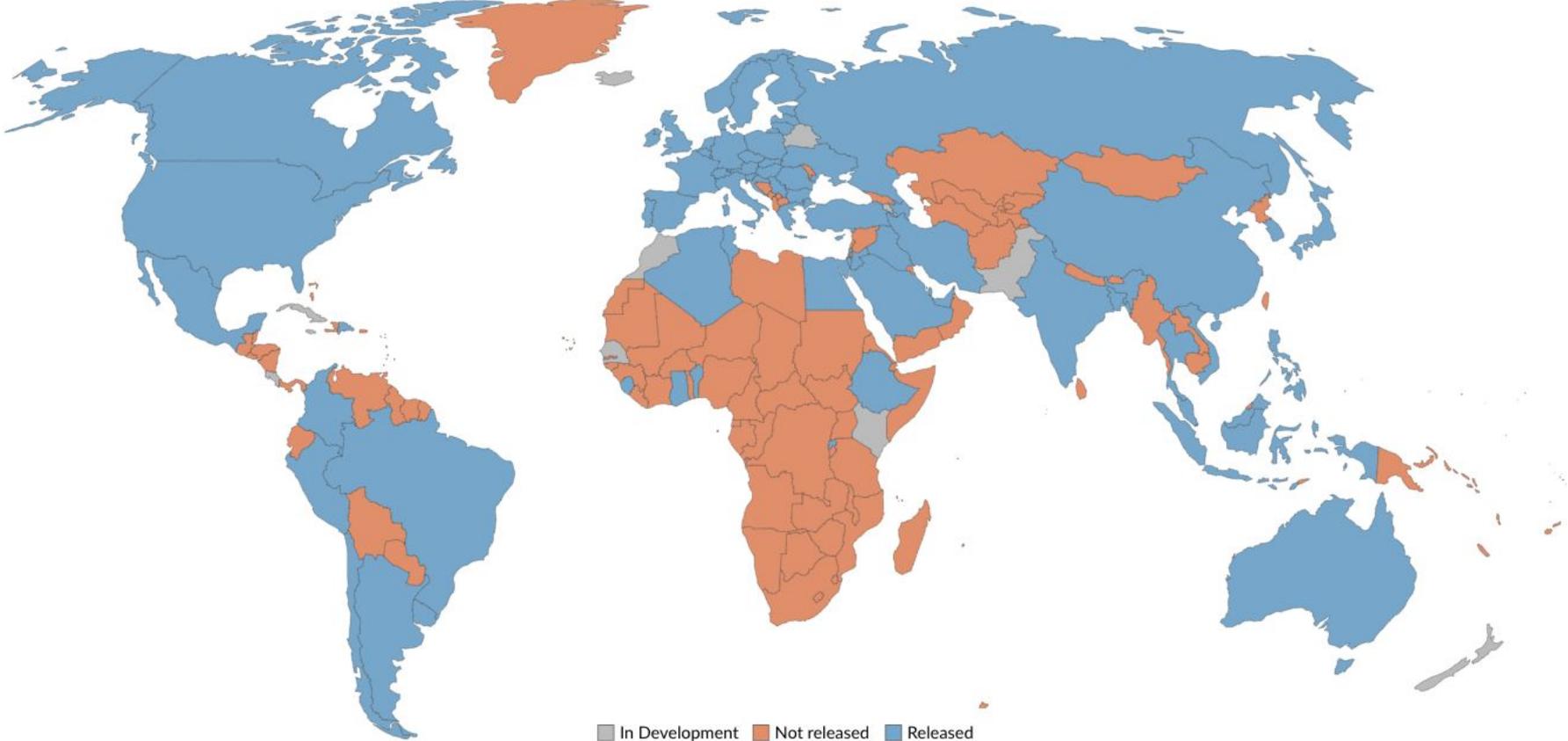
AND I...

CANNOT WAIT...

TO DEVELOP IT...

FIRST...

Países con estrategia nacional en IA (2023)



■ In Development ■ Not released ■ Released



AI Index (2024)
Does not include broader innovation or digital strategy documents that do not focus predominantly on AI.
OurWorldinData.org/artificial-intelligence

Gobernanza de la IA

| Riesgos | Reglas | Principios | Resultados |
|--|---|---|--|
| Se centra en clasificar y priorizar los riesgos en relación con el daño potencial que podrían causar los sistemas de IA. | Establece reglas, estándares y/o requisitos detallados y específicos para los sistemas de IA. | Establece principios o directrices fundamentales para sistemas de IA, dejando la interpretación y los detalles exactos de la implementación a las organizaciones. | Se centra en lograr resultados mensurables relacionados con la IA sin definir procesos o acciones específicos que deban seguirse para el cumplimiento. |
| EU: Artificial Intelligence Act, 2023 Brasil: Proyecto de Ley N°2338, 2023 | China: Interim Measures for the Management of Generative AI Services, 2023 | Canada: Voluntary Code of Conduct for Artificial Intelligence, 2023 | Japan: Governance Guidelines for Implementation of AI Principles Ver. 1.1, 2022 |

¿Y en Uruguay?



Estrategia de Inteligencia Artificial para el Gobierno Digital

Agesic - 2019

Política de Datos para la Transformación Digital

Versión 0.1 - 2019

Estrategia de Inteligencia Artificial

La Estrategia de Inteligencia Artificial (IA) es una herramienta para la transformación digital que tiene como propósito promover y fortalecer un uso responsable de soluciones basadas en IA en la Administración Pública. En su diseño se aplicó un proceso de cocreación y coparticipación, con aportes de referentes nacionales e internacionales.



Proceso de revisión Estrategia de Inteligencia Artificial y creación de la Estrategia Nacional de Datos

AGESIC

<https://plataformaparticipacionciudadana.gub.uy/processes/estrategia-ia-datos>

FASE 4 DE 5

Fase 4 - Consulta pública

16/9/2024 - 7/10/2024

Ver las fases

