IA: LA RESPONSABILIDAD EN UN MUNDO CAMBIANTE

Prof. Dra. Virginia Dignum

Cátedra de IA responsable – Directora del AI Policy Lab, Universidad de Umeå

Profesora invitada SIAS, Universidad de Ámsterdam

Correo: virginia@cs.umu.se

UMEÅ UNIVERSITY

AI Policy Lab

at Umeå University

LA IA EN 2026: LOS RETOS

¿CONCIENCIA?



¿LA IA PRIMERO?



AI Policy Lab
at Umeå University

EL ENFOQUE «LA IA PRIMERO»

- Pasa DE la gobernanza AL despliegue sin control
- Ignora la evidencia de que la IA aporta ganancias de productividad limitadas
- **Faltan las preguntas** «¿Por qué IA?» **y** «¿IA sí o no?»
- Déficit democrático: no hay consulta a las partes interesadas
- Genera presión para adoptar sin evaluar (Dignum et al., 2025)

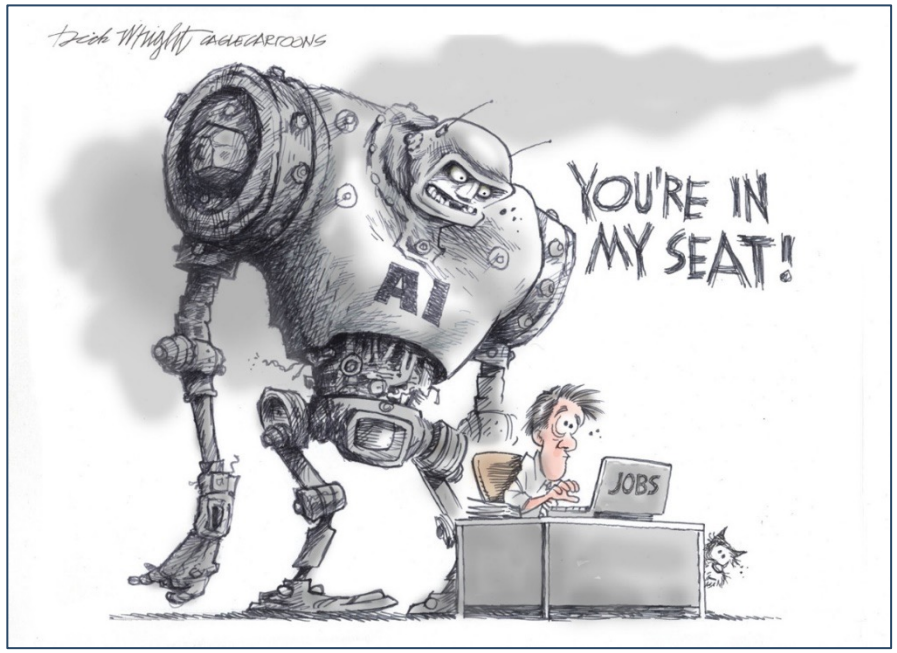
Los incentivos que promueven el uso de la IA al margen de su idoneidad serán contraproducentes: municipios y regiones necesitan competencia para valorar cuándo y dónde es adecuada la IA, no estructuras que impulsen su adopción indiscriminada

Se basa en la idea errónea de que la regulación frena los servicios digitales

Adoptar la IA sin invertir en I+D nacional aumenta la dependencia de unos pocos proveedores extranjeros y socava la misma soberanía digital que la estrategia dice impulsar.

La responsabilidad sigue al control. Los productores diseñan el sistema; deben asumir la responsabilidad principal. Trasladar la verificación a los usuarios solo desplaza la responsabilidad sin reducir el riesgo.

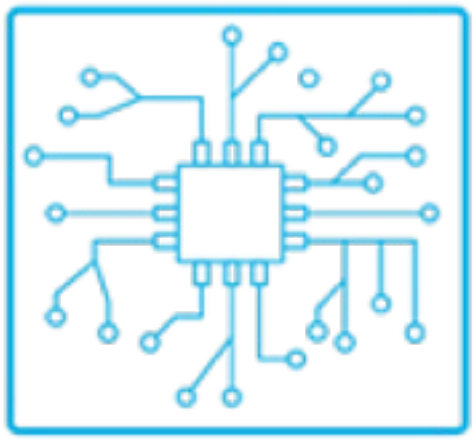
¿MOLDEAMOS LA IA O LA IA NOS MOLDEA A NOSOTROS?





- La IA plantea tantos problemas como resuelve
- La IA no es un medio para sustituir a las personas
- Hace falta la mirada humana para abordar las consecuencias éticas, sociales e imprevistas de la IA

IA: ¿MÁS ES MEJOR?



Computing Power



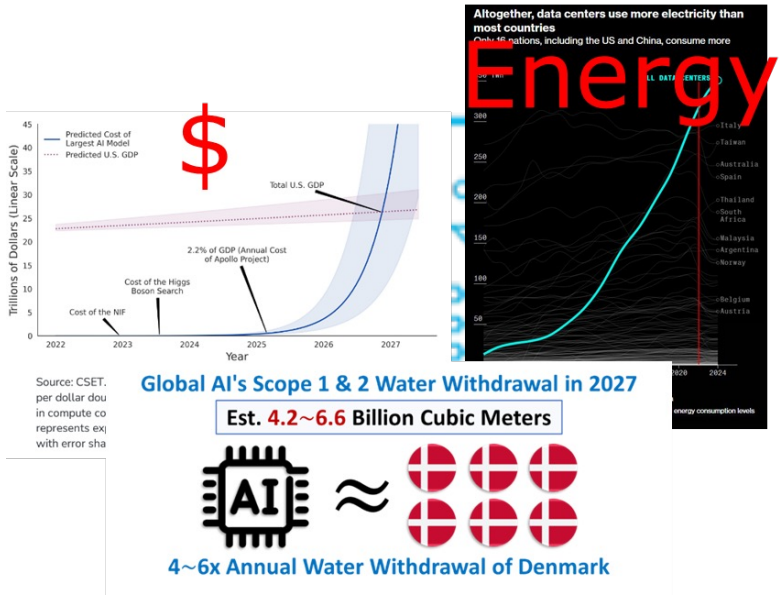
Algorithm Power



Data Availability

AI Policy Lab
at Umeå University

IA: ¿MÁS ES MEJOR?



Global AI's Scope 1 & 2 Water Withdrawal in 2027

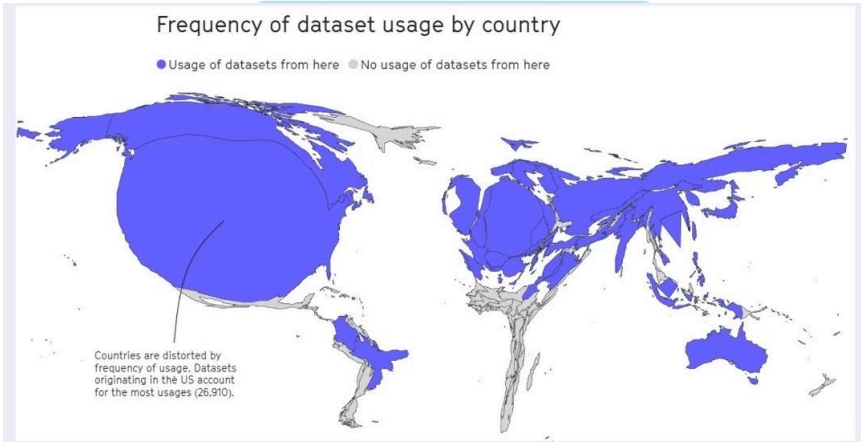
Est. 4.2~6.6 Billion Cubic Meters

4~6x Annual Water Withdrawal of Denmark

Computing Power



Algorithm Power



Data Availability

AI Policy Lab
at Umeå University

EL IMPACTO DE LA IA IMPLICA RESPONSABILIDAD HUMANA

Overcoming Racial Bias In AI Systems And Startlingly Even In AI Self-Driving Cars

Racial
patien

AI expert calls for end to UK use of 'racially biased' algorithms

AI E
Live
Reg

Gender bias in AI: building fairer algorithms

Bias in AI: A problem recogn
still unresolved

Amazon, Apple
transcribing bl
AI voice recogn

Millions of black people affected by racial bias in health-care algorithms

When It Comes to Gori

Google exploited homeless black people to develop the Pixel 4's facial recognition AI

Russia Tests New Disinformation Tactics in Africa to Expand Influen

Amazon's facial recognition matched 28 members of Congress to criminal mugshots

Flawed Algori





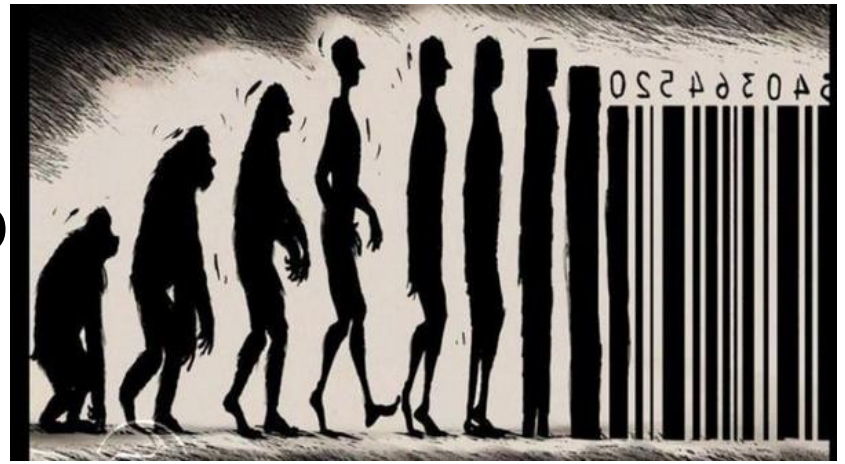




AI
at Umeå University

IA RESPONSABLE: ¿POR QUÉ IMPORTA?

- Datificación
  - La realidad es más que dato
  - Los datos se construyen
  - Los datos están sesgados
  - Ricos en datos, pobres en conocimiento



- Poder
- Sostenibilidad

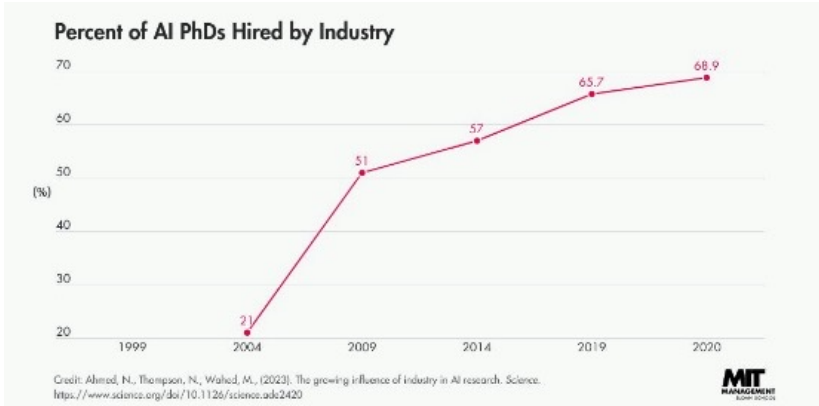


Revista Time, enero de 2023

IA RESPONSABLE: ¿POR QUÉ IMPORTA?

- Datificación
- Poder
  - ¿Quién desarrolla la IA?
  - ¿Quién decide?
- Sostenibilidad

At Tech's Leading Edge, Worry About a Concentration of Power



- **18% researchers at conferences are women**
- **80% professors are men**
- **Workforce**
  - **Google: 2,5% black, 3,6% Latino, 10% women**
  - **Facebook: 3,8% black, 5% Latino, 15% women**

AI Policy Lab
at Umeå University

IA RESPONSABLE: ¿POR QUÉ IMPORTA?

- Datificación
- Poder
- Sostenibilidad
  - Coste computacional de la
  - Costes humanos y sociales



INDEPENDENT PREMIUM

THE LONGER READ

Behind the AI boom, the armies of overseas workers in 'digital sweatshops'



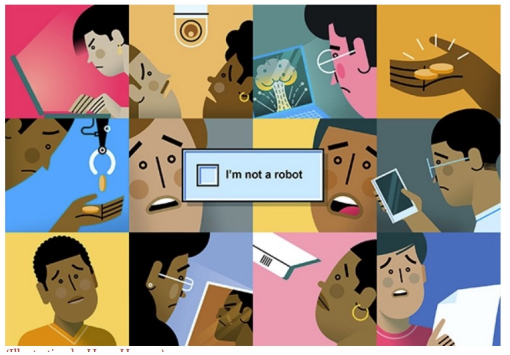
Technology

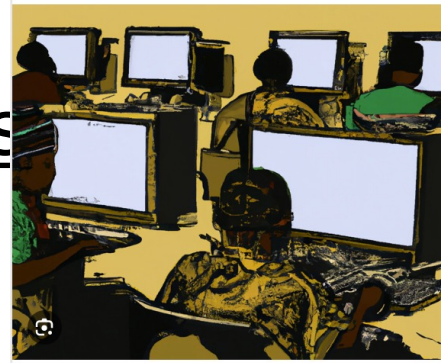
Turk Wars: How AI Threatens the Workers Who Fuel It

The much-hyped AI tools of the future are being built by a globally dispersed army of data workers.

CITE SHARE COMMENT PRINT ORDER REPRINTS

By Krystal Kauffman & Adrienne Williams | Oct. 11, 2023





OpenAI Used Kenyan Workers on Less Than $2 Per Hour: Exclusive | TIME

AI Policy Lab
at Umeå University



¿PODEMOS SER RESPONSABLES ?

Lone Thomasky & Bits&Bäume /
https://betterimagesofai.org /
https://creativecommons.org/licenses/by/4.0

LA IA NO ES NEUTRAL

- La IA puede promover la dignidad humana, o socavarla
- encarna decisiones humanas
- Gobernar la IA = gobernarnos a nosotros mismos



IA – ÉTICA DE LA IA – IA RESPONSABLE – GOBERNANZA DE LA IA…

- Primera etapa
  - Rendimiento técnico y eficiencia de los sistemas de IA
- Creciente toma de conciencia
  - Reconocimiento de riesgos como sesgos, discriminación, vulneración de la privacidad, falta de rendición de cuentas
- Giro hacia la ética
  - Llamamientos a la equidad, la transparencia, la supervisión humana y la explicabilidad
- IA responsable
  - Pasar de los principios a la práctica integrando la ética en el diseño, el despliegue y el uso
- Necesidad de gobernanza
  - Garantizar la confianza social, la sostenibilidad a largo plazo y la alineación con los valores democráticos

¿DE QUÉ ESTAMOS HABLANDO?

- **Sesgo**: desviación sistemática, no sinónimo de error.
  - Distinción importante: el sesgo estadístico (muestreo, medición) no es lo mismo que el sesgo social (el sistema reproduce un patrón injusto).
- **Equidad**: una reivindicación distributiva disputada, no una métrica.
  - Los criterios formales (paridad, igualdad de oportunidades, calibración) son incompatibles entre sí salvo en casos degenerados, así que elegir uno es un acto normativo.

(Kleinberg et al. 2016; Chouldechova 2017)

- **La transparencia** (sobre el proceso y la procedencia) es distinta de la explicabilidad (sobre un resultado concreto).
  - Son obligaciones distintas con públicos distintos.

¿QUÉ IMPORTANCIA TIENE PARA EL CLIMA?

- **Elección y delimitación del problema**
  - Qué cuenta como problema, y de quién es, se decide antes de recoger ningún dato.
  - P. ej. «predecir el rendimiento» y «apoyar la decisión de un pequeño agricultor» no son el mismo problema.
- **Variable objetivo y proxies**
  - La etiqueta rara vez es lo que realmente importa: reclamaciones de daños como proxy del impacto de las inundaciones, demanda conectada a la red como proxy de la necesidad de energía.
- **Cobertura de datos**
  - Los modelos se validan donde ya hay datos y se infieren donde no los hay.
  - En gran parte de África la cobertura de estaciones

La solución no es un modelo mejor, sino mejores preguntas, antes.

AI Policy Lab
at Umeå University

IA RESPONSABLE PARA LA IA CLIMÁTICA – EJERCICIO 1

Escenario: una ciudad encarga un modelo de riesgo de inundación. Se entrena con veinte años de reclamaciones de seguros por daños: cada reclamación registrada marca un lugar como inundado, y los lugares sin reclamación se consideran no inundados. El modelo usa imágenes satelitales, altitud y precipitación, y alcanza un 91 % de exactitud en datos reservados.

¿Dónde entra primero el sesgo?

- La resolución de las imágenes satelitales
- La elección de las reclamaciones de seguros como

Join at
slido.com
#4275 548

AI Policy Lab
at Umeå University

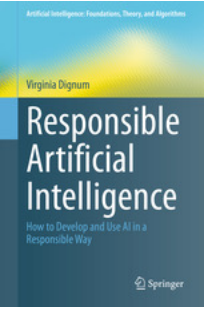
IA RESPONSABLE PARA LA IA CLIMÁTICA – EJERCICIO 1

Escenario: una ciudad encarga un modelo de riesgo de inundación. Se entrena con veinte años de reclamaciones de seguros por daños: cada reclamación registrada marca un lugar como inundado, y los lugares sin reclamación se consideran no inundados. El modelo usa imágenes satelitales, altitud y precipitación, y alcanza un 91 % de exactitud en datos reservados.

¿Dónde entra primero el sesgo?

- La resolución de las imágenes satelitales
- La elección de las reclamaciones de seguros como etiqueta
- La arquitectura del modelo

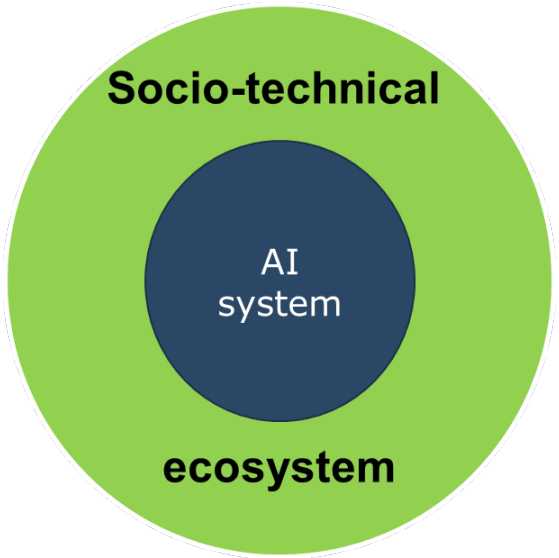
Las reclamaciones registran quién estaba asegurado y quién reclamó, no quién se inundó. Los hogares sin seguro y los asentamientos informales no están infrarrepresentados: figuran como secos. El 91 % se mide contra la misma etiqueta defectuosa, así que no puede detectar el problema. Todo lo que sigue lo hereda.



RESPONSABLES ACERCA DE LA IA

¡La IA no nos sucede sin más!

- La IA se diseña. Nosotros tomamos las decisiones
  - ¿Quién debe decidir?
  - ¿Qué valores deben tenerse en cuenta?
  - ¿Los valores de quién? ¿Cómo priorizar?
- ¡Haga la Pregunta Cero!
  - ¿Es la IA la mejor opción aquí?
  - Pregunte por qué antes de preguntar cómo
- La IA no existe en el vacío
  - ¡No hay arreglo tecnológico para los efectos nocivos!
  - La ética, la regulación y la gobernanza conciernen al ecosistema



¡Las soluciones de IA responsable deben ser sociales más que técnicas!

AI Policy Lab
at Umeå University

LA ALTERNATIVA: EL PROPÓSITO PRIMERO (PREGUNTA CERO)

El enfoque «el propósito primero» **pregunta: «**¿Si la IA es realmente necesaria y beneficiosa, y dónde?»

«¿Debería siquiera construirse este sistema de IA?»

Principios clave:

- Partir de las **necesidades sociales,** no de las capacidades tecnológicas
- Preguntar sistemáticamente: ¿Por qué? ¿Para quién? ¿Qué alternativas hay?
- Incluir voces diversas en la toma de decisiones
- Centrarse en la dirección y el valor público, no en la velocidad

La Pregunta Cero:

- Pasa de la VELOCIDAD a la **DIRECCIÓN**
- Empodera a actores diversos (pymes, sociedad civil, países pequeños)
- Ofrece un marco práctico y de bajo coste
- Basa las decisiones en una **identificación real del problema** (Dignum et al., 2025)

Descarga: https://aipolicylab.se/wp-content/uploads/2026/04/Q0-Assessment-Tool-v3_draft_march-2026.pdf

LA PREGUNTA CERO VA MÁS ALLÁ DEL SÍ/NO

El ciclo de la Pregunta Cero:

- **Planteamiento**: de quién es el problema, quién lo definió, cuál es la referencia sin IA, quién asume el coste de una respuesta errónea.
- **Datos**: procedencia, consentimiento, lagunas de cobertura, quién es invisible en los datos, en qué condiciones se obtuvieron.
- **Modelización**: qué se optimiza, qué se sacrifica, qué incertidumbre se comunica.
- **Evaluación**: rendimiento desagregado, no agregado; modos de fallo identificados, no solo la exactitud.
- **Despliegue**: quién lo opera, qué recursos existen, qué pasa cuando se equivoca.

EJERCICIO 2: IA PARA PLANIFICAR LA ADAPTACIÓN A LAS INUNDACIONES

Escenario: una ciudad quiere introducir un sistema de IA que clasifique los barrios para financiar la adaptación a las inundaciones

- El sistema predice el riesgo de inundación a partir de imágenes satelitales, datos de altitud y reclamaciones de daños pasadas, y clasifica las zonas para invertir.
- El sistema estima la vulnerabilidad de los hogares combinando teledetección con datos de telefonía móvil y de servicios públicos.
- El sistema genera resúmenes y recomendaciones para el ayuntamiento, que tiene tres semanas para decidir.

PREGUNTA CERO – EJERCICIO 2

- ¿Es la IA adecuada para este fin?

- Opciones
  - A) Sí, la alternativa es un proceso manual más lento e igual de sesgado.
  - B) Sí, la alternativa es un proceso manual más lento e igual de sesgado.
  - C) No, los datos no pueden respaldar la decisión para la que se usan.
  - D) No, los datos no pueden respaldar la decisión para la que se usan.



Join at
slido.com
#4275 548

EL DISEÑO ES POLÍTICO

- Elecciones
- Formulación
- Información
- Implicación
- Legitimidad
- Agregación







¿QUÉ ES LA IA?

¿QUÉ QUEREMOS QUE SEA LA IA?

- ¿Similar al humano?
  - ¿Por qué?
  - ¿Qué significa esto?

- ¿Herramienta?
  - ¿Para qué? ¿Para quién?

- ¿Simulación u operación?
  - Entender la inteligencia construyendo inteligencia, o
  - Intervención activa en el mundo real

- ¿Normativa o descriptiva?
  - ¿Haz lo que decimos o haz lo que hacemos?



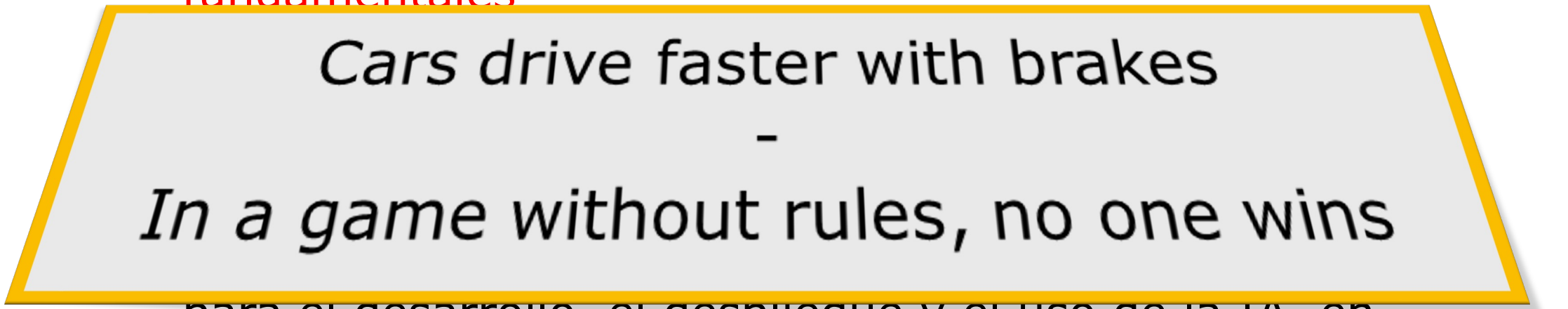
PRINCIPIOS, DIRECTRICES, ÉTICA

AI Policy Lab
at Umeå University

Shady Sharify / https://betterimagesofai.org
https://creativecommons.org/licenses/by/4.0

GOBERNANZA: ¿POR QUÉ? ¿PARA QUÉ?

- La regulación como incentivo para la innovación responsable, la sostenibilidad y los derechos humanos fundamentales



para el desarrollo, el despliegue y el uso de la IA, en especial de modelos de IA generativa con riesgos diversos

- Exigencias de responsabilidad, rendición de cuentas y gobernanza
  - Regular a los actores organizativos más que los resultados tecnológicos
  - Confianza pública y responsabilidad por los errores de la

AI Policy Lab
at Umeå University

PRINCIPIOS Y DIRECTRICES

EU HLEG	OECD	IEEE EAD
• Human agency and oversight • **Technical robustness and safety** • Privacy and data governance • **Transparency** • **Diversity**, non-discrimination and fairness • **Societal and environmental well-being** • **Accountability**	• benefit people and the planet • respects the rule of law, **human rights**, democratic values and **diversity**, • include appropriate safeguards (e.g. human intervention) to ensure a **fair and just society**. • **transparency** and responsible disclosure • **robust, secure and safe** • Hold organisations and individuals **accountable** for proper functioning of AI	• How can we ensure that A/IS do not infringe **human rights**? • effect of A/IS technologies on **human well-being**. • How can we assure that designers, manufacturers, owners and operators of A/IS are responsible and **accountable**? • How can we ensure that A/IS are **transparent**? • How can we extend the benefits and minimize the risks of AI/AS technology being misused?

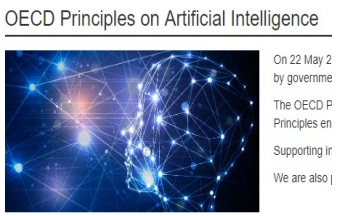
Nivel	Marco y alcance
Mundial	UNESCO (194 países), OCDE (>70 jurisdicciones), GPAI (25+ miembros), tratado del Consejo de Europa (50+ países), principios del G7
Regional	Ley de IA de la UE (27 Estados miembros), Declaración de Santiago (América Latina/Caribe)
Nacional	930+ iniciativas en 71 países
Otros	200+ directrices de ONG, universi… entidades privadas



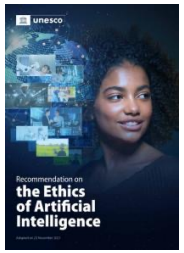
https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence



https://ethicsinaction.ieee.org



https://www.oecd.org/going-digital/ai/principles/



https://www.unesco.org/en/artificial-intelligence/recommendation-ethics



https://www.un.org/ai-advisory-body

¡Más de 1000 publicadas en el mundo!

AI Policy Lab
at Umeå University

PROTEGER LA DEMOCRACIA

- Reforzar la regulación y la supervisión
  - Aprobar leyes claras sobre IA, reconocimiento facial, criptomonedas y programas espía.
  - Reforzar la transparencia y la rendición de cuentas en el despliegue tecnológico.
  - Reforzar las instituciones democráticas para supervisar y orientar el desarrollo tecnológico.
- Promover la participación pública
  - Democratizar el desarrollo de la IA implicando a actores sociales diversos.
  - Alinear los sistemas tecnológicos con los valores democráticos y el interés público.
- Mejorar la alfabetización digital
  - Fomentar mediante la educación una comprensión crítica de la IA y las tecnologías digitales.
  - Capacitar a la ciudadanía para defender sus derechos en

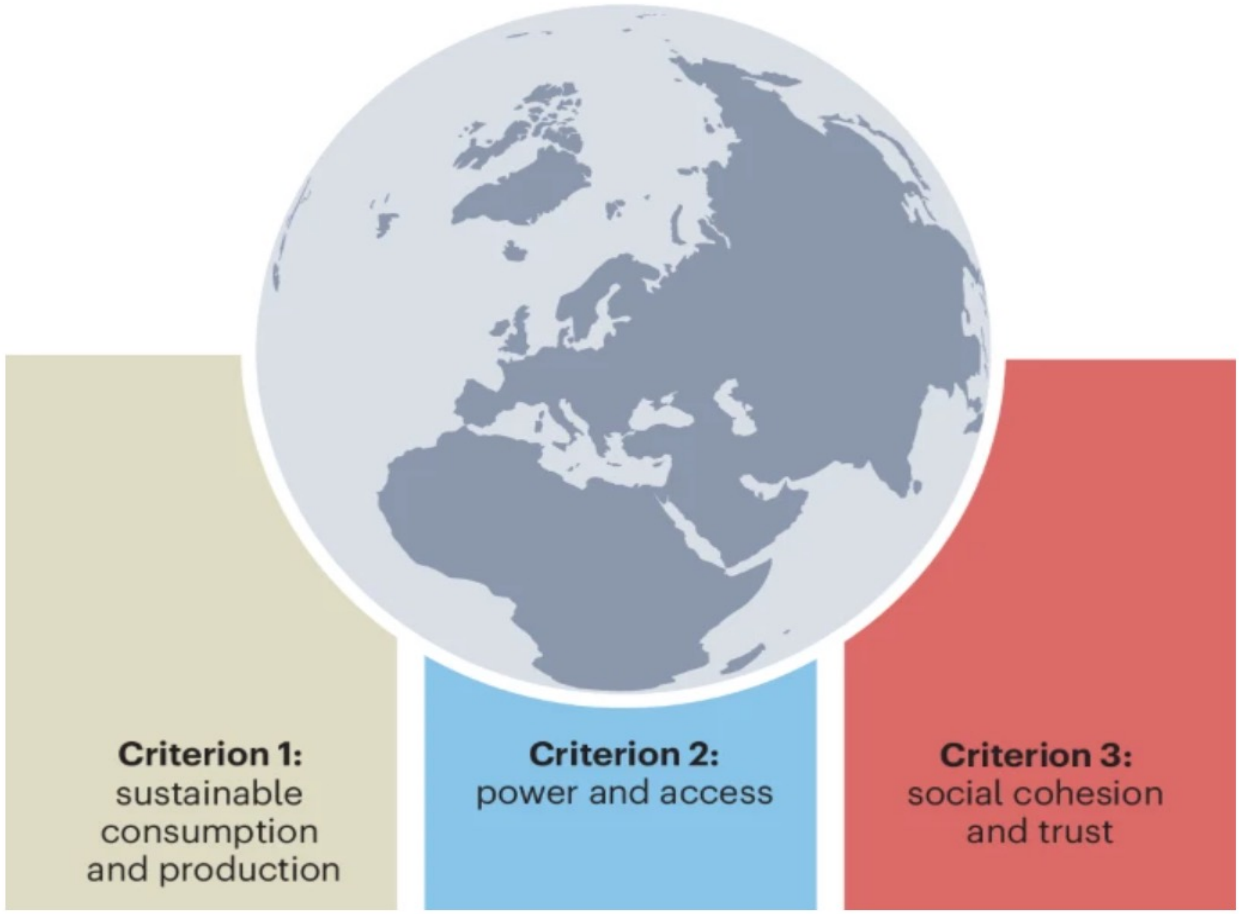
PROTEGER EL PLANETA

El impacto de la IA exige una custodia sistémica, más allá de los impactos ambientales:

1. ayudar a acelerar la **transición hacia una producción** y un consumo sostenibles que respeten los límites planetarios, o al menos no obstaculicen estos objetivos

2. desarrollarse, desplegarse y usarse de forma que **garantice un acceso equitativo** a las herramientas de IA para la sostenibilidad mundial y evite concentraciones de poder

3. desarrollarse, desplegarse y usarse



Owen Gaffney, Amy Luers, Franklin Carrero-Martinez, Berna Oztekin-Gunaydin, Felix Creutzig, Virginia Dignum, Victor Galaz, Naoko Ishii, Francesca Larosa, Maria Leptin & Ken Takahashi Guevara. The Earth alignment principle for artificial intelligence. *Nat Sustain* (2025). https://doi.org/10.1038/s41893-025-01536-6

MEDIR EL IMPACTO:
QUÉ CUENTA Y QUIÉN RESPONDE DE ELLO

Gobernamos lo que contamos. La mayor parte del impacto de la IA no se cuenta.

Tres niveles de impacto

(Kaack et al., 2022)

- Cómputo
- Aplicación
- Sistema

La consecuencia para la gobernanza

- La rendición de cuentas sigue a la medición
  - el cómputo tiene propietario; los efectos sistémicos a menudo no
- Lo que se informa se convierte en aquello para lo que se diseña
  - Elegir la métrica es un acto político, no técnico
- Alineación con la Tierra
  - límites planetarios, acceso equitativo, cohesión social.
- La Pregunta Cero también se aplica a las métricas
  - acordar qué cuenta, y quién lo audita, antes de construir

Kaack et al., Nature Climate Change 12, 518-527 (2022) · Gaffney et al., Nature Sustainability 8, 467-469 (2025)

AI Policy Lab at Umeå University

JUSTICIA CLIMÁTICA

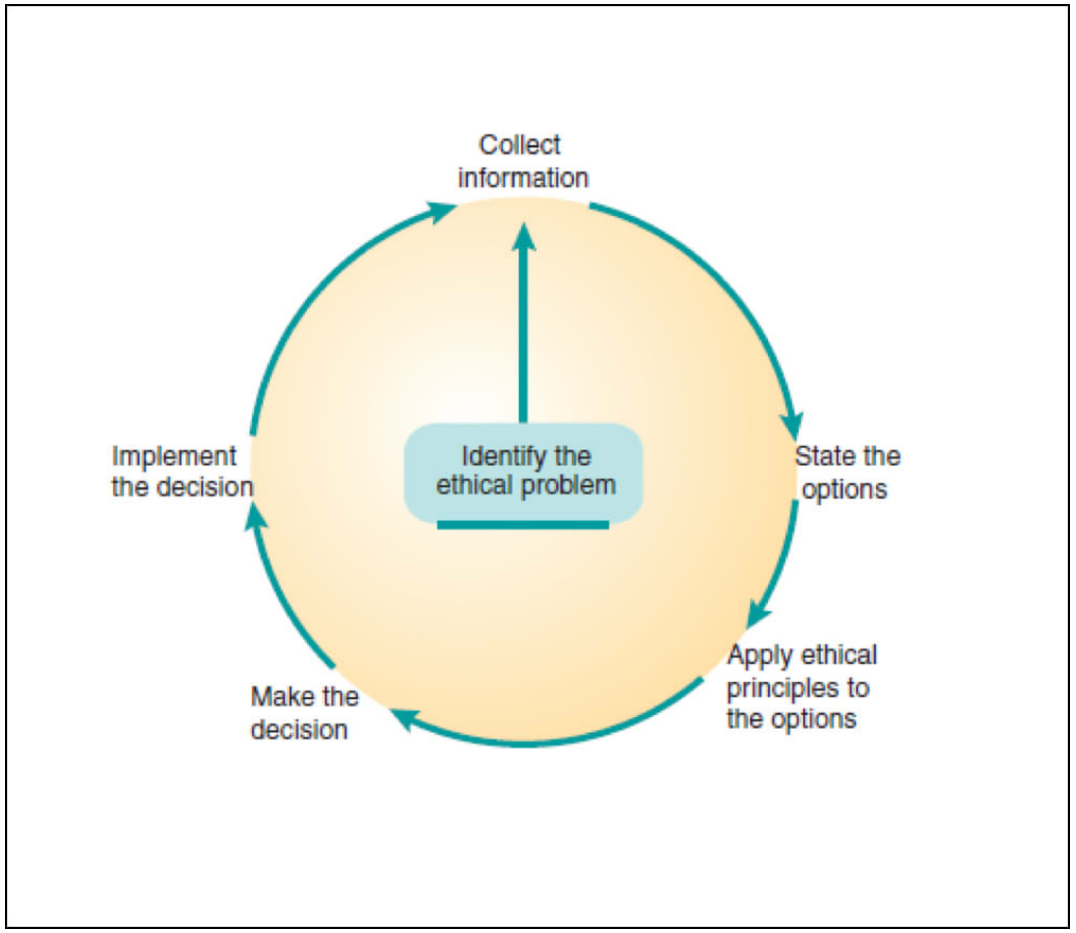
- distributiva (quién obtiene el beneficio y quién soporta el coste en cómputo, agua, tierra y trabajo),
- procedimental (quién estaba en la sala),
- de reconocimiento (qué conocimientos cuentan como evidencia).

- ¿El fin justifica los medios? ¿Compromisos?
  - P. ej. ¿es un uso responsable de la IA mejorar la extracción de combustibles fósiles?
  - ¿Qué alternativas hay? ¿Quién gana? ¿Quién pierde?

PUESTA EN PRÁCTICA: REGULACIÓN Y MÁS

- Regulación
  - Ley de IA: enfoque centrado en las personas y basado en el riesgo
- Normas
  - gobernanza blanda; cumplimiento no obligatorio
  - demostrar la diligencia debida y limitar la responsabilidad
  - integración sencilla entre productos
- Estructuras y procedimientos organizativos
  - Consejos asesores y responsables de ética
  - Fijar y supervisar directrices éticas
- Evaluación de una IA fiable
  - la IA responsable es más que marcar casillas
  - Hacen falta medios para evaluar la madurez
- Concienciación y participación

RAZONAMIENTO ÉTICO



- No se trata de la respuesta, sino de reconocer el problema
- ¡No hay una sola respuesta! ¡No hay solución mágica!
  - Enfoque utilitarista
  - Enfoque de derechos
  - Enfoque de justicia
  - Enfoque del bien común
  - Enfoque de las virtudes
  - ...

VIRGINIA DIGNUM; CORREO: VIRGINIA@CS.UMU.SE - TWITTER: @VDIGNUM

LA ÉTICA EN LA VIDA REAL

- Equidad
  - Subsidio de apoyo infantil
  - Permiso parental

- Seguridad
  - Radares de velocidad
  - Clases de educación vial para niños

- ¿Equidad o seguridad?
  - Recursos limitados

- equidad

- igualdad

– corto plazo

– largo plazo

DILEMAS – EJERCICIO 3

- **Subvenciones de adaptación climática**: le piden desarrollar un sistema de IA para filtrar las solicitudes a un fondo municipal de adaptación climática y conceder las subvenciones según la probabilidad de cada solicitud de lograr un «alto impacto climático por euro», usando datos de proyectos financiados anteriormente. Esto sustituye la revisión manual y las visitas sobre el terreno de los técnicos.
- Compromisos: cómo repartir sus recursos:
  - ¿Cuánto coste (computacional) vale un 5 % más de exactitud?
  - ¿Cuánto coste (computacional) vale incluir una explicación?

Join at
slido.com
#4275 548

AI Policy Lab
at Umeå University

ÉTICA DE LA IA

- ¿Es la IA tan diferente de otras tecnologías?
  - Solapamiento con la ética de la tecnología, la ética empresarial, la ética médica…

- Algunos retos específicos de la IA, derivados de:
  - La autonomía. Si se delega el control ejecutivo, ¿quién es responsable de las acciones? ¿Cómo mantener el respeto a los valores humanos en la toma de decisiones autónoma?
  - El despliegue en entornos abiertos. Si no podemos prever el contexto que encontrará el sistema, ¿cómo garantizamos un comportamiento ético?
  - La generalización. Si una decisión se toma generalizando a partir de datos, ¿qué pasa con nuestro derecho a ser vistos como individuos? ¿Cómo evitar la reproducción y amplificación de sesgos?

IA RESPONSABLE: MÁS QUE ÉTICA

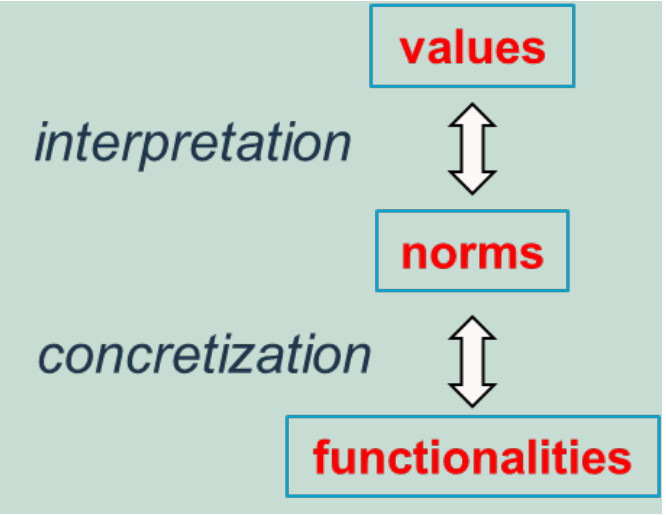
- No se trata de filosofar sobre ética
  - La ética no trata de la respuesta, sino de reconocer el problema
  - La ética es un proceso (social), no una solución

- No a la tecnificación de la ética
  - Su implementación no «resuelve» la ética

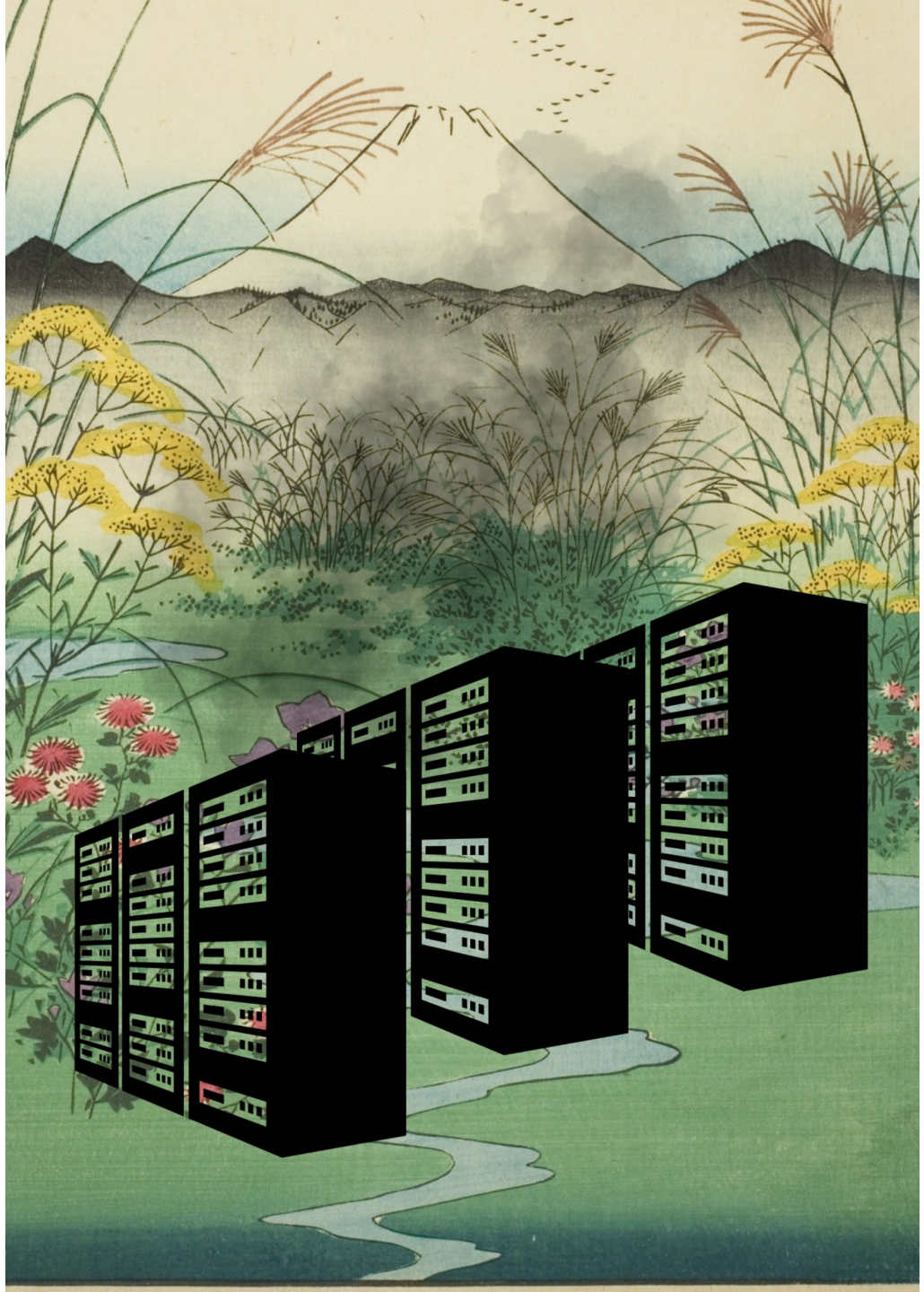
- En el fondo, se trata de elecciones, prioridades y compromisos
  - Exactitud / explicación
  - Exactitud / recursos de cómputo
  - Seguridad / privacidad
  - Equidad / igualdad
  - Beneficio a largo plazo / corto plazo
  - ...

LA IA RESPONSABLE CONSISTE EN SER EXPLÍCITO

- Diseño guiado por valores
  - ¡Los aspectos jurídicos y éticos no son un añadido!
- Gobernanza
  - Supervisión y control externos
  - Acuerdos, contratos, normas
- Diseño
  - Cuestione sus opciones y elecciones
  - Justifique sus elecciones
  - Documente sus elecciones y opciones



https://medium.com/@virginiadignum/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence-bcde64f59f5b



UN NUEVO PARADIGMA DE IA

Deborah Lupton /
https://betterimagesofai.org /
https://creativecommons.org/licenses/by/4.0

¿UN NUEVO PARADIGMA DE IA?

- La IA no es una solución universal
  - reconocer sus límites ante retos complejos
  - Pregunta cero: ¿debe usarse aquí la IA?
- Abordar los riesgos inherentes de sesgo y discriminación
  - mejorar la transparencia de los datos y los algoritmos
  - tener en cuenta contextos y valores diversos
- Más allá de las disciplinas
  - integrar conocimiento social y técnico en el diseño de la IA
  - abordar retos sociales sistémicos
  - implicar a actores sociales diversos, no solo a tecnólogos
- normas técnicas sólidas
  - Verificable, sostenible, participativa

ENFOQUE DE IA RESPONSABLE

Comprometerse con principios de IA responsable y fiable, para una IA innovadora, ética y respetuosa de los derechos humanos

Reconocer las posibles tensiones entre principios éticos y la necesidad de priorizarlos

Desarrollar una cultura de la responsabilidad en las organizaciones

Priorizar la interacción y la comunicación con las partes interesadas

UNA GOBERNANZA FIABLE

- Desarrollo de la IA
  - ¿Qué? Requisitos, necesidades, posibilidades, capacidades
- Uso de la IA
  - ¿Cómo? Licencia para operar
- Contexto en que se desarrolla y usa la IA
  - ¿Dónde/cuándo? Reglas del juego

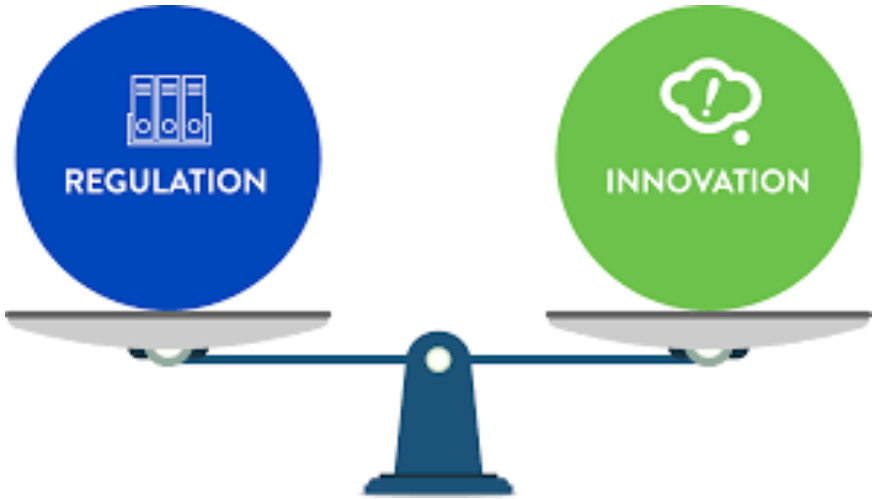
- Esfuerzos mundiales
  - Sostenibilidad
  - Inclusión y participación
  - Diversidad
  - Reparto de beneficios y costes
  - Agenda 2030 – ODS



LA IA RESPONSABLE ES INNOVACIÓN

- Innovación necesaria
  - Requisitos técnicos
  - Requisitos de uso
  - Contexto de operación

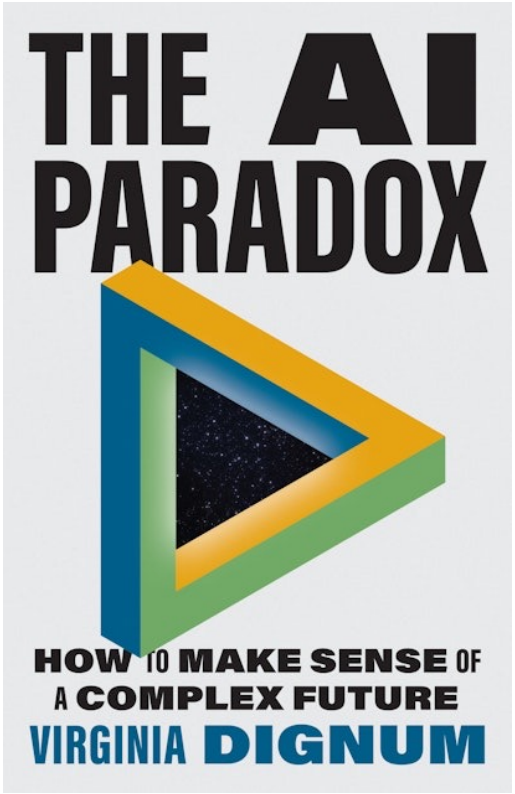
- ¡Innovación multidisciplinar!
  - Innovación tecnológica
  - Innovación organizativa
  - Innovación regulatoria
  - Innovación en gobernanza
  - Innovación social



La IA no es una fuerza imparable; la moldean nuestras decisiones

- las decisiones que tomemos hoy determinarán si se convierte en una herramienta de empoderamiento o en una fuente de control y desigualdad.

La gran paradoja de la IA: cuanto más inteligente es, más sabiduría humana necesitamos



Disponible en febrero de 2026

TEST

- Verdadero o falso: un modelo que cumple la paridad demográfica es justo.
- El sesgo puede entrar en un sistema de IA para el clima a través de
  - (a) la cobertura de los datos de entrenamiento
  - (b) la elección de la variable proxy
  - (c) la delimitación del problema
  - (d) todo lo anterior
- En un sistema de alerta temprana, bajar el umbral de alerta supone sobre todo un compromiso entre:
  - (a) exactitud y cómputo
  - (b) falsas alarmas y eventos no detectados
  - (c) transparencia y privacidad
  - (d) latencia y resolución

LECTURAS RECOMENDADAS

- Dignum, *Responsible Artificial Intelligence* (Springer, 2019)
- Nota informativa sobre la Pregunta Cero del AI Policy Lab
- Kaack et al., "Aligning artificial intelligence with climate change mitigation", *Nature Climate Change* 12, 518–527 (2022)
- Gaffney et al., "The Earth alignment principle for artificial intelligence", *Nature Sustainability* (2025)

MÁS LECTURAS

- **Mi trabajo y los materiales de la Pregunta Cero**
  - Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer.
  - Dignum, V. (2026). *The AI Paradox*. Princeton University Press.
  - Dignum, V., Carli, R., Ericson, P., Titareva, T. and Tucker, J. (2025). "AI First" to "Purpose First": Rethinking Europe's AI strategy. AI Policy Lab, Umeå University. https://aipolicylab.se/2025/11/05/ai-first-to-purpose-first-rethinking-europes-ai-strategy/
  - AI Policy Lab (2026). *Question Zero (Q0) Self-Assessment Tool for Responsible AI*, v3 draft. https://aipolicylab.se/wp-content/uploads/2026/04/Q0-Assessment-Tool-v3_draft_march-2026.pdf
- **IA y clima**
  - Gaffney, O., Luers, A., Carrero-Martinez, F., Oztekin-Gunaydin, B., Creutzig, F., Dignum, V., Galaz, V., Ishii, N., Larosa, F., Leptin, M. and Takahashi Guevara, K. (2025). The Earth alignment principle for artificial intelligence. *Nature Sustainability*, 8, 467–469. https://doi.org/10.1038/s41893-025-01536-6
  - Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F. and Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 12, 518–527. https://doi.org/10.1038/s41558-022-01377-7
  - Dinku, T. (2019). Challenges with availability and quality of climate data in Africa. In A. M. Melesse, W. Abtew and G. Senay (Eds.), *Extreme Hydrology and Climate Variability* (pp. 71–80). Elsevier. https://doi.org/10.1016/B978-0-12-815998-9.00007-5
  - Nakalembe, C. and Kerner, H. (2023). Considerations for AI-EO for agriculture in Sub-Saharan Africa. *Environmental Research Letters*, 18(4), 041002. https://doi.org/10.1088/1748-9326/acc476
  - Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A. and Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870), 80–86. https://doi.org/10.1038/s41586-021-03695-w
  - Li, P., Yang, J., Islam, M. A. and Ren, S. (2025). Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. Communications of the ACM, 68, 54–63. (Preprint: arXiv:2304.03271, 2023.)
- **Sesgo, equidad, privacidad, rendición de cuentas**
  - Chouldechova, A. (2017). Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. https://doi.org/10.1089/big.2016.0047
  - Kleinberg, J., Mullainathan, S. and Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *ITCS 2017*, LIPIcs 67, 43:1–43:23. Preprint arXiv:1609.05807 (2016). https://doi.org/10.4230/LIPIcs.ITCS.2017.43
  - Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119–158.
  - Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
  - Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
  - Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S. and Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *FAT* '19*, 59–68. https://doi.org/10.1145/3287560.3287598
  - Koch, B., Denton, E., Hanna, A. and Foster, J. G. (2021). Reduced, reused and recycled: the life of a dataset in machine learning research. *NeurIPS Datasets and Benchmarks Track*. arXiv:2112.01716. *(source for the dataset concentration figures on your "more is better" slide)*
- **Práctica, marcos y documentación**
  - Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J. and Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), 43. https://doi.org/10.5334/dsj-2020-043
  - Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H. and Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. https://doi.org/10.1145/3458723
  - Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D. and Gebru, T. (2019). Model cards for model reporting. *FAT* '19*, 220–229. https://doi.org/10.1145/3287560.3287596
  - Sloane, M., Moss, E., Awomolo, O. and Forlano, L. (2022). Participation is not a design fix for machine learning. *EAAMO '22*, 1–6. https://doi.org/10.1145/3551624.3555285
  - Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18
  - Kaplan, J. et al. (2020). Scaling laws for neural language models. arXiv:2001.08361. *(for the "what does 5% more accuracy cost" point)*
- **Instrumentos políticos y directrices**
  - Council of Europe (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. CETS No. 225.
  - European Commission High-Level Expert Group on AI (2019). *Ethics Guidelines for Trustworthy AI (*https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai); and (2020) *Assessment List for Trustworthy AI (ALTAI) (*https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment)
  - European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act).
  - IEEE (2019). *Ethically Aligned Design*, 1st edition. https://ethicsinaction.ieee.org

¡GRACIAS!

AI Policy Lab
at Umeå University

SOLUCIONES DEL TEST

- Verdadero o falso: un modelo que cumple la paridad demográfica es justo.
- El sesgo puede entrar en un sistema de IA para el clima a través de
  - (a) la cobertura de los datos de entrenamiento
  - (b) la elección de la variable proxy
  - (c) la delimitación del problema
  - (d) todo lo anterior
- En un sistema de alerta temprana, bajar el umbral de alerta supone sobre todo un compromiso entre:
  - (a) exactitud y cómputo
  - (b) falsas alarmas y eventos no detectados
  - (c) transparencia y privacidad
  - (d) latencia y resolución