IA: LA RESPONSABILITÀ IN UN MONDO CHE CAMBIA

Prof.ssa Virginia Dignum

Cattedra di IA responsabile – Direttrice dell'AI Policy Lab, Università di Umeå

Professoressa ospite SIAS, Università di Amsterdam

E-mail: virginia@cs.umu.se

UMEÅ UNIVERSITY

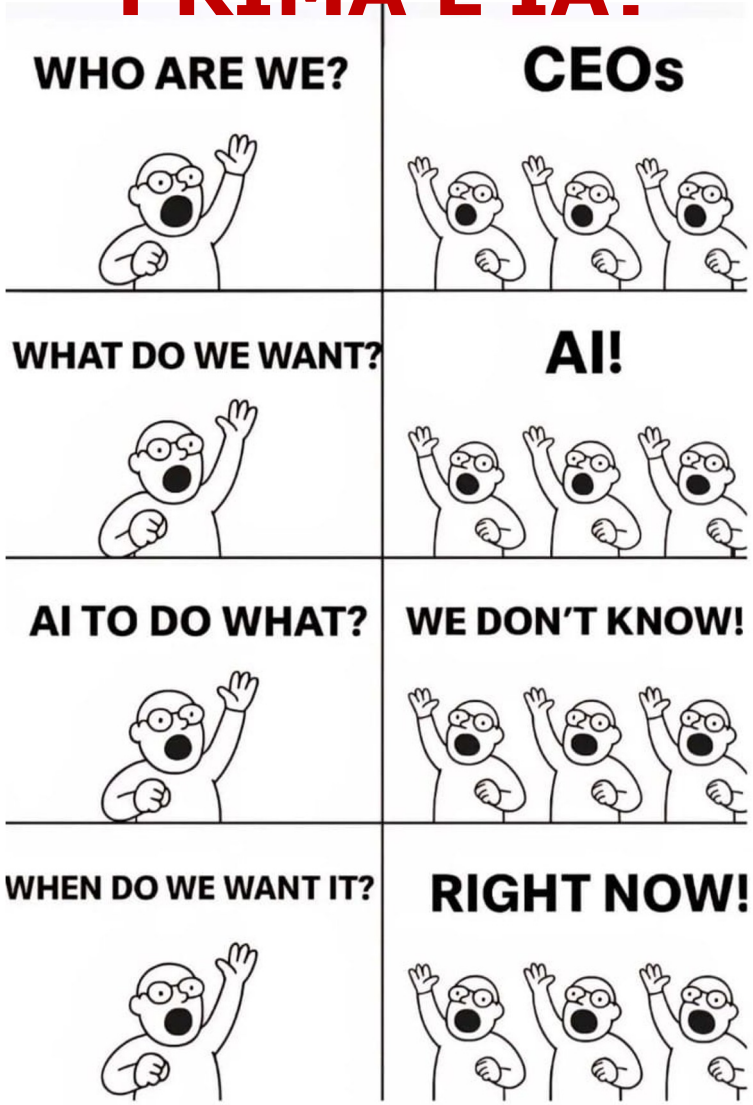
AI Policy Lab at Umeå University

L’IA NEL 2026: LE QUESTIONI APERTE

COSCIENZA?



PRIMA L’IA?



AI Policy Lab
at Umeå University

L'APPROCCIO «PRIMA L'IA»

- Passa DALLA governance ALLA diffusione senza controlli
- Ignora le evidenze di guadagni di produttività limitati grazie all'IA
- **Mancano le domande** «Perché l'IA?» **e** «IA sì o no?»
- Deficit democratico: manca la consultazione degli stakeholder
- Crea pressione ad adottare senza valutare (Dignum et al., 2025)

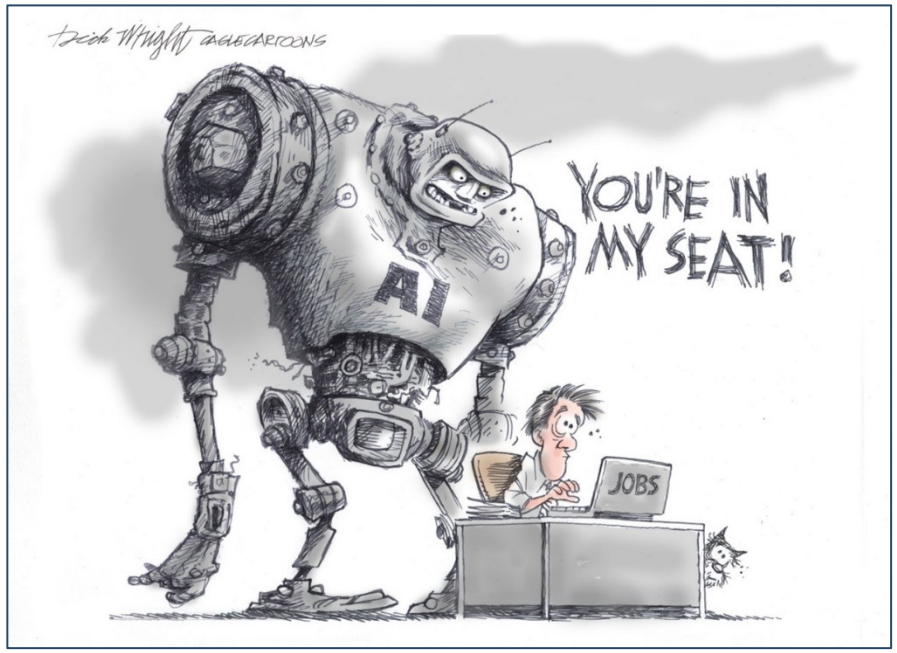
Incentivi che promuovono l'uso dell'IA a prescindere dall'idoneità si ritorceranno contro: comuni e regioni hanno bisogno di competenze per valutare quando e dove l'IA è appropriata, non di strutture che ne spingano l'adozione indiscriminata

Si basa sull'idea sbagliata che la regolamentazione ostacoli i servizi digitali

Adottare l'IA senza investire in R&S nazionale aumenta la dipendenza da pochi fornitori stranieri, minando proprio la sovranità digitale che la strategia dice di promuovere.

La responsabilità segue il controllo. I produttori progettano il sistema; devono portarne la responsabilità principale. Scaricare la verifica sugli utenti sposta la responsabilità senza ridurre il rischio.

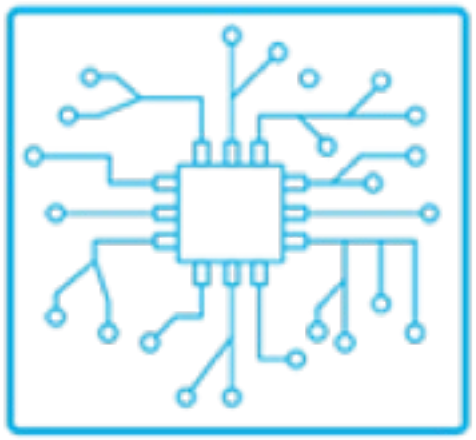
SIAMO NOI A PLASMARE L'IA, O L'IA A PLASMARE NOI?





- L'IA solleva tanti problemi quanti ne risolve
- L'IA non è un mezzo per sostituire le persone
- Serve lo sguardo umano per affrontare le conseguenze etiche, sociali e impreviste dell'IA

IA: PIÙ È MEGLIO?



Computing Power



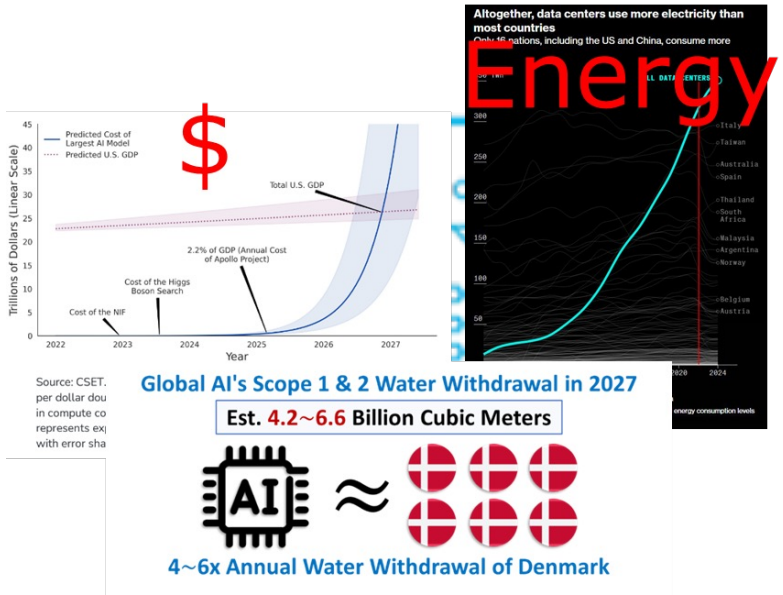
Algorithm Power



Data Availability

AI Policy Lab
at Umeå University

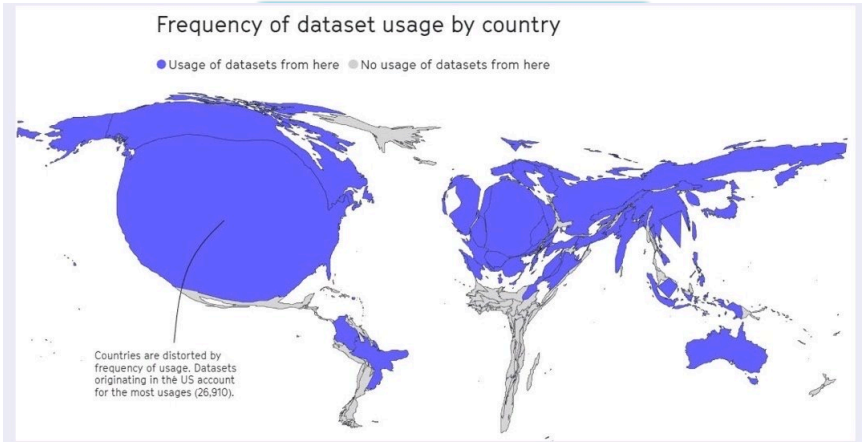
IA: PIÙ È MEGLIO?



Computing Power



Algorithm Power



Data Availability

AI Policy Lab
at Umeå University

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

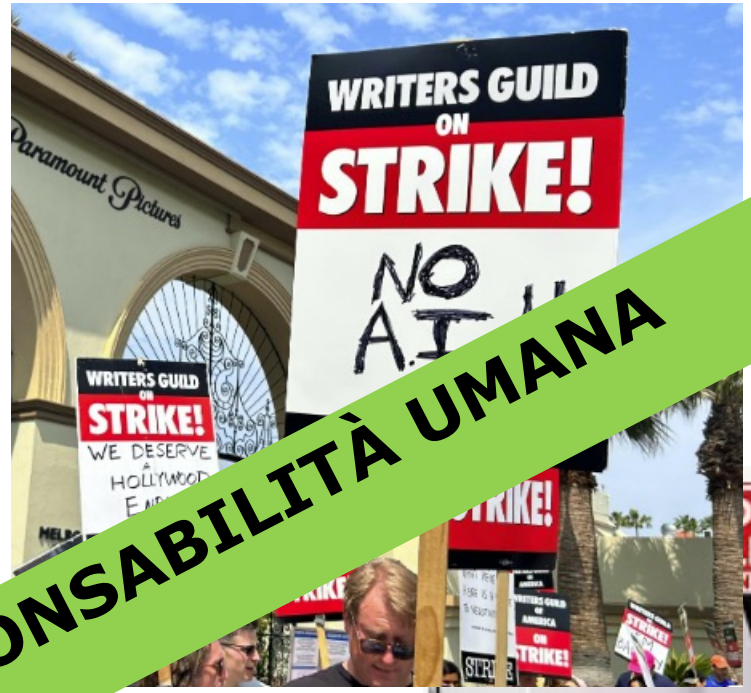
Google exploited homeless black people to develop the Pixel 4's facial recognition AI

Russia Tests New Disinformation Tactics in Africa to Expand Influen

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

Flawed Algori





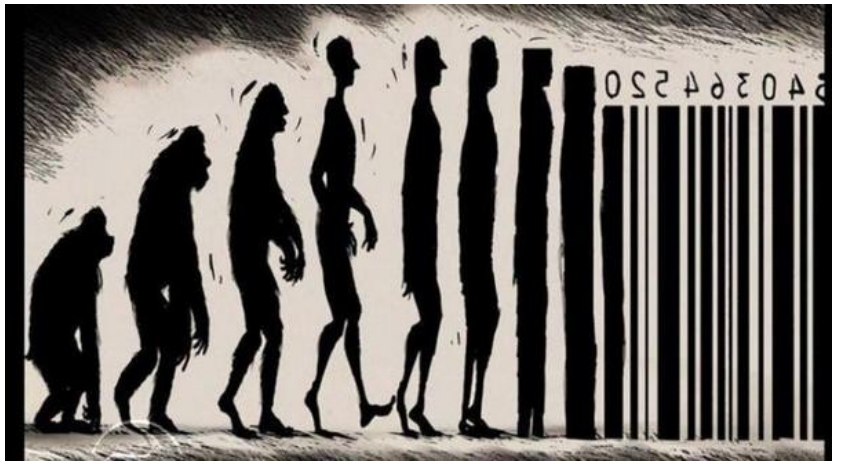




L'IMPATTO DELL'IA RICHIEDE RESPONSABILITÀ UMANA

AI
at Umeå University

IA RESPONSABILE: PERCHÉ IMPORTA?



- Datificazione
  - La realtà è più dei dati
  - I dati sono costruiti
  - I dati sono distorti
  - Ricchi di dati, poveri di comprensione
- Potere
- Sostenibilità

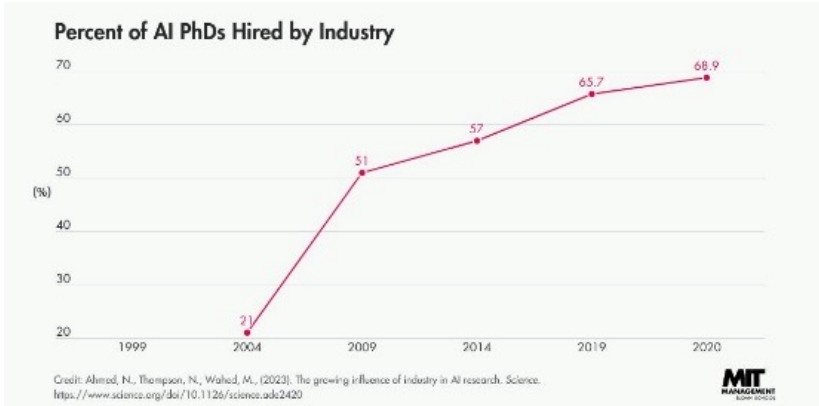


Rivista Time, gen. 2023

IA RESPONSABILE: PERCHÉ IMPORTA?

- Datificazione
- Potere
  - Chi sviluppa l'IA?
  - Chi decide?
- Sostenibilità

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 RESPONSABILE: PERCHÉ IMPORTA?

- Datificazione
- Potere
- Sostenibilità
  - Costo computazionale dell'IA
  - Costi umani e sociali



INDEPENDENT PREMIUM

THE LONGER READ

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



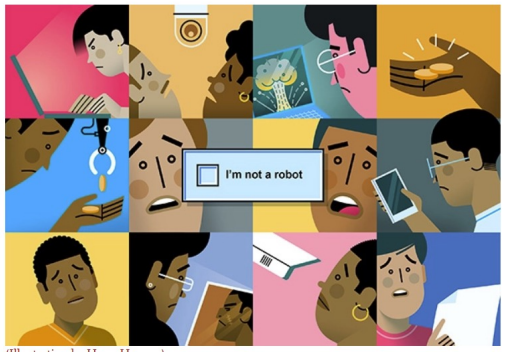
Technology

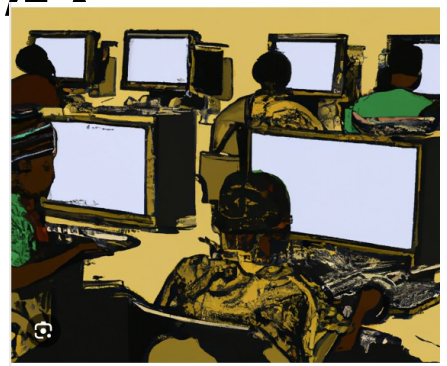
Turk Wars: How AI Threatens the Workers Who Fuel It

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



POSSIAMO ESSERE RESPONSABILI ?

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

L’IA NON È NEUTRALE

- L’IA può promuovere la dignità umana, o minarla
- incarna scelte umane
- Governare l’IA = governare noi stessi



IA – ETICA DELL'IA – IA RESPONSABILE – GOVERNANCE DELL'IA…

- Fase iniziale
  - Prestazioni tecniche ed efficienza dei sistemi di IA
- Crescente consapevolezza
  - Riconoscimento di rischi come bias, discriminazione, violazioni della privacy, mancanza di responsabilità
- Svolta etica
  - Richieste di equità, trasparenza, supervisione umana e spiegabilità
- IA responsabile
  - Passare dai principi alla pratica integrando l'etica in progettazione, implementazione e uso
- Bisogno di governance
  - Garantire fiducia sociale, sostenibilità a lungo termine e allineamento ai valori democratici

DI CHE COSA STIAMO PARLANDO?

- **Bias**: deviazione sistematica, non sinonimo di errore.
  - Distinzione importante: il bias statistico (campionamento, misura) non è il bias sociale (il sistema riproduce uno schema ingiusto).
- **Equità**: una rivendicazione distributiva contesa, non una metrica.
  - I criteri formali (parità, pari opportunità, calibrazione) sono incompatibili tra loro salvo casi degeneri, quindi sceglierne uno è un atto normativo.

(Kleinberg et al. 2016; Chouldechova 2017)

- **La trasparenza** (su processo e provenienza) è diversa dalla spiegabilità (su un output specifico).
  - Sono obblighi diversi, per pubblici diversi.
- **La privacy** riguarda l'integrità contestuale, non solo la conformità alla protezione dei dati

CHE COSA CAMBIA PER IL CLIMA?

- **Scelta e delimitazione del problema**
  - Che cosa conta come problema, e di chi è il problema, si decide prima di raccogliere qualsiasi dato.
  - Es. «prevedere la resa» e «supportare la decisione di un piccolo agricoltore» non sono lo stesso problema.
- **Variabile obiettivo e proxy**
  - L'etichetta raramente è ciò che interessa davvero: le richieste di risarcimento come proxy dell'impatto delle alluvioni, la domanda allacciata alla rete come proxy del fabbisogno energetico.
- **Copertura dei dati**
  - I modelli vengono validati dove i dati esistono già e inferiti dove mancano.
  - In gran parte dell'Africa la copertura delle stazioni meteo

La soluzione non è un modello migliore, ma domande migliori, prima.

IA RESPONSABILE PER L'IA CLIMATICA – ESERCIZIO 1

Scenario: una città commissiona un modello di rischio di alluvione. È addestrato su vent'anni di richieste di risarcimento assicurativo: ogni richiesta registrata indica un luogo come allagato, e i luoghi senza richieste sono considerati non allagati. Il modello usa immagini satellitari, altimetria e precipitazioni, e raggiunge il 91% di accuratezza sui dati di test.

Dove entra per primo il bias?

- La risoluzione delle immagini satellitari
- La scelta delle richieste di risarcimento come etichetta





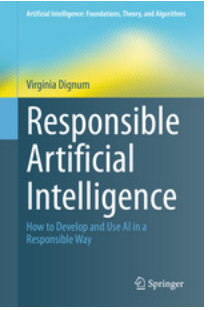
IA RESPONSABILE PER L'IA CLIMATICA – ESERCIZIO 1

Scenario: una città commissiona un modello di rischio di alluvione. È addestrato su vent'anni di richieste di risarcimento assicurativo: ogni richiesta registrata indica un luogo come allagato, e i luoghi senza richieste sono considerati non allagati. Il modello usa immagini satellitari, altimetria e precipitazioni, e raggiunge il 91% di accuratezza sui dati di test.

Dove entra per primo il bias?

- La risoluzione delle immagini satellitari
- La scelta delle richieste di risarcimento come etichetta
- L'architettura del modello

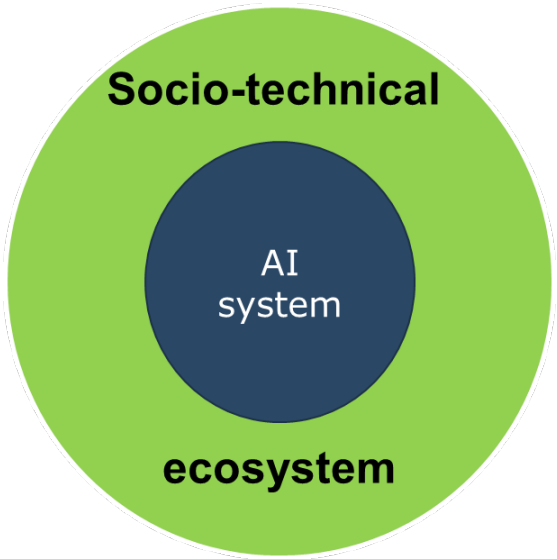
Le richieste registrano chi era assicurato e chi ha presentato domanda, non chi è stato allagato. Le famiglie non assicurate e gli insediamenti informali non sono sottorappresentati: risultano asciutti. Il 91% è misurato rispetto alla stessa etichetta difettosa, quindi non può rilevare il problema. Tutto ciò che segue lo eredita.



RESPONSABILI RIGUARDO ALL'IA

L'IA non ci capita addosso!

- L'IA è progettata. Le scelte le facciamo noi
  - Chi dovrebbe decidere?
  - Quali valori considerare?
  - I valori di chi? Come stabilire le priorità?
- Ponete la Domanda Zero!
  - L'IA è l'opzione migliore qui?
  - Chiedete perché prima di chiedere come
- L'IA non esiste nel vuoto
  - Non esiste una soluzione tecnologica agli effetti negativi!
  - Etica, regolamentazione e governance riguardano l'ecosistema



Le soluzioni di IA responsabile devono essere sociali più che tecniche!

AI Policy Lab
at Umeå University

L’ALTERNATIVA: PRIMA LO SCOPO (DOMANDA ZERO)

L’approccio «prima lo scopo» **chiede: «**Se e dove l’IA è davvero necessaria e utile?»

«Questo sistema di IA andrebbe costruito?»

Principi chiave:

- Partire dai **bisogni della società,** non dalle capacità tecnologiche
- Chiedere sistematicamente: Perché? Per chi? Quali alternative?
- Includere voci diverse nel processo decisionale
- Puntare sulla direzione e sul valore pubblico, non sulla velocità

La Domanda Zero:

- Sposta l’attenzione dalla VELOCITÀ alla **DIREZIONE**
- Dà potere ad attori diversi (PMI, società civile, piccoli paesi)
- Offre un quadro pratico e a basso costo
- Fonda le decisioni su una **reale identificazione del problema** (Dignum et al., 2025)

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

AI Policy Lab at Umeå University

LA DOMANDA ZERO VA OLTRE IL SÌ/NO

Il ciclo della Domanda Zero:

- **Impostazione**: di chi è il problema, chi l'ha definito, qual è il riferimento senza IA, chi sopporta il costo di una risposta sbagliata.
- **Dati**: provenienza, consenso, lacune di copertura, chi è invisibile nei dati, a quali condizioni sono stati ottenuti.
- **Modellazione**: che cosa si ottimizza, che cosa si sacrifica, quale incertezza si comunica.
- **Valutazione**: prestazioni disaggregate, non aggregate; modalità di errore nominate, non solo l'accuratezza.
- **Implementazione**: chi lo gestisce, quali rimedi esistono, che cosa succede quando sbaglia.
- **Monitoraggio e uscita**: chi lo mantiene dopo la fine del

ESERCIZIO 2: L'IA PER PIANIFICARE L'ADATTAMENTO ALLE ALLUVIONI

Scenario: una città vuole introdurre un sistema di IA che classifica i quartieri per finanziare l'adattamento alle alluvioni

- Il sistema prevede il rischio di alluvione da immagini satellitari, dati altimetrici e richieste di risarcimento passate, e classifica le aree per gli investimenti.
- Il sistema stima la vulnerabilità delle famiglie combinando telerilevamento con dati di telefonia mobile e dei servizi pubblici.
- Il sistema genera sintesi e raccomandazioni per il consiglio comunale, che ha tre settimane per decidere.

DOMANDA ZERO – ESERCIZIO 2

- L’IA è adatta a questo scopo?

- Opzioni
  - A) Sì, l’alternativa è un processo manuale più lento e altrettanto distorto.
  - B) Sì, l’alternativa è un processo manuale più lento e altrettanto distorto.
  - C) No, i dati non possono sostenere la decisione per cui vengono usati.
  - D) No, i dati non possono sostenere la decisione per cui vengono usati.





LA PROGETTAZIONE È POLITICA

- Scelte
- Formulazione
- Informazione
- Coinvolgimento
- Legittimità
- Aggregazione







CHE COSA VOGLIAMO CHE SIA L'IA?

CHE COS'È L'IA?

- Simile all'uomo?
  - Perché?
  - Che cosa significa?

- Strumento?
  - Per che cosa? Per chi?

- Simulazione o operatività?
  - Capire l'intelligenza costruendo intelligenza, oppure
  - Intervento attivo nel mondo reale

- Normativa o descrittiva?
  - Fate come diciamo o come facciamo?



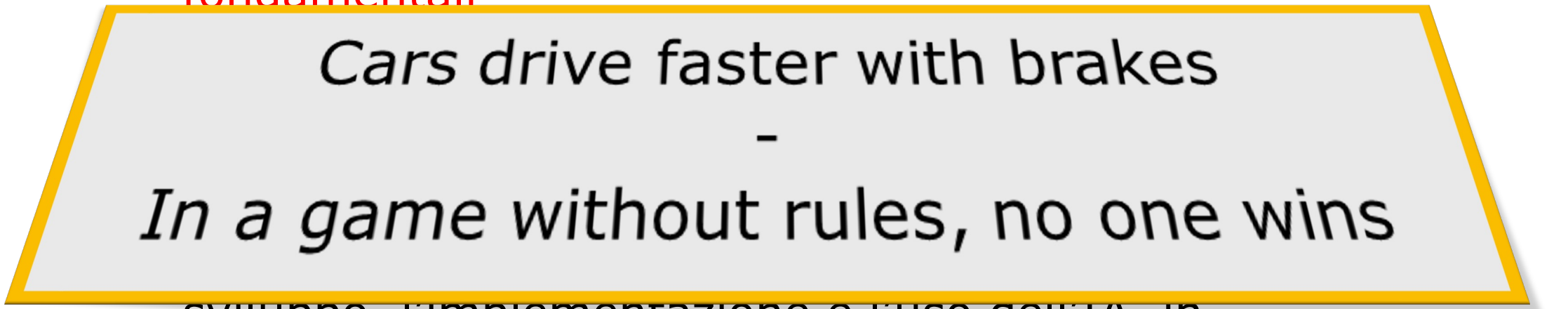
PRINCIPI, LINEE GUIDA, ETICA

AI Policy Lab
at Umeå University

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

GOVERNANCE: PERCHÉ? PER CHE COSA?

- La regolamentazione come incentivo all'innovazione responsabile, alla sostenibilità e ai diritti umani fondamentali



  - sviluppo, l'implementazione e l'uso dell'IA, in particolare dei modelli di IA generativa con rischi diversi
- Esigenze di responsabilità, rendicontazione e governance
  - Regolare gli attori organizzativi più che i risultati tecnologici
  - Fiducia pubblica e responsabilità per gli errori delle

AI Policy Lab
at Umeå University

PRINCIPI E LINEE GUIDA

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?

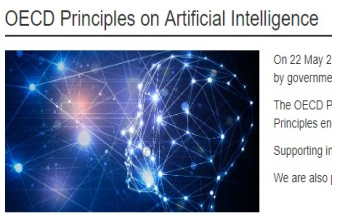
Livello	Quadro e portata
Globale	UNESCO (194 paesi), OCSE (>70 giurisdizioni), GPAI (25+ membri), trattato del Consiglio d'Europa (50+ paesi), principi del G7
Regionale	AI Act dell'UE (27 Stati membri), Dichiarazione di Santiago (America Latina/Caraibi)
Nazionale	930+ iniziative in 71 paesi
Altri	200+ linee guida di ONG, universi… privati



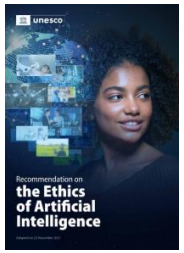
https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence



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

Ben oltre 1.000 pubblicate nel mondo!

AI Policy Lab
at Umeå University

TUTELARE LA DEMOCRAZIA

- Rafforzare regolamentazione e vigilanza
  - Adottare leggi chiare su IA, riconoscimento facciale, criptovalute e spyware.
  - Migliorare trasparenza e responsabilità nell'uso delle tecnologie.
  - Rafforzare le istituzioni democratiche per vigilare e orientare lo sviluppo tecnologico.
- Promuovere la partecipazione pubblica
  - Democratizzare lo sviluppo dell'IA coinvolgendo attori sociali diversi.
  - Allineare i sistemi tecnologici ai valori democratici e all'interesse pubblico.
- Migliorare l'alfabetizzazione digitale
  - Promuovere con l'istruzione una comprensione critica dell'IA e delle tecnologie digitali.
  - Mettere i cittadini in grado di difendere i propri diritti

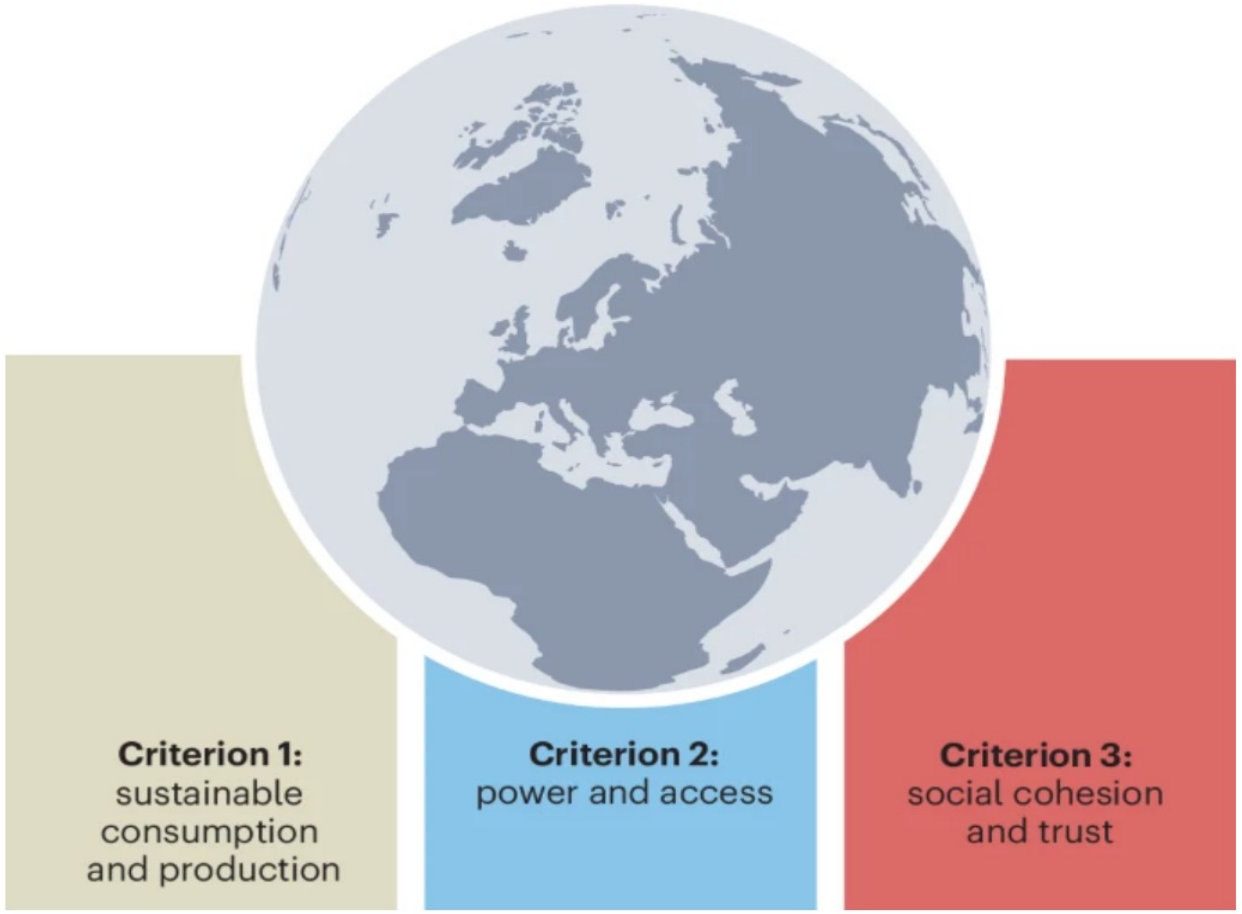
TUTELARE IL PIANETA

L’impatto dell’IA richiede una custodia sistemica, oltre agli impatti ambientali:

1. contribuire ad accelerare la **transizione verso produzione** e consumo sostenibili nel rispetto dei limiti planetari, o almeno senza ostacolare questi obiettivi

2. essere sviluppata, implementata e usata in modo da **garantire un accesso equo** agli strumenti di IA per la sostenibilità globale ed evitare concentrazioni di potere

3. essere sviluppata, implementata e usata per **sostenere una maggiore**



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

MISURARE L'IMPATTO:
CHE COSA CONTA, E CHI NE RISPONDE

Governiamo ciò che contiamo. Gran parte dell'impatto dell'IA non viene contata.

Tre livelli di impatto

(Kaack et al., 2022)

- Calcolo
- Applicazione
- Sistema

La conseguenza per la governance

- La responsabilità segue la misurazione
  - il calcolo ha un proprietario, gli effetti di sistema spesso nessuno
- Ciò che si rendiconta diventa ciò per cui si progetta
  - Scegliere la metrica è un atto politico, non tecnico
- Allineamento con la Terra
  - limiti planetari, accesso equo, coesione sociale.
- La Domanda Zero vale anche per le metriche
  - concordare che cosa conta, e chi lo verifica, prima di costruire

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

GIUSTIZIA CLIMATICA

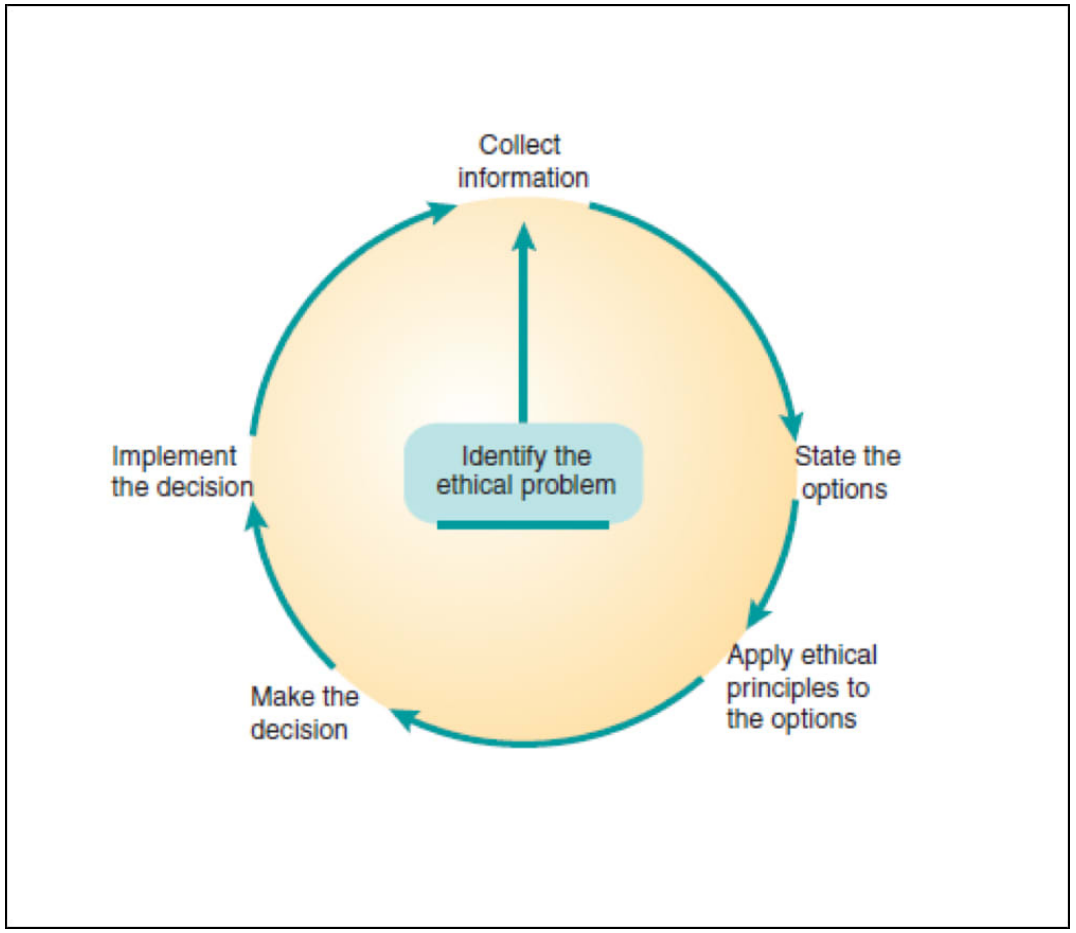
- distributiva (chi ottiene il beneficio e chi sostiene i costi di calcolo, acqua, terra e lavoro),
- procedurale (chi era nella stanza),
- di riconoscimento (quali conoscenze contano come evidenza).

- Il fine giustifica i mezzi? Compromessi?
  - Es. migliorare l'estrazione di combustibili fossili è un uso responsabile dell'IA?
  - Quali sono le alternative? Chi ci guadagna? Chi ci perde?

ATTUAZIONE: REGOLAMENTAZIONE E ALTRO

- Regolamentazione
  - AI Act: approccio antropocentrico e basato sul rischio
- Standard
  - soft governance; adesione non obbligatoria
  - dimostrare la due diligence e limitare la responsabilità
  - integrazione agevole tra prodotti
- Strutture e procedure organizzative
  - Comitati consultivi e responsabili per l'etica
  - Definire e monitorare linee guida etiche
- Valutazione di un'IA affidabile
  - l'IA responsabile è più che spuntare caselle
  - Servono strumenti per valutare la maturità
- Consapevolezza e partecipazione
  - Istruzione e formazione

RAGIONAMENTO ETICO



- Non si tratta della risposta, ma di riconoscere il problema
- Non c'è una sola risposta! Nessuna soluzione miracolosa!
  - Prospettiva utilitarista
  - Prospettiva dei diritti
  - Prospettiva della giustizia
  - Prospettiva del bene comune
  - Prospettiva delle virtù
  - ...

L’ETICA NELLA VITA REALE

- Equità
  - Assegno per i figli
  - Congedo parentale

- Sicurezza
  - Autovelox
  - Lezioni di educazione stradale per bambini

- Equità o sicurezza
  - Risorse limitate

- equità

- uguaglianza

– breve termine

– lungo termine

DILEMMI – ESERCIZIO 3

- **Contributi per l'adattamento climatico**: vi è stato chiesto di sviluppare un sistema di IA per vagliare le domande a un fondo comunale per l'adattamento climatico, assegnando i contributi in base alla probabilità di ogni domanda di generare un «alto impatto climatico per euro», usando i dati dei progetti finanziati in passato. Ciò sostituisce l'esame manuale e i sopralluoghi dei funzionari.
- Compromessi: come ripartire le risorse:
  - Quanto costo (computazionale) vale un 5% di accuratezza in più?
  - Quanto costo (computazionale) vale includere una spiegazione?

Join at
slido.com
#4275 548

ETICA DELL'IA

- L'IA è così diversa dalle altre tecnologie?
  - Sovrapposizioni con l'etica della tecnologia, degli affari, l'etica medica…

- Alcune sfide specifiche dell'IA, legate a:
  - Autonomia. Se il controllo esecutivo è delegato, chi è responsabile delle azioni? Come mantenere il rispetto dei valori umani nelle decisioni autonome?
  - Uso in ambienti aperti. Se non possiamo prevedere il contesto che il sistema incontrerà, come garantiamo un comportamento etico?
  - Generalizzazione. Se una decisione viene presa generalizzando dai dati, che ne è del nostro diritto a essere visti come individui? Come evitare la riproduzione e l'amplificazione dei bias?

IA RESPONSABILE: PIÙ CHE ETICA

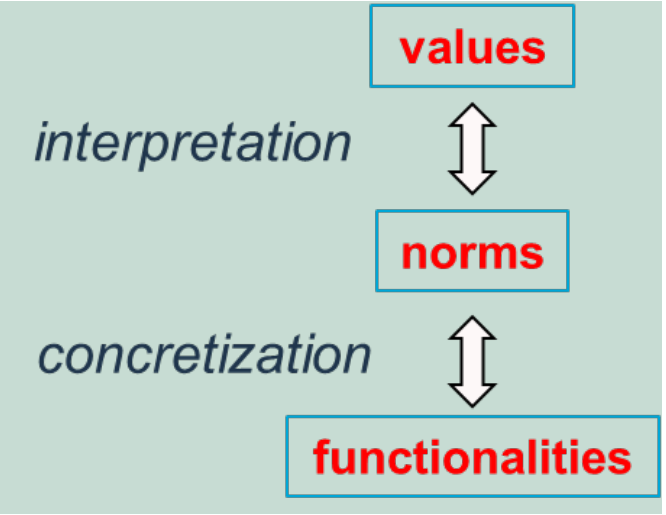
- Non filosofeggiare sull'etica
  - L'etica non riguarda la risposta, ma il riconoscimento del problema
  - L'etica è un processo (sociale), non una soluzione

- Non una tecnicizzazione dell'etica
  - La vostra implementazione non «risolve» l'etica

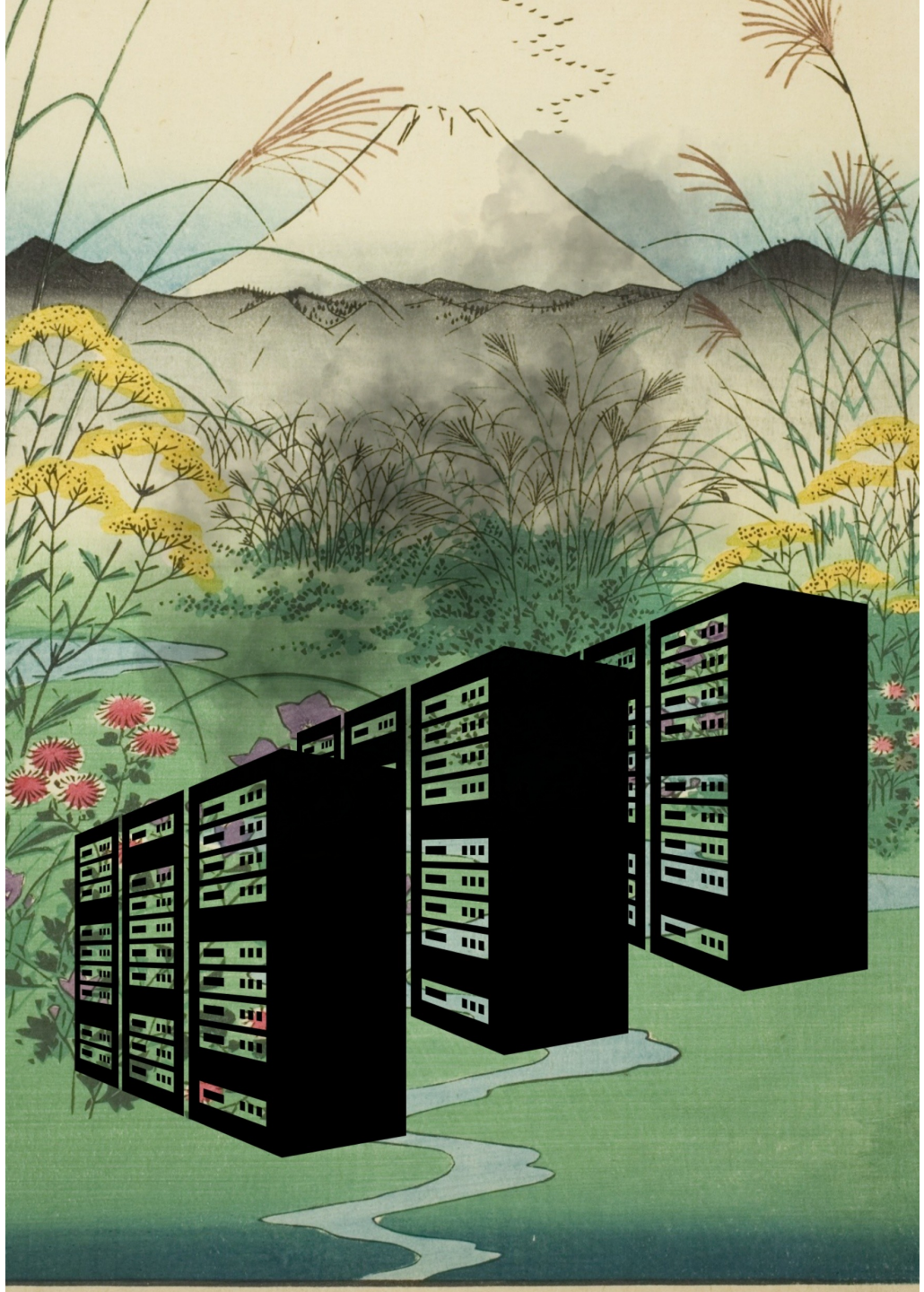
- In sostanza, si tratta di scelte, priorità, compromessi
  - Accuratezza / spiegazione
  - Accuratezza / risorse di calcolo
  - Sicurezza / privacy
  - Equità / uguaglianza
  - Beneficio a lungo termine / breve termine
  - ...

L'IA RESPONSABILE SIGNIFICA ESSERE ESPLICITI

- Progettazione guidata dai valori
  - Gli aspetti giuridici ed etici non sono un'aggiunta!
- Governance
  - Monitoraggio e controllo esterni
  - Accordi, contratti, norme
- Progettazione
  - Mettete in discussione opzioni e scelte
  - Motivate le vostre scelte
  - Documentate scelte e opzioni



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



UN NUOVO PARADIGMA DELL'IA

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

UN NUOVO PARADIGMA DELL'IA?

- L'IA non è una soluzione universale
  - riconoscerne i limiti di fronte a sfide complesse
  - Domanda zero: l'IA va usata qui?
- Affrontare i rischi intrinseci di bias e discriminazione
  - migliorare la trasparenza di dati e algoritmi
  - tenere conto di contesti e valori diversi
- Oltre le discipline
  - integrare competenze sociali e tecniche nella progettazione dell'IA
  - affrontare sfide sociali sistemiche
  - coinvolgere attori sociali diversi, non solo tecnologi
- standard tecnici solidi
  - Verificabile, sostenibile, partecipativo

APPROCCIO ALL'IA RESPONSABILE

Impegnarsi su principi di IA responsabile e affidabile, per un'IA innovativa, etica e rispettosa dei diritti umani

Riconoscere le possibili tensioni tra principi etici e la necessità di stabilire priorità

Sviluppare una cultura della responsabilità nelle organizzazioni

Dare priorità all'interazione e alla comunicazione con gli stakeholder

UNA GOVERNANCE AFFIDABILE

- Sviluppo dell'IA
  - Che cosa? Requisiti, bisogni, possibilità, capacità
- Uso dell'IA
  - Come? Licenza a operare
- Contesto di sviluppo e uso dell'IA
  - Dove/quando? Regole del gioco

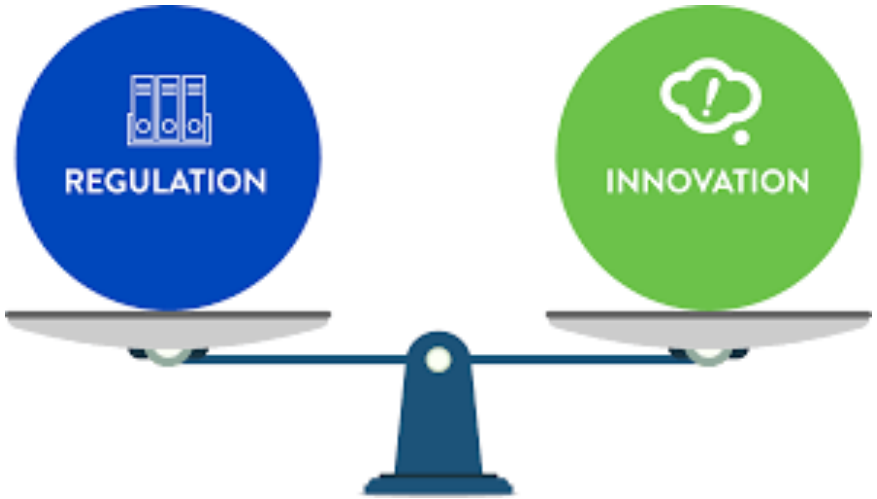
- Sforzi globali
  - Sostenibilità
  - Inclusione e partecipazione
  - Diversità
  - Distribuzione di benefici e costi
  - Agenda 2030 – SDG



L’IA RESPONSABILE È INNOVAZIONE

- Innovazione necessaria
  - Requisiti tecnici
  - Requisiti d’uso
  - Contesto operativo

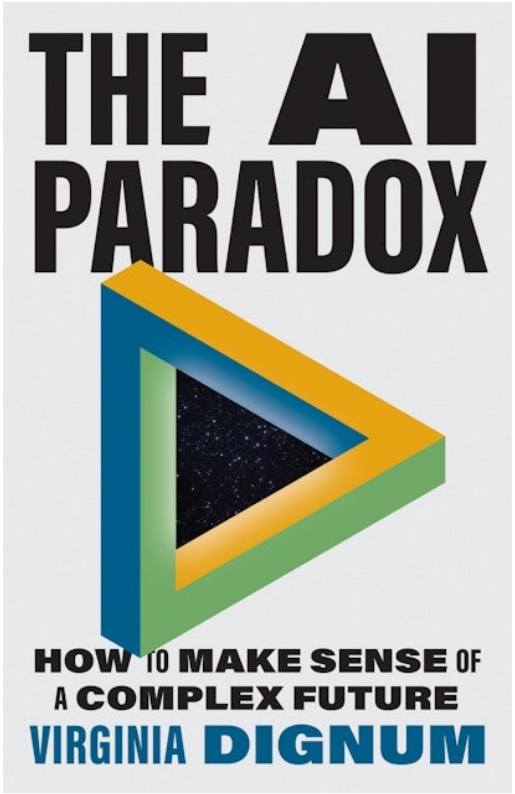
- Innovazione multidisciplinare!
  - Innovazione tecnologica
  - Innovazione organizzativa
  - Innovazione normativa
  - Innovazione nella governance
  - Innovazione sociale



L'IA non è una forza inarrestabile; è plasmata dalle nostre scelte

- le decisioni che prendiamo oggi determineranno se diventerà uno strumento di emancipazione o una fonte di controllo e disuguaglianza.

Il grande paradosso dell'IA: più diventa intelligente, più abbiamo bisogno di saggezza umana



In uscita a febbraio 2026

QUIZ

- Vero o falso: un modello che soddisfa la parità demografica è equo.
- Il bias può entrare in un sistema di IA per il clima attraverso
  - (a) la copertura dei dati di addestramento
  - (b) la scelta della variabile proxy
  - (c) la delimitazione del problema
  - (d) tutte le precedenti
- In un sistema di allerta precoce, abbassare la soglia di allerta comporta soprattutto un compromesso tra:
  - (a) accuratezza e calcolo
  - (b) falsi allarmi ed eventi mancati
  - (c) trasparenza e privacy
  - (d) latenza e risoluzione

LETTURE CONSIGLIATE

- Dignum, *Responsible Artificial Intelligence* (Springer, 2019)
- Policy brief sulla Domanda Zero dell'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)

ALTRE LETTURE

- **Il mio lavoro e i materiali sulla Domanda Zero**
  - 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 e 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.)
- **Bias, equità, privacy, responsabilità**
  - 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)*
- **Pratica, quadri e documentazione**
  - 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)*
- **Strumenti politici e linee guida**
  - 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

GRAZIE!

AI Policy Lab
at Umeå University

SOLUZIONI DEL QUIZ

- Vero o falso: un modello che soddisfa la parità demografica è equo.
- Il bias può entrare in un sistema di IA per il clima attraverso
  - (a) la copertura dei dati di addestramento
  - (b) la scelta della variabile proxy
  - (c) la delimitazione del problema
  - (d) tutte le precedenti
- In un sistema di allerta precoce, abbassare la soglia di allerta comporta soprattutto un compromesso tra:
  - (a) accuratezza e calcolo
  - (b) falsi allarmi ed eventi mancati
  - (c) trasparenza e privacy
  - (d) latenza e risoluzione