IA: RESPONSABILIDADE EM UM MUNDO EM TRANSFORMAÇÃO

Profa. Dra. Virginia Dignum

Cátedra de IA Responsável – Diretora do AI Policy Lab, Universidade de Umeå

Professora visitante SIAS, Universidade de Amsterdã

E-mail: virginia@cs.umu.se

UMEÅ UNIVERSITY

AI Policy Lab at Umeå University

IA EM 2026: AS QUESTÕES

CONSCIÊNCIA?



IA EM PRIMEIRO LUGAR?



AI Policy Lab
at Umeå University

A ABORDAGEM "IA EM PRIMEIRO LUGAR"

- Passa DA governança PARA a implantação sem controle
- Ignora evidências de ganhos de produtividade limitados com a IA
- **Faltam as perguntas** "Por que IA?" **e** "IA ou não?"
- Déficit democrático: falta consulta às partes interessadas
- Cria pressão para adotar sem avaliar (Dignum et al., 2025)

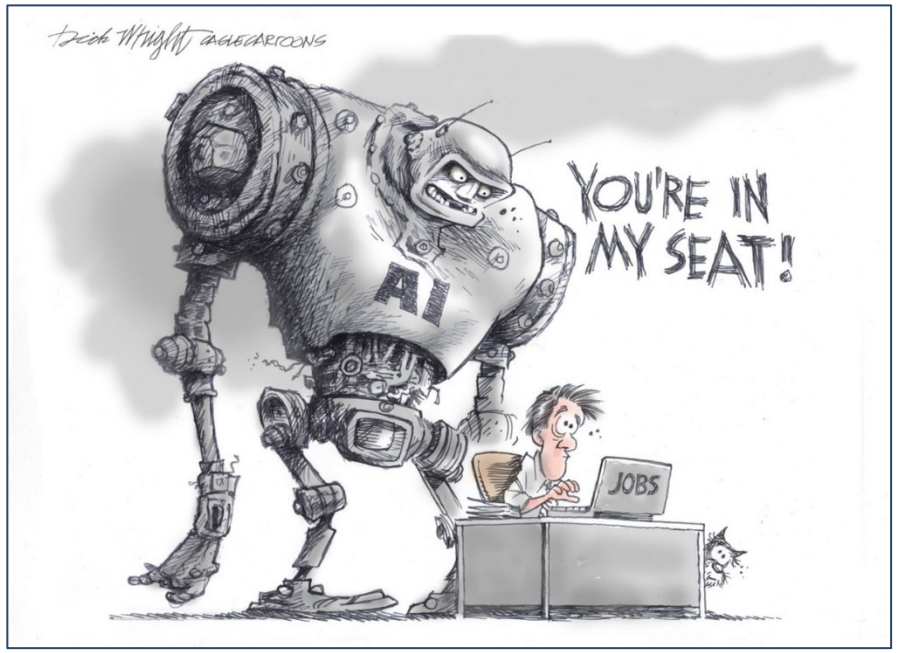
Incentivos que promovem o uso da IA sem considerar sua adequação vão sair pela culatra: municípios e regiões precisam de competência para avaliar quando e onde a IA é apropriada, não de estruturas que impulsionem a adoção indiscriminada

Baseia-se na ideia equivocada de que a regulação atrapalha os serviços digitais

Adotar IA sem investir em P&D nacional aprofunda a dependência de poucos fornecedores estrangeiros, minando a própria soberania digital que a estratégia diz promover.

A responsabilização acompanha o controle. Os produtores projetam o sistema; devem arcar com a responsabilidade principal. Transferir a verificação aos usuários só realoca a responsabilidade sem reduzir o risco.

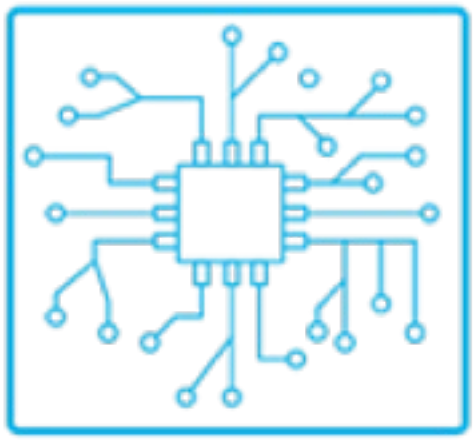
ESTAMOS MOLDANDO A IA, OU A IA ESTÁ NOS MOLDANDO?





- A IA levanta tantos problemas quanto resolve
- A IA não é um meio de substituir pessoas
- É preciso o olhar humano para lidar com as consequências éticas, sociais e não intencionais da IA

IA: QUANTO MAIS, MELHOR?



Computing Power



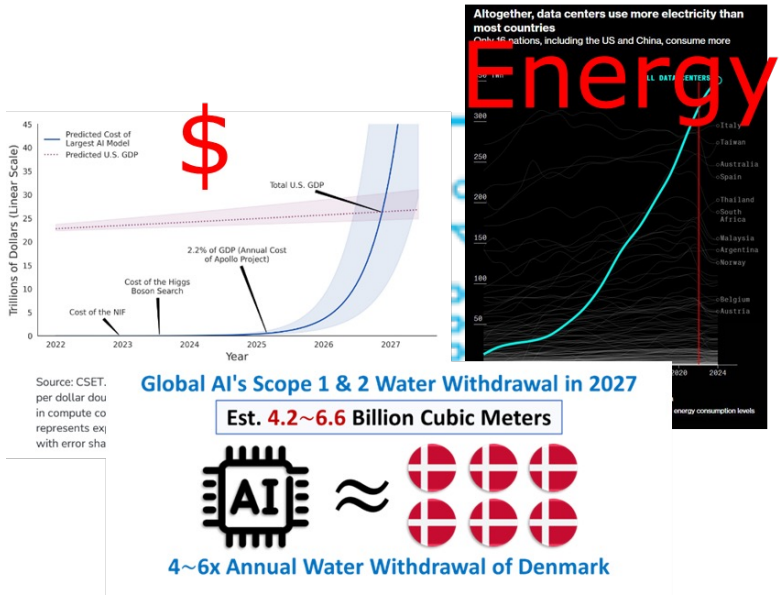
Algorithm Power



Data Availability

AI Policy Lab
at Umeå University

IA: QUANTO MAIS, MELHOR?



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

Est. 4.2~6.6 Billion Cubic Meters

AI ≈

4~6x Annual Water Withdrawal of Denmark

Computing Power



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

Russia Tests New Disinfor… Tactics in Africa to Expand…

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

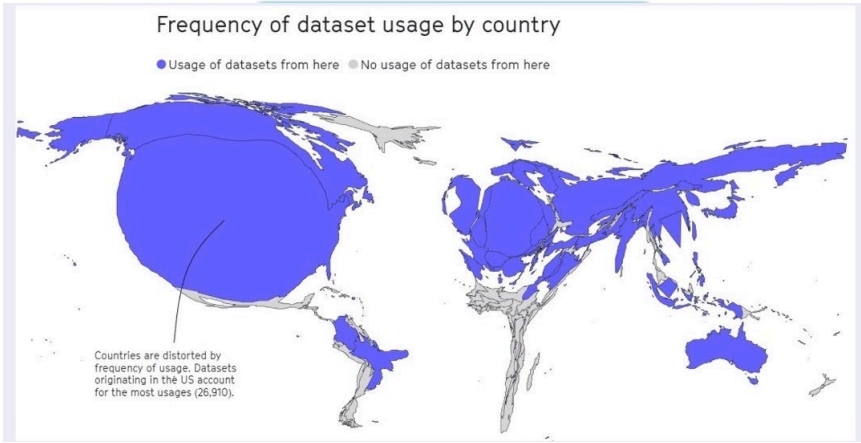
Flawed Algorithms Are… Millions of Students' Es…

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

Indigenous elder slams 'hollow a… tokenistic' consultation by Sidew…



Algorithm Power



Data Availability

AI Policy Lab
at Umeå University

O IMPACTO DA IA EXIGE RESPONSABILIDADE 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 bi 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





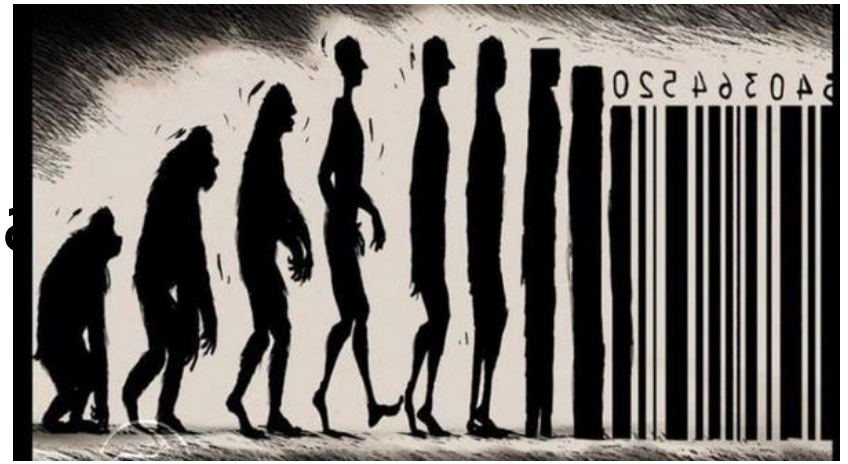




AI
at Umeå University

IA RESPONSÁVEL: POR QUE SE IMPORTAR?

- Dataficação
  - A realidade é mais do que da
  - Os dados são construídos
  - Os dados são enviesados
  - Ricos em dados, pobres em compreensão



- Poder
- Sustentabilidade

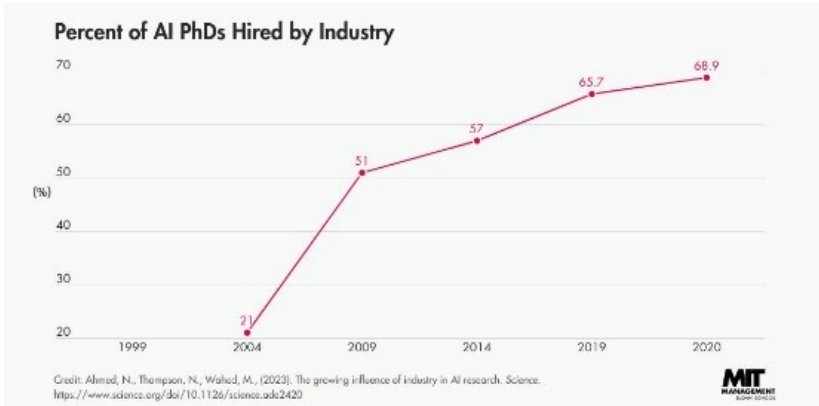


Revista Time, jan. de 2023

IA RESPONSÁVEL: POR QUE SE IMPORTAR?

- Datificação
- Poder
  - Quem desenvolve a IA?
  - Quem decide?
- Sustentabilidade

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**

IA RESPONSÁVEL: POR QUE SE IMPORTAR?

- Datificação
- Poder
- Sustentabilidade
  - Custo computacional da IA
  - Custos humanos e sociais



INDEPENDENT PREMIUM

THE LONGER READ

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



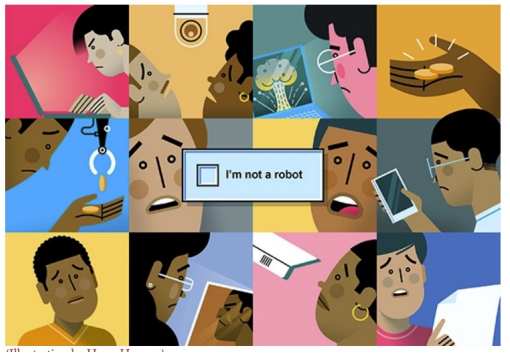
Technology

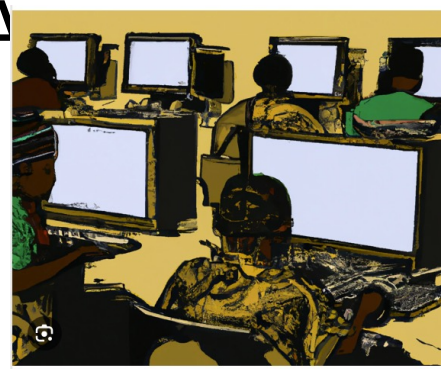
Turk Wars: How AI Threatens the Workers Who Fuel It

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 RESPONSÁVEIS ?

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

A IA NÃO É NEUTRA

- A IA pode promover a dignidade humana, ou miná-la
- materializa escolhas humanas
- Governar a IA = governar a nós mesmos



IA – ÉTICA DA IA – IA RESPONSÁVEL – GOVERNANÇA DA IA...

- Foco inicial
  - Desempenho técnico e eficiência dos sistemas de IA
- Consciência crescente
  - Reconhecimento de riscos como viés, discriminação, violações de privacidade, falta de responsabilização
- Virada para a ética
  - Pedidos de equidade, transparência, supervisão humana e explicabilidade
- IA responsável
  - Passar dos princípios à prática incorporando a ética ao projeto, à implantação e ao uso
- Necessidade de governança
  - Garantir a confiança da sociedade, a sustentabilidade de longo prazo e o alinhamento com valores democráticos

DO QUE ESTAMOS FALANDO?

- **Viés**: desvio sistemático, não sinônimo de erro.
  - Distinção importante: o viés estatístico (amostragem, medição) não é o mesmo que o viés social (o sistema reproduz um padrão injusto).
- **Equidade**: uma reivindicação distributiva contestada, não uma métrica.
  - Os critérios formais (paridade, igualdade de chances, calibração) são incompatíveis entre si, salvo em casos degenerados; escolher um é um ato normativo.

(Kleinberg et al. 2016; Chouldechova 2017)

- **A transparência** (sobre processo e procedência) é diferente da explicabilidade (sobre um resultado específico).
  - São obrigações diferentes, com públicos diferentes.
- **A privacidade** envolve integridade contextual, não só

QUE DIFERENÇA ISSO FAZ PARA O CLIMA?

- **Escolha e delimitação do problema**
  - O que conta como problema, e de quem é o problema, é decidido antes da coleta de qualquer dado.
  - P. ex. “prever a produtividade” e “apoiar a decisão de um pequeno agricultor” não são o mesmo problema.
- **Variável-alvo e proxies**
  - O rótulo raramente é o que realmente importa: pedidos de indenização como proxy do impacto de inundações, demanda conectada à rede como proxy da necessidade de energia.
- **Cobertura de dados**
  - Os modelos são validados onde já existem dados e interessados onde não existem.
  - Em grande parte da África a cobertura de estações

A solução não é um modelo melhor, e sim perguntas melhores, mais cedo.

AI Policy Lab
at Umeå University

IA RESPONSÁVEL PARA IA CLIMÁTICA – EXERCÍCIO 1

Cenário: uma cidade encomenda um modelo de risco de inundação. Ele é treinado com vinte anos de pedidos de indenização de seguros: cada pedido registrado marca um local como inundado, e locais sem pedido são tratados como não inundados. O modelo usa imagens de satélite, altitude e chuva, e atinge 91% de acurácia em dados de teste.

Onde o viés entra primeiro?

- A resolução das imagens de satélite
- A escolha dos pedidos de indenização como rótulo
- A arquitetura do modelo



Join at
slido.com
#4275 548

IA RESPONSÁVEL PARA IA CLIMÁTICA – EXERCÍCIO 1

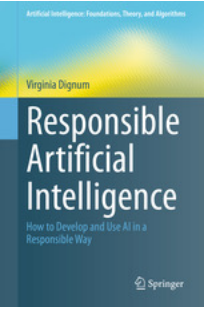
Cenário: uma cidade encomenda um modelo de risco de inundação. Ele é treinado com vinte anos de pedidos de indenização de seguros: cada pedido registrado marca um local como inundado, e locais sem pedido são tratados como não inundados. O modelo usa imagens de satélite, altitude e chuva, e atinge 91% de acurácia em dados de teste.

Onde o viés entra primeiro?

- A resolução das imagens de satélite
- A escolha dos pedidos de indenização como rótulo
- A arquitetura do modelo

Os pedidos registram quem tinha seguro e quem acionou, não quem foi inundado. Domicílios sem seguro e assentamentos informais não estão sub-representados: são registrados como secos. Os 91% são medidos contra o mesmo rótulo falho, então não conseguem detectar o problema. Tudo o que vem depois herda isso.

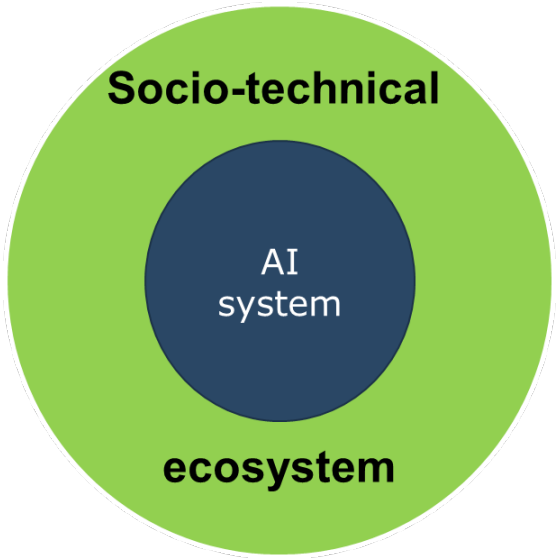
AI Policy Lab
at Umeå University



RESPONSÁVEIS SOBRE A IA

A IA não simplesmente acontece com a gente!

- A IA é projetada. Nós fazemos as escolhas
  - Quem deve decidir?
  - Quais valores devem ser considerados?
  - Os valores de quem? Como priorizar?
- Faça a Pergunta Zero!
  - A IA é a melhor opção aqui?
  - Pergunte por que antes de perguntar como
- A IA não existe no vácuo
  - Não há solução tecnológica para efeitos nocivos!
  - Ética, regulação e governança dizem respeito ao ecossistema



Soluções de IA responsável precisam ser mais sociais do que técnicas!

AI Policy Lab
at Umeå University

A ALTERNATIVA: PROPÓSITO EM PRIMEIRO LUGAR (PERGUNTA ZERO)

A abordagem "propósito em primeiro lugar" **pergunta: "**Se e onde a IA é de fato necessária e benéfica?"

"Este sistema de IA deveria sequer ser construído?"

Princípios-chave:

- Partir das **necessidades da sociedade**, não das capacidades tecnológicas
- Perguntar sistematicamente: Por quê? Para quem? Quais são as alternativas?
- Incluir vozes diversas na tomada de decisão
- Foco na direção e no valor público, não na velocidade

A Pergunta Zero:

- Desloca o foco da VELOCIDADE para a **DIREÇÃO**
- Fortalece atores diversos (PMEs, sociedade civil, países pequenos)
- Oferece um arcabouço prático e de baixo custo
- Ancora as decisões em uma **identificação real do 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

A PERGUNTA ZERO VAI ALÉM DO SIM/NÃO

O ciclo da Pergunta Zero:

- **Enquadramento**: de quem é o problema, quem o definiu, qual é a referência sem IA, quem arca com o custo de uma resposta errada.
- **Dados**: procedência, consentimento, lacunas de cobertura, quem é invisível nos dados, em que termos foram obtidos.
- **Modelagem**: o que é otimizado, o que é sacrificado, que incerteza é comunicada.
- **Avaliação**: desempenho desagregado, não agregado; modos de falha nomeados, não só a acurácia.
- **Implantação**: quem o opera, que recursos existem, o

EXERCÍCIO 2: IA PARA O PLANEJAMENTO DA ADAPTAÇÃO A INUNDAÇÕES

Cenário: uma cidade quer introduzir um sistema de IA que classifique bairros para financiar a adaptação a inundações

- O sistema prevê o risco de inundação a partir de imagens de satélite, dados de altitude e pedidos de indenização anteriores, e classifica as áreas para investimento.
- O sistema estima a vulnerabilidade dos domicílios combinando sensoriamento remoto com dados de telefonia móvel e de serviços públicos.
- O sistema gera resumos e recomendações para a câmara municipal, que tem três semanas para decidir.

PERGUNTA ZERO – EXERCÍCIO 2

- A IA é adequada para esse fim?

- Opções
  - A) Sim, a alternativa é um processo manual mais lento e igualmente enviesado.
  - B) Sim, a alternativa é um processo manual mais lento e igualmente enviesado.
  - C) Não, os dados não sustentam a decisão para a qual estão sendo usados.
  - D) Não, os dados não sustentam a decisão para a qual estão sendo usados.



Join at
slido.com
#4275 548

O PROJETO É POLÍTICO

- Escolhas
- Formulação
- Informação
- Envolvimento
- Legitimidade
- Agregação







O QUE É IA?

O QUE QUEREMOS QUE A IA SEJA?

- Semelhante ao humano?
  - Por quê?
  - O que isso significa?

- Ferramenta?
  - Para quê? Para quem?

- Simulação ou operação?
  - Entender a inteligência construindo inteligência, ou
  - Intervenção ativa no mundo real

- Normativa ou descritiva?
  - Faça o que dizemos ou faça o que fazemos?



PRINCÍPIOS, DIRETRIZES, ÉTICA

AI Policy Lab
at Umeå University

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

GOVERNANÇA: POR QUÊ? PARA QUÊ?

- A regulação como incentivo à inovação responsável, à sustentabilidade e aos direitos humanos fundamentais
  - um poderoso trampolim para a inovação com benefícios

Cars drive faster with brakes

-

In a game without rules, no one wins

especialmente modelos de IA generativa com riscos variados

- Exigências de responsabilidade, responsabilização e governança
  - Regular os atores organizacionais em vez dos resultados tecnológicos
  - Confiança pública e responsabilização por erros na tomada de decisão automatizada, independentemente da

AI Policy Lab
at Umeå University

PRINCÍPIOS E DIRETRIZES

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?

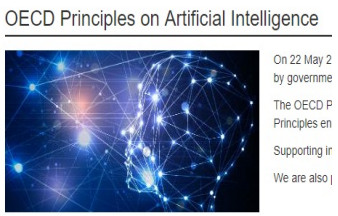
Nível	Marco e alcance
Global	UNESCO (194 países), OCDE (>70 jurisdições), GPAI (25+ membros), tratado do Conselho da Europa (50+ países), princípios do G7
Regional	AI Act da UE (27 Estados-membros), Declaração de Santiago (América Latina/Caribe)
Nacional	930+ iniciativas em 71 países
Outros	200+ diretrizes de ONGs, academia e privadas



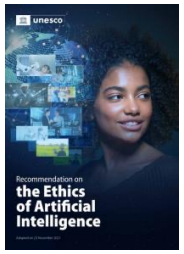
https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence



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

Bem mais de 1.000 publicadas no mundo!

AI Policy Lab
at Umeå University

PROTEGER A DEMOCRACIA

- Fortalecer a regulação e a supervisão
  - Adotar leis claras sobre IA, reconhecimento facial, criptomoedas e spyware.
  - Aumentar a transparência e a responsabilização na implantação de tecnologias.
  - Fortalecer as instituições democráticas para supervisionar e orientar o desenvolvimento tecnológico.
- Promover a participação pública
  - Democratizar o desenvolvimento da IA envolvendo atores sociais diversos.
  - Alinhar os sistemas tecnológicos aos valores democráticos e ao interesse público.
- Ampliar o letramento digital
  - Promover pela educação uma compreensão crítica da IA e das tecnologias digitais.
  - Capacitar os cidadãos a defender seus direitos na era

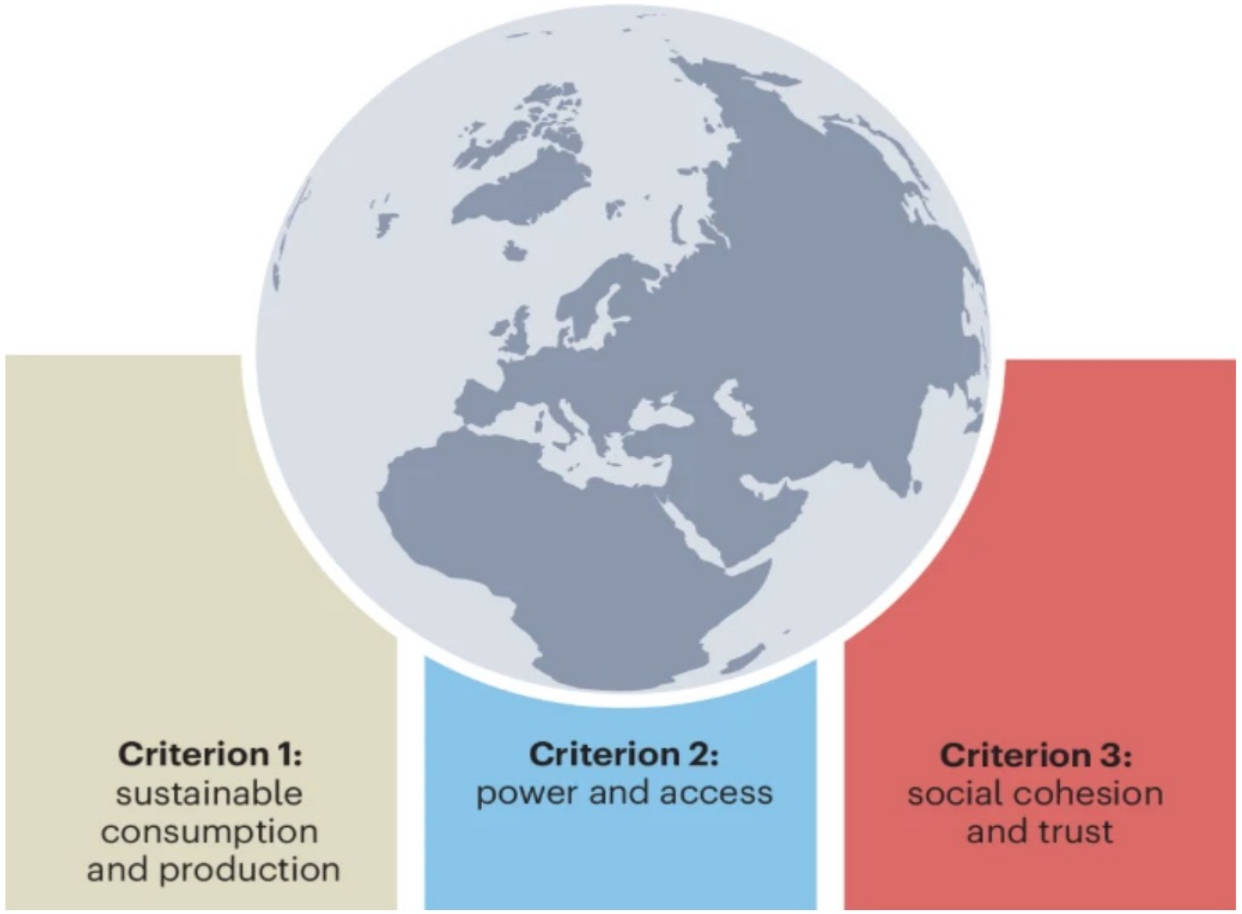
PROTEGER O PLANETA

O impacto da IA exige um cuidado sistêmico, para além dos impactos ambientais:

1. ajudar a acelerar a **transição para produção** e consumo sustentáveis que respeitem os limites planetários, ou ao menos não obstruam esses objetivos

2. ser desenvolvida, implantada e usada de modo a **garantir acesso equitativo** às ferramentas de IA para a sustentabilidade global e evitar concentrações de poder

3. ser desenvolvida, implantada e usada



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 O IMPACTO:
O QUE CONTA, E QUEM RESPONDE POR ISSO

Governamos o que contamos. A maior parte do impacto da IA não é contada.

Três níveis de impacto

(Kaack et al., 2022)

- Computação
- Aplicação
- Sistema

A consequência para a governança

- A responsabilização acompanha a medição
  - a computação tem um dono; os efeitos sistêmicos muitas vezes não têm
- O que é reportado vira aquilo para o que se projeta
  - Escolher a métrica é um ato político, não técnico
- Alinhamento com a Terra
  - limites planetários, acesso equitativo, coesão social.
- A Pergunta Zero também vale para as métricas
  - combinar o que conta, e quem audita, antes de construir

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

JUSTIÇA CLIMÁTICA

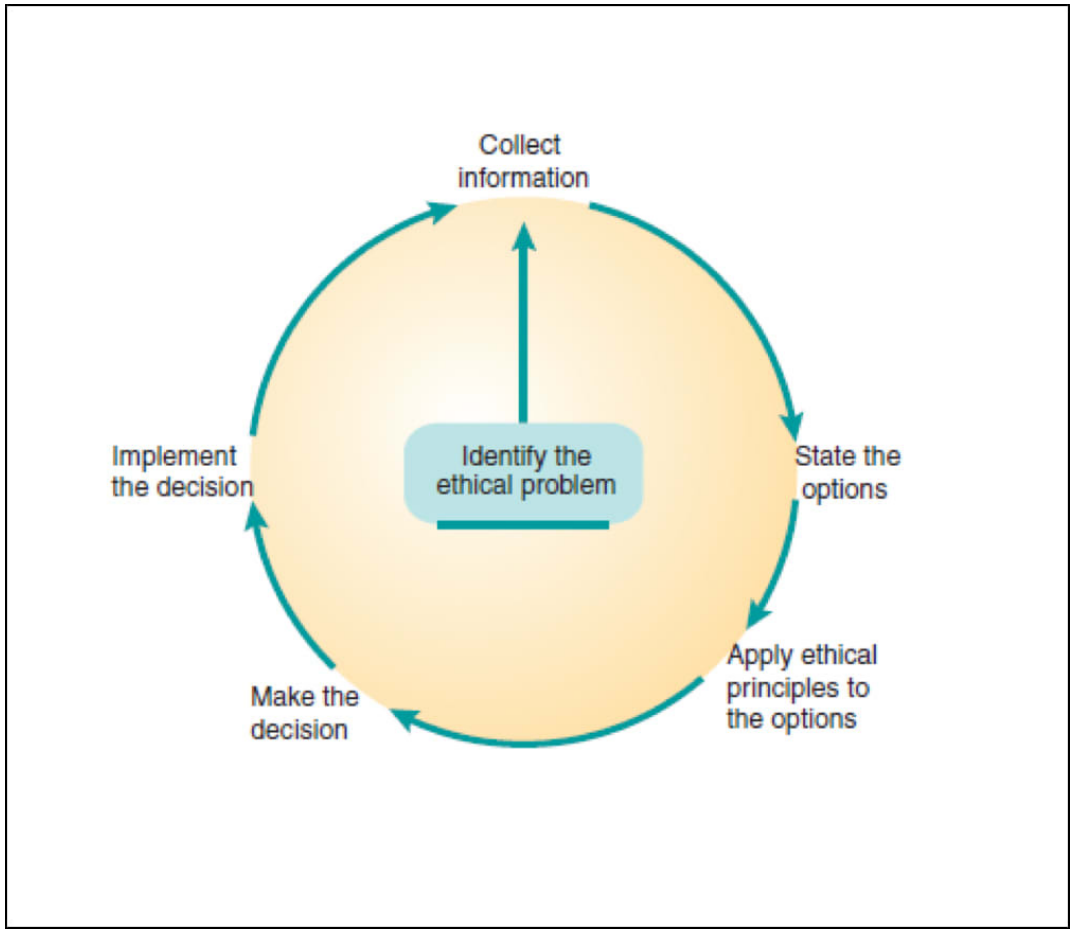
- distributiva (quem recebe o benefício e quem arca com os custos de computação, água, terra e trabalho),
- procedimental (quem estava na sala),
- de reconhecimento (que conhecimentos contam como evidência).

- Os fins justificam os meios? Trade-offs?
  - P. ex. melhorar a extração de combustíveis fósseis é um uso responsável da IA?
  - Quais são as alternativas? Quem ganha? Quem perde?

OPERACIONALIZAÇÃO: REGULAÇÃO E MAIS

- Regulação
  - AI Act: abordagem centrada no ser humano e baseada em risco
- Normas técnicas
  - governança branda; adesão não obrigatória
  - demonstrar diligência e limitar a responsabilidade
  - integração amigável entre produtos
- Estruturas e procedimentos organizacionais
  - Conselhos consultivos e responsáveis por ética
  - Definir e monitorar diretrizes éticas
- Avaliação de IA confiável
  - IA responsável é mais do que marcar caixinhas
  - São necessários meios de avaliar a maturidade
- Conscientização e participação

RACIOCÍNIO ÉTICO



- Não se trata da resposta, mas de reconhecer a questão
- Não há uma única resposta! Não há solução mágica!
  - Visão utilitarista
  - Visão dos direitos
  - Visão da justiça
  - Visão do bem comum
  - Visão das virtudes
  - ...

ÉTICA NA VIDA REAL

- Equidade
  - Auxílio-criança
  - Licença parental

- equidade

- igualdade

- Segurança
  - Radares de velocidade
  - Aulas de trânsito para crianças

– curto prazo

– longo prazo

- Equidade ou segurança
  - Recursos limitados

DILEMAS – EXERCÍCIO 3

- **Subsídios para adaptação climática**: pediram que você desenvolvesse um sistema de IA para triar candidaturas a um fundo municipal de adaptação climática, concedendo subsídios conforme a probabilidade de cada proposta gerar "alto impacto climático por euro", com dados de projetos financiados anteriormente. Isso substitui a análise manual e as visitas de campo dos técnicos.
- Trade-offs: como dividir seus recursos:
  - Quanto custo (computacional) vale 5% a mais de acurácia?
  - Quanto custo (computacional) vale incluir uma explicação?

Join at
slido.com
#4275 548

ÉTICA DA IA

- A IA é tão diferente de outras tecnologias?
  - Sobreposição com ética da tecnologia, ética empresarial, ética médica...

- Alguns desafios específicos da IA, trazidos por:
  - Autonomia. Se o controle executivo é delegado, quem é responsável pelas ações? Como manter a adesão aos valores humanos na tomada de decisão autônoma?
  - Implantação em ambientes abertos. Se não podemos prever o contexto que o sistema vai encontrar, como garantir um comportamento ético?
  - Generalização. Se uma decisão é tomada generalizando a partir de dados, o que acontece com nosso direito de ser visto como indivíduo? Como evitar a reprodução e a amplificação de vieses?

IA RESPONSÁVEL: MAIS QUE ÉTICA

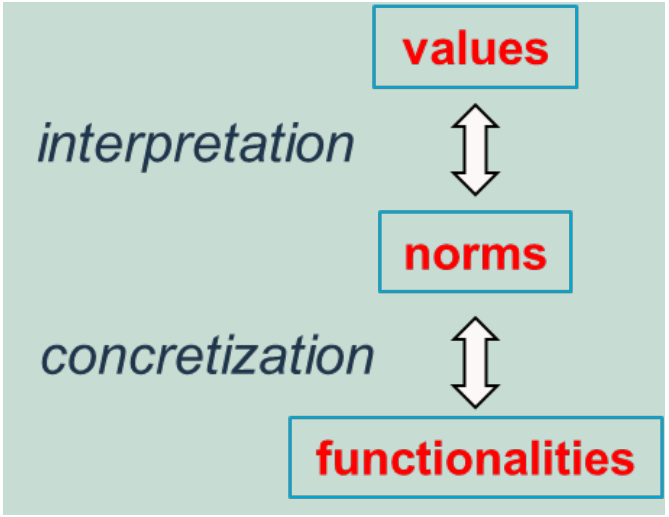
- Não é filosofar sobre ética
  - Ética não é sobre a resposta, mas sobre reconhecer a questão
  - Ética é um processo (social), não uma solução

- Não é tecnificar a ética
  - Sua implementação não “resolve” a ética

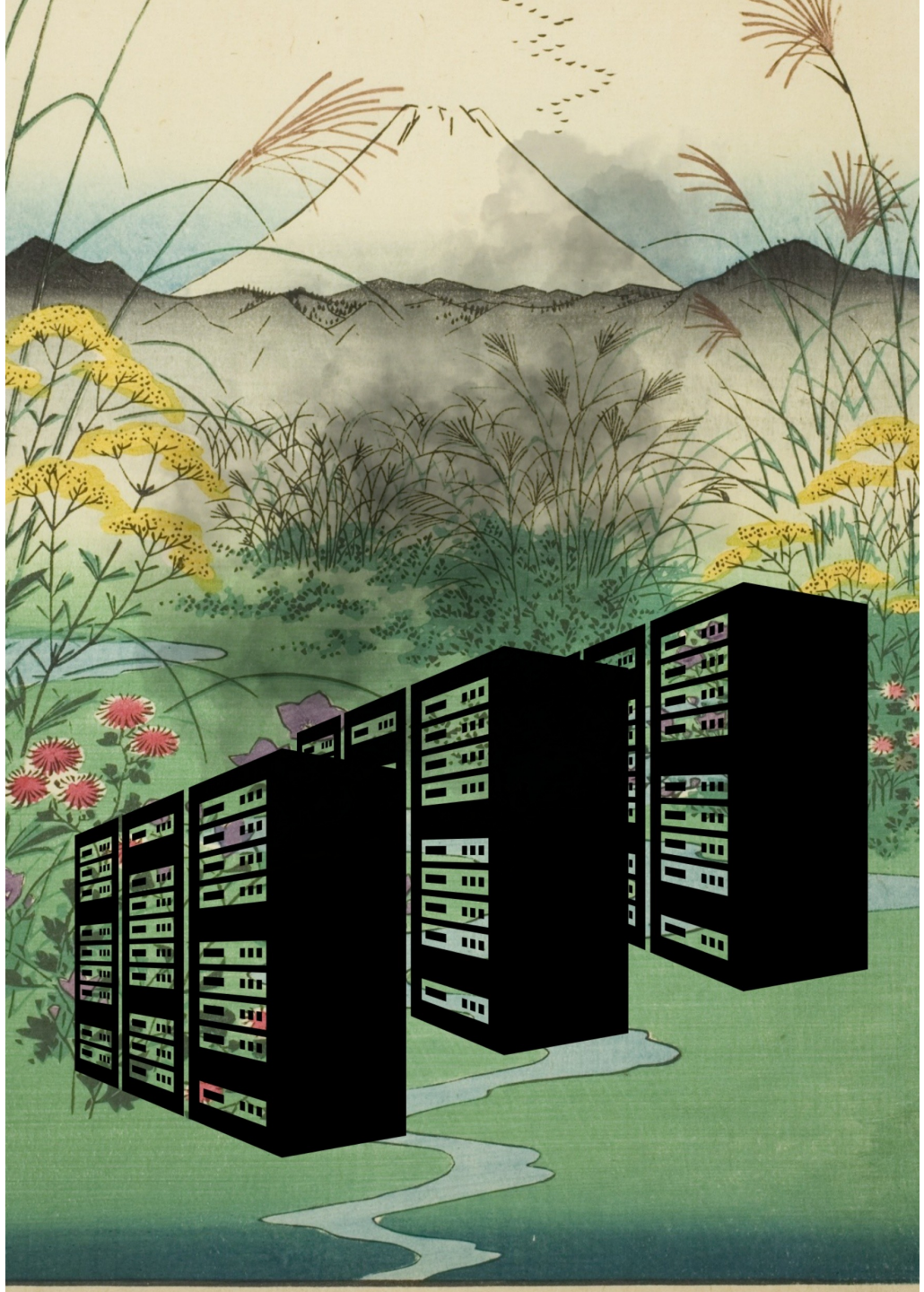
- No fundo, trata-se de escolhas, prioridades, trade-offs
  - Acurácia / explicação
  - Acurácia / recursos computacionais
  - Segurança / privacidade
  - Equidade / igualdade
  - Benefício de longo prazo / curto prazo
  - ...

IA RESPONSÁVEL É SER EXPLÍCITO

- Projeto orientado por valores
  - Aspectos legais e éticos não são um acessório!
- Governança
  - Monitoramento e controle externos
  - Acordos, contratos, normas
- Projeto
  - Questione suas opções e escolhas
  - Justifique suas escolhas
  - Documente suas escolhas e opções



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



UM NOVO PARADIGMA DE IA

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

UM NOVO PARADIGMA DE IA?

- A IA não é uma solução universal
  - reconhecer seus limites diante de desafios complexos
  - Pergunta zero: a IA deve ser usada aqui?
- Enfrentar os riscos inerentes de viés e discriminação
  - melhorar a transparência dos dados e algoritmos
  - acomodar contextos e valores diversos
- Para além das disciplinas
  - integrar conhecimento social e técnico no projeto da IA
  - enfrentar desafios sociais sistêmicos
  - envolver atores sociais diversos, não só tecnólogos
- normas técnicas robustas
  - Verificável, sustentável, participativo

ABORDAGEM DE IA RESPONSÁVEL

Comprometer-se com princípios de IA responsável e confiável, para uma IA inovadora, ética e que respeite os direitos humanos

Reconhecer possíveis tensões entre princípios éticos e a necessidade de priorizá-los

Desenvolver uma cultura de responsabilidade nas organizações

Priorizar a interação e a comunicação com as partes interessadas

GOVERNANÇA CONFIÁVEL

- Desenvolvimento da IA
  - O quê? Requisitos, necessidades, possibilidades, capacidades
- Uso da IA
  - Como? Licença para operar
- Contexto em que a IA é desenvolvida e usada
  - Onde/quando? Regras do jogo

- Esforços globais
  - Sustentabilidade
  - Inclusão e participação
  - Diversidade
  - Distribuição de benefícios e custos
  - Agenda 2030 – ODS

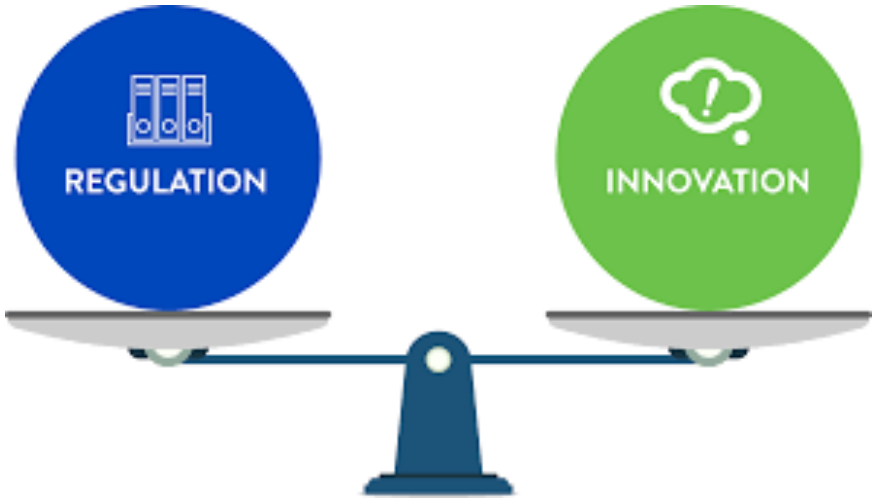


AI Policy Lab
at Umeå University

IA RESPONSÁVEL É INOVAÇÃO

- Inovação necessária
  - Requisitos técnicos
  - Requisitos de uso
  - Contexto de operação

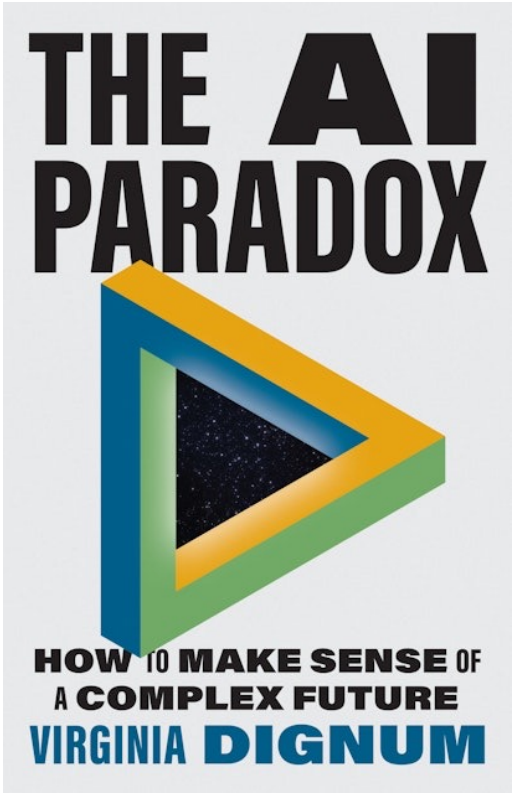
- Inovação multidisciplinar!
  - Inovação tecnológica
  - Inovação organizacional
  - Inovação regulatória
  - Inovação em governança
  - Inovação social



A IA não é uma força imparável; ela é moldada por nossas escolhas

- as decisões que tomamos hoje vão determinar se ela se tornará uma ferramenta de emancipação ou uma fonte de controle e desigualdade.

O grande paradoxo da IA: quanto mais inteligente ela fica, mais precisamos de sabedoria humana



Lançamento em fevereiro de 2026

QUIZ

- Verdadeiro ou falso: um modelo que satisfaz a paridade demográfica é justo.
- O viés pode entrar em um sistema de IA para o clima por meio
  - (a) da cobertura dos dados de treinamento
  - (b) da escolha da variável proxy
  - (c) da delimitação do problema
  - (d) todas as anteriores
- Em um sistema de alerta precoce, baixar o limiar de alerta troca principalmente:
  - (a) acurácia por computação
  - (b) alarmes falsos por eventos perdidos
  - (c) transparência por privacidade
  - (d) latência por resolução

LEITURAS RECOMENDADAS

- Dignum, *Responsible Artificial Intelligence* (Springer, 2019)
- Nota técnica sobre a Pergunta Zero do 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)

MAIS LEITURAS

- **Meu trabalho e o material da Pergunta 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.)
- **Viés, equidade, privacidade, responsabilização**
  - 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)*
- **Prática, marcos e documentação**
  - 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 de política e diretrizes**
  - 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

OBRIGADA!

AI Policy Lab
at Umeå University

RESPOSTAS DO QUIZ

- Verdadeiro ou falso: um modelo que satisfaz a paridade demográfica é justo.
- O viés pode entrar em um sistema de IA para o clima por meio
  - (a) da cobertura dos dados de treinamento
  - (b) da escolha da variável proxy
  - (c) da delimitação do problema
  - (d) todas as anteriores
- Em um sistema de alerta precoce, baixar o limiar de alerta troca principalmente:
  - (a) acurácia por computação
  - (b) alarmes falsos por eventos perdidos
  - (c) transparência por privacidade
  - (d) latência por resolução