L'IA : LA RESPONSABILITÉ DANS UN MONDE QUI CHANGE

Pr Dr Virginia Dignum

Chaire IA responsable – Directrice de l'AI Policy Lab, Université d'Umeå

Professeure invitée SIAS, Université d'Amsterdam

Courriel : virginia@cs.umu.se

UMEÅ UNIVERSITY

AI Policy Lab

at Umeå University

L’IA EN 2026 – LES ENJEUX

CONSCIENCE ?



L’IA D’ABORD ?



AI Policy Lab
at Umeå University

L'APPROCHE « IA D'ABORD »

- Passe DE la gouvernance AU déploiement sans contrôle
- Ignore les données montrant des gains de productivité limités de l'IA
- **Oublie les questions** « Pourquoi l'IA ? » **et** « L'IA, oui ou non ? »
- Déficit démocratique : pas de consultation des parties prenantes
- Crée une pression à adopter sans évaluer (Dignum et al., 2025)

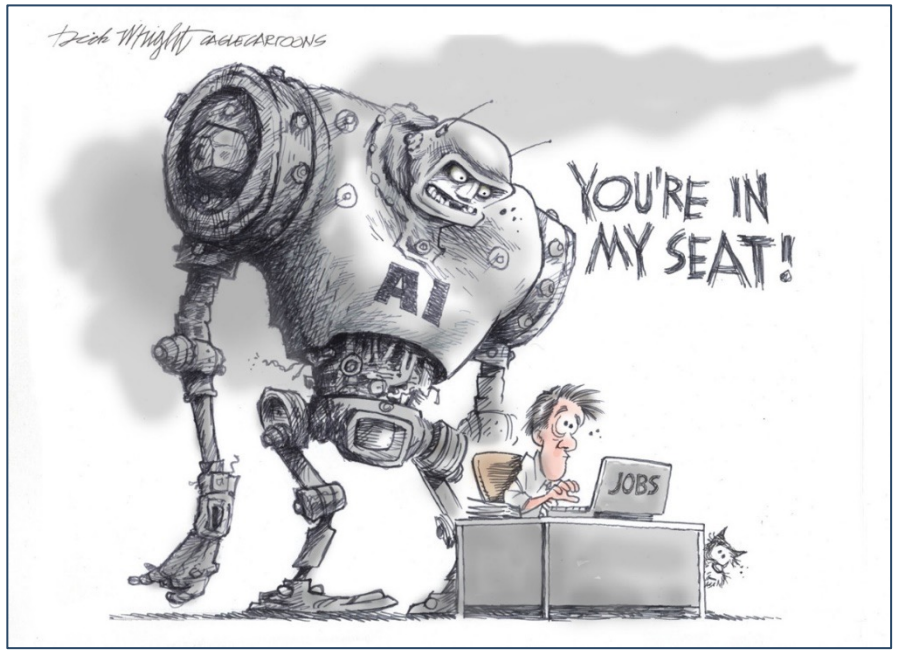
Des incitations qui encouragent l'usage de l'IA sans égard à sa pertinence se retourneront contre leurs auteurs : communes et régions ont besoin de compétences pour juger quand et où l'IA est appropriée, pas de dispositifs qui poussent à l'adopter sans discernement

Repose sur l'idée fausse que la réglementation freine les services numériques

Adopter l'IA sans investir dans la R-D nationale accroît la dépendance envers un petit nombre de fournisseurs étrangers, et sape la souveraineté numérique que la stratégie prétend renforcer.

La responsabilité suit le contrôle. Les producteurs conçoivent le système ; ils doivent en porter la responsabilité principale. Reporter la vérification sur les utilisateurs déplace la responsabilité sans réduire le risque.

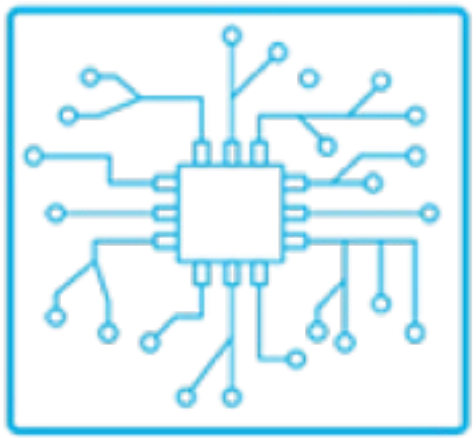
FAÇONNONS-NOUS L'IA, OU L'IA NOUS FAÇONNE-T-ELLE ?





- L'IA soulève autant de problèmes qu'elle en résout
- L'IA n'est pas un moyen de remplacer les humains
- Il faut le regard humain pour traiter les conséquences éthiques, sociales et imprévues de l'IA

L’IA : PLUS, C’EST MIEUX ?



Computing Power

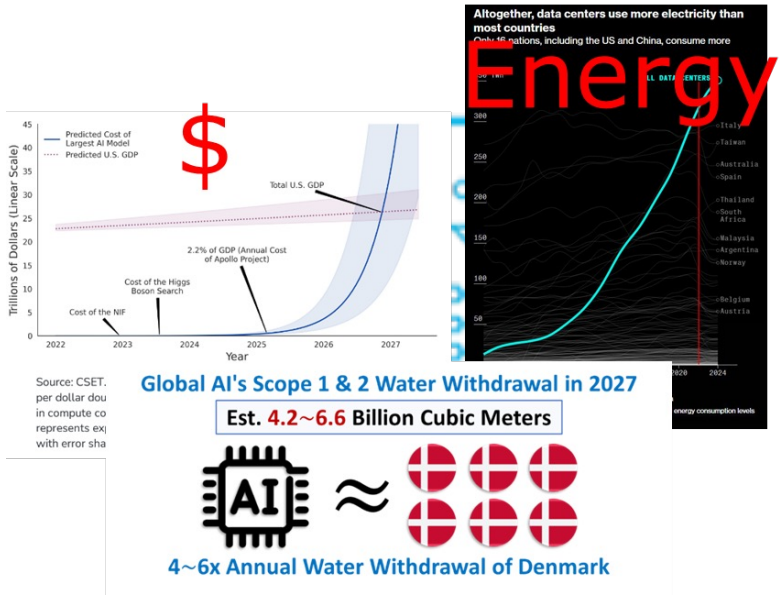


Algorithm Power



Data Availability

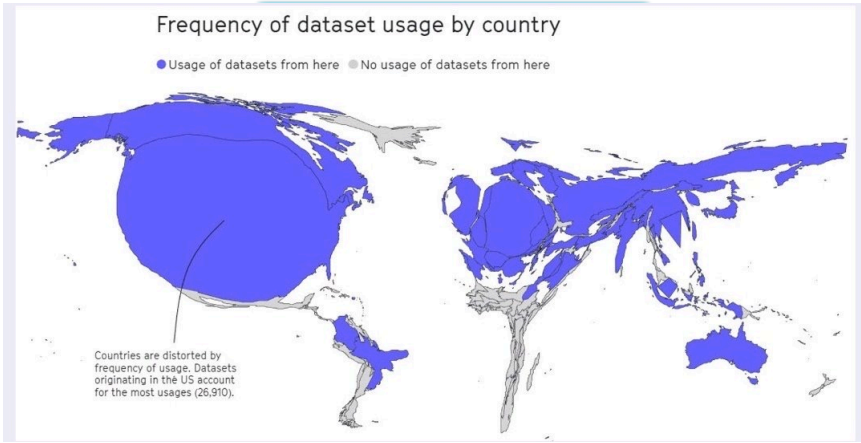
L’IA : PLUS, C’EST MIEUX ?



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





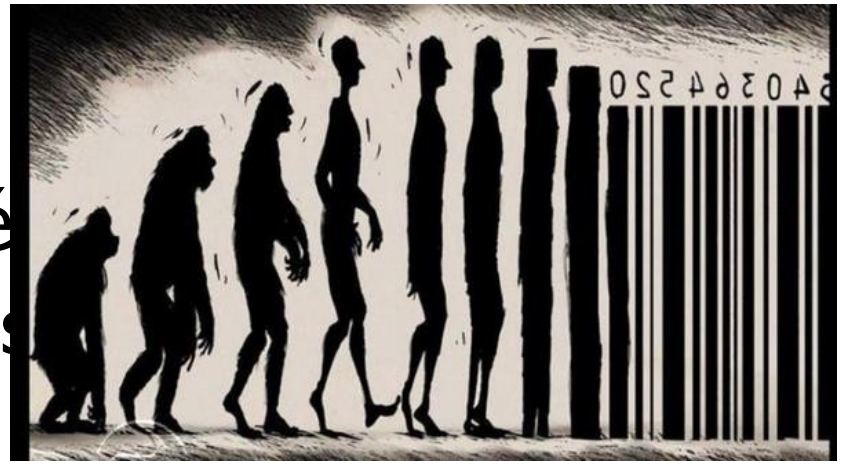




L'IMPACT DE L'IA SUPPOSE LA RESPONSABILITÉ HUMAINE

AI
at Umeå University

L'IA RESPONSABLE : POURQUOI S'EN SOUCIER ?



- Mise en données
  - La réalité dépasse les donné
  - Les données sont construites
  - Les données sont biaisées
  - Riche en données, pauvre en compréhension



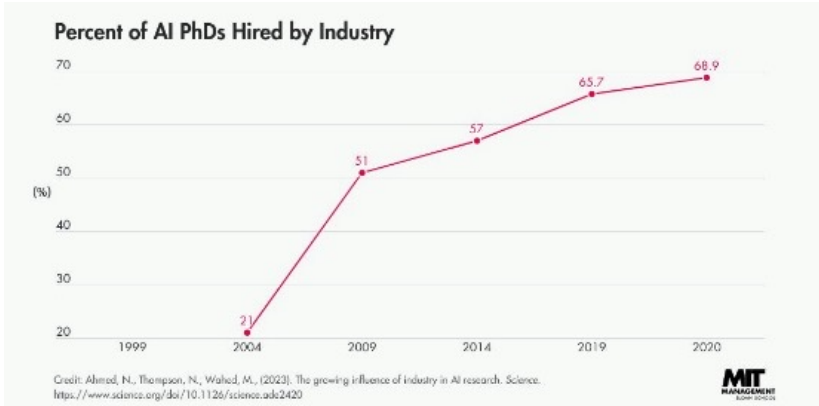
Time magazine, janv. 2023

- Pouvoir
- Durabilité

L'IA RESPONSABLE : POURQUOI S'EN SOUCIER ?

- Mise en données
- Pouvoir
  - Qui développe l'IA ?
  - Qui décide ?
- Durabilité

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**

L'IA RESPONSABLE : POURQUOI S'EN SOUCIER ?

- Mise en donné
- Pouvoir
- Durabilité
  - Coût de calcul de l'IA
  - Coûts humains et sociaux



INDEPENDENT PREMIUM

THE LONGER READ

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



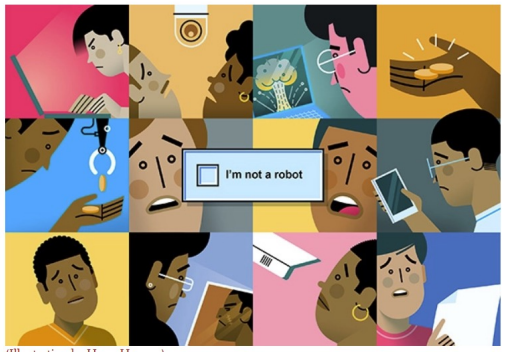
Technology

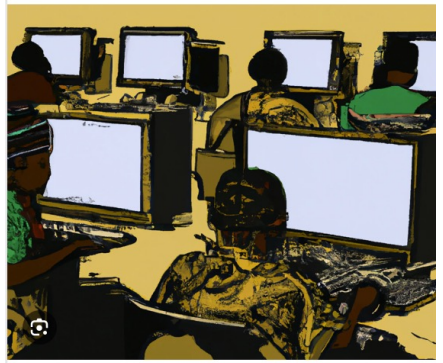
Turk Wars: How AI Threatens the Workers Who Fuel It

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



POUVONS-NOUS ÊTRE RESPONSABLES ?

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

L’IA N’EST PAS NEUTRE

- L’IA peut servir la dignité humaine, ou la saper
- incarne des choix humains
- Gouverner l’IA = nous gouverner nous-mêmes



IA – ÉTHIQUE DE L'IA – IA RESPONSABLE – GOUVERNANCE DE L'IA…

- Les débuts
  - Performance technique et efficacité des systèmes d'IA
- Prise de conscience
  - Reconnaissance des risques : biais, discrimination, atteintes à la vie privée, manque de redevabilité
- Virage éthique
  - Appels à l'équité, la transparence, le contrôle humain et l'explicabilité
- IA responsable
  - Passer des principes à la pratique en intégrant l'éthique à la conception, au déploiement et à l'usage
- Besoin de gouvernance
  - Garantir la confiance de la société, la durabilité à long terme et l'alignement sur les valeurs démocratiques

DE QUOI PARLONS-NOUS ?

- **Biais** : un écart systématique, pas un synonyme d'erreur.
  - Distinction importante : le biais statistique (échantillonnage, mesure) n'est pas le biais sociétal (le système reproduit un schéma injuste).
- **Équité** : une revendication distributive contestée, pas une métrique.
  - Les critères formels (parité, égalité des chances, calibration) sont mutuellement incompatibles sauf cas dégénérés : en choisir un est un acte normatif.

(Kleinberg et al. 2016; Chouldechova 2017)

- **La transparence** (sur le processus et la provenance) diffère de l'explicabilité (sur un résultat précis).
  - Ce sont des obligations différentes pour des publics différents.

QU'EST-CE QUE CELA CHANGE POUR LE CLIMAT ?

- **Choix et cadrage du problème**
  - Ce qui compte comme problème, et à qui il appartient, est décidé avant toute collecte de données.
  - P. ex. « prédire le rendement » et « aider un petit exploitant à décider » ne sont pas le même problème.
- **Variable cible et proxys**
  - L'étiquette est rarement ce qui vous intéresse vraiment : demandes d'indemnisation comme proxy de l'impact des inondations, demande raccordée au réseau comme proxy du besoin d'énergie.
- **Couverture des données**
  - Les modèles sont validés là où les données existent déjà, et exrapolés là où elles manquent.
  - Dans une grande partie de l'Afrique, les stations météo

La solution n'est pas un meilleur modèle, mais de meilleures questions, plus tôt.

IA RESPONSABLE POUR L'IA CLIMATIQUE – EXERCICE 1

Scénario : une ville commande un modèle de risque d'inondation. Il est entraîné sur vingt ans de demandes d'indemnisation d'assurance : chaque demande enregistrée marque un lieu comme inondé, et les lieux sans demande sont considérés comme non inondés. Le modèle utilise images satellitaires, altitude et précipitations, et atteint 91 % d'exactitude sur des données réservées.

Où le biais entre-t-il en premier ?

- La résolution des images satellitaires
- Le choix des demandes d'indemnisation comme





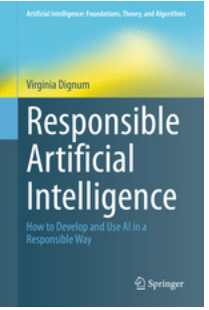
IA RESPONSABLE POUR L'IA CLIMATIQUE – EXERCICE 1

Scénario : une ville commande un modèle de risque d'inondation. Il est entraîné sur vingt ans de demandes d'indemnisation d'assurance : chaque demande enregistrée marque un lieu comme inondé, et les lieux sans demande sont considérés comme non inondés. Le modèle utilise images satellitaires, altitude et précipitations, et atteint 91 % d'exactitude sur des données réservées.

Où le biais entre-t-il en premier ?

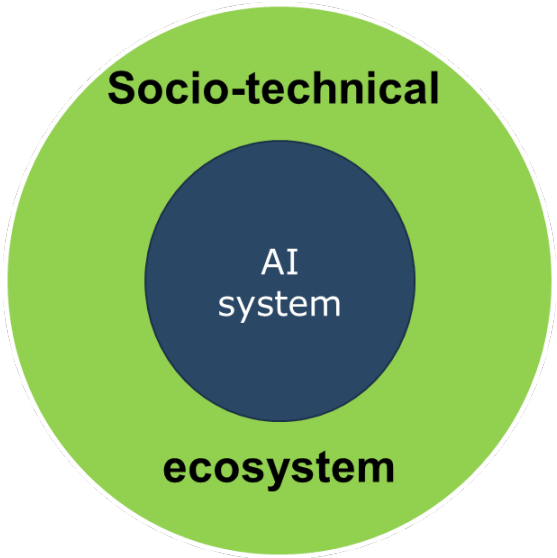
- La résolution des images satellitaires
- Le choix des demandes d'indemnisation comme étiquette
- L'architecture du modèle
- La métrique d'évaluation

Les demandes d'indemnisation indiquent qui était assuré et qui a déclaré, pas qui a été inondé. Les ménages non assurés et les quartiers informels ne sont pas sous-représentés : ils sont enregistrés comme secs. Les 91 % sont mesurés par rapport à la même étiquette biaisée, donc ils ne peuvent pas détecter le problème. Tout ce qui suit en hérite.



RESPONSABLES À PROPOS DE L'IA

L'IA ne nous tombe pas dessus !



- L'IA est conçue. C'est nous qui faisons les choix
  - Qui doit décider ?
  - Quelles valeurs prendre en compte ?
  - Les valeurs de qui ? Comment prioriser ?
- Posez la Question Zéro !
  - L'IA est-elle la meilleure option ici ?
  - Demandez pourquoi avant de demander comment
- L'IA n'existe pas en vase clos
  - Il n'y a pas de solution technique aux effets néfastes !
  - Éthique, réglementation et gouvernance concernent l'écosystème

Les solutions d'IA responsable doivent être sociales plutôt que techniques !

AI Policy Lab
at Umeå University

L'ALTERNATIVE : LA FINALITÉ D'ABORD (QUESTION ZÉRO)

L'approche « la finalité d'abord » **demande : «** Si et où l'IA est-elle réellement nécessaire et bénéfique ? »

« Faut-il seulement construire ce système d'IA ? »

Principes clés :

- Partir des **besoins de la société,** pas des capacités technologiques
- Demander systématiquement : Pourquoi ? Pour qui ? Quelles alternatives ?
- Inclure des voix diverses dans la prise de décision
- Viser l'orientation et la valeur publique, pas la vitesse

La Question Zéro :

- Fait passer de la VITESSE à **l'ORIENTATION**
- Donne du pouvoir à des acteurs divers (PME, société civile, petits pays)
- Offre un cadre pratique et peu coûteux
- Ancre les décisions dans une **identification réelle du problème** (Dignum et al., 2025)

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

AI Policy Lab at Umeå University

LA QUESTION ZÉRO VA AU-DELÀ DU OUI/NON

Le cycle de la Question Zéro :

- **Cadrage** : à qui appartient le problème, qui l'a défini, quelle est la référence sans IA, qui supporte le coût d'une mauvaise réponse.
- **Données** : provenance, consentement, lacunes de couverture, qui est invisible dans les données, à quelles conditions ont-elles été obtenues.
- **Modélisation** : ce qui est optimisé, ce qui est sacrifié, quelle incertitude est communiquée.
- **Évaluation** : performances désagrégées, pas agrégées ; modes de défaillance nommés, pas seulement l'exactitude.
- **Déploiement** : qui l'exploite, quels recours existent, que se passe-t-il quand il se trompe.

EXERCICE 2 : L'IA POUR PLANIFIER L'ADAPTATION AUX INONDATIONS

Scénario : une ville veut introduire un système d'IA qui classe les quartiers pour financer l'adaptation aux inondations

- Le système prédit le risque d'inondation à partir d'images satellitaires, de l'altitude et des demandes d'indemnisation passées, et classe les zones pour l'investissement.
- Le système estime la vulnérabilité des ménages en combinant télédétection, données de téléphonie mobile et données des services publics.
- Le système génère des synthèses et des recommandations pour le conseil municipal, qui a trois semaines pour décider.

QUESTION ZÉRO – EXERCICE 2

- L’IA est-elle adaptée à cet usage ?

- Options
  - A) Oui, l’alternative est un processus manuel plus lent et tout aussi biaisé.
  - B) Oui, l’alternative est un processus manuel plus lent et tout aussi biaisé.
  - C) Non, les données ne peuvent pas étayer la décision pour laquelle on les utilise.
  - D) Non, les données ne peuvent pas étayer la décision pour laquelle on les utilise.



Join at
slido.com
#4275 548

LA CONCEPTION EST POLITIQUE

- Choix
- Formulation
- Information
- Implication
- Légitimité
- Agrégation







QU'EST-CE QUE L'IA ?

QUE VOULONS-NOUS QUE L'IA SOIT ?

- Comme un humain ?
  - Pourquoi ?
  - Qu'est-ce que cela signifie ?
- Un outil ?
  - Pour quoi ? Pour qui ?
- Simulation ou action ?
  - Comprendre l'intelligence en construisant de l'intelligence, ou
  - Intervention active dans le monde réel
- Normatif ou descriptif ?



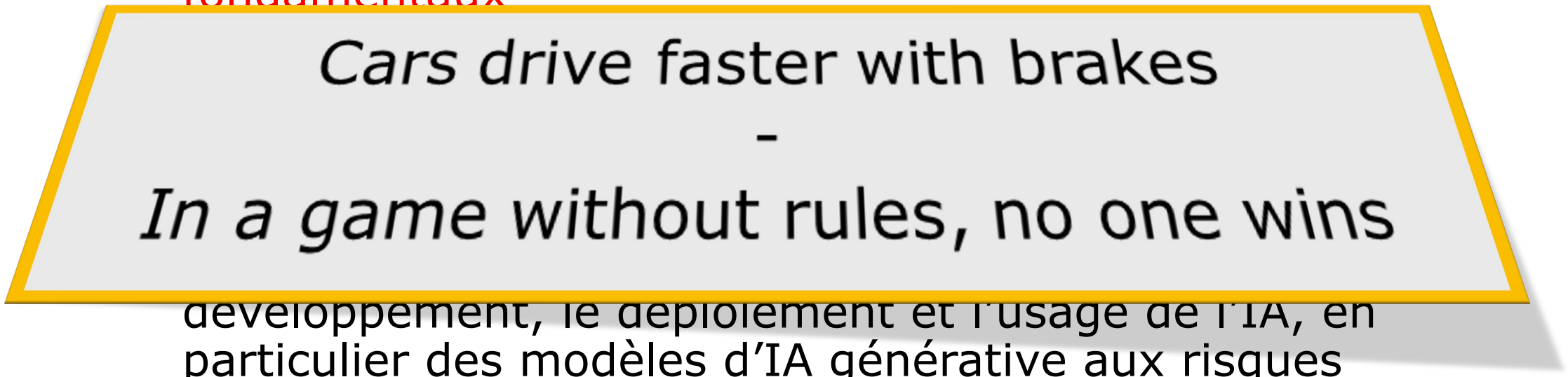
PRINCIPES, LIGNES DIRECTRICES, ÉTHIQUE

AI Policy Lab
at Umeå University

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

GOUVERNANCE : POURQUOI ? POUR QUOI ?

- La réglementation comme incitation à l'innovation responsable, à la durabilité et aux droits humains fondamentaux



développement, le déploiement et l'usage de l'IA, en particulier des modèles d'IA générative aux risques variables

- Exigences de responsabilité, de redevabilité et de gouvernance
  - Encadrer les acteurs organisationnels plutôt que les résultats technologiques
  - Confiance du public et responsabilité des erreurs de la

AI Policy Lab
at Umeå University

PRINCIPES ET LIGNES 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?

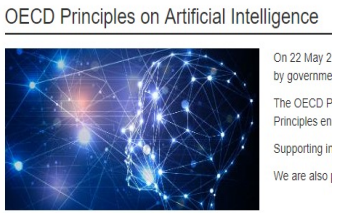
Niveau	Cadre et portée
Mondial	UNESCO (194 pays), OCDE (>70 juridictions), PMIA (25+ membres), traité du Conseil de l'Europe (50+ pays), principes du G7
Régional	AI Act de l'UE (27 États membres), Déclaration de Santiago (Amérique latine/Caraïbes)
National	930+ initiatives dans 71 pays
Autres	200+ lignes directrices d'ONG, d'univ… d'organismes privés



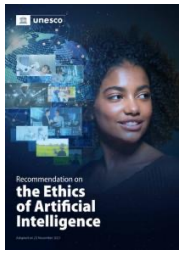
https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence



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

Bien plus de 1 000 publiées dans le monde !

AI Policy Lab
at Umeå University

PROTÉGER LA DÉMOCRATIE

- Renforcer la réglementation et le contrôle
  - Adopter des lois claires sur l'IA, la reconnaissance faciale, les cryptomonnaies et les logiciels espions.
  - Renforcer la transparence et la redevabilité dans le déploiement des technologies.
  - Renforcer les institutions démocratiques pour encadrer et orienter le développement technologique.
- Promouvoir la participation du public
  - Démocratiser le développement de l'IA en associant des acteurs divers de la société.
  - Aligner les systèmes technologiques sur les valeurs démocratiques et l'intérêt général.
- Renforcer la culture numérique
  - Développer par l'éducation une compréhension critique de l'IA et des technologies numériques.
  - Donner aux citoyens les moyens de défendre leurs droits

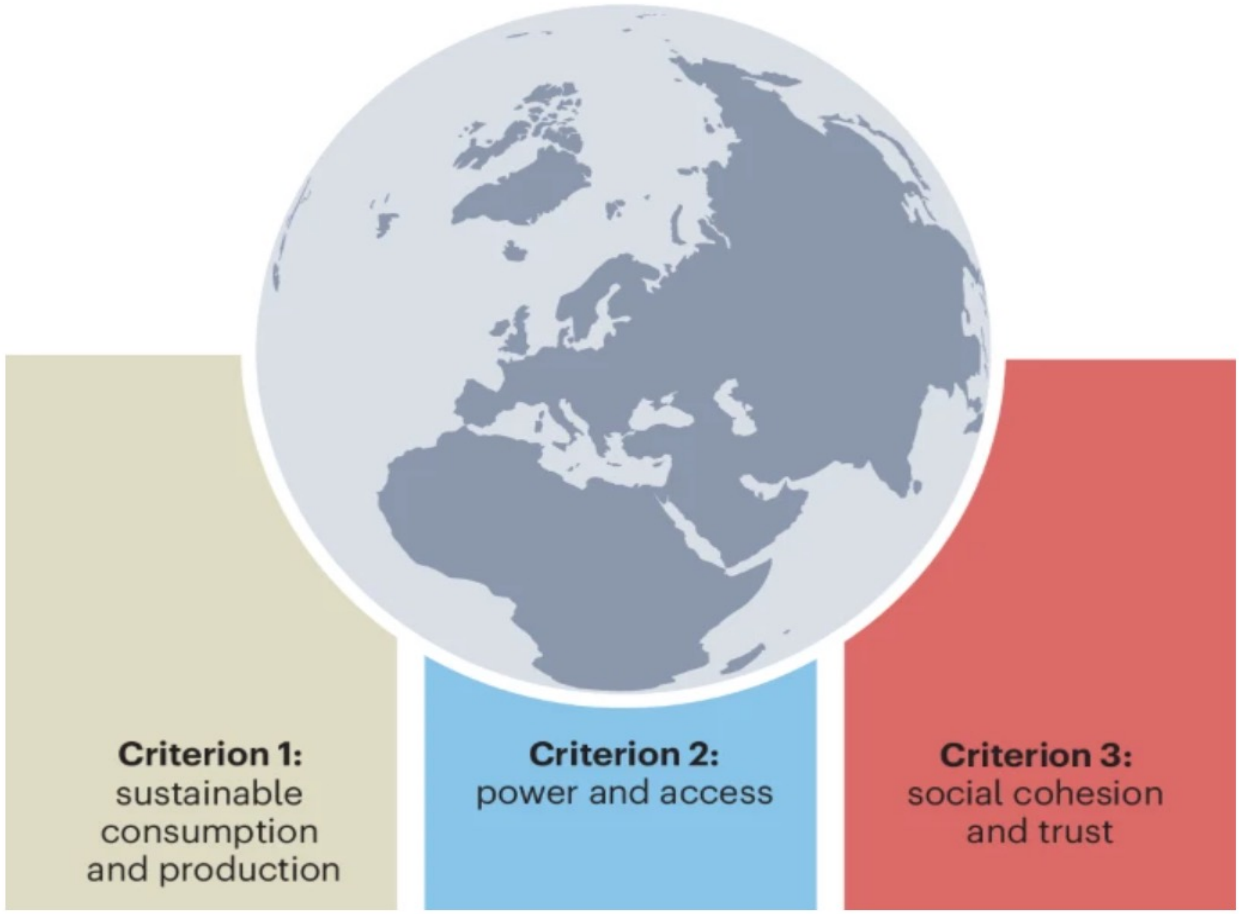
PROTÉGER LA PLANÈTE

L’impact de l’IA exige une gestion responsable systémique, au-delà des impacts environnementaux :

1. contribuer à accélérer la **transition vers une production** et une consommation durables, dans le respect des limites planétaires, ou du moins sans entraver ces objectifs

2. être développée, déployée et utilisée de façon à **garantir un accès équitable** aux outils d’IA pour la durabilité mondiale et à éviter les concentrations de pouvoir

3. être développée, déployée et utilisée



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

MESURER L'IMPACT : CE QUI COMPTE, ET QUI EN RÉPOND

Nous gouvernons ce que nous comptons. La majeure partie de l'impact de l'IA n'est pas comptée.

Trois niveaux d'impact

(Kaack et al., 2022)

- Calcul
- Application
- Système

La conséquence pour la gouvernance

- La redevabilité suit la mesure
  - le calcul a un propriétaire, les effets systémiques souvent aucun
- Ce qui est déclaré devient ce pour quoi on conçoit
  - Choisir la métrique est un acte politique, pas technique
- Alignement sur la Terre
  - limites planétaires, accès équitable, cohésion sociale.
- La Question Zéro s'applique aussi aux métriques
  - convenir de ce qui compte, et de qui l'audite, avant de construire

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

JUSTICE CLIMATIQUE

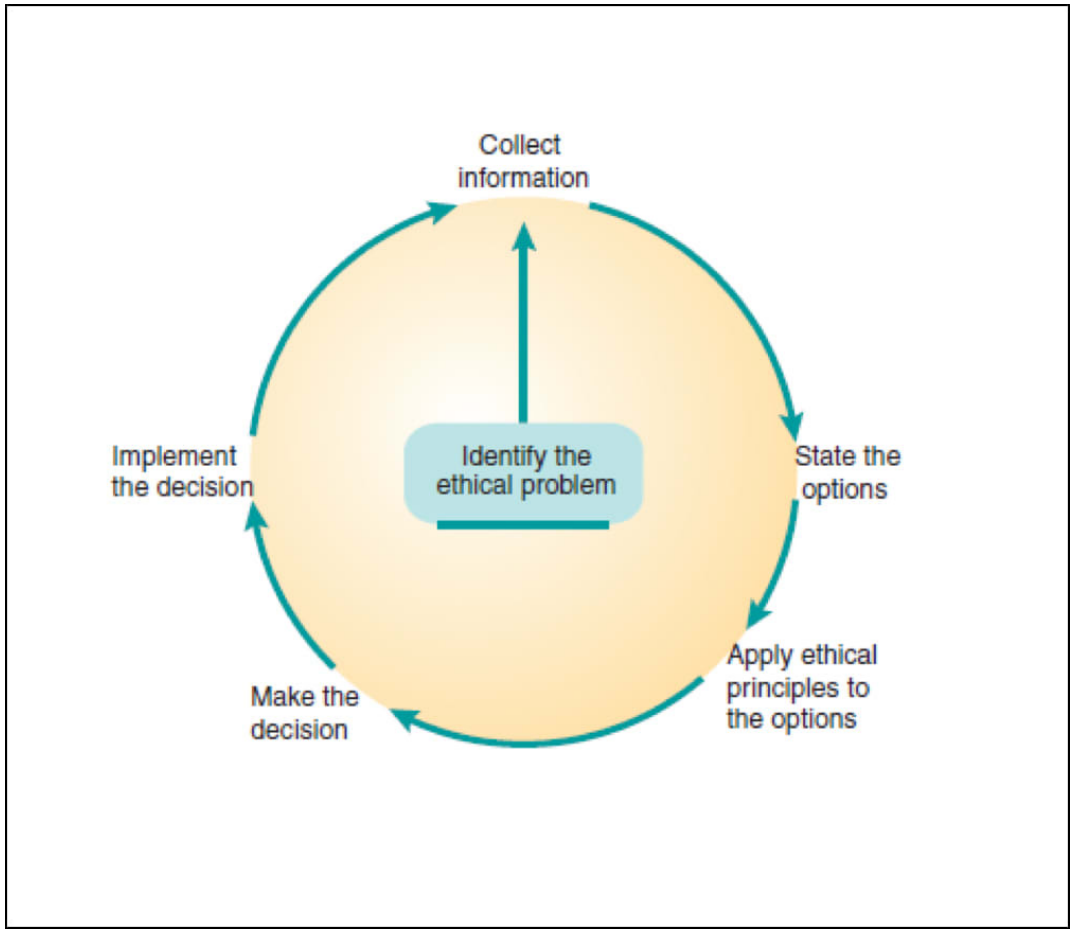
- distributive (qui reçoit le bénéfice et qui supporte le coût en calcul, eau, terres et travail),
- procédurale (qui était dans la salle),
- de reconnaissance (quels savoirs comptent comme preuve).

- La fin justifie-t-elle les moyens ? Compromis ?
  - P. ex. améliorer l'extraction de combustibles fossiles est-il un usage responsable de l'IA ?
  - Quelles alternatives ? Qui gagne ? Qui perd ?

MISE EN ŒUVRE : RÉGLEMENTATION ET PLUS ENCORE

- Réglementation
  - AI Act : approche centrée sur l'humain et fondée sur les risques
- Normes
  - gouvernance souple ; application non obligatoire
  - démontrer la diligence raisonnable et limiter la responsabilité
  - intégration conviviale entre produits
- Structures et procédures organisationnelles
  - Comités consultatifs et responsables de l'éthique
  - Fixer et suivre des lignes directrices éthiques
- Évaluation d'une IA digne de confiance
  - l'IA responsable, c'est plus que cocher des cases
  - Il faut des moyens d'évaluer la maturité
- Sensibilisation et participation

RAISONNEMENT ÉTHIQUE



- Il ne s'agit pas de la réponse, mais de reconnaître le problème
- Pas une seule réponse ! Pas de solution miracle !
  - Approche utilitariste
  - Approche par les droits
  - Approche par la justice
  - Approche par le bien commun
  - Approche par les vertus
  - ...

L'ÉTHIQUE AU QUOTIDIEN

- Équité
  - Allocation de soutien aux enfants
  - Congé parental

- Sécurité
  - Radars de vitesse
  - Cours de sécurité routière pour les enfants

- Équité ou sécurité
  - Ressources limitées

- équité

- égalité

– court terme

– long terme

DILEMMES – EXERCICE 3

- **Subventions à l'adaptation climatique** : on vous demande de développer un système d'IA pour trier les candidatures à un fonds municipal d'adaptation au climat, en attribuant les subventions selon la probabilité de chaque dossier d'obtenir un « fort impact climatique par euro », à partir des données des projets financés précédemment. Cela remplace l'examen manuel et les visites sur site des chargés de dossier.
- Compromis : comment répartir vos ressources :
  - Combien de coût (de calcul) vaut 5 % d'exactitude en plus ?
  - Combien de coût (de calcul) vaut l'ajout d'une explication ?

Join at
slido.com
#4275 548

ÉTHIQUE DE L'IA

- L'IA est-elle si différente des autres technologies ?
  - Recoupements avec l'éthique des technologies, des affaires, l'éthique médicale…

- Quelques défis propres à l'IA, liés à :
  - L'autonomie. Si le contrôle exécutif est délégué, qui est responsable des actions ? Comment maintenir le respect des valeurs humaines dans la prise de décision autonome ?
  - Le déploiement en environnement ouvert. Si l'on ne peut pas prévoir le contexte que le système rencontrera, comment garantir un comportement éthique ?
  - La généralisation. Si une décision est prise en généralisant à partir de données, que devient notre droit d'être vu comme un individu ? Comment éviter la reproduction et l'amplification des biais ?

IA RESPONSABLE : PLUS QUE L'ÉTHIQUE

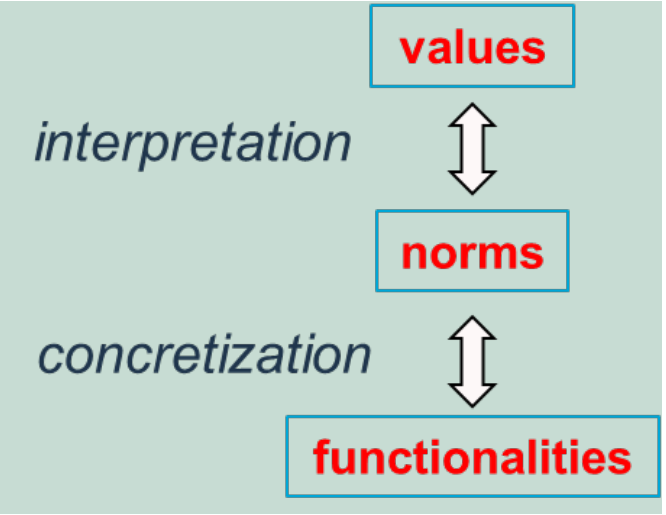
- Pas philosopher sur l'éthique
  - L'éthique ne porte pas sur la réponse, mais sur la reconnaissance du problème
  - L'éthique est un processus (social), pas une solution

- Pas une technicisation de l'éthique
  - Votre implémentation ne « résout » pas l'éthique

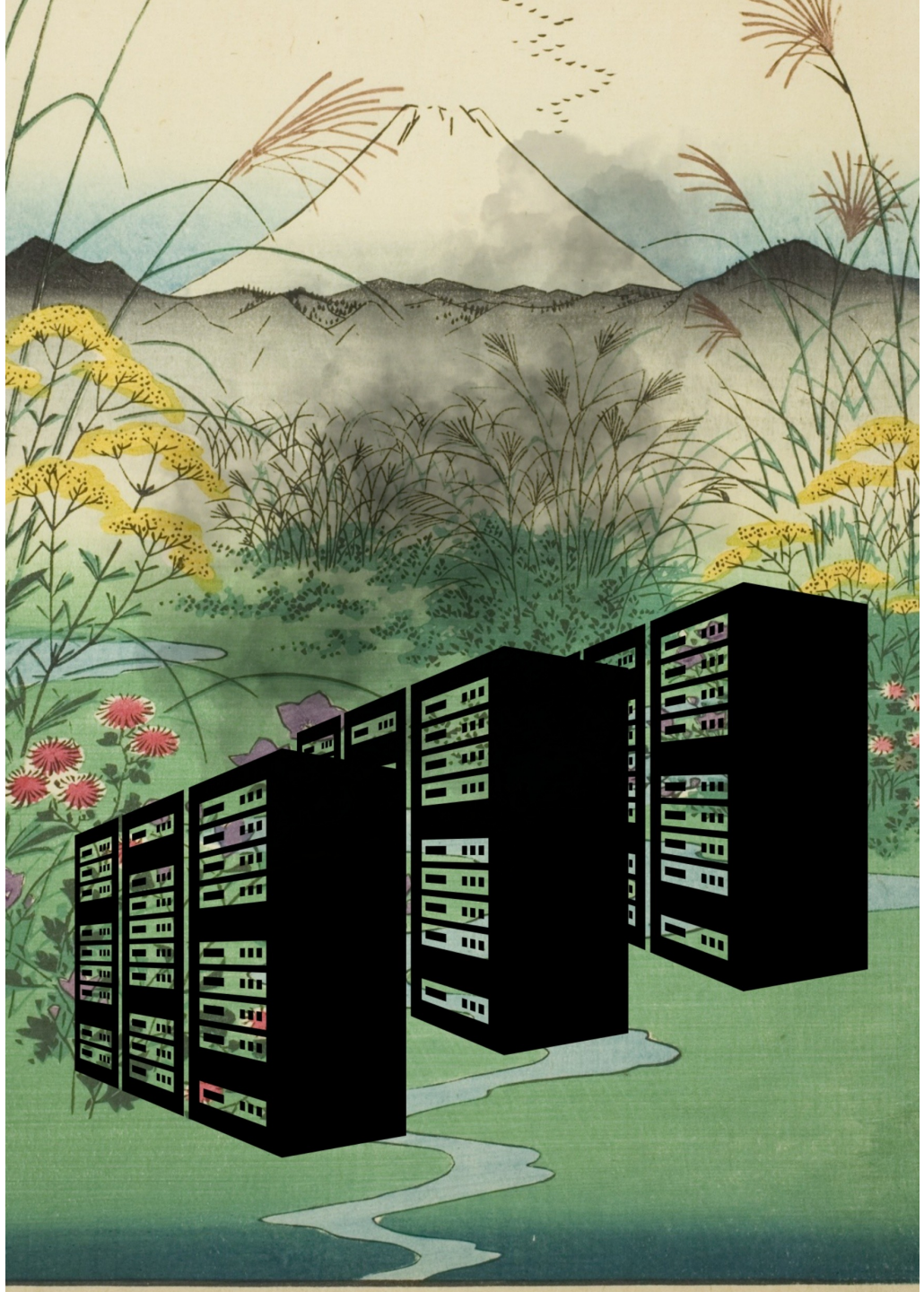
- Fondamentalement, il s'agit de choix, de priorités, de compromis
  - Exactitude / explication
  - Exactitude / ressources de calcul
  - Sécurité / vie privée
  - Équité / égalité
  - Bénéfice à long terme / court terme

L'IA RESPONSABLE, C'EST ÊTRE EXPLICITE

- Conception guidée par les valeurs
  - Les aspects juridiques et éthiques ne sont pas un supplément !
- Gouvernance
  - Suivi et contrôle externes
  - Accords, contrats, normes
- Conception
  - Interrogez vos options et vos choix
  - Justifiez vos choix
  - Documentez vos choix et options



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



UN NOUVEAU PARADIGME DE L'IA

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

UN NOUVEAU PARADIGME DE L'IA ?

- L'IA n'est pas une solution universelle
  - reconnaître ses limites face aux défis complexes
  - Question zéro : faut-il utiliser l'IA ici ?
- Traiter les risques inhérents de biais et de discrimination
  - améliorer la transparence des données et des algorithmes
  - tenir compte de contextes et de valeurs divers
- Au-delà des disciplines
  - intégrer expertises sociales et techniques dans la conception de l'IA
  - traiter les défis sociétaux systémiques
  - associer des acteurs divers de la société, pas seulement des technologues
- normes techniques solides
  - Vérifiable, durable, participatif

APPROCHE DE L'IA RESPONSABLE

S'engager sur des principes d'IA responsable et digne de confiance, pour une IA innovante, éthique et respectueuse des droits humains

Reconnaître les tensions possibles entre principes éthiques et la nécessité de les hiérarchiser

Développer une culture de la responsabilité dans les organisations

Donner la priorité au dialogue et à la communication avec les parties prenantes

UNE GOUVERNANCE DIGNE DE CONFIANCE

- Développement de l'IA
  - Quoi ? Exigences, besoins, possibilités, capacités
- Usage de l'IA
  - Comment ? Licence d'exploitation
- Contexte de développement et d'usage de l'IA
  - Où/quand ? Règles du jeu

- Efforts mondiaux
  - Durabilité
  - Inclusion et participation
  - Diversité
  - Répartition des bénéfices et des coûts
  - Agenda 2030 – ODD

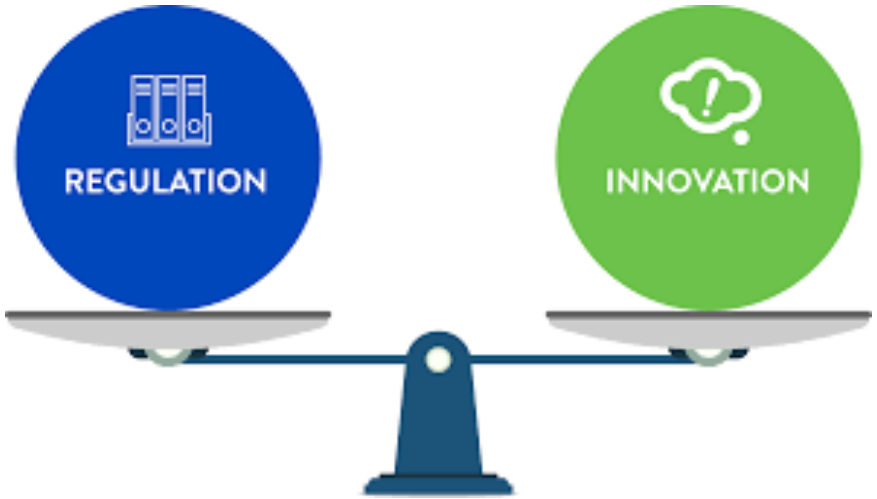


AI Policy Lab
at Umeå University

L’IA RESPONSABLE, C’EST DE L’INNOVATION

- Innovation nécessaire
  - Exigences techniques
  - Exigences d’usage
  - Contexte d’exploitation

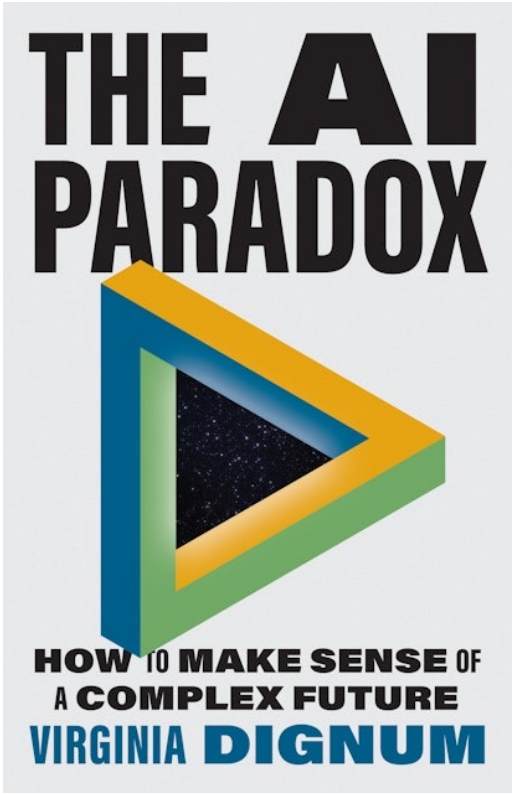
- Innovation multidisciplinaire !
  - Innovation technologique
  - Innovation organisationnelle
  - Innovation réglementaire
  - Innovation de gouvernance
  - Innovation sociale



L'IA n'est pas une force irrésistible ; elle est façonnée par nos choix

- les décisions que nous prenons aujourd'hui détermineront si elle devient un outil d'émancipation ou une source de contrôle et d'inégalités.

Le grand paradoxe de l'IA : plus elle devient intelligente, plus nous avons besoin de sagesse humaine



À paraître en février 2026

QUIZ

- Vrai ou faux : un modèle qui satisfait la parité démographique est équitable.
- Le biais peut entrer dans un système d'IA pour le climat par
  - (a) la couverture des données d'entraînement
  - (b) le choix de la variable proxy
  - (c) le cadrage du problème
  - (d) tout ce qui précède
- Dans un système d'alerte précoce, abaisser le seuil d'alerte arbitre principalement entre :
  - (a) exactitude et calcul
  - (b) fausses alertes et événements manqués
  - (c) transparence et vie privée
  - (d) latence et résolution

LECTURES RECOMMANDÉES

- Dignum, *Responsible Artificial Intelligence* (Springer, 2019)
- Note d’orientation Question Zéro de l’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)

AUTRES LECTURES

- **Mes travaux et les ressources Question Zéro**
  - 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 et climat**
  - 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.)
- **Biais, équité, vie privée, redevabilité**
  - 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)*
- **Pratique, cadres et documentation**
  - 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)*
- **Instruments politiques et lignes 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

MERCI !

AI Policy Lab
at Umeå University

SOLUTIONS DU QUIZ

- Vrai ou faux : un modèle qui satisfait la parité démographique est équitable.
- Le biais peut entrer dans un système d'IA pour le climat par
  - (a) la couverture des données d'entraînement
  - (b) le choix de la variable proxy
  - (c) le cadrage du problème
  - (d) tout ce qui précède
- Dans un système d'alerte précoce, abaisser le seuil d'alerte arbitre principalement entre :
  - (a) exactitude et calcul
  - (b) fausses alertes et événements manqués
  - (c) transparence et vie privée
  - (d) latence et résolution