# « Ethique et Intelligence artificielle »

## Webconférence, 2 octobre 2024

Thierry Ménissier
prof. « sciences humaines et innovation »
Chaire éthique&IA MIAI, IphiG
thierry.menissier@univ-grenoble-alpes.fr

# Chaire « éthique & IA »
# 2019-2024

ÉTHIQUE & IA

https://miai.univ-grenoble-alpes.fr/en/

*2019-2024 (62 mois)*

ÉTHIQUE & IA

*Axe 4 : IA & société*

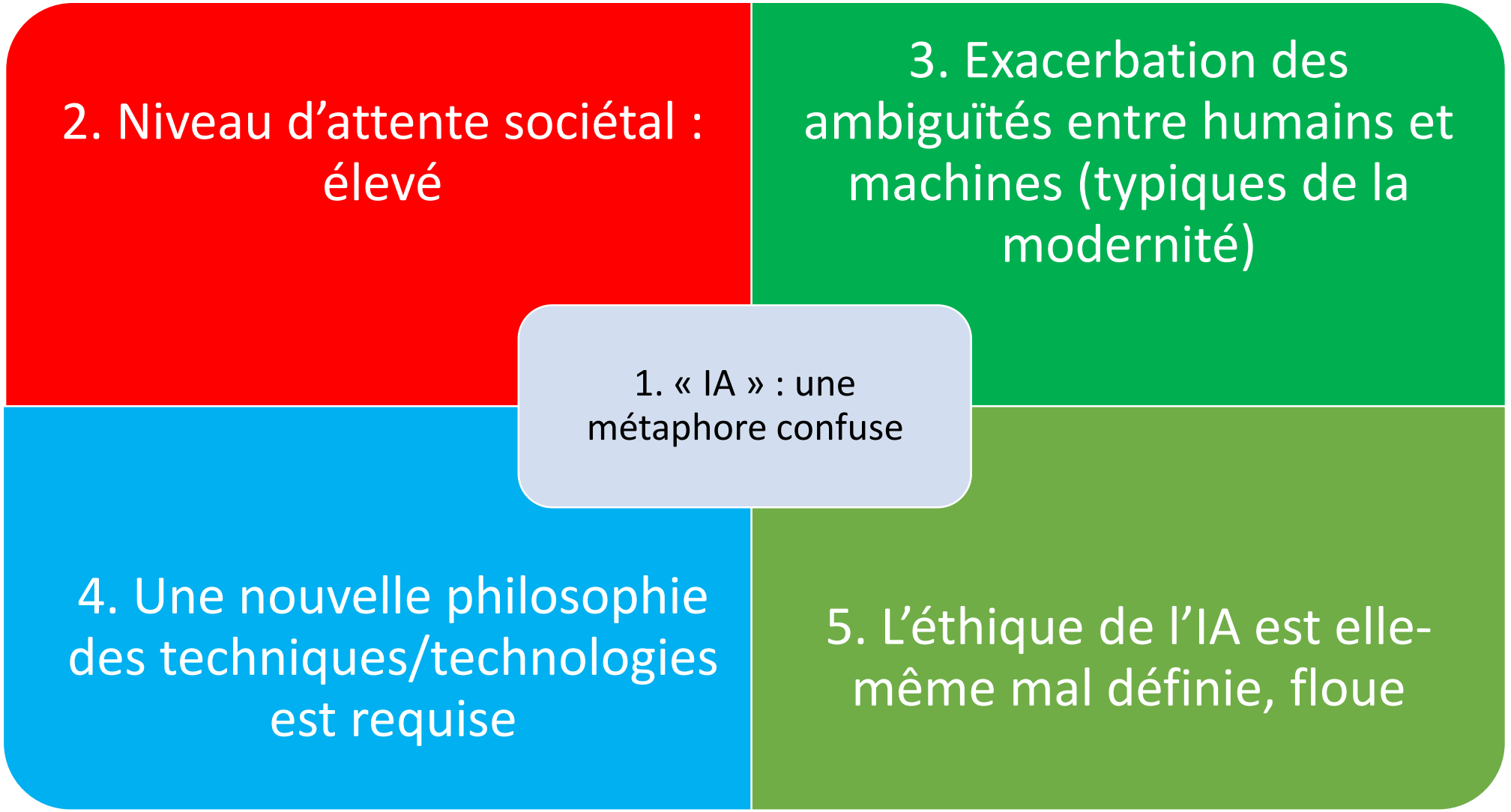
https://www.ethics-ai.fr/la-chaire/

- *12 contributrices/teurs,*
- *3 post-doctorant.es*
- *6 doctorant.es*
- *Soutiens internationaux & partenariats privilégiés : Sherbrooke & Montréal, Naples, Bruxelles*
- *Dialogue de 7 disciplines*
  - *Philosophie, mathématiques-infomatique, robotique, sociologie, psychologie sociale et clinique, sciences de l'information et de la communication, marketing, droit*
- *6 unités de recherche UGA impliquées*

*« Faire dialoguer l'informatique-robotique, la philosophie et les sciences humaines et sociales en vue de la compréhension des enjeux psycho-sociaux, moraux et politiques du déploiement de l'IA, ainsi que la détermination de règles éthiques en vue de solutions compatibles les valeurs de la démocratie. »*

CERAG Université Grenoble Alpes · CRJ Université Grenoble Alpes · GRESEC Université Grenoble Alpes · IPhiG Université Grenoble Alpes · Pacte Laboratoire de sciences sociales · TIMC

# *5 hypothèses de recherche*



CONFÉRENCE

L'intelligence artificielle, le défi éthique.

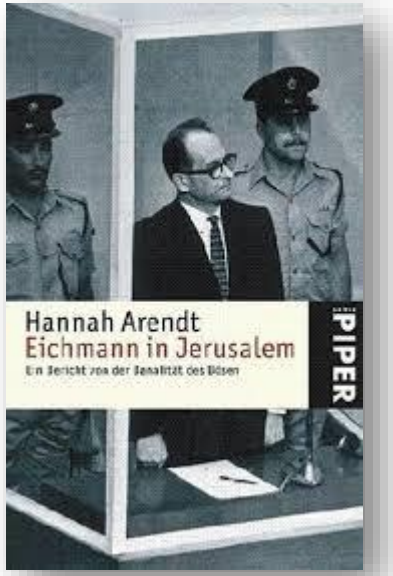
12 janvier 2022
De 17h à 18h30
Amphis 1 et 3

Par
**Thierry Menissier**
Responsable de la chaire éthique et IA, Grenoble.
Philosophe français, spécialisé en philosophie politique et en histoire des idées.

INP PURPAN ÉCOLE D'INGÉNIEURS
Sciences du vivant | Agriculture
Agroalimentaire | Marketing | Management

# ***Une nouvelle forme de « banalité du mal »*** (Chomsky et al., 8/03/2023)

OpenAI

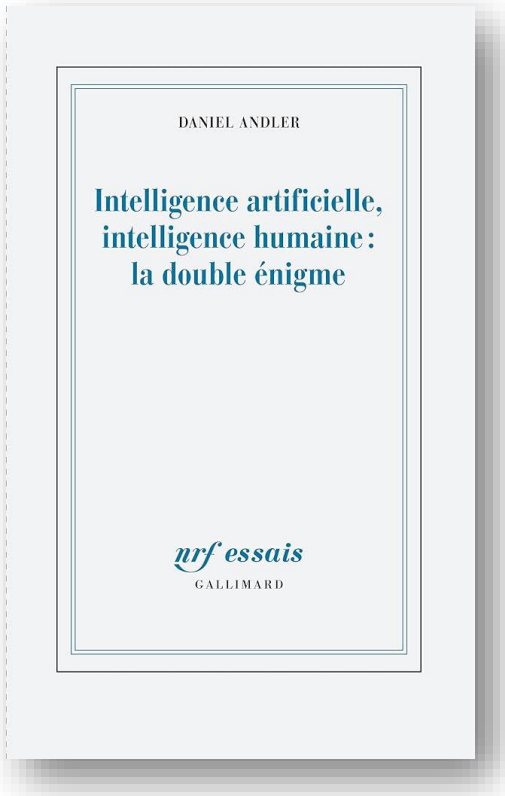



Arendt, 1963

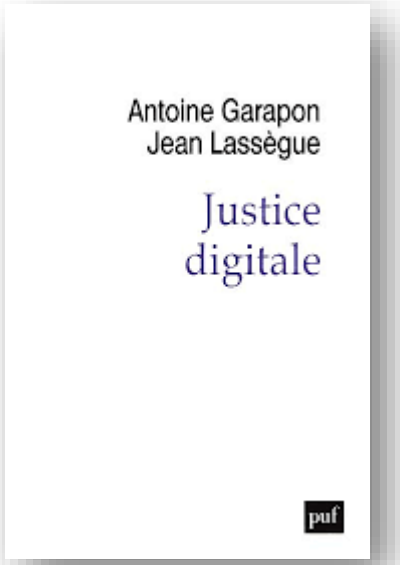
The New York Times

- Une représentation qui interroge…
  - Auguste Rodin, *Le Penseur*, 1902

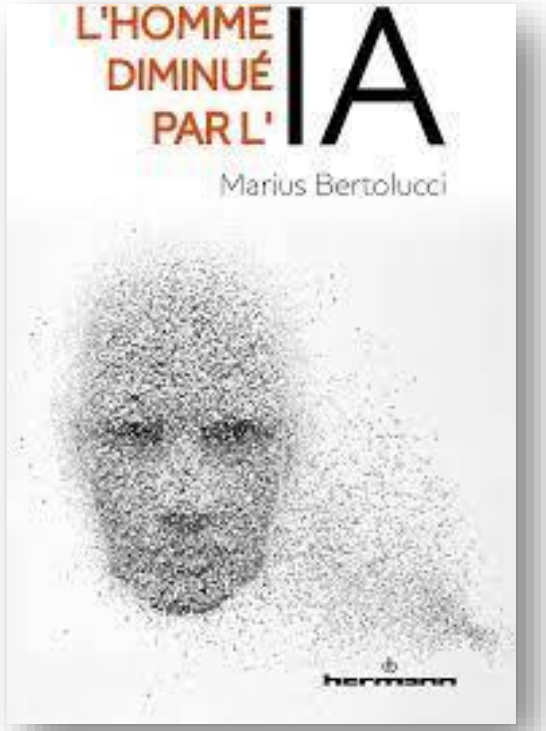
# Comment qualifier la situation actuelle ?












Fascinante : Andler, 2023

Catastrophique : Garapon & Lassègue, 2018, 2021

Totalement désespérée : Bertolucci, 2023

Grave mais pas désespérée
(à condition d'avoir de la culture) : Grinbaum, 2023

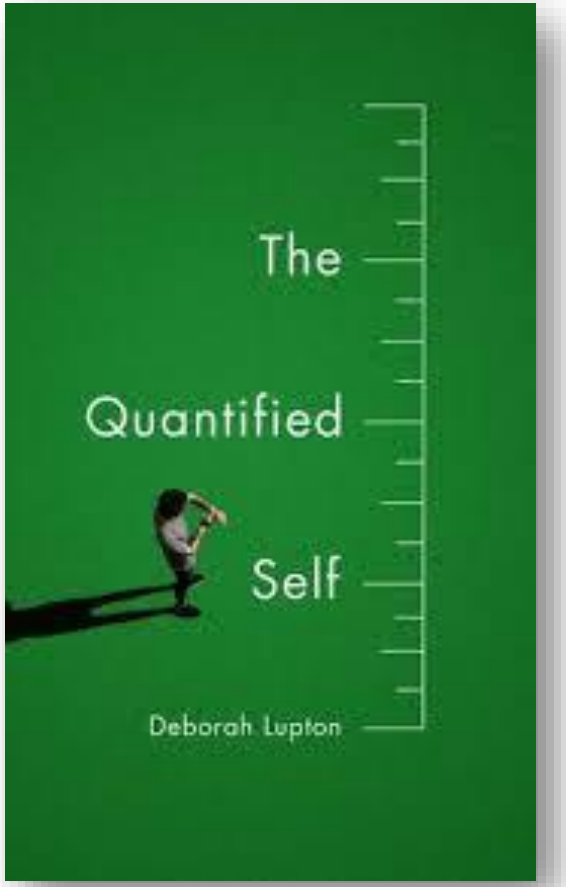
# Plusieurs tentatives pour créer de nouveaux concepts : caractériser les technologies mobilisant l'IA, le numérique & les données

- « *Catopticon* » & « sous-veillance » (Ganascia, 2009, recension Zetlaoui, 2010)
  - Servitude librement consentie (Martin-Juchat & Pierre, 2011 )
  - Servitude volontaire dans la société algorithmique (Ménissier, 2021)
- « Gouvernementalité algorithmique » (Rouvroy & Berns, 2013),
  - Cf. les manières de « **gouverner sans gouverner** » (Berns, 2009) : « archéologie politique de la statistique »
- « Datacratie » (*Pouvoirs*, 2018)
- « *Algocraty* » (Danaher, 2016, 2022)
- « *Stack* » (« empilement ») (Bratton, 2016, trad. 2019 ; recension Giroud, 2020)
- « *Age of Surveillance Capitalism* » (Zuboff, 2019)
- « Empire du signal » (Chardel, 2020)
- « Sociétés du profilage » (Huneman, 2023)
- « Société de contrôle » / Deleuze & Foucault (Razac, 2023)

« Hypnose technologique »

Alombert & Giraud, 2024

# *LA TECHNOLOGIE COMME QUASI-MILIEU VITAL*

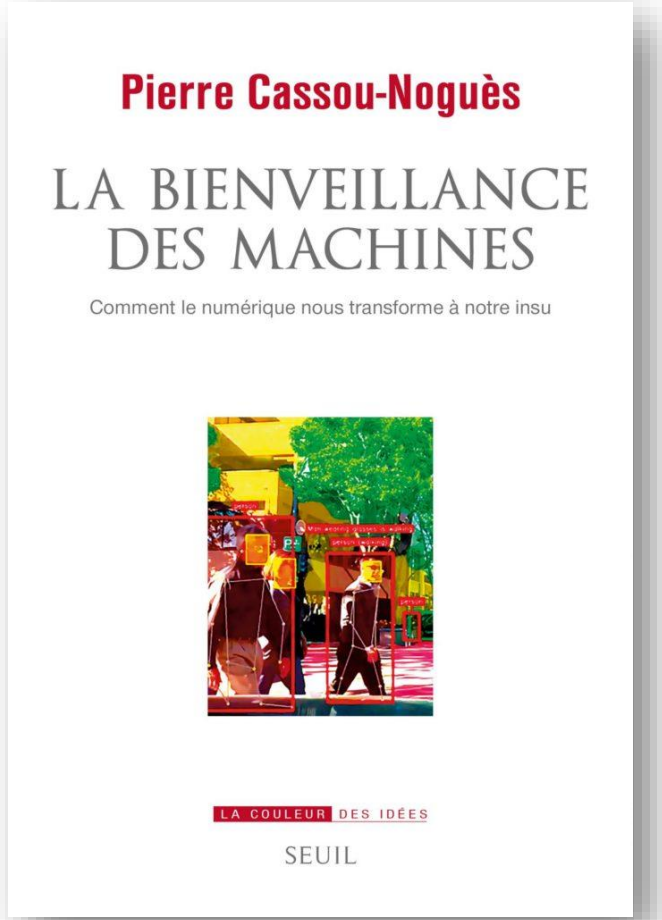



Lupton, D., 2016 : *The Quantified Self. A Sociology of Self-tracking*

*Quantified Self* : toutes les manifestations par lesquelles un individu mesure ce qui peut l'être dans son activité corporelle *via* des capteurs variés (pouls et tension artérielle, sudation, Nombre de pas effectués, etc.) et rythme sn activité par des « alertes » qui contribuent à soumettre son existence à la double normativité biologique et technique. Syn. : « métronomie du moi »

Les machines nous « bienveillent » (néol.).

- sollicitude *by design*
- pouvoir croissant d'automatisation
- influence qui s'étend sans limitation claire ni définissable *a priori*.
- On peut émettre le soupçon qu'elles manifestent un « intérêt » à étendre sans fin leur pouvoir d'action, et même qu'elles utilisent les humains pour y parvenir. Suite ici.



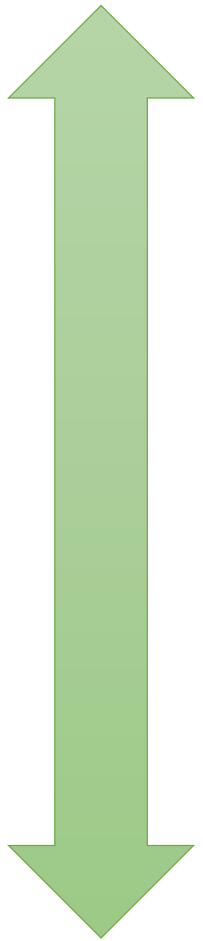

Cassou-Noguès, P., 2022 : *La Bienveillance des machines. Comment le numérique nous transforme à notre insu*

*Que signifie, dans ce contexte,*
*œuvrer à l'éthique de l'IA ?*

# “Les 4 éthiques de l’IA” : Ménissier, 2023

REVUE d’ANTHROPOLOGIE des CONNAISSANCES

1. Computer ethics : éthique informatique
   - Design of AI systems: computer scientists
2. Artificial/algorithmic/robotic ethics
   - Design, programming & uses of autonomous intelligent machines: roboticists, users
3. Digital & data ethics
   - Design and use of platforms in the innovation dynamic of the digital economy: economic players, users
   - Creation and data management: users, public authorities, economic players
4. UX AI ethics : éthique des usages de l’IA
   - Relationships between AIS and their meaning/values in a society: philosophers, users/citizens, public authorities
   - Co-design, participation, stakeholders’ involvement



***Primauté implicite du raisonnement utilitariste***

***Mobiliser d’autres formes de raisonnement éthique***

# *Computer ethics* : Travailler sur/avec les valeurs qui devraient régir l'écriture du code

## Quelles valeurs ?

A/ valeurs épistemiques

- Explicabilité
- Interprétabilité
- *Accountability :* redevabilité
- Transparence

B/ valeurs sociales

- *Trustworthiness & Trust :* fiabilité & confiance
- *Responsible Artificial Intelligence*
- *Fairness :* équité

## Activation des valeurs ?

Très nombreuses chartes & declarations !

IBM Meta GENERALI

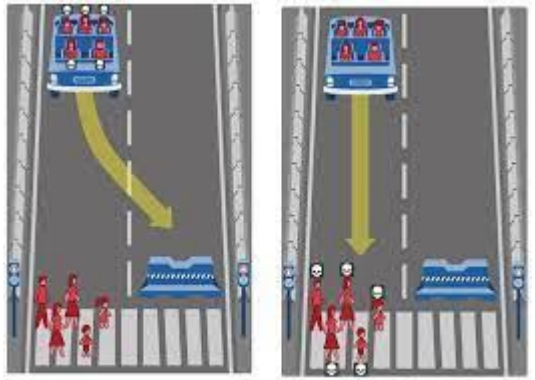
Une experience remarquable : la Déclaration de **Montréal pour un développement responsable de l'IA** (2018)

acm Association for Computing Machinery

A. Barredo Arietta *et alii*, « Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI », 2019

# Éthique artificielle

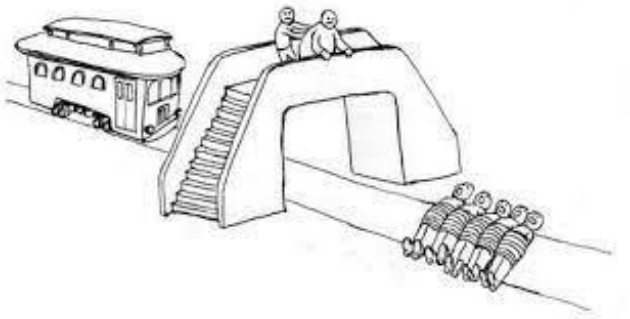
## Le fallacieux paradigme du véhicule “autonome”



**MIT Moral Machines experiment**

https://www.moralmachine.net/

Awad *et al.* (2018)




nature
International weekly journal of science

Toutefois,

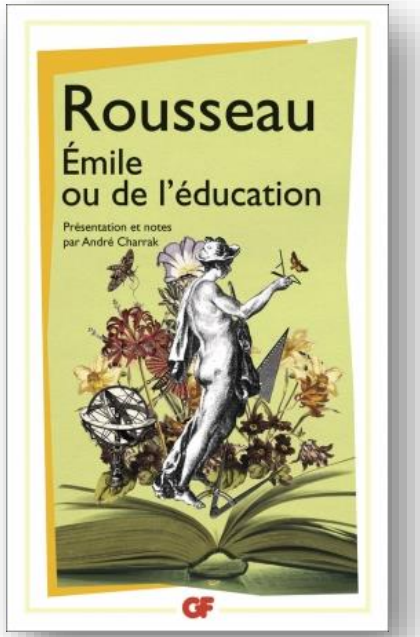
- pour de nombreux éthiciens, le dilemme du tramway **n'a pas de signification éthique :** Philippa Foot, « The Problem of Abortion and the Doctrine of the Double Effect », 1967
- Aucune compréhension des contextes culturels : les informaticiens ne dialoguent pas encore avec les anthropologues

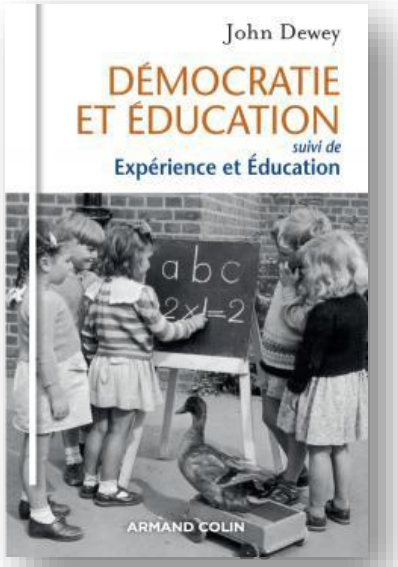
# Le défi spécifique de l'éthique artificielle

## Éduquer TAY !

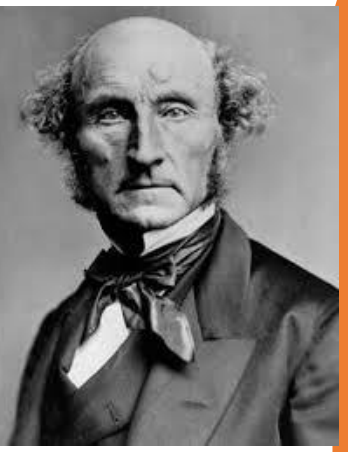
- Mars 2016 : l'agent conversationnel Tay, au bout de 24 h d'autoapprentissage sur Twitter, tient des propos inappropriés (sexiste, raciste, antisémite, complotiste)
  - https://www.lemonde.fr/pixels/article/2016/03/24/a-peine-lancee-une-intelligence-artificielle-de-microsoft-derape-sur-twitter_4889661_4408996.html
  - https://www.latribune.fr/technos-medias/internet/tay-l-intelligence-artificielle-raciste-et-sexiste-de-microsoft-559646.html


- *Comment éduquer des êtres à la fois artificiels, nouveau-nés, « à fort potentiel » quant à leurs facultés de calcul & en interaction sociale permanente ?*
  - Une question pédagogique classique



# Pluraliser les formes du raisonnement éthique



## Conséquentialisme

Calculer l'intérêt individuel & collectif

## Déontologisme

Se donner des règles

éthique

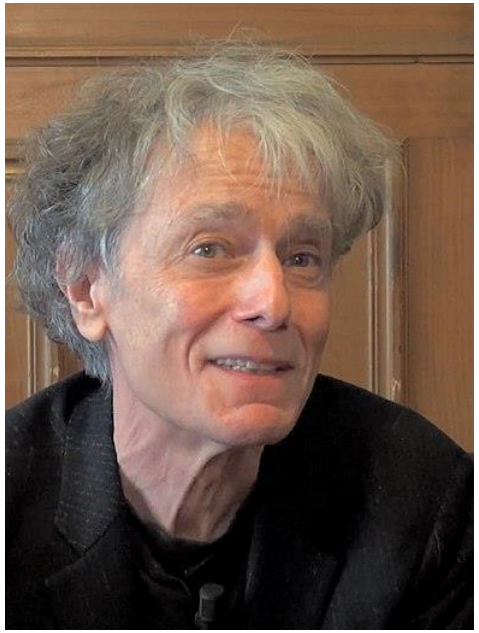
## Arétaïsme ou Perfectionnisme moral

Devenir une personne meilleure

## Axiologisme

Se conformer à des valeurs estimées supérieures

# Minimalisme *vs*. « maximalisme » éthique



*1947-2017*

- Ruwen Ogien, 2007, 2013
- Minimalisme : éthique conséquentialiste, d'inspiration libérale, antipaternaliste
  - Principe de non-nuisance envers autrui
- *Vs*. Supposé maximalisme

(donc impossible à réaliser)

  - Devoir envers soi-même
  - Recherche de l'excellence : devenir moralement meilleur(s)
  - Éducation morale


# Une approche non étroitement pragmatique des technologies (contrer le “technosolutionnisme”)

**Est-ce possible ?**

**Oui**

*Résister aux éventuels effets néfastes des technologies / soutenir leur développement pour de meilleures pratiques*

Deontologisme

Arétaïsme

Vallor, S. (2016).*Technology and the Virtues. A Philosophical Guide to a Future Worth Wanting*

Axiologisme



**Est-ce pertinent ?**

**Ça dépend…**

Haut degré d’ambition de la démarche

1. *Veiller à ce que les technologies ne nous empêchent pas de mener une vie morale*
2. *Devenir meilleur sur le plan éthique grâce à la technologie*

*Ethics by design*

*Design sensitive values*

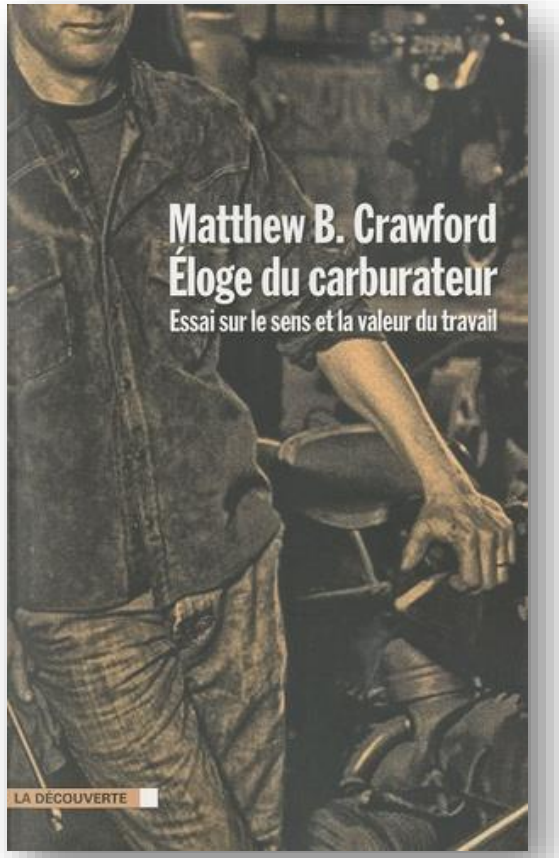
Critique du **« monopole radical »**

Sur-outillage individuel & consumérisme industriel

*Vertu supérieure :* ***joie & austérité***

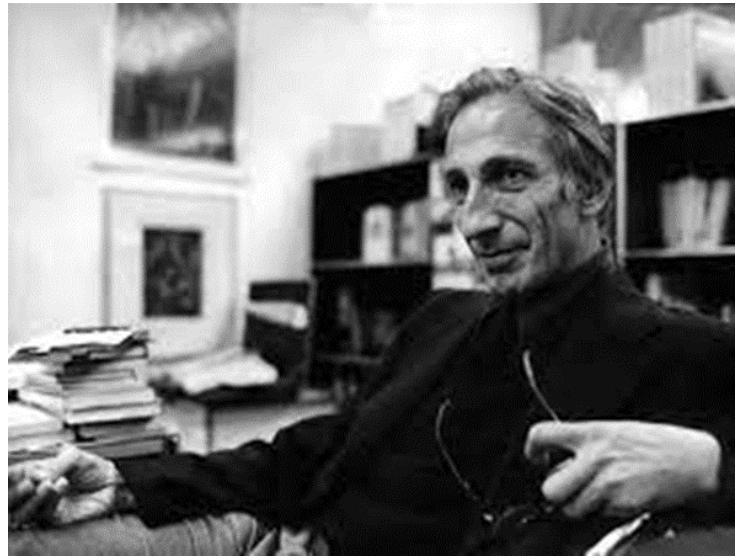
Cf. Thomas d'Aquin, *Somme Théologique,* IIa IIae, q. 164, art. 4 *:* le défaut de jeu confine paradoxalement au péché

Cf. *eutrapélia* [εὐτραπελία], Aristote, *Eth. Nic.*, 1108 a 24



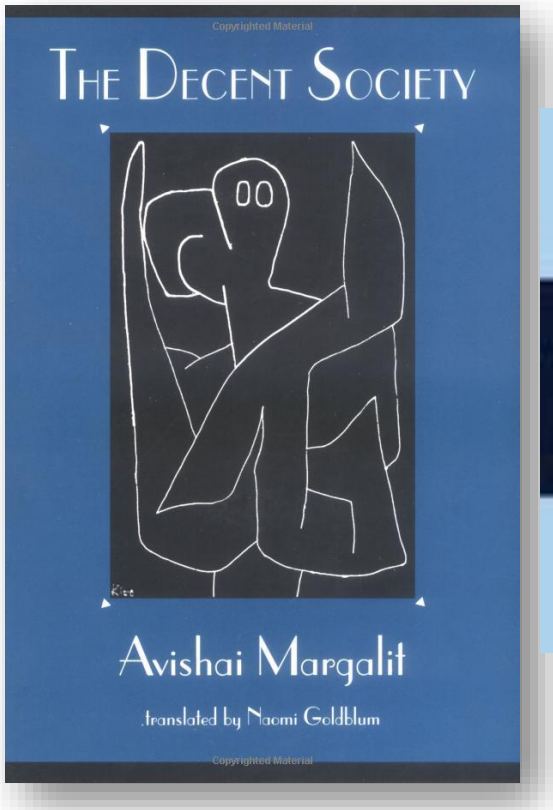

Crawford, 2009

« CONVIVIALITÉ »
Ivan Illich, 1973

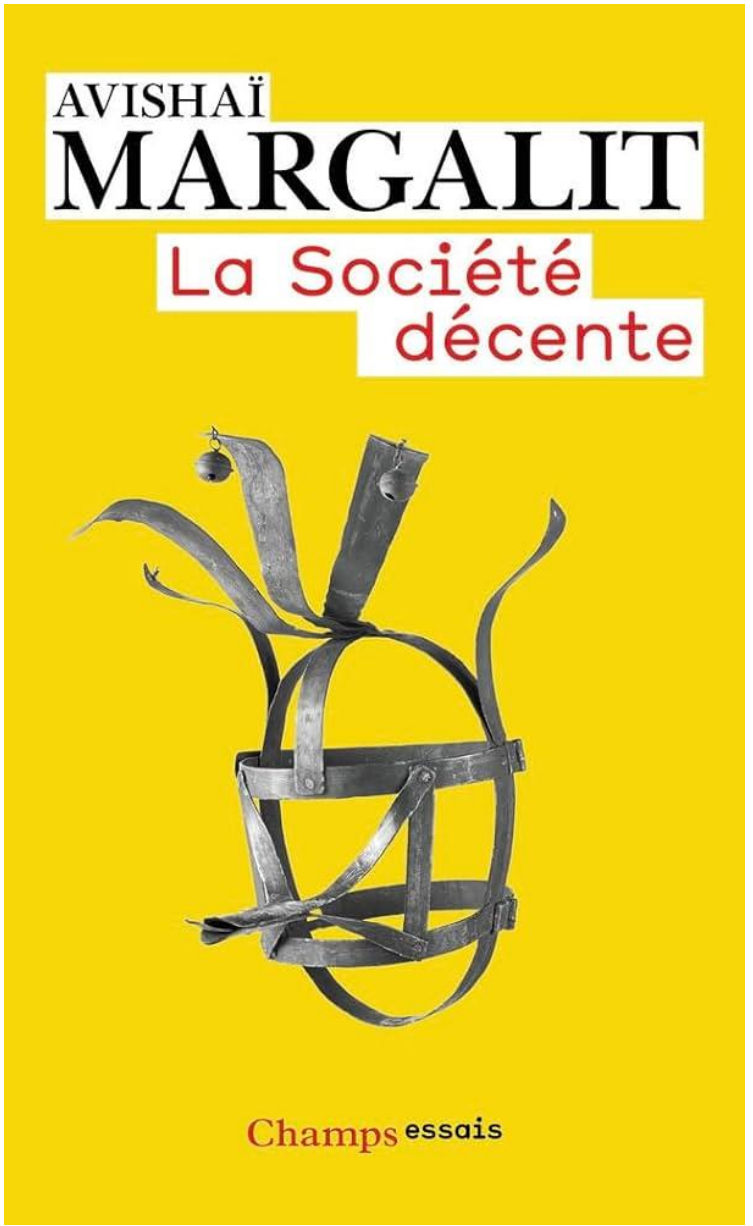


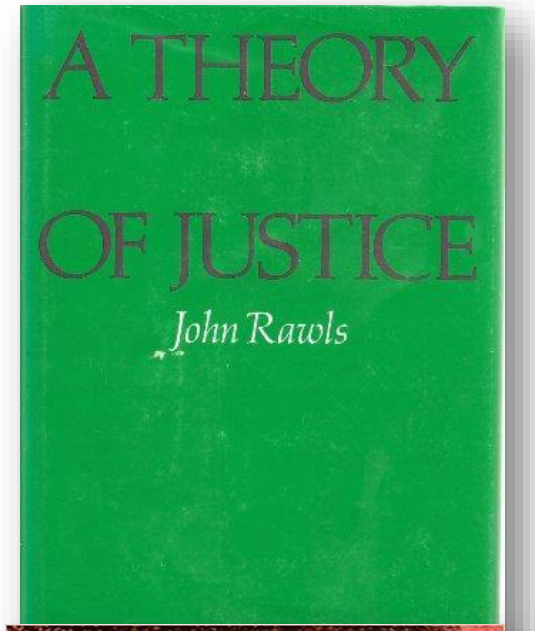
*« J'appelle société conviviale une société où l'outil moderne est au service de la personne intégrée à la collectivité, et non au service d'un corps de spécialistes »*

# « Société décente » (Margalit, 1996)

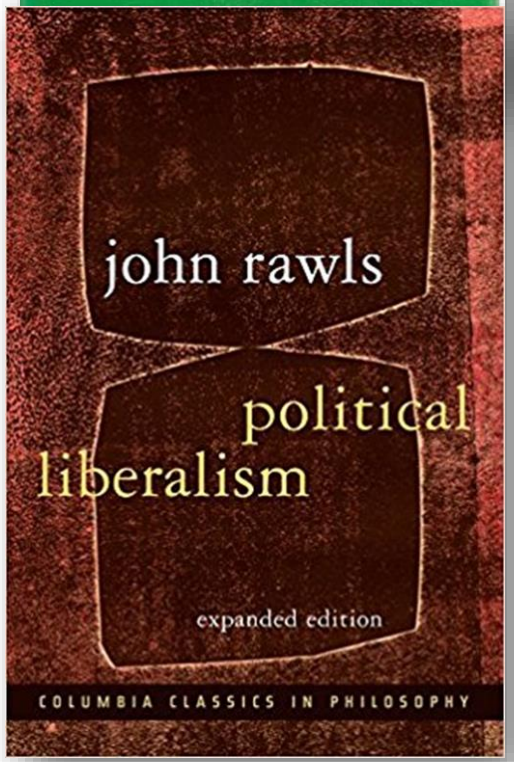




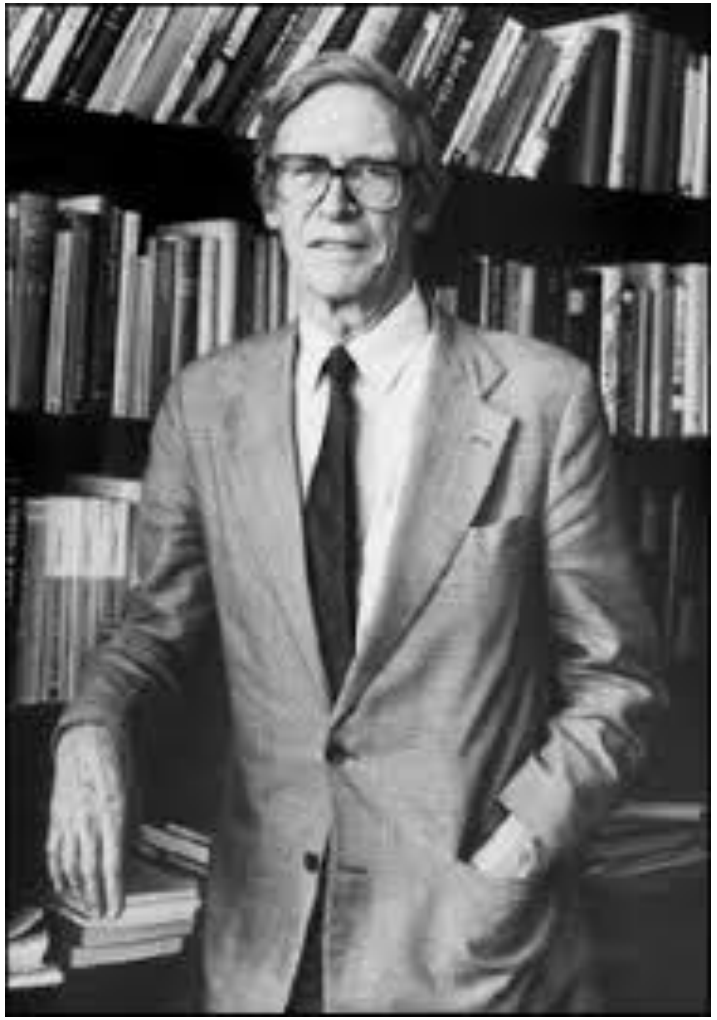





*A Theory of Justice* (1971)




*Political Liberalism* (1993)



John Rawls (1921-2002)

# Questions de **philosophie pratique** : éthique, politique, technologique

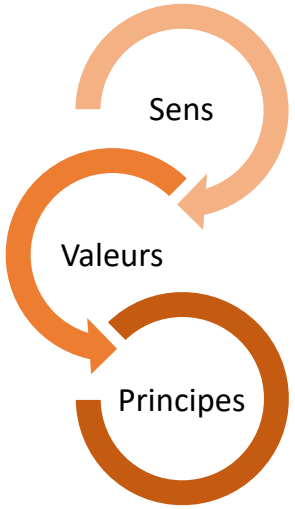
Rawls : *Quelle théorie de la justice distributive pour la société démocratique ? = A quelles conditions une distribution de biens peut-elle être équitable et acceptée par des participants raisonnables ?*

Margalit : *Comment faire en sorte de parvenir à une société décente, à savoir, une société dont les institutions ne sont pas susceptibles d'humilier les/certains citoyens ?*

Nous : ***Comment développer des systèmes d'IA efficients, fiables, utiles et susceptibles de ne pas être nocifs ou dangereux pour ceux qui les utilisent ou les subissent et qui ne dégradent pas leur dignité sociale et humaine ?***

# Tâches à accomplir tout au long du processus de co-design

**Dimensions éthiques**




- Construire une définition commune de l'objet technique
- Comprendre les critères d'évaluation des différentes parties prenantes
- Faire dialoguer les différentes formes de raisonnement éthique
- Mettre en œuvre la participation des différentes parties prenantes

**Exemple**

RealWorld4Clinic

Identification des futurs utilisateurs
Identification des risques et des bénéfices du point de vue des patients
Définition du parcours et de l'équipe de soin idéale

Utilitarisme : amélioration des indicateurs médicaux-économiques
Déontologisme : respect des principes de la bioéthique
éthique du care
Axiologisme : conformité à un idéal démocratique
respect de l'intégrité physique
Arétaïsme : éducation thérapeutique du patient

18 séances de co-design avec le groupe de patients partenaires
26 entretiens complémentaires avec des patients
2 ateliers
1 journée de conférence

**Pratiques de co-conception impliquant des procédures de formation et d'évaluation éthiques**

1. There is **no universal use** for AI
2. All ethical assessment is **culturally determined**
3. Workshop participants must demonstrate **ethical & cultural tolerance**

*The ethics of/for co-design: a problematic triangle*

Common definition of the technical system/device to be assessed


Participation implementation

Choice of ethical assessment principles

Davat, Martin-Juchat, Ménissier, **Co-design with affect stories and applied ethics for health technologies**. 2024

*A safe space for inventing* ***common sense*** *about technology*

- Expérimenter collectivement en incluant des formes enrichies et approfondies de raisonnement éthique
- Civiliser la société algorithmique

*Merci pour votre attention !*