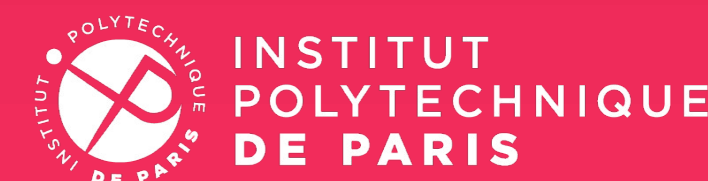# AI as a sociotechnical system

Tiphaine Viard – Associate professor
Numérique, Organisation et Société,
Computational Social Science @ IP Paris

tiphaine.viard@telecom-paris.fr
github.com/TiphaineV
https://tiphainev.github.io
Mastodon: tviard@lipn.info

Before we start, a few questions...

Scan the QR code or
go to wooclap.com

Enter code: XDANPG

Answers are anonymous and
aggregated

Teaching team

Tiphaine Viard
Sociology of AI, computational social sciences
PhD in computer science
3A319 – tiphaine.viard@telecom-paris.fr

Thomas Le Goff
Digital law, regulation of AI
PhD in law
3A321 – thomas.legoff@telecom-paris.fr

Mathilde Abel
Sociology of AI, data workers, natural langage processing
PhD in economics
3A342 – mathilde.abel@telecom-paris.fr

Please copy all of us + put the course code (5DA06) in exchanges

This maximises the probability of a fast reply and minimises loss of information :)

What is AI?

- One of the 5 **promises of technology** (WEF)

- A booming **research** and **development** field

- A **prism** for **bias, inequalities and power struggles**

- **Successes** and **failures** are both **flashy** and **silent**

  - AlphaGo**,** spam filters, COMPAS, SyRI…

- One challenge: separating **hype** from **reality**
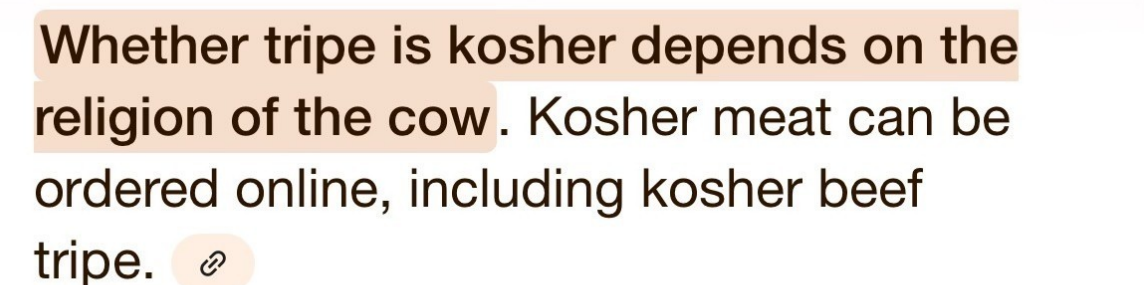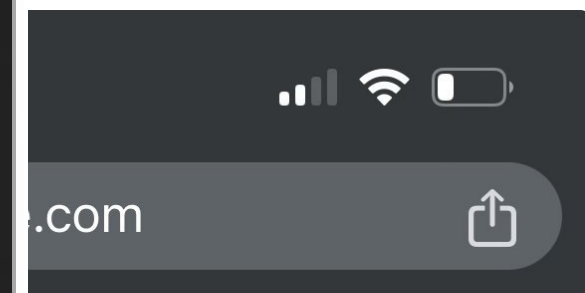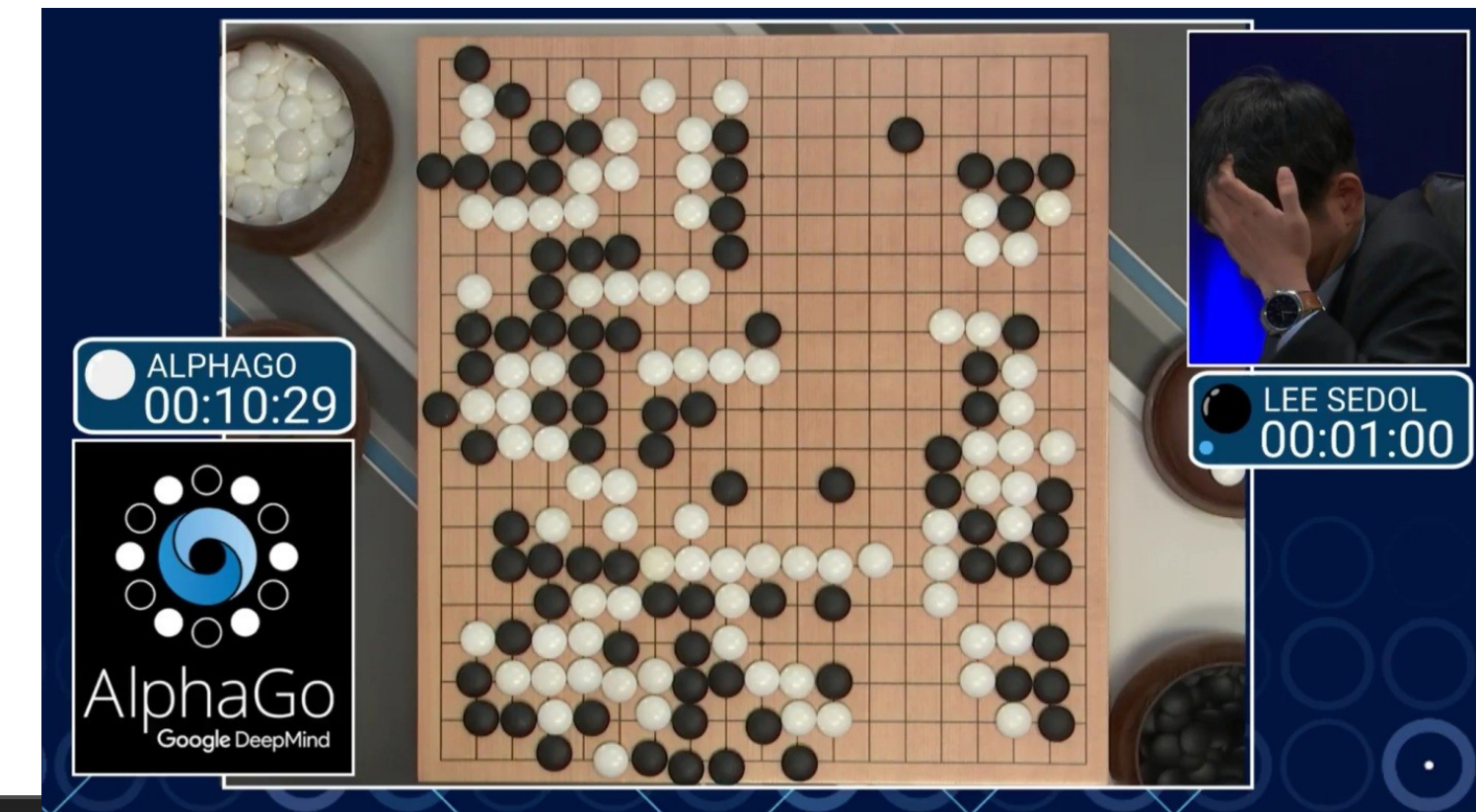
What is AI?

- "Which metric should we use, micro- or macro-F1 score?"

- "When is deep learning more suited than non-deep models?"

- "Which rules should govern 'AI' use in high-risk systems?"

- "Will AI exterminate humanity?"

- "We use 'AI' to determine recidivism risk"

- "How do scientists conciliate AI's energy consumption with a sustainable world?"

- "Who talks about 'AI' and what do they say?"

- "We want fair AI, but fair to whom?"

**AI = the general concept that encompasses systems
that learn inferences from data**

What we will not talk about

- Methods for debiasing
- Methods for explainable AI
- Fairness metrics and algorithms

Lectures in this course

- Today: introduction and **AI as a sociotechnical system** (Tiphaine)

- The **european regulation of AI: GDPR and the AI Act** (Thomas)

- The regulation of **generative AI** (Thomas)

- The materiality and specificities of **generative AI** (Tiphaine)

- The **users of AI** (Tiphaine)

- **04/11**: Case study presentation / jury questions

- **13/11**: Table exam
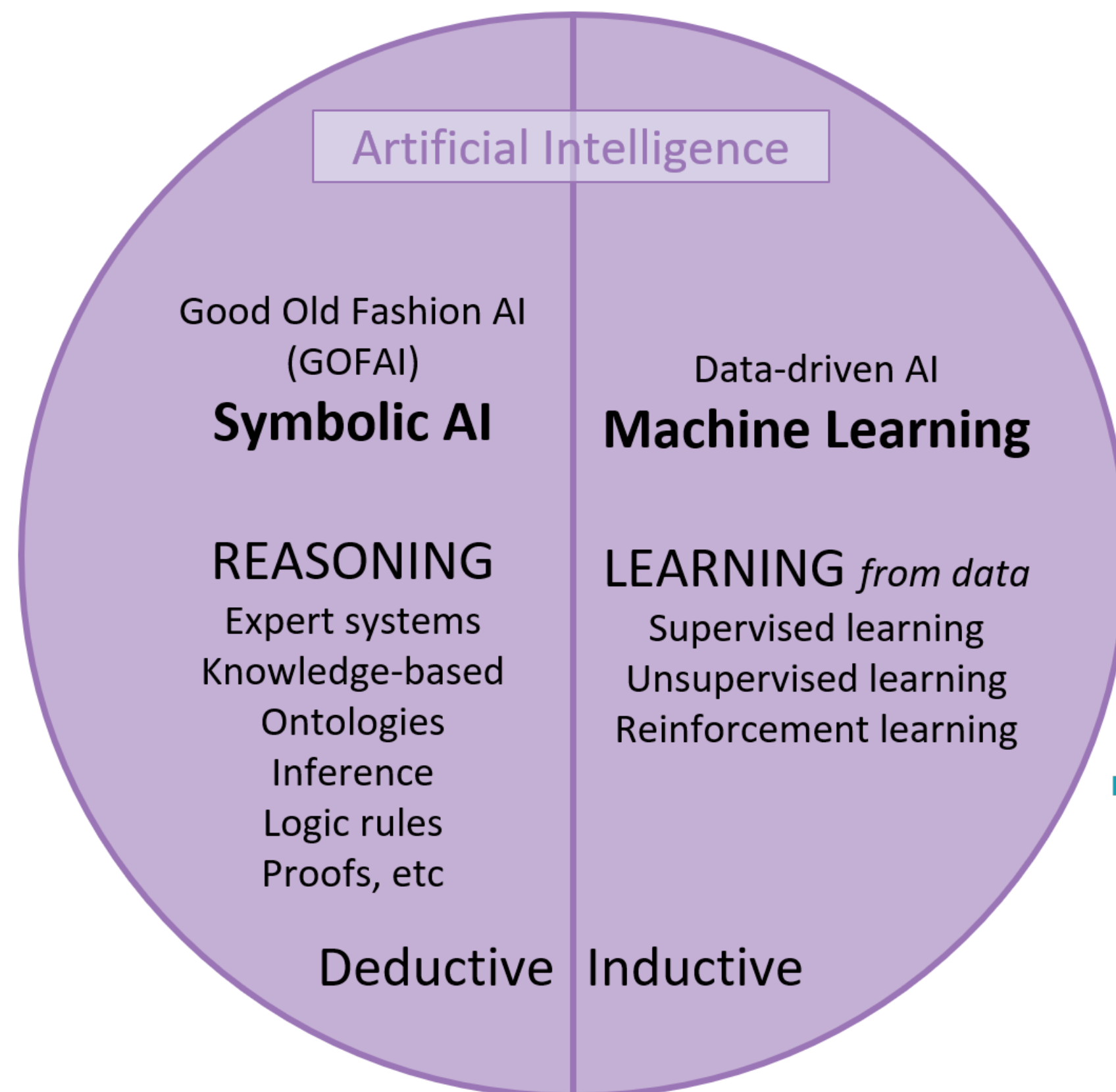
About the evaluation of the course

- **Table-based exam** (50%); topics covered in lectures and case analysis

- **Case study** (50%):

  o On a real-case topic: CAF, Passenger Name Records, facial recognition...

  o Work in groups of 5 students on a case study with **different roles** (compliance, computer science, cybersecurity, ethics...)

  o Come up with a **proposal** and defend it

  o Restitution: a **7mn-video** + a **defense** in front of a jury

      o For each group, 5 "opponents" will contribute to asking questions and clarifications

  We are interested in how and why you make trade-offs, rather than the solution itself
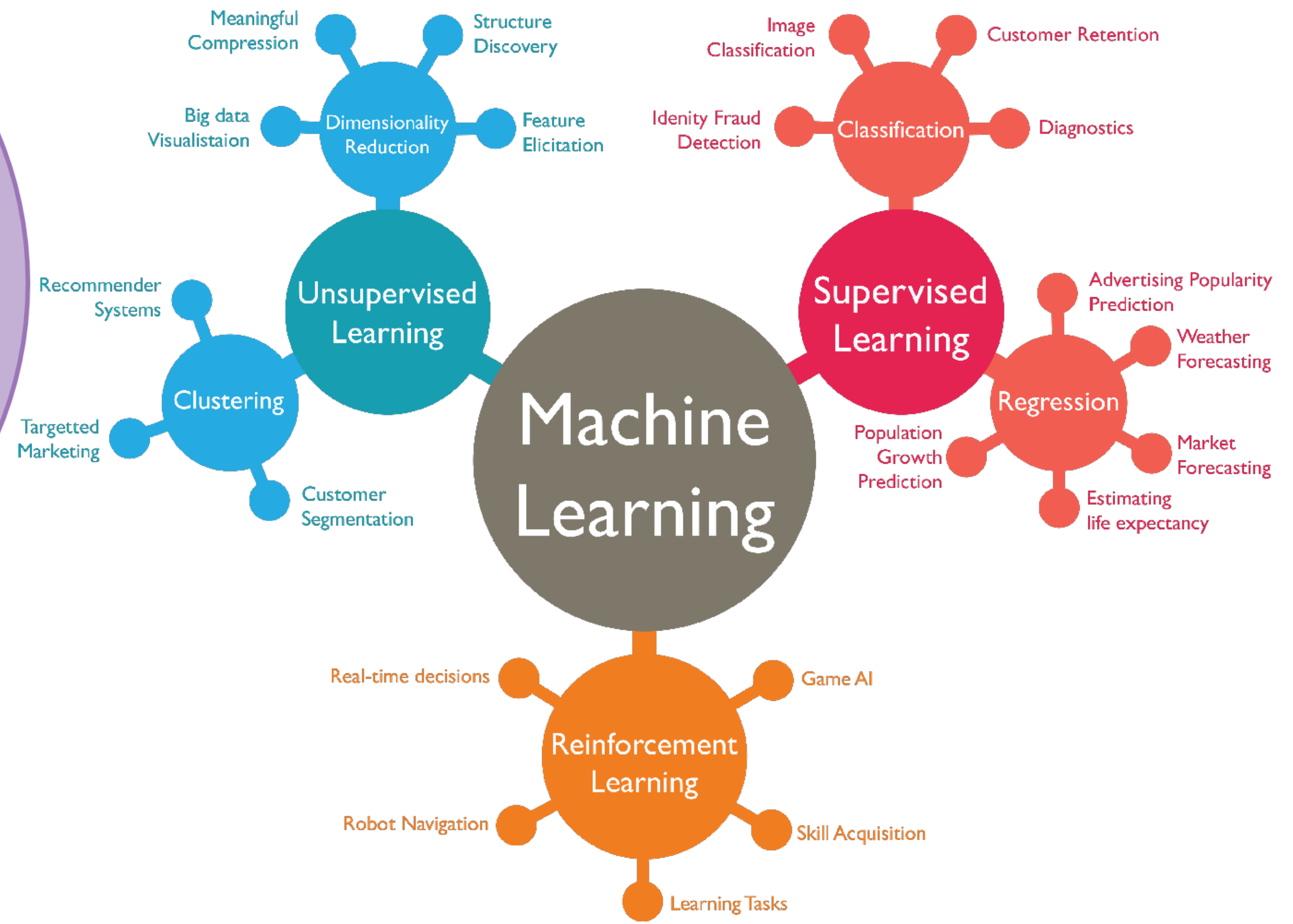
This lecture

- Exhibit the **tension** between "artificial intelligence" and "machine learning"

- Get a broad overview of the **sociotechnical system** that is AI and its **history**

- Get an idea of the **promises and critiques** related to AI technology

- Understand the **definitional conflicts** around AI

- Understand how the **landscape** of AI is shaped: which topics are discussed? Which tenets are central? How do actors **interact** and **negotiate meaning**?

- I will speak at the crossroads of **computer science** and **science and technology studies**

Some AI classifications

Artificial Intelligence

Good Old Fashion AI (GOFAI)
Symbolic AI

REASONING
Expert systems
Knowledge-based
Ontologies
Inference
Logic rules
Proofs, etc

Deductive

Data-driven AI
Machine Learning

LEARNING *from data*
Supervised learning
Unsupervised learning
Reinforcement learning

Inductive

Useful but **not sufficient**

Machine Learning

Unsupervised Learning

Dimensionality Reduction
- Meaningful Compression
- Structure Discovery
- Big data Visualisation
- Feature Elicitation

Clustering
- Recommender Systems
- Targetted Marketing
- Customer Segmentation

Supervised Learning

Classification
- Image Classification
- Customer Retention
- Idenity Fraud Detection
- Diagnostics

Regression
- Advertising Popularity Prediction
- Weather Forecasting
- Market Forecasting
- Estimating life expectancy
- Population Growth Prediction

Reinforcement Learning
- Real-time decisions
- Game AI
- Robot Navigation
- Skill Acquisition
- Learning Tasks

The social space defined by AI

The social space defined by AI is shaped by the political and regulatory contexts, technical realities, our common imaginaries

- Study the interplay between **objects and persons as a system**

- Actors shape the **social space** in which they are embedded

- Identify actors and their roles

"it is essential to clarify the role of actors who may contribute to the development of AI systems, notably high-risk AI systems" (AI Act)

- How do actors **position themselves** in this conceptual space? charters, manifestos, regulations...

- Study the operations of variation and selection that lead to innovative **stabilisation**, while allowing for **interpretive flexibility** and **closure**

# Sociotechnical systems and their analysis

- Studying technical objects in society: the focus of **Science and Technology Studies (STS)**

- **Framing** is important: where do technical objects start and end?

- The **reception** of technical objects cannot be ignored: uses emerge, evolve and disappear

- Citizens' attention, policy makers' attention are limited resources

- Decisions are taken in **sub-political spaces** (Beck, 1986)

Methods for sociotechnical system analysis

- **Framing** is important: where do technical objects start and end?

- **Ethnography** and **on-site interviews**

- **Discourse analysis**

- **Computational** methods

- Creating **sociotechnical imaginaries** (utopias and dystopias)

- Focus on **controversies** and **points of tension**

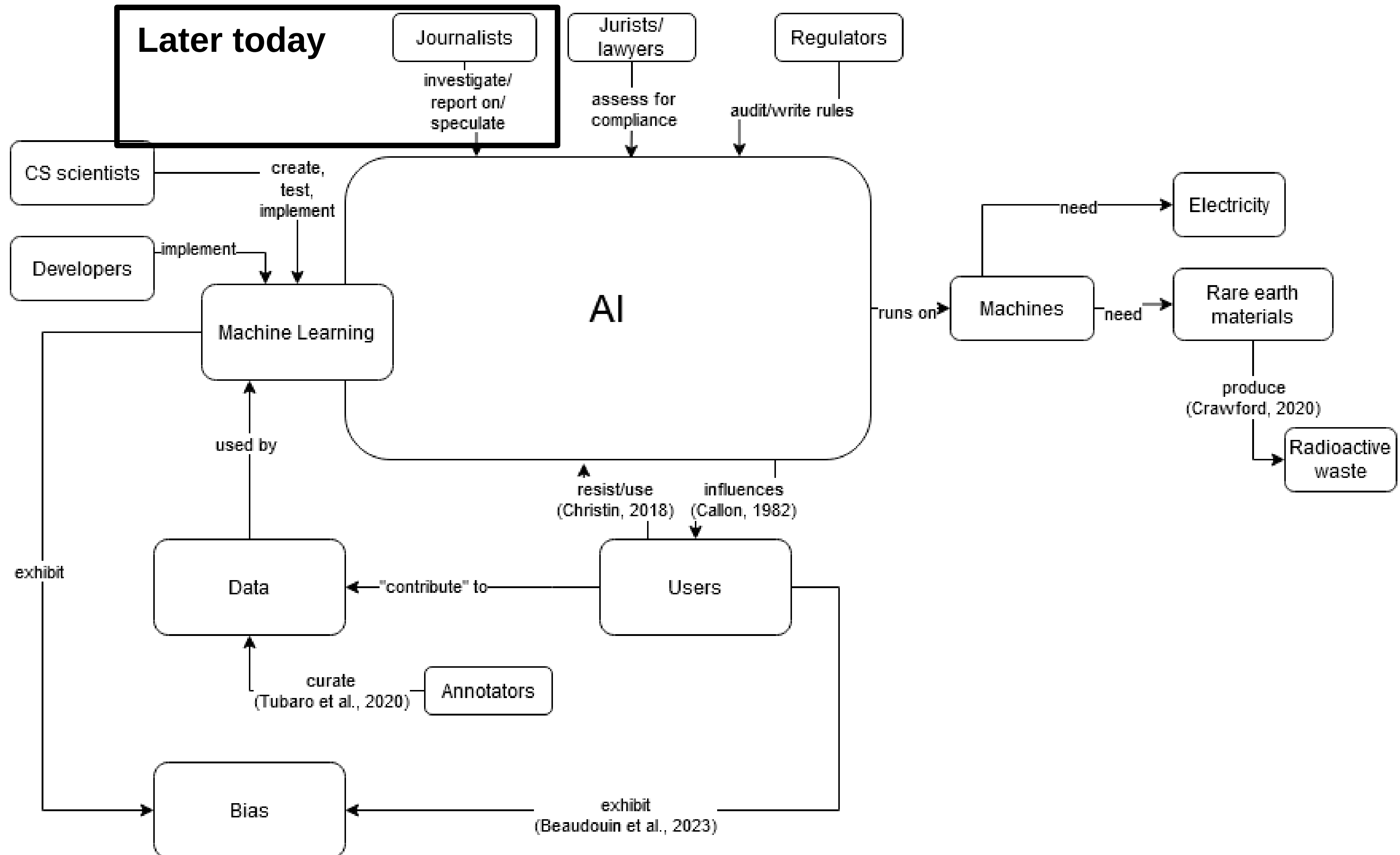- Comparing **heterogeneous spheres of information**: press, science, regulation, civil society
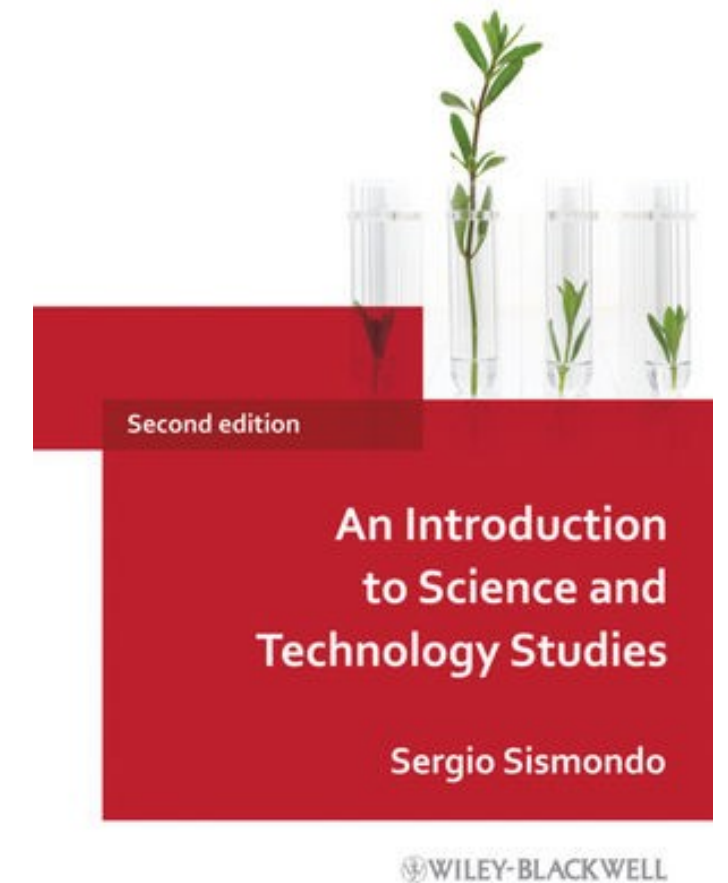
The (incomplete) AI sociotechnical system
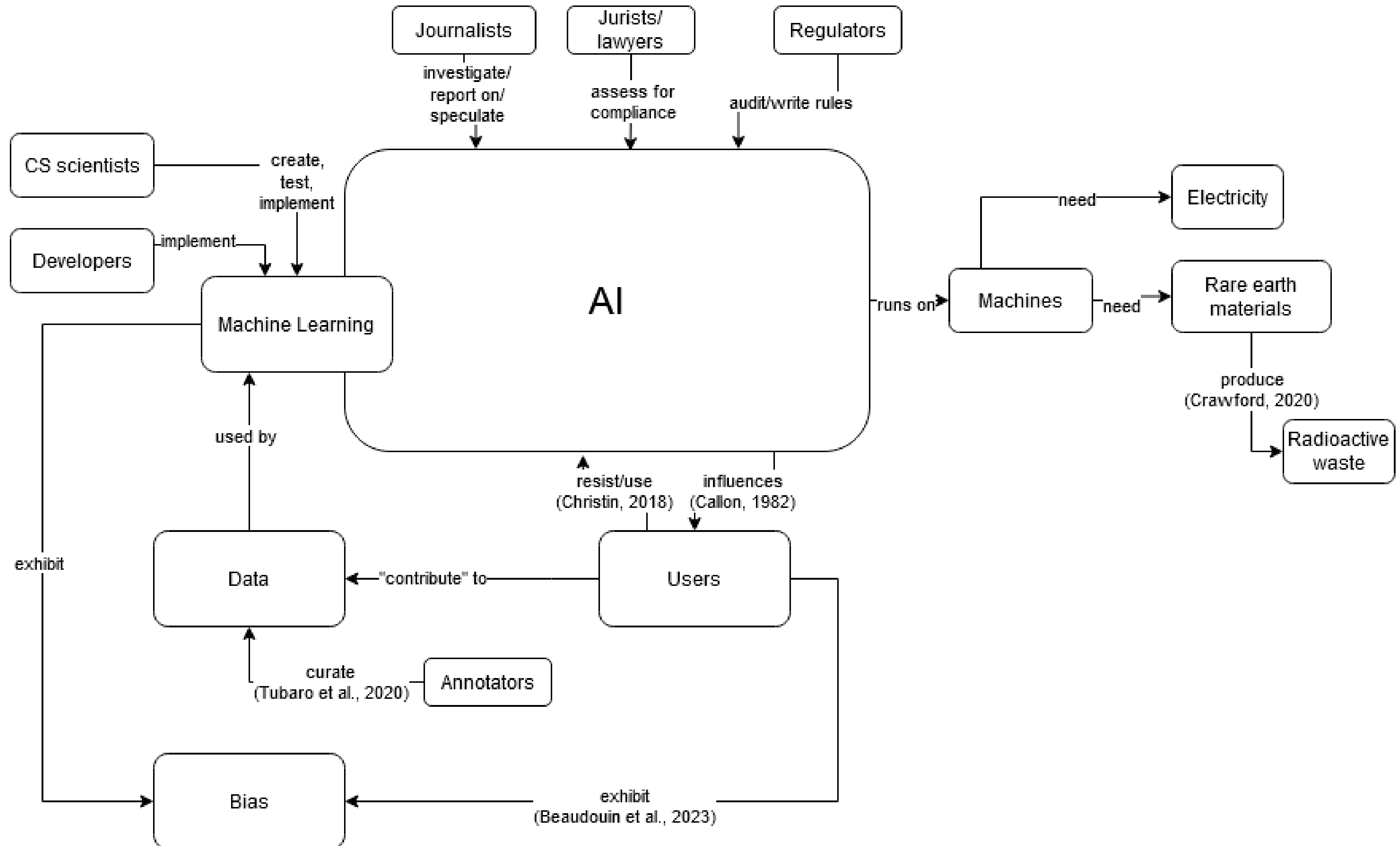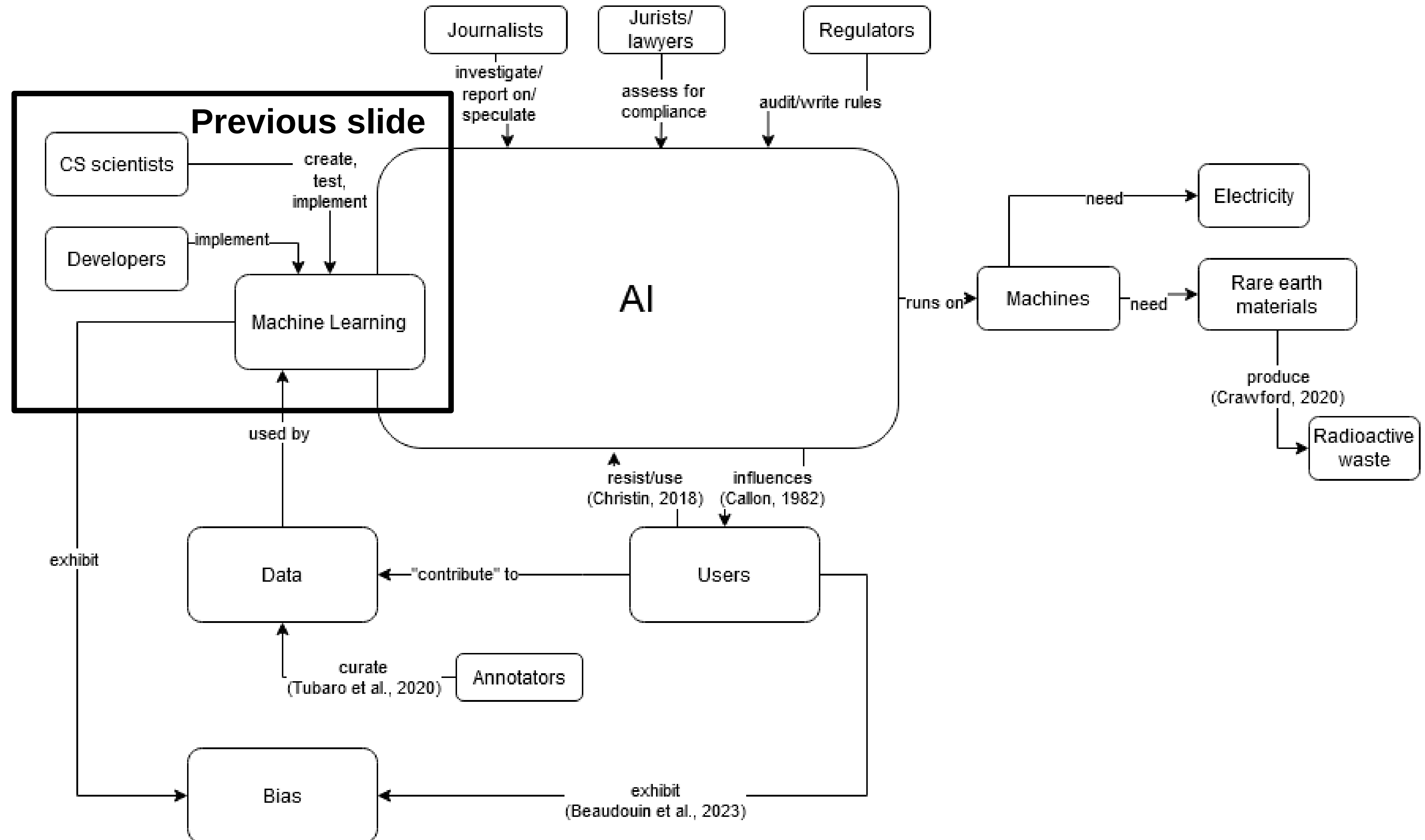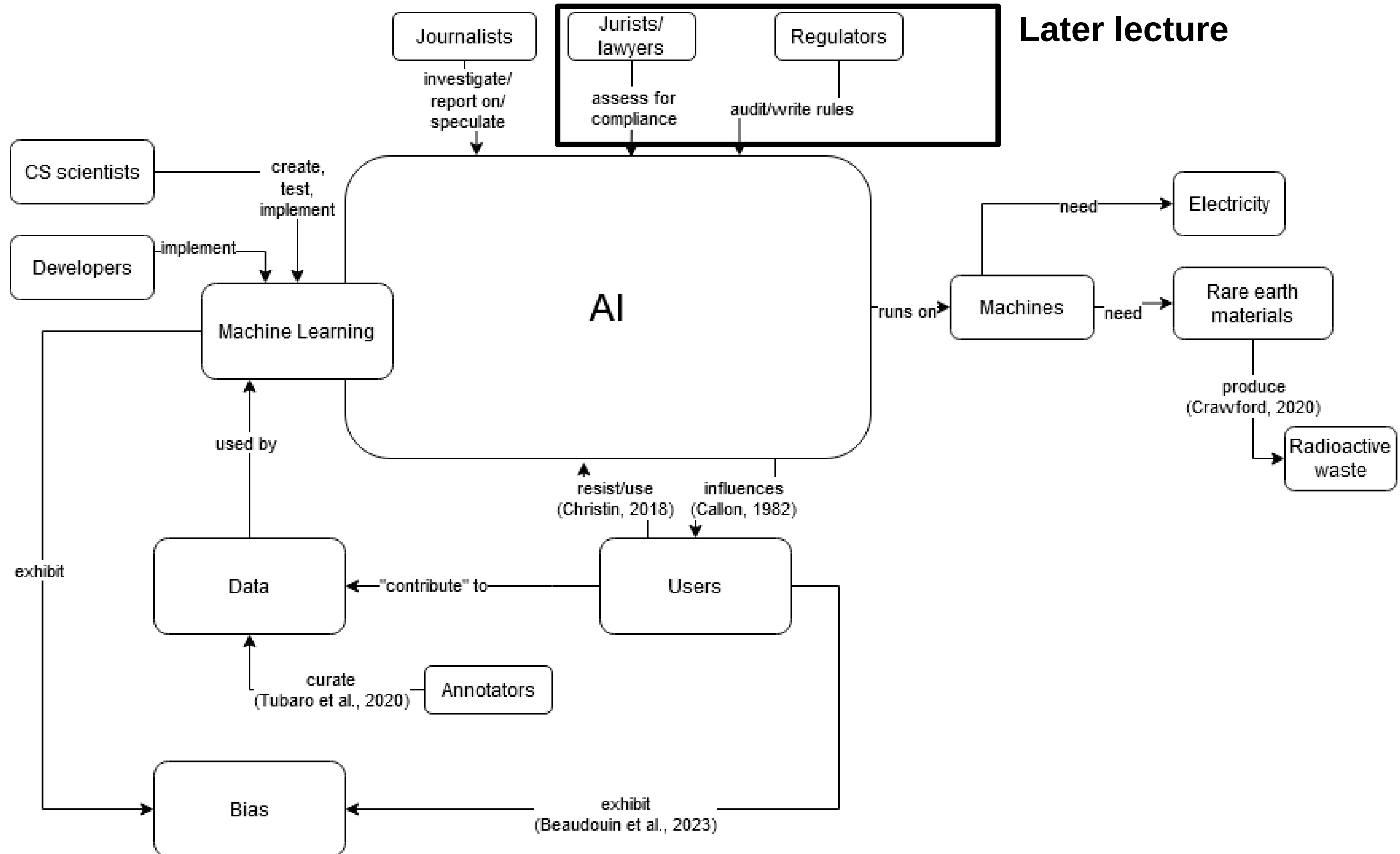
The (incomplete) AI sociotechnical system

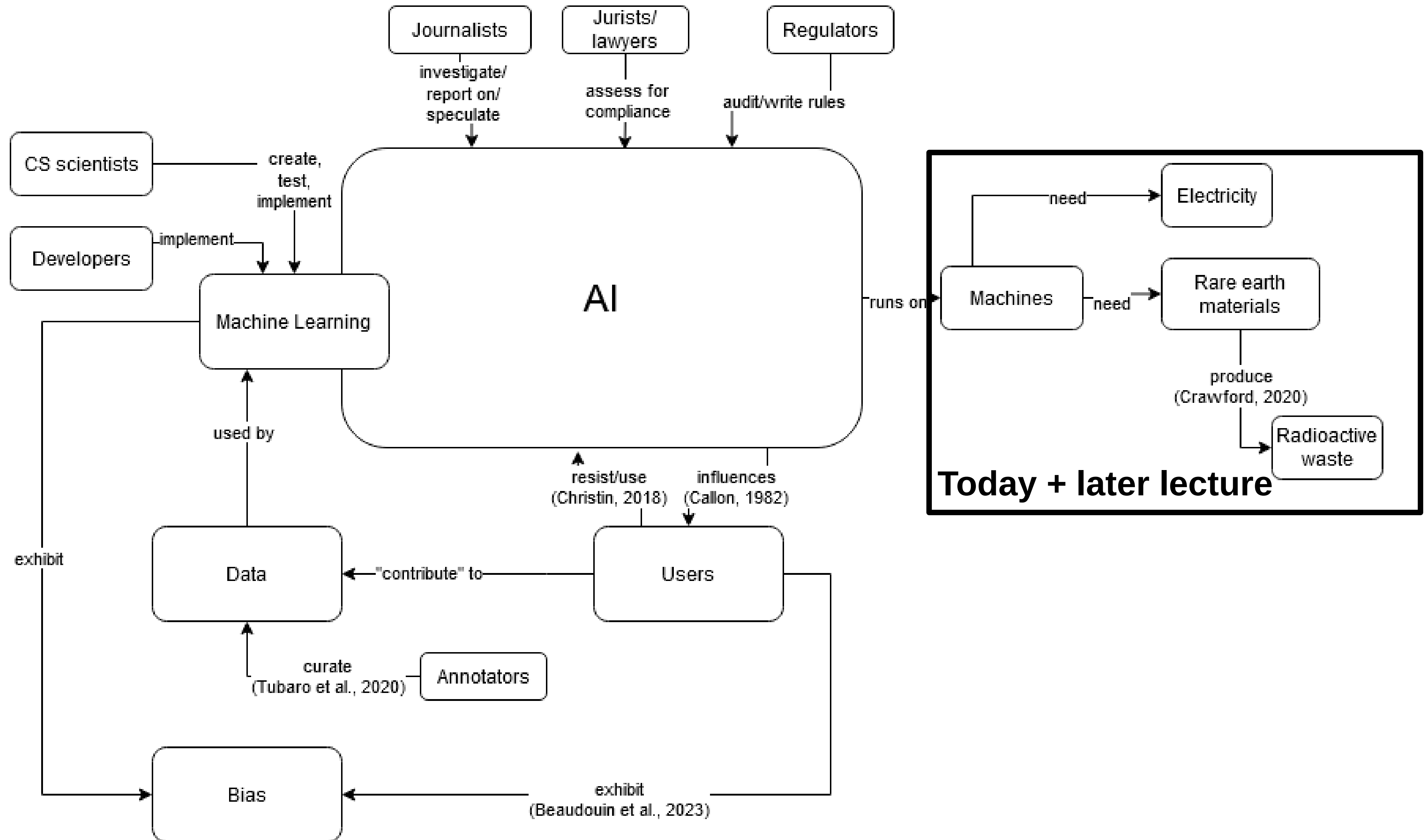# The (incomplete) AI sociotechnical system

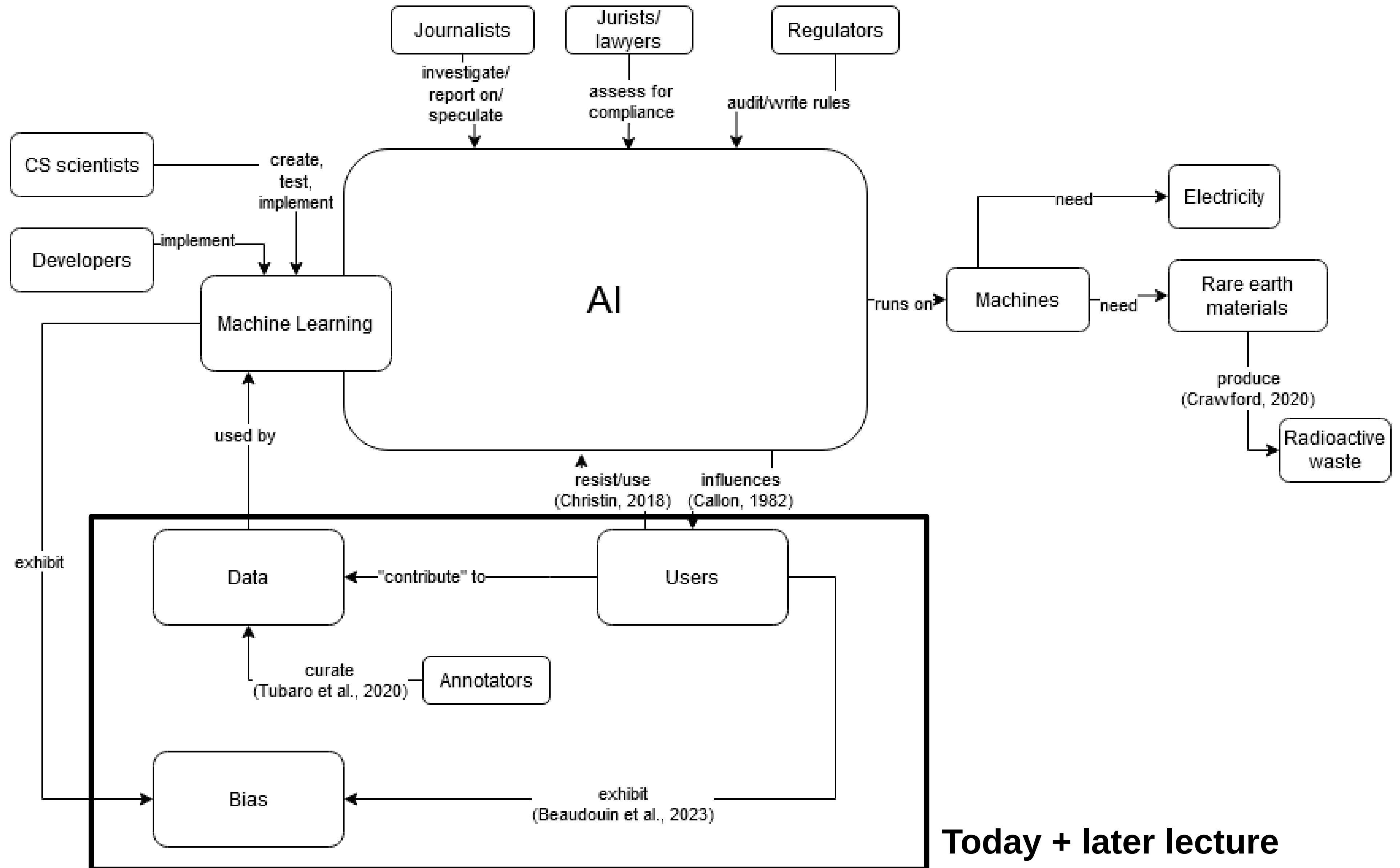# The (incomplete) AI sociotechnical system

The (incomplete) AI sociotechnical system

The (incomplete) AI sociotechnical system

What is, and what was AI ?

- AI is **not recent**: the term was coined in the 1950s
  - But AI today is nothing like AI in 1950

- Let us walk through two definitions:
  - A **socio-historical** one: contextualise the initial scientific promises of AI, and how/why they evolved in time
  - A **grounded/empirical** one: performing a distant reading of academic papers related to AI
- **Preconception**: AI followed a linear development cycle, with symbolic methods in the past, and neural methods now

Preconception: the linear development of technology

Symbolic AI, expert systems, decision trees: semi-manual learning from data and expert knowledge

Neural AI: statistical inference from data, bypassing the need for experts at the cost of large amounts of data/training

Generative AI? Creating its own training examples, bypassing the need for data/training
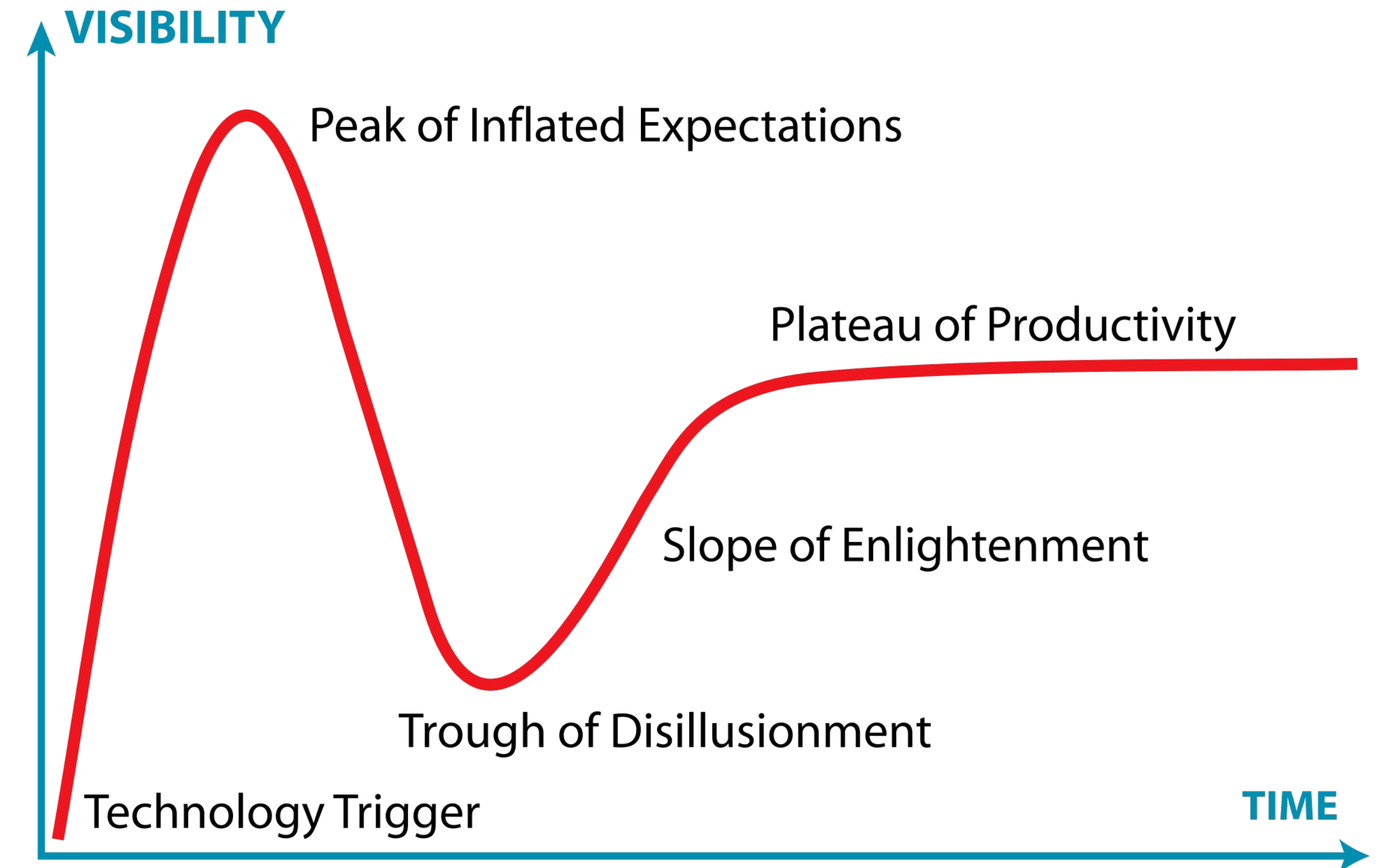
Gartner's (1955) cycle of hype is commonly seen but a **rarity** in practice

Sociohistory of AI

Cardon, D., Cointet, J. P., & Mazières, A. (2018). La revanche des neurones. *Réseaux*, *211*(5), 173-220.

Base hypothesis: the scientific proposal of AI is a paradigmatic change of computing

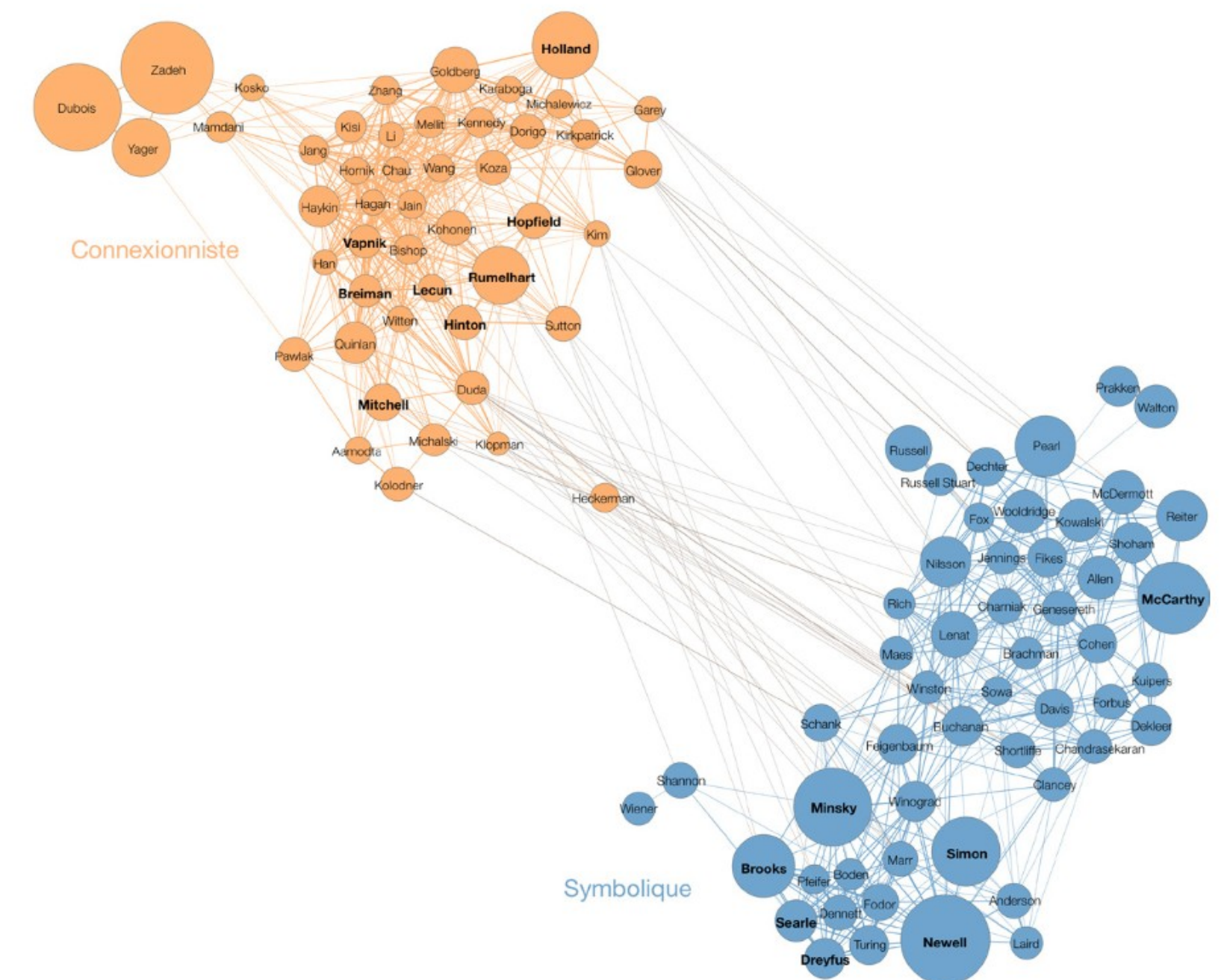Figure 1. Machine hypothético-déductive (1) et machine inductive (2)

Sociohistory of AI

Cardon, D., Cointet, J. P., & Mazières, A. (2018). La revanche des neurones. *Réseaux*, *211*(5), 173-220.

This hypothesis is supported by an inductive **bi-clustering of research works** (obtained by building a bibliographical network and a Louvain clustering)

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Figure 1. Machine hypothético-déductive (1) et machine inductive (2)
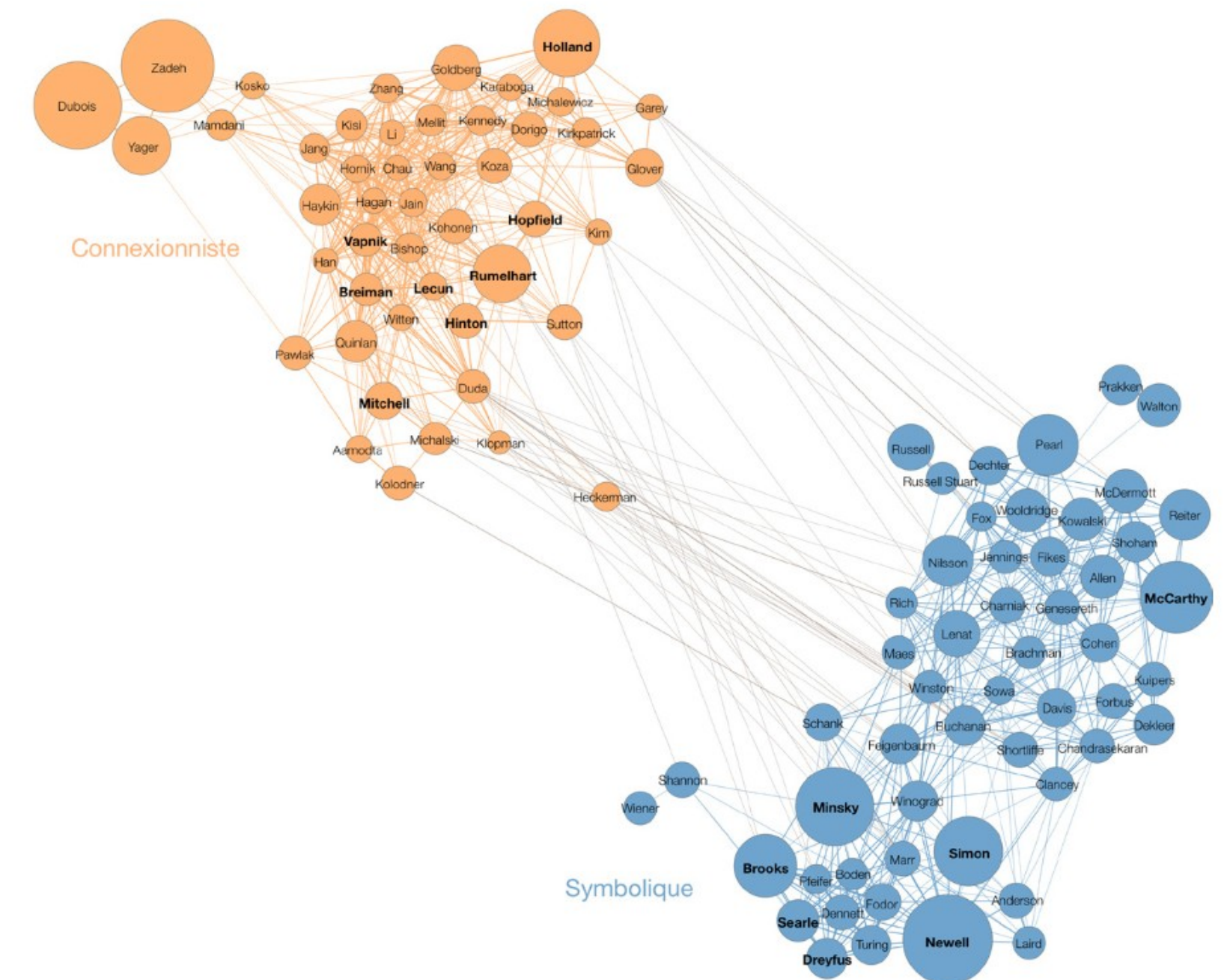
Sociohistory of AI

Cardon, D., Cointet, J. P., & Mazières, A. (2018). La revanche des neurones. *Réseaux*, *211*(5), 173-220.
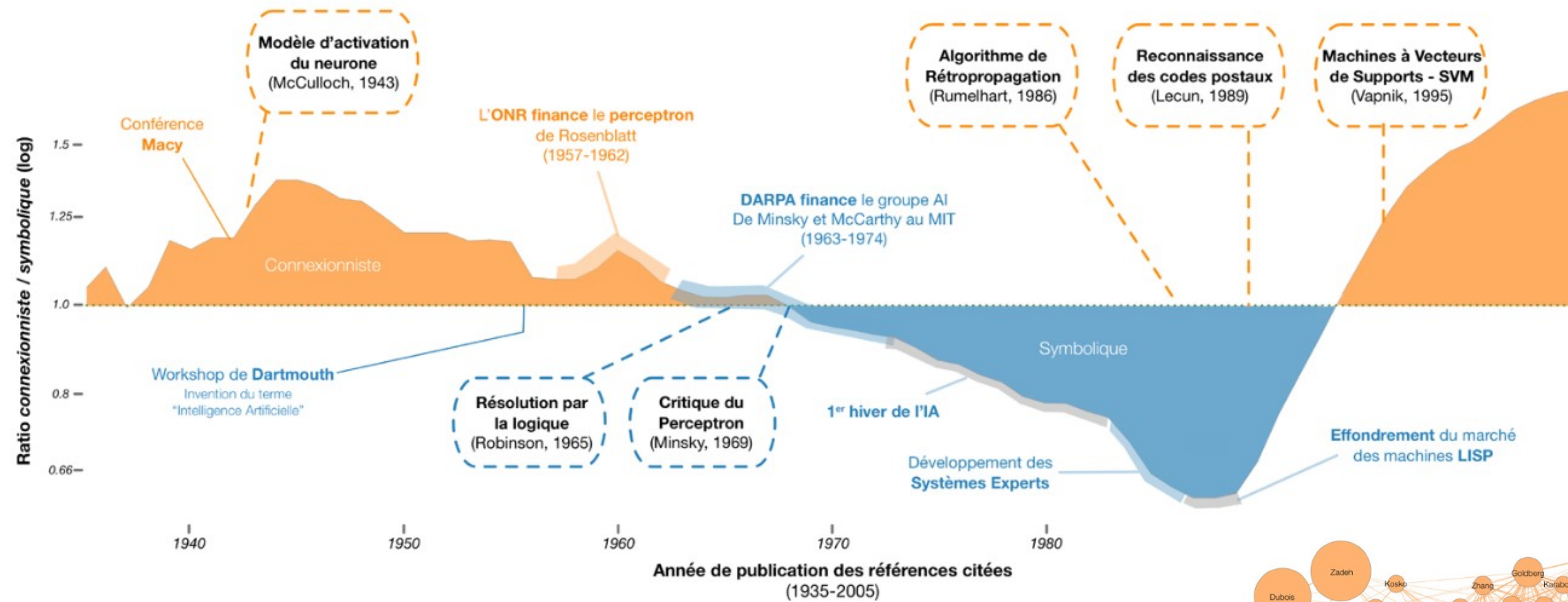
The bi-clustering sheds light on the tensions in the scientific field of AI

Figure 1. Machine hypothético-déductive (1) et machine inductive (2)

The four ages of machines

Table 1. The four ages of predictive machines

Machine	World	Calculator	Target
Cybernetics (connectionist)	Environment	"Black box"	Negative feedback
Symbolic AI (symbolic)	"Toy" world	Logical reasoning	Problem-solving
Expert systems (symbolic)	World of expert knowledge	Selection of hypotheses	Examples/counterexamples
Deep learning (connectionist)	The world as a vector of big data	Deep neural network	Objective-based error optimization

Connectionnist tenets:

- It is better to sacrifice intelligibility and rigour for **better perceptions of the complexity** of dimensions in the data

- There are enough redundancies and symmetries in the data to allow for **robustness**

- The "real-world", like connectionists, is **messy** and **non-linear**

- Data must be **atomized** and **granularized**

What is AI in academic papers

Quantitative analysis of the **Microsoft Academic Graph**

139 million scientific articles

What is AI in academic papers

Quantitative analysis of the **Microsoft Academic Graph**

139 million scientific articles

What is AI in academic papers
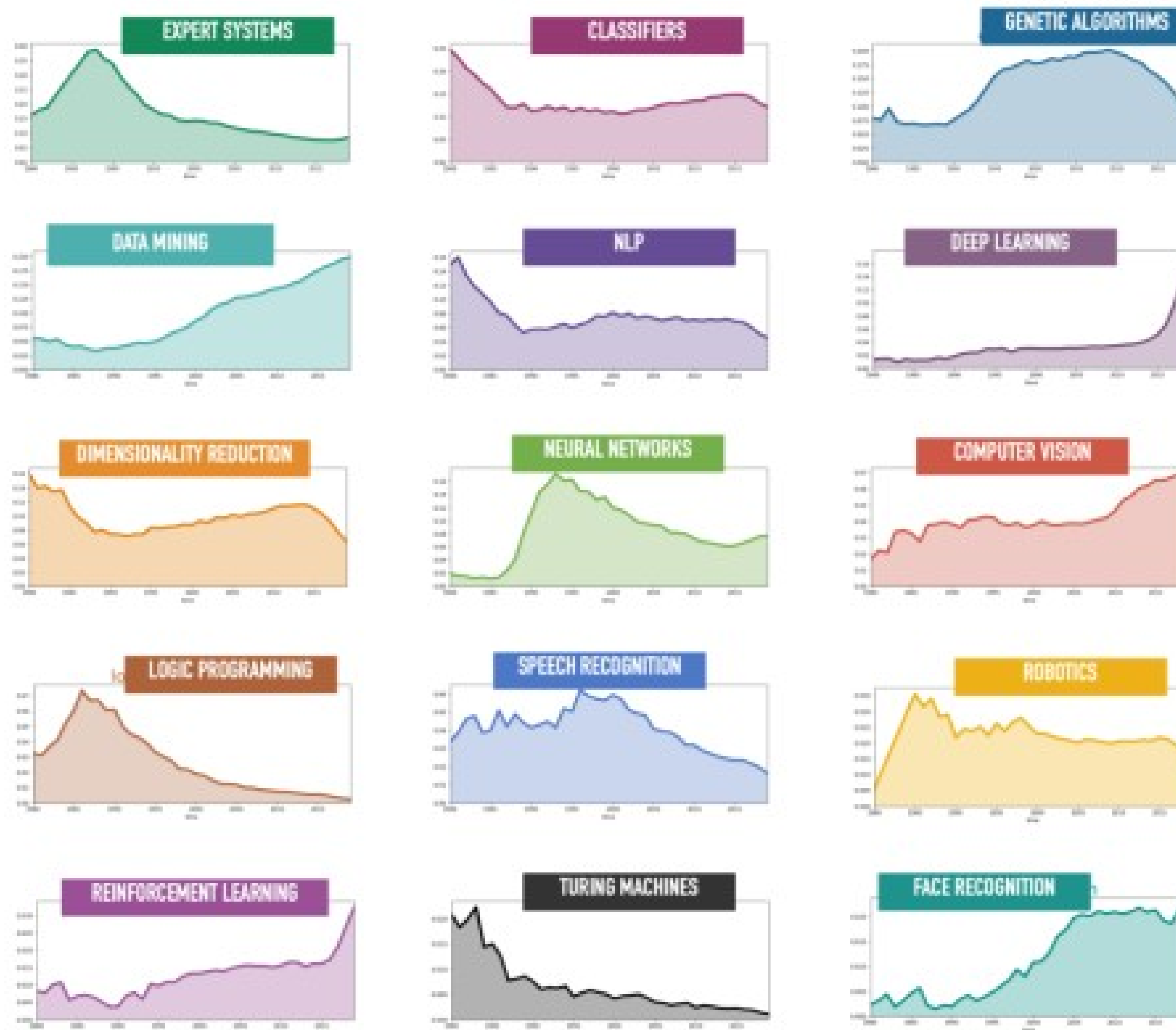
Gargiulo, F., Fontaine, S., Dubois, M., & Tubaro, P. (2023). A meso-scale cartography of the AI ecosystem. *Quantitative Science Studies, 4*(3), 574-593.
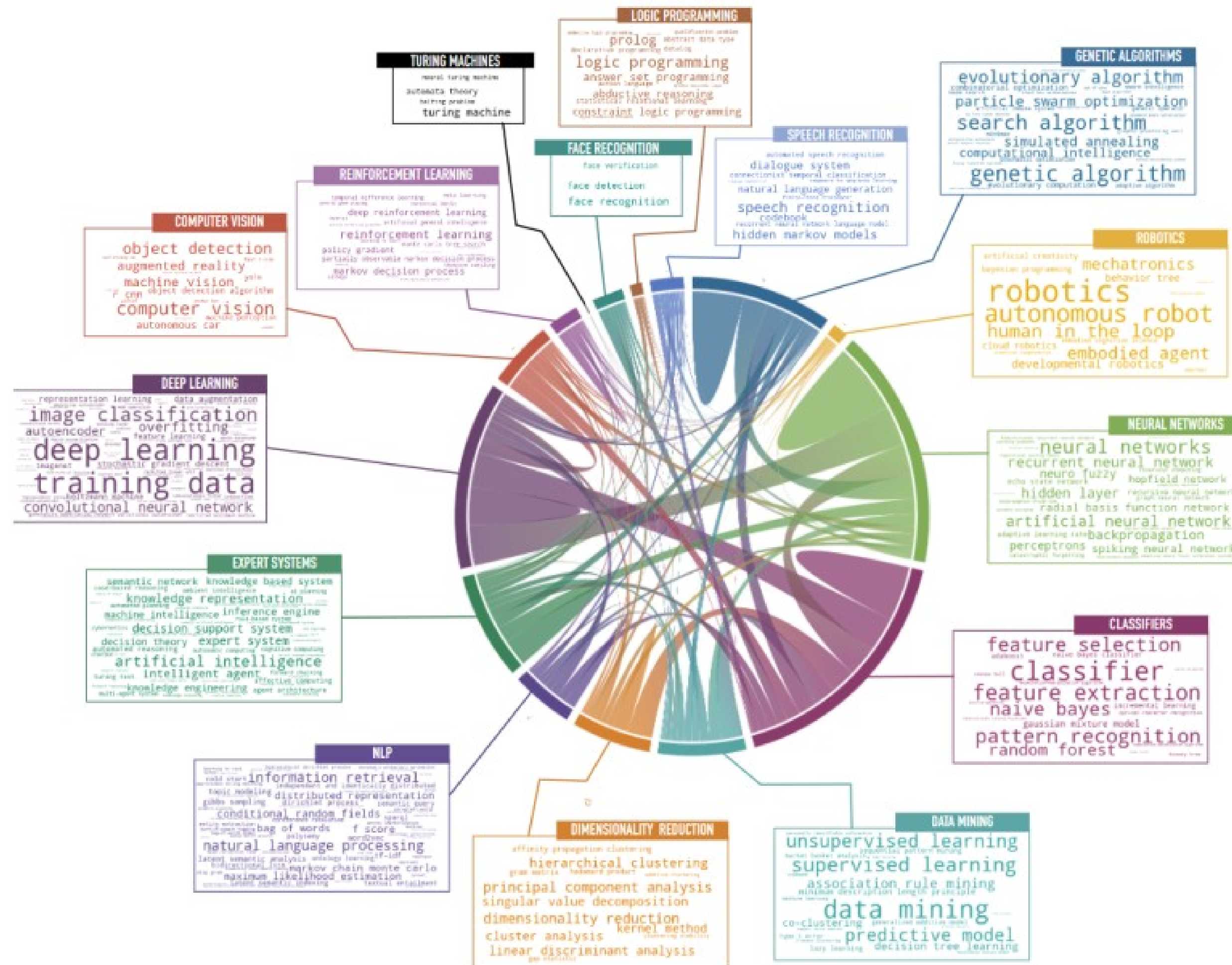
Quantitative analysis of the **Microsoft Academic Graph**

139 million scientific articles

- A multifaceted technical image, with sub-disciplines that heavily depend on each other

- A space where **symbolic**, **neural**, and **neuro-symbolic** methods are developed

- A scientific field split between **theoretical advances and applications**

- A field **influenced by press and fiction discourses**

- A field that is entering **regulated territory**

We need to reframe AI as a sociotechnical object

The materiality of AI: GPUs and servers
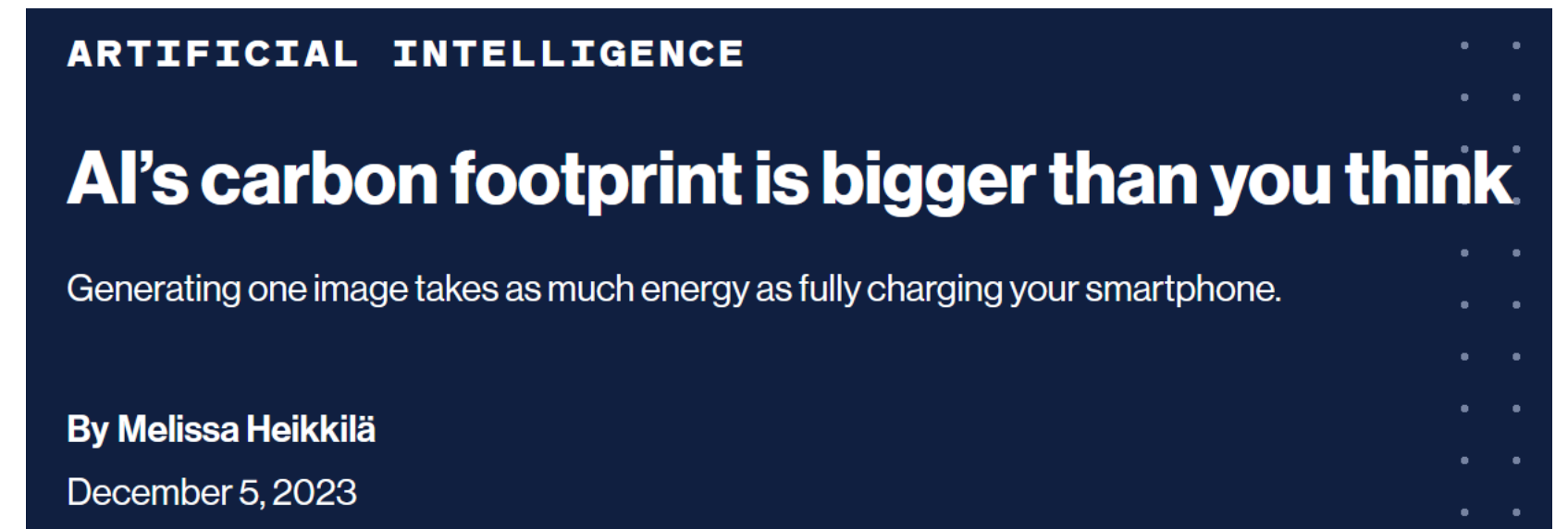
Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

Luccioni, A. S., Viguier, S., & Ligozat, A. L. (2023). Estimating the carbon footprint of bloom, a 176b parameter language model. Journal of machine learning research, 24(253), 1-15.

Delarue, S., Bonald, T., & Viard, T. (2024, October). Link Prediction Without Learning. In European Conference on Artificial Intelligence.

- Rematerialising the "cloud"

- Dependent on mine sites (esp. lithium),

- 99.8% of earth is treated as waste,

- Work conditions are abhorrent,

- 1 ton of rare earth elements = 75,000L of acidic water, 1 ton radioactive residue

- Deep learning is **carbon-intensive** and

labour intensive

- Deep learning is not always better (Delarue et al., 2024)

ARTIFICIAL INTELLIGENCE

AI's carbon footprint is bigger than you think

Generating one image takes as much energy as fully charging your smartphone.

By Melissa Heikkilä
December 5, 2023

Common carbon footprint benchmarks

in lbs of CO2 equivalent

Roundtrip flight b/w NY and SF (1 passenger)	1,984
Human life (avg. 1 year)	11,023
American life (avg. 1 year)	36,156
US car including fuel (avg. 1 lifetime)	126,000
Transformer (213M parameters) w/ neural architecture search	626,155

Chart: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

The materiality of AI

- Trends in AI research focus on **environmental concerns**

- Some trends: **frugal AI, carbon footprint estimates, "green AI", no AI**
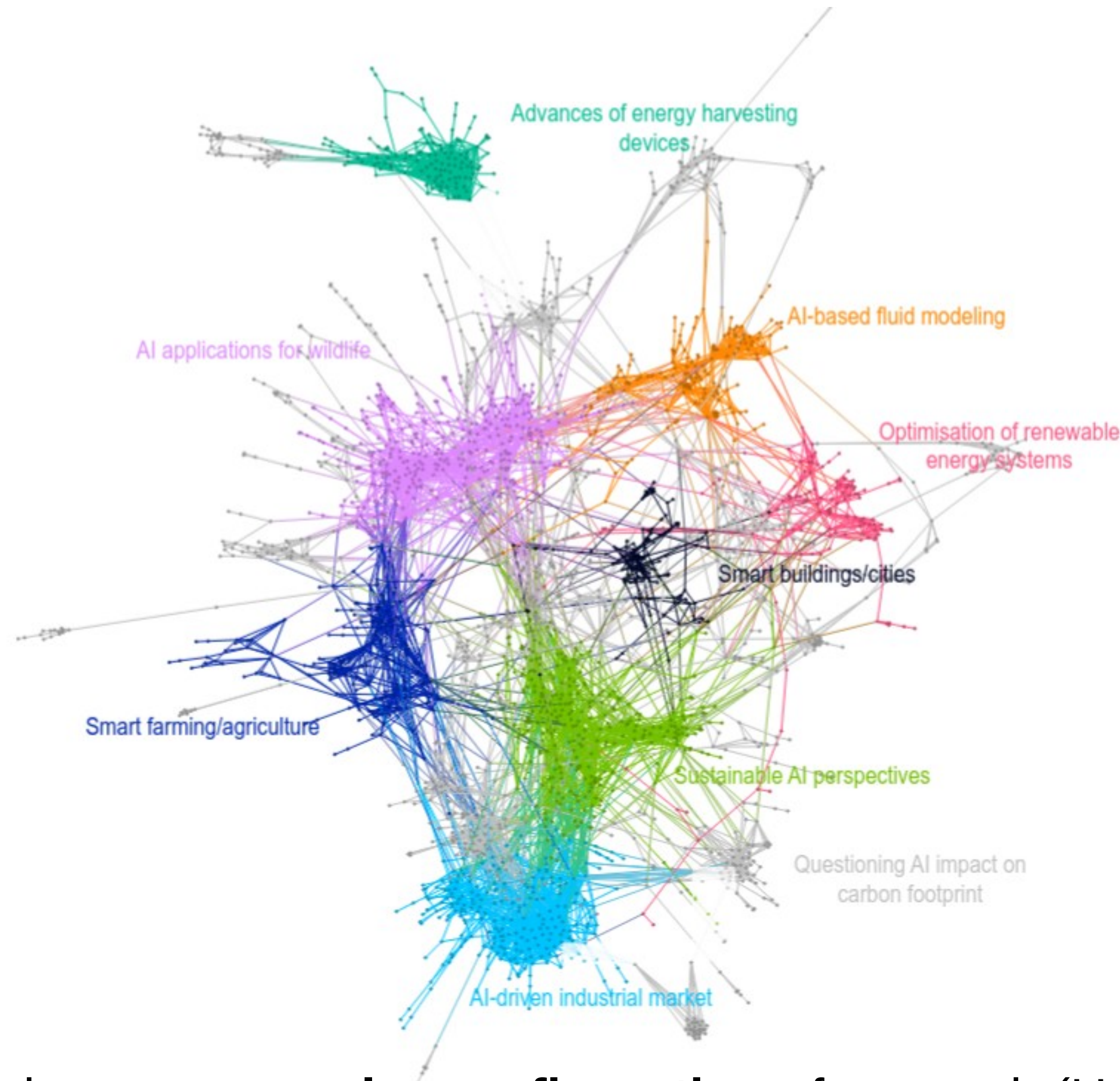
We observe a **moral reconfiguration** of research (Hardy, 2023): an embrace of the ideological pendants of scientific research

The materiality of AI: human labour

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
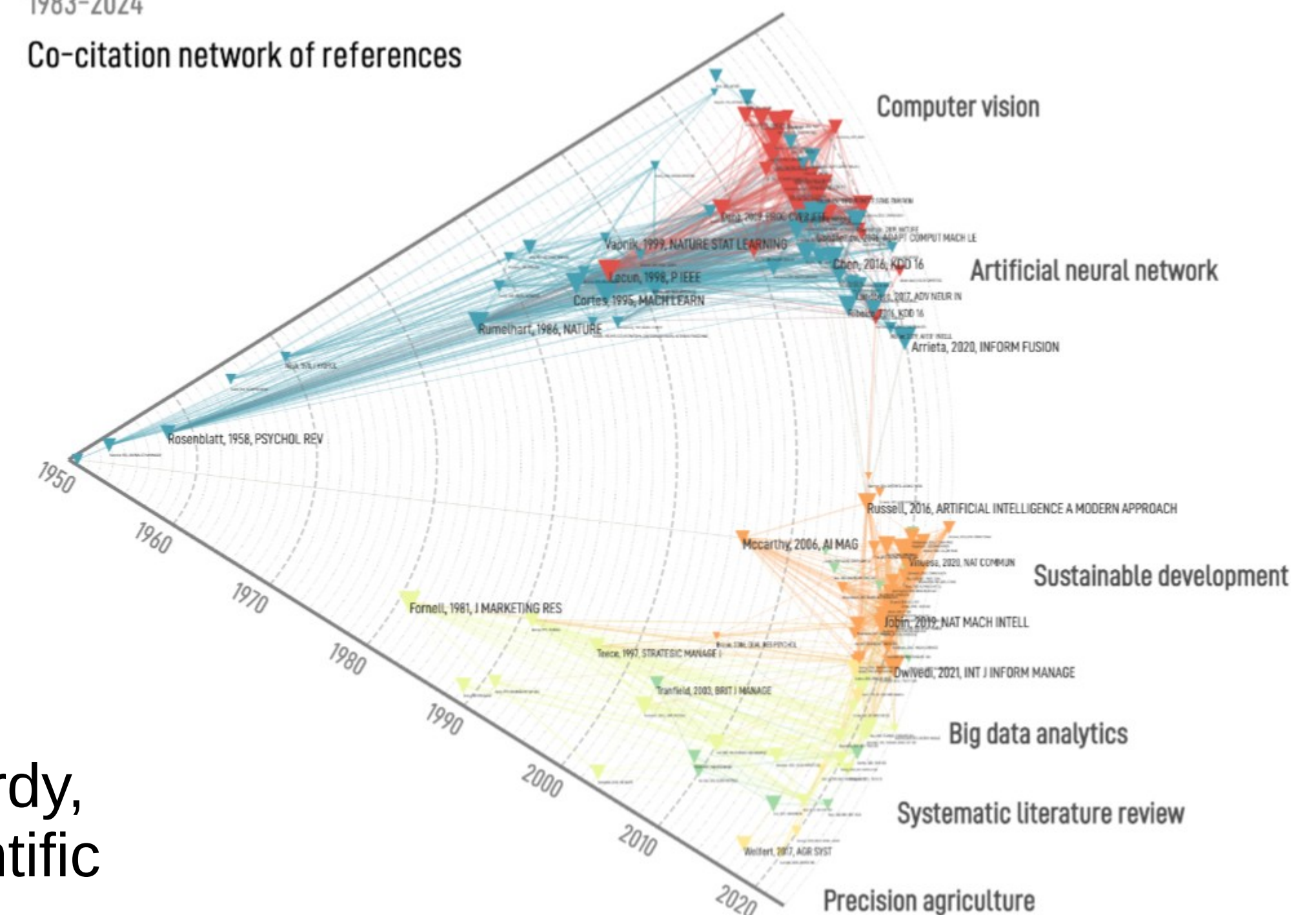
Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society.*

Le Ludec, C., Cornet, M., & Casilli, A. A. (2023). The problem with annotation. Human labour and outsourcing between France and Madagascar. *Big Data & Society*

- Clever Hans or the illusion of intelligence

- Understanding the **shaping of labour**:

  - *e.g.* Amazon's "matrix" algorithm **minimises breakage** but robs workers of **habituation**,

  - The **promise of transitory human work before automation is mostly false** (see later lecture)

- Digital labour reinforces North-South dynamics, with **cascades of delocalisation**.

- Multiple **hidden human roles**: the trainer, the verifier, the imitator (Tubaro et al., 2020)
  - The trainer **trains** AI systems
  - The verifier **checks** AI systems
  - The imitator **pretends** being an AI system

Power and data

Espeland, W. N., &
Sauder, M. (2007).
Rankings and
reactivity: How
public measures
recreate social
worlds. American
journal of sociology,
113(1), 1-40.

Kalluri, P. (2020).
Don't ask if artificial
intelligence is good
or fair, ask how it
shifts power.
Nature, 583(7815),
169-169.

Data colonialism: extracting value from data
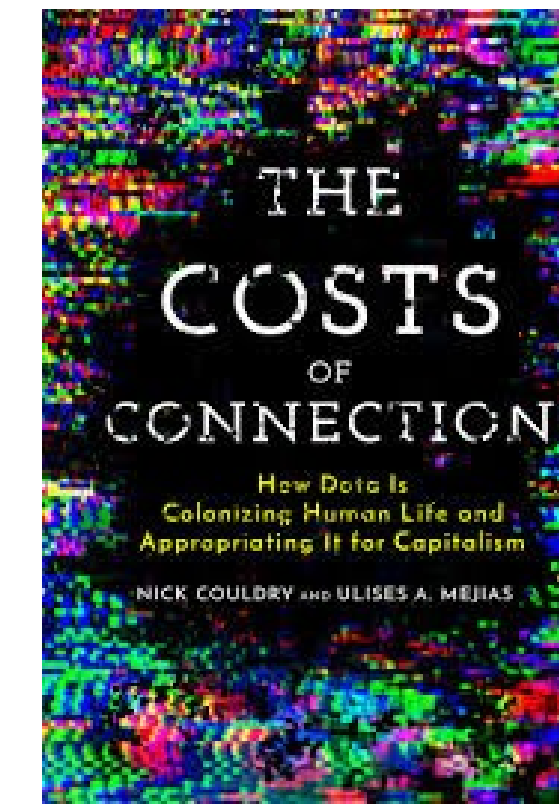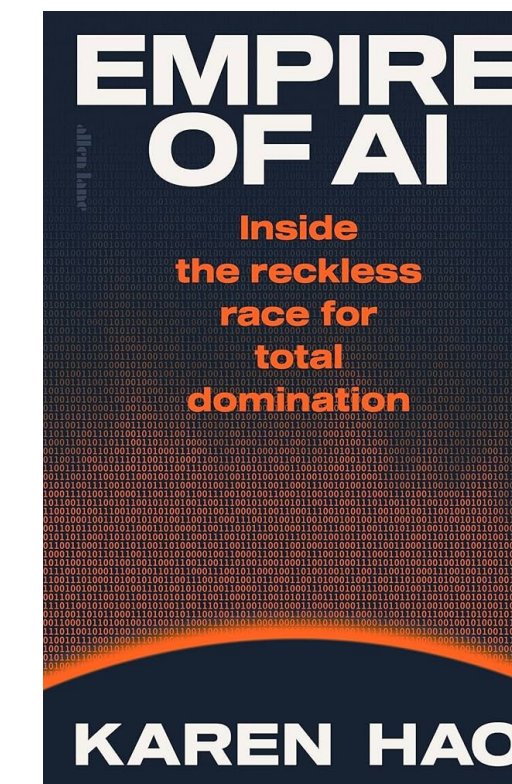
Not new (S. Zuboff) but an extension of **colonialism** and **capitalism**

- Extraction of data to **maximise profit + concentration of wealth**,

- Theory of **social quantification:** do we find what we measure?

- Showcased as a **basic need**,

- Reinforces existing **North-South inequalities**

- The example of language: communities **give their linguistic resources**, that is then **sold back to them as a service**
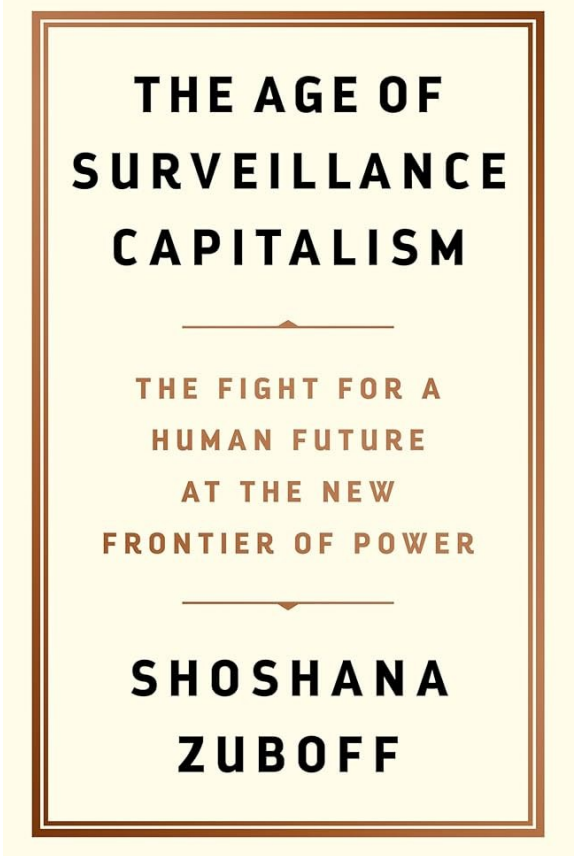
Challenging AI ideologies

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33,* 659-684.

Baradaran, A. (2024). Towards a decolonial I in AI: mapping the pervasive effects of artificial intelligence on the art ecosystem. Ai & Society, 39(1), 7-19.

The data ecosystem has technical, economical and practical developments, as well as **ideologies**

Challenge the idea that **data is a 'natural resource'**, requiring **unique skills** and **shaping essential needs** (*community building, democratising, personalisation*), and that is **conveys dematerialisation**

Use **media literacy**, **regulation**, challenge data collection as *inherently rational*

Build alternate presents and futures with **genuine pluralism**

Studies from (de-)colonialism shows that:

- Individual action is not sufficient (*i.e.* the issue is **systemic**),

- Nor are alternative platforms (*e.g.* IRC vs Messenger, Mastodon vs X/Twitter)

The media landscape of AI
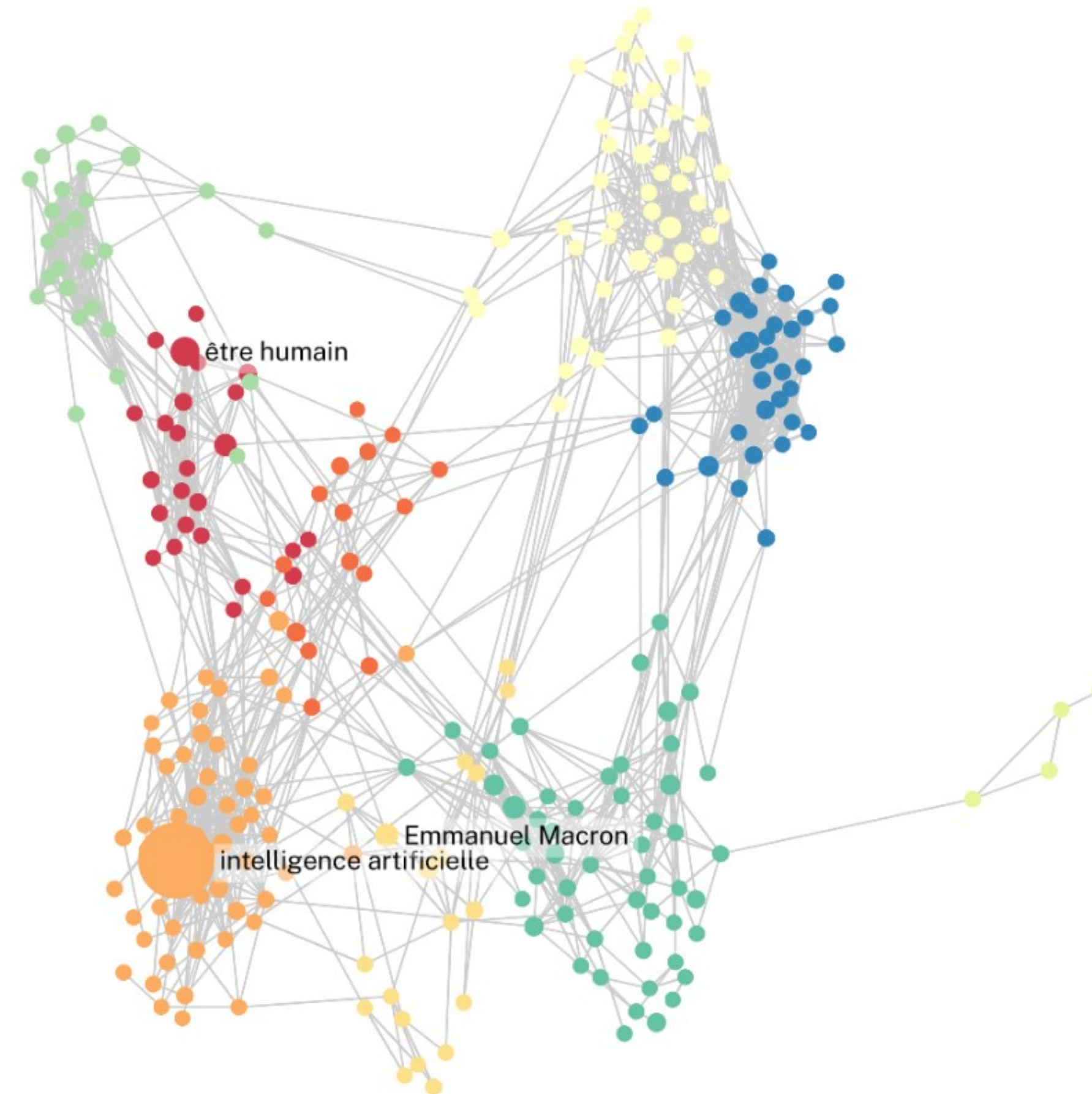
- Nodes are n-grams
- Edge between two nodes = they co-occur statistically significantly often (χ^2 test)
- Colour is obtained through Louvain clustering

- Pre-2023, a corpus of 1000 articles in French mentioning AI and ethics

- Let's explore it online

The media landscape of AI

Technical cluster

The media landscape of AI

AI private sector cluster

Transhumanist cluster

The media landscape of AI

AI regulation cluster

The media landscape of AI

AI & bioethics cluster

The media landscape of AI (1990–2025)
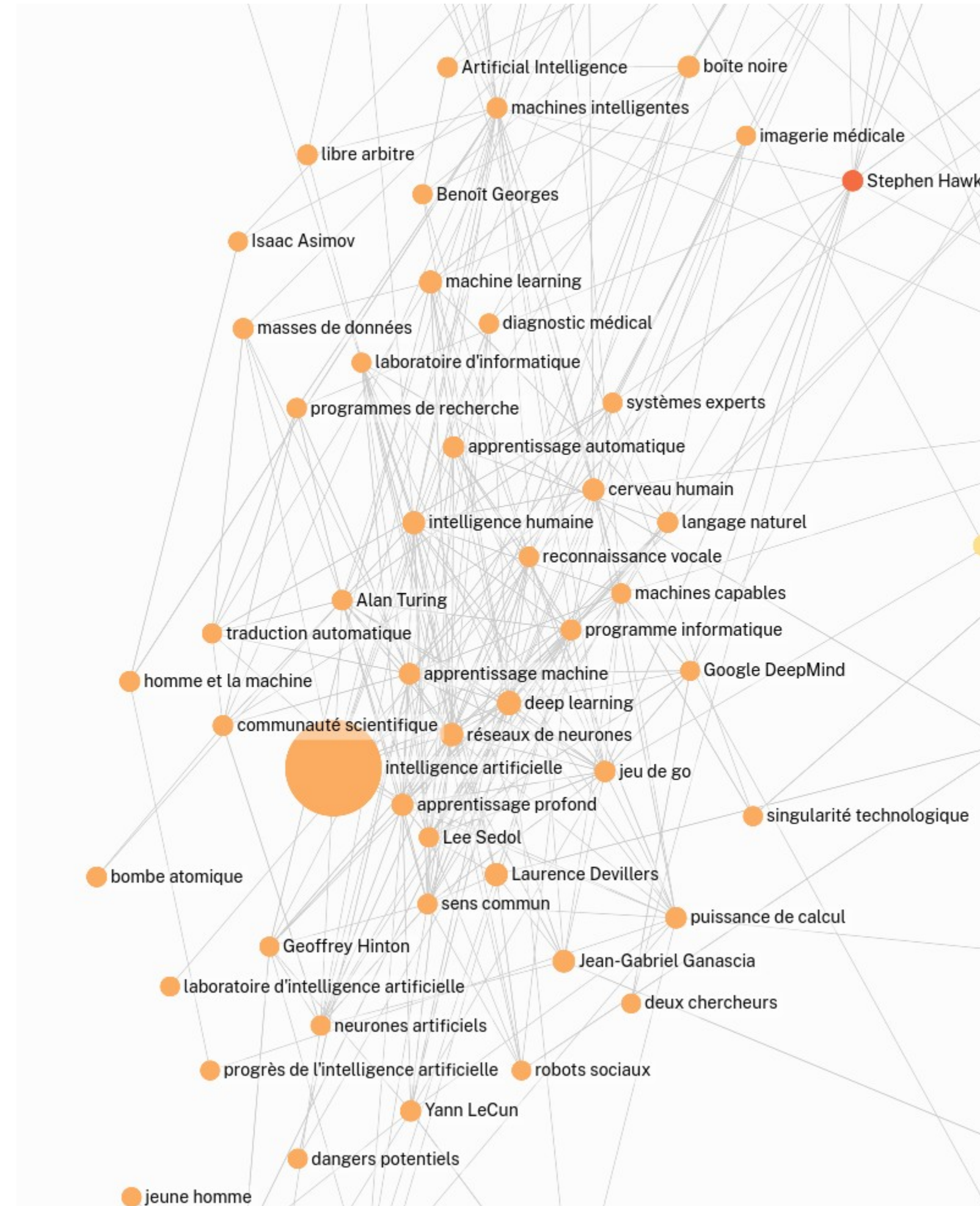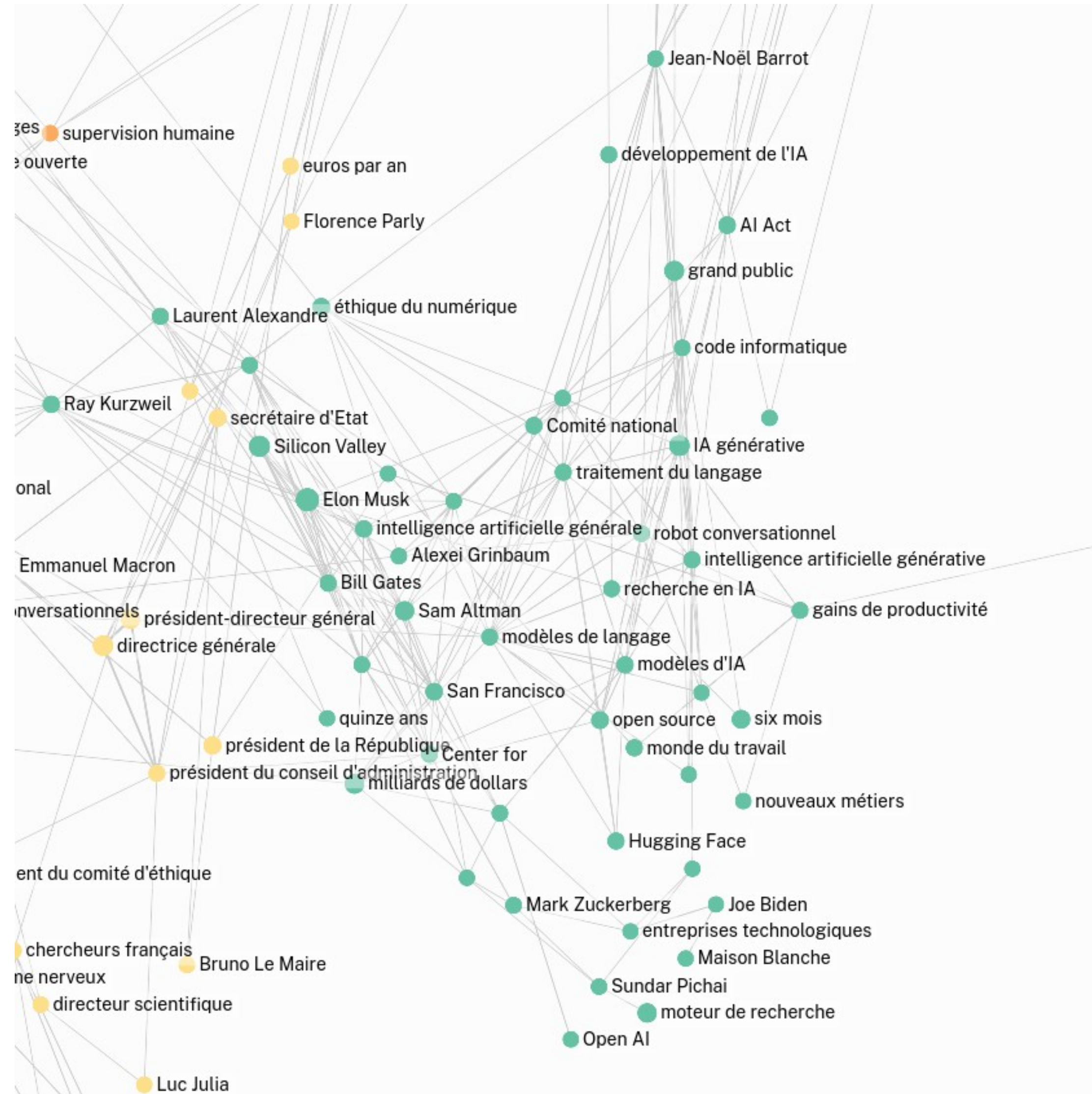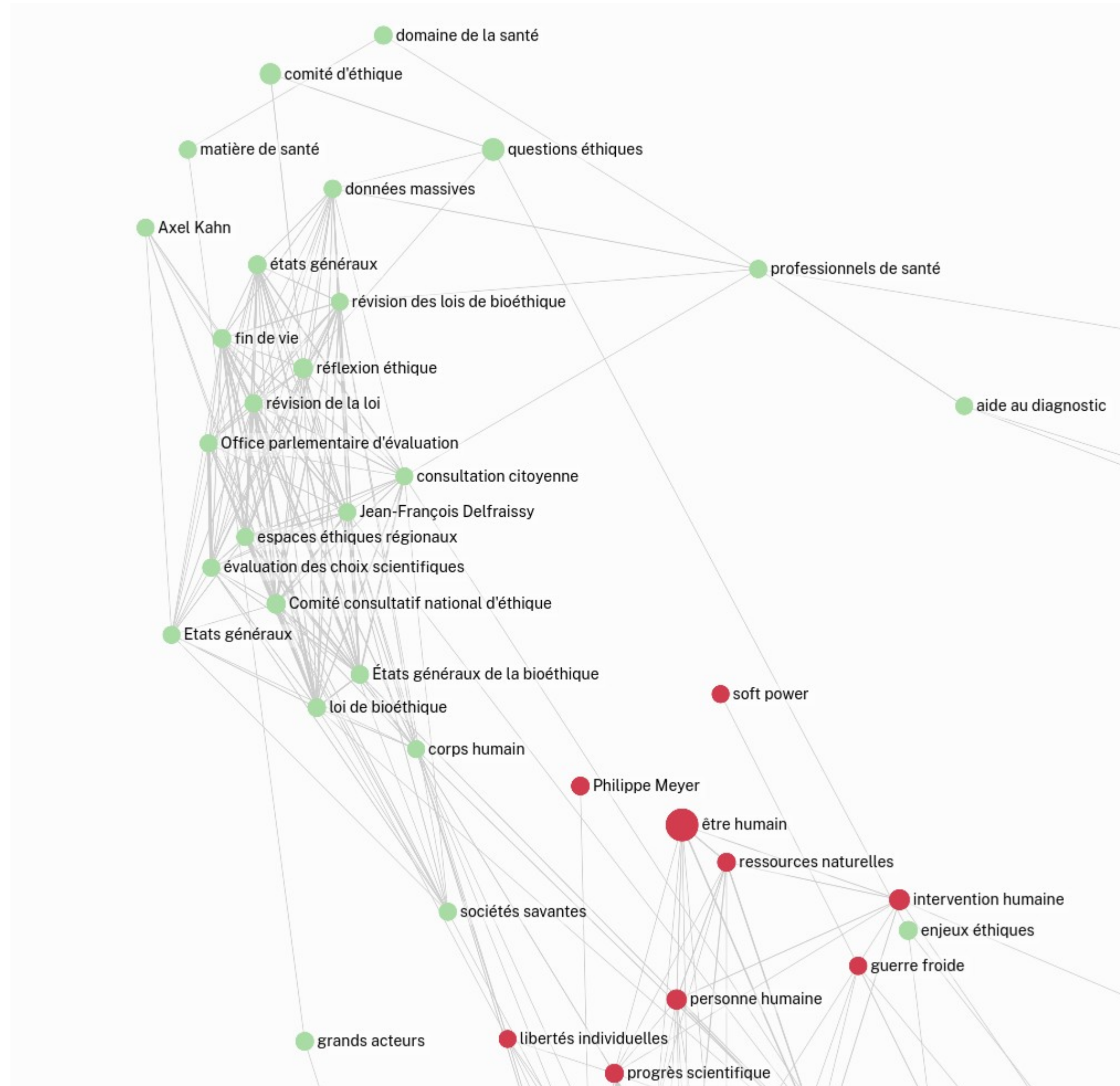
Source: T. Viard
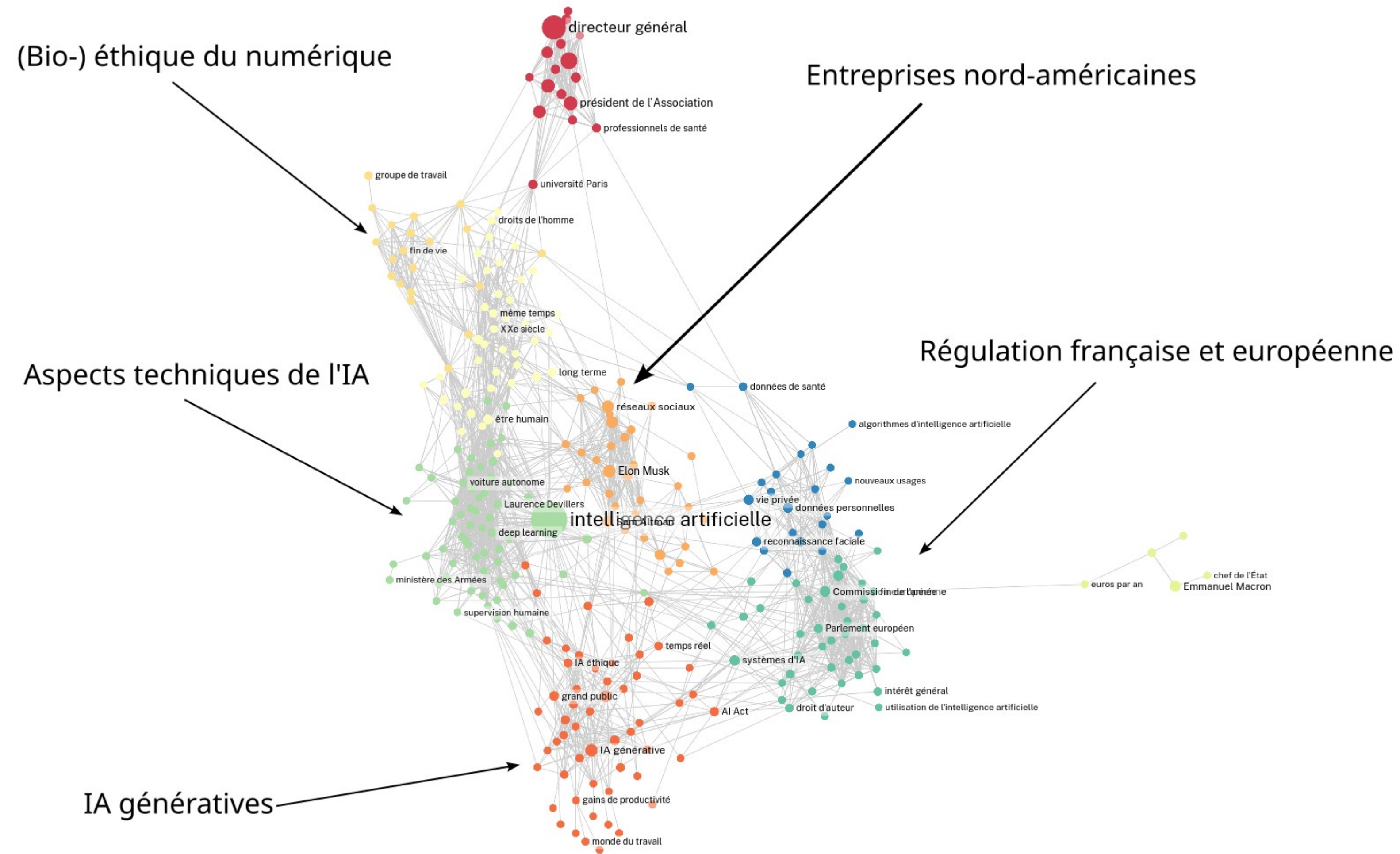
The media landscape of AI (2023-2025)
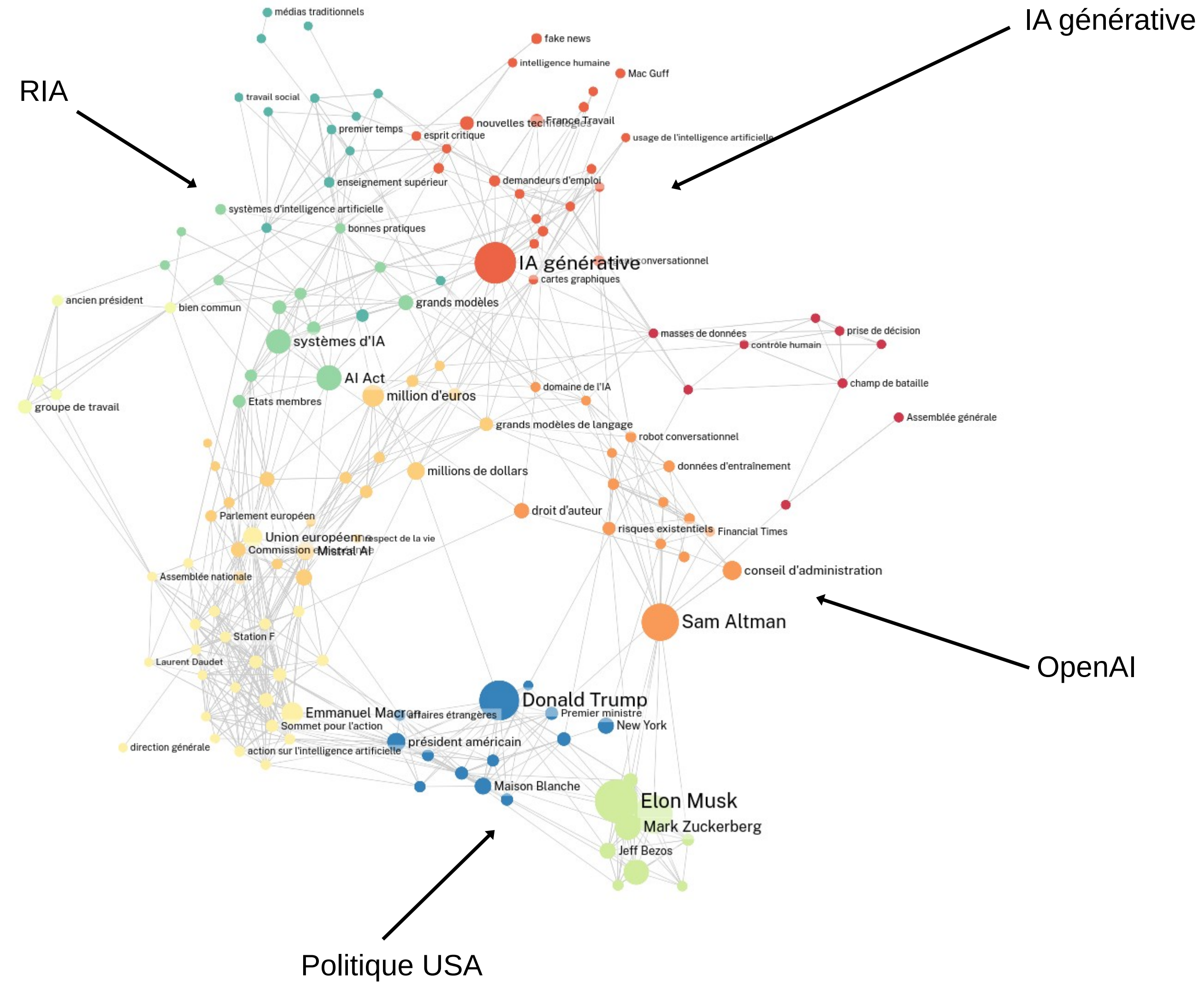
The hype around AI

- **AI** (like crypto, 'Web3') is a **broad term**, rather than a **cohesive technology**
- **Crypto** also has benefitted from large marketing investments, and has **enormous costs**
- Crypto has shown **no socially beneficial use**
- AI has defined, beneficial uses: the hype comes from the **mismatch between claims and reality**
- The AI field has been **fueled by waves of hope/hype**
- **AI** (and computer science) largely does not consider working with industry a conflict of interest
- Focus in AI conferences is still on **performance** rather than **conceptual understanding**
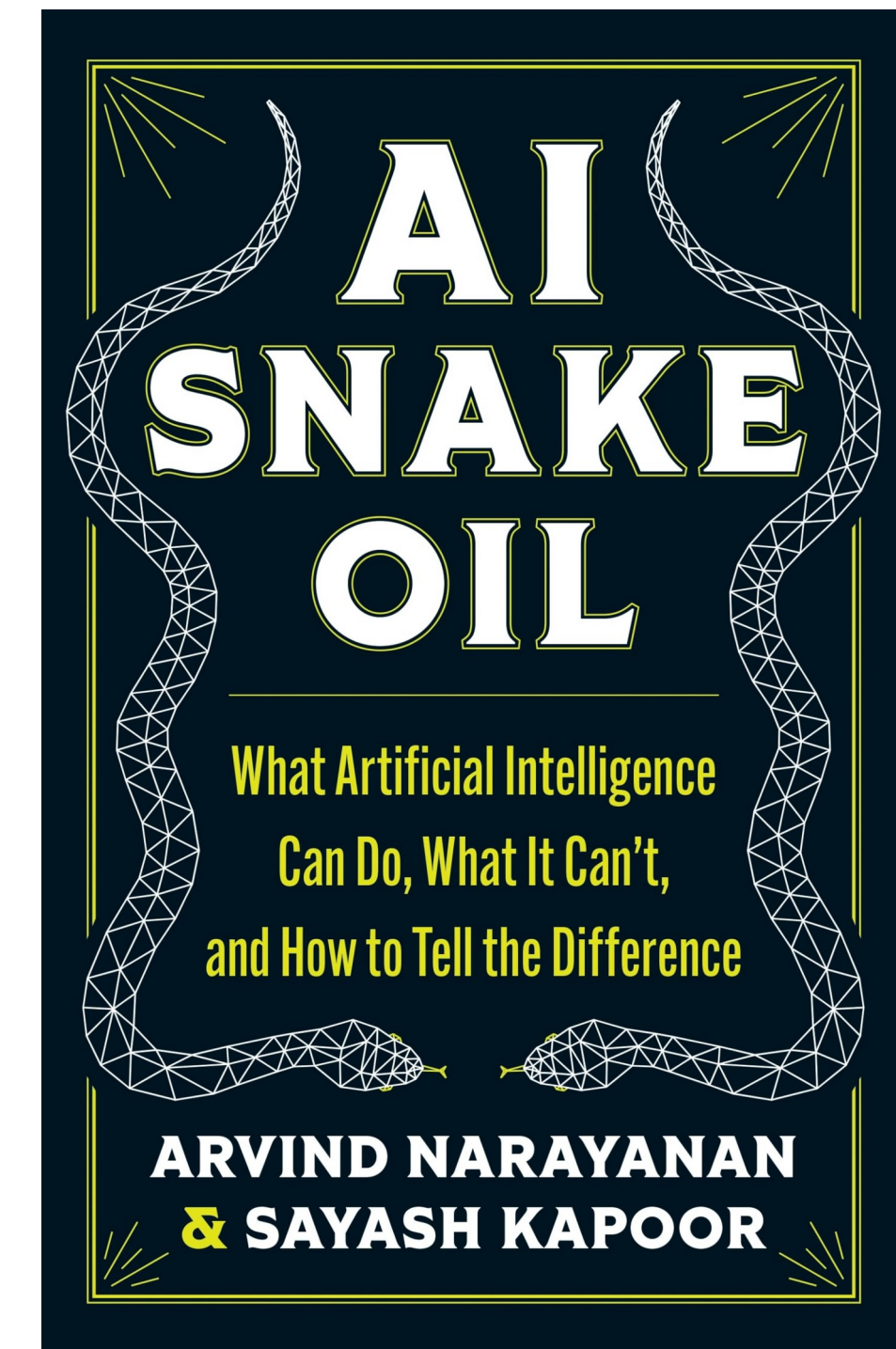
'If you are a radiologist, you are like the coyote above the cliff [...] it is just completely obvious that in five years, deep learning is going to do better than radiologists' – Hinton, 2016

"AI ethics": what is it and what role has it?

Growing field of **interdisciplinary** research: computer science, economy, law, mathematics, sociology...

The International Library of Ethics, Law and Technology 41

Francisco Lara
Jan Deckers *Editors*

Ethics of Artificial Intelligence

Springer
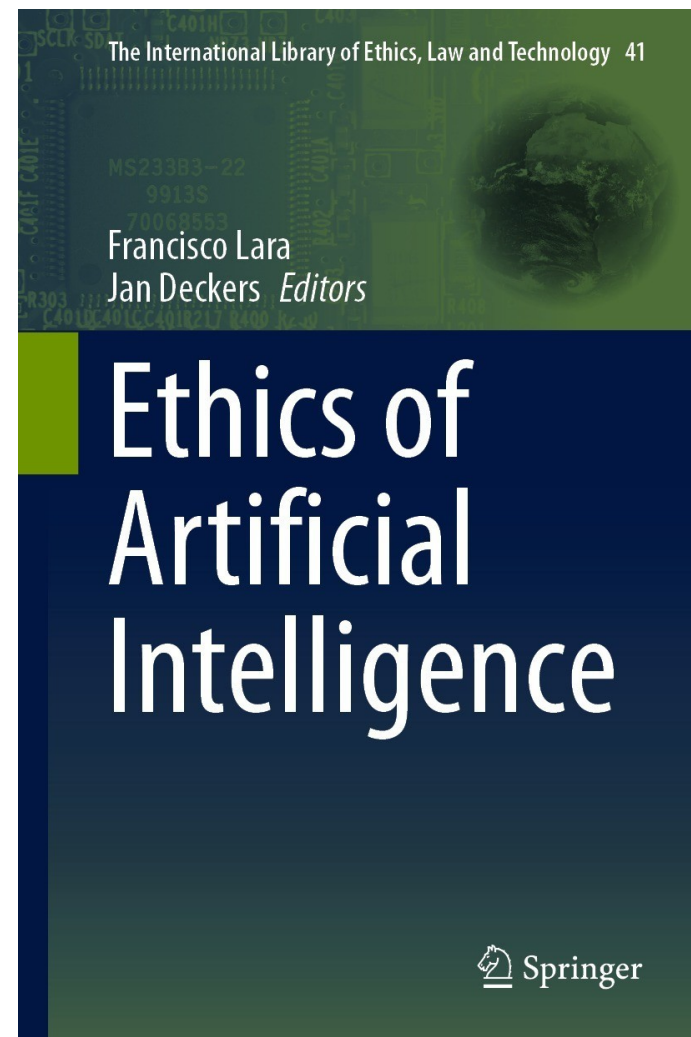
Comment | Published: 09 August 2019

AI's social sciences deficit

Mona Sloane ✉ & Emanuel Moss

Nature Machine Intelligence 1, 330–331 (2019) | Cite this article

1504 Accesses | 27 Citations | 179 Altmetric | Metrics

To create less harmful technologies and ignite positive social change, AI engineers need to enlist ideas and expertise from a broad range of social science disciplines, including those embracing qualitative methods, say Mona Sloane and Emanuel Moss.

AI ETHICS

MARK COECKELBERGH

THE MIT PRESS ESSENTIAL KNOWLEDGE SERIES
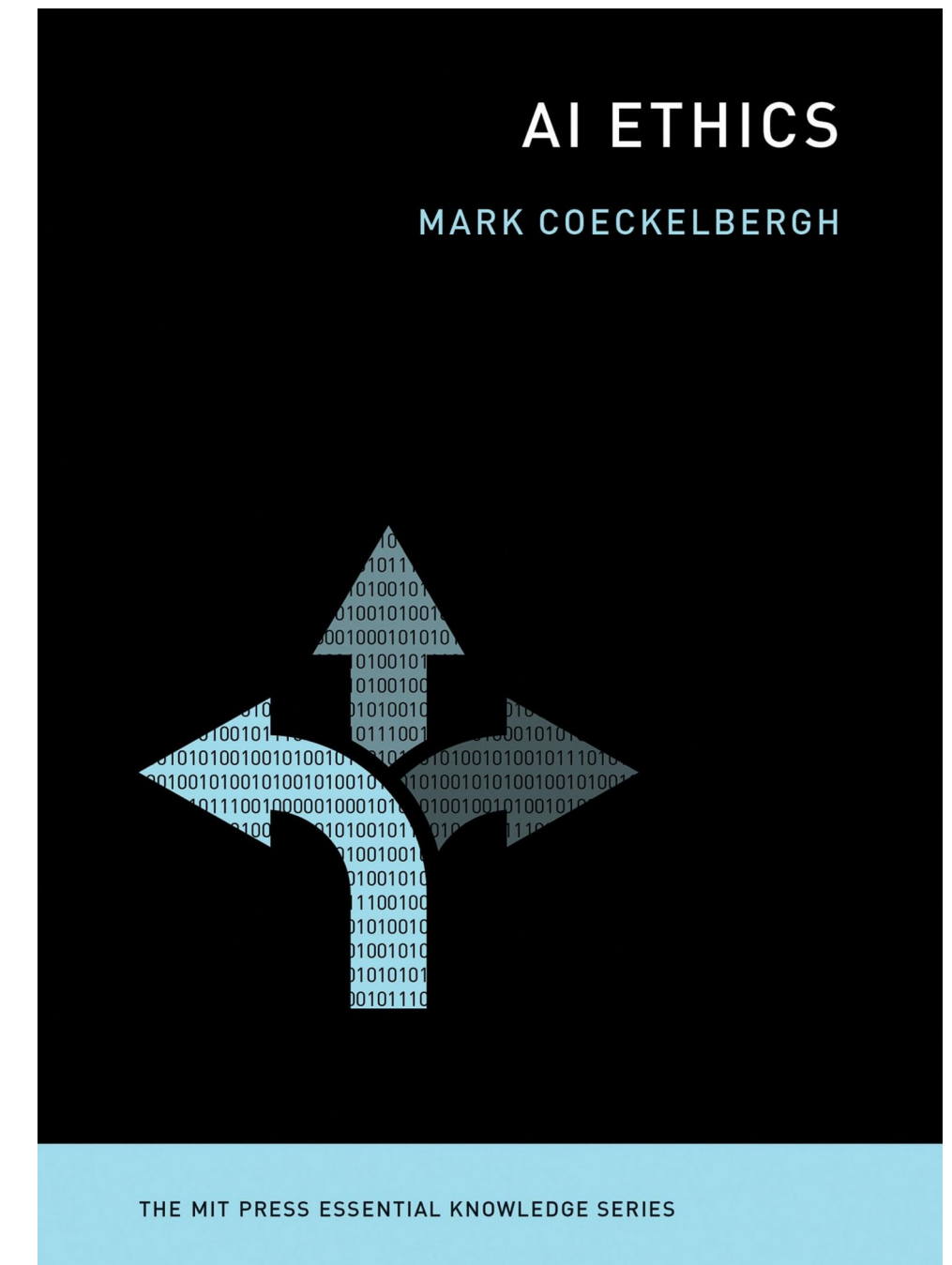
Some goals:

- Separate **hype from reality**

- Understand how **discourses, arguments and actors** articulate themselves

- Highlight important areas for **regulation**

- Give **operational** guidelines

What AI ethics says for itself

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399.

The Manifesto for Ethical AI

The commitment of the actors who design, develop, deploy and administer artificial intelligence solutions allied with the educators who provide job training in this field.

Download the document

Artificial Intelligence (AI) promises so much: : to help us
way companies operate, to take the strain and discomf
faster and better, to make our daily lives easier, etc.

Ethics of Artificial Intelligence

The Recommendation

Global AI Ethics and Governance Observatory

Getting AI governance right is one of the most consequential challenges of our time, calling for mutual learning based on the lessons and good practices emerging from the different jurisdictions around the world.

The aim of the **Global AI Ethics and Governance Observatory** is to provide a global resource for policymakers, regulators, academics, the private sector and civil society to find solutions to the most

iness of countries to adopt AI ethically and

athers contributions, impactful research,

Government of Canada   Gouvernement du Canada

Search Canada.ca

MENU ▾

Canada.ca > Innovation, Science and Economic Development Canada

AI PRINCIPLES

Our AI Principles

Our approach to developing and harnessing the potential of AI is grounded in our founding mission — to organize the world's information and make it universally accessible and useful.

We believe our approach to AI must be bot
rapidly innovating and deploying AI in grou
benefiting people everywhere, contributing
deepen our understanding of the world, an
most pressing challenges and opportunitie
and deploying AI that addresses both user
responsibilities, while safeguarding user sa

We approach this work together, by collabo
partners to make breakthroughs and maxin
empowering others to build their own bold

Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems

From: **Innovation, Science and Economic Development Canada**

September 2023

Advanced AI systems capable of generating content — such as ChatGPT, DALL·E 2, and Midjourney — have captured the world's attention. The general-purpose capabilities of these advanced AI systems offer enormous potential for innovation in a number of fields, and they are already being adopted and put to use in a variety of contexts. These advanced systems may be used to perform many different kinds of tasks — such as writing emails, answering complex questions, generating realistic images or videos, or writing software code.

What AI ethics says for itself

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399.

(Jobin et al., 2019) identifies 11 common themes in 84 documents.

Themes obtained through qualitative reading

- Transparency , Justice & fairness, Non-maleficence, Responsibility, Privacy, Beneficence, Freedom & autonomy, Trust, Sustainability, Dignity, Solidarity

In a varying number of documents, with **high polysemy**

Preconception: Documents tend to be **short** and **vague** (i.e. **useless)**

Next: a **quantitative analysis** of 436 manifestos (mapaie.telecom-paris.fr)

AI ethics through its manifestos

- A **tri-partition of discourses**: technical, regulatory and business/innovation
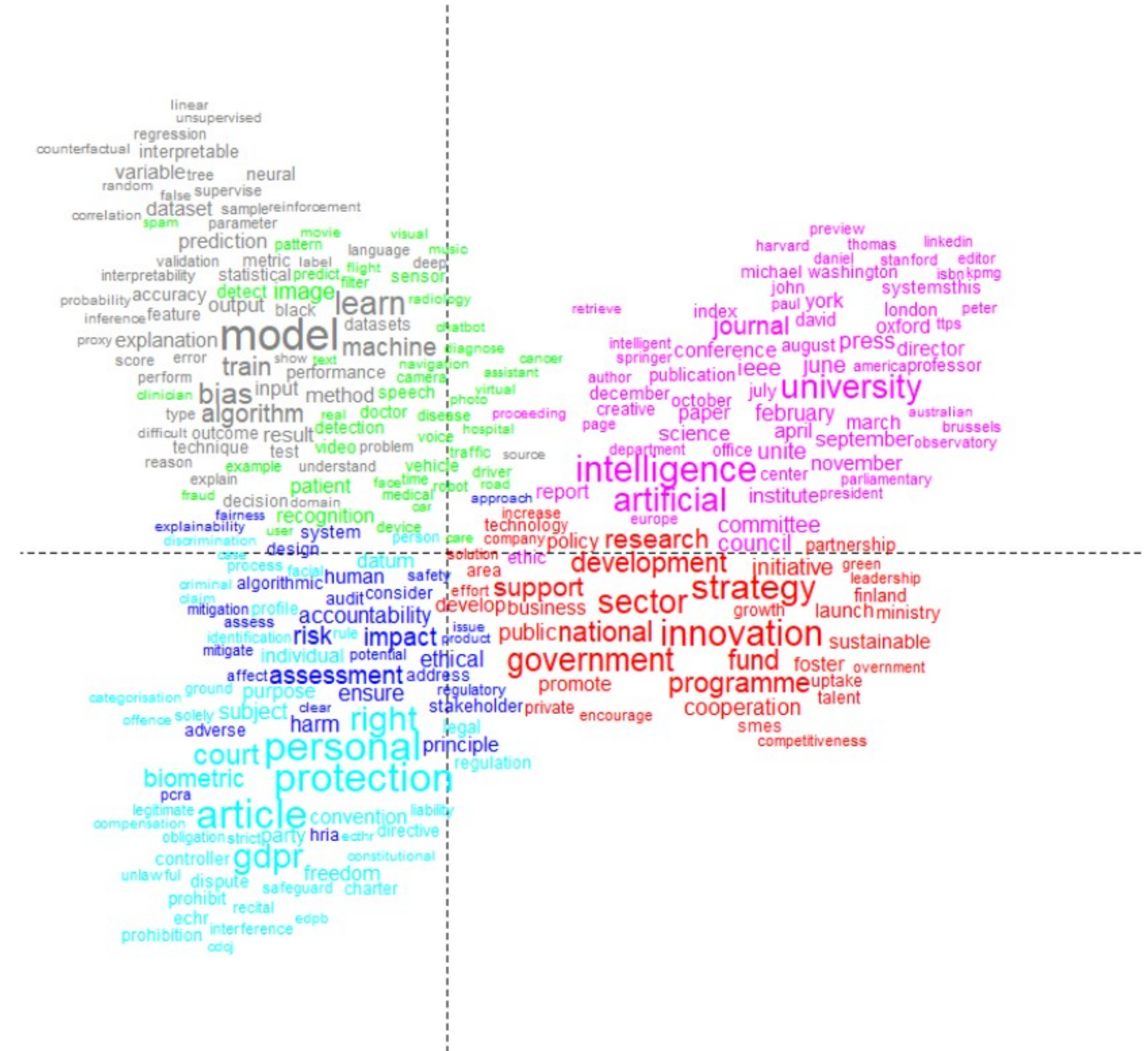
AI ethics through its manifestos

- A **tri-partition of discourses**: technical, regulatory and business/innovation

- On **explainable AI**: more technical and less regulation

Gornet, M., Delarue, S., Boritchev, M., & Viard, T. (2024, June). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 127-140).

AI ethics through its manifestos

- A **tri-partition of discourses**: technical, regulatory and business/innovation

- On **explainable AI**: more scientific/technical and less regulation

- On **fairness**: no significant change

Gornet, M., Delarue, S., Boritchev, M., & Viard, T. (2024, June). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 127-140).

AI ethics through its manifestos

- A **tri-partition of discourses**: technical, regulatory and business/innovation

- On **explainable AI**: more scientific/technical and less regulation

- On **fairness**: no significant change

- On **AGI**: little scientific/technical, little regulation

Gornet, M., Delarue, S., Boritchev, M., & Viard, T. (2024, June). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 127-140).

AI ethics through its manifestos

- A **tri-partition of discourses**: technical, regulatory and business/innovation

- On **explainable AI**: more scientific/technical and less regulation

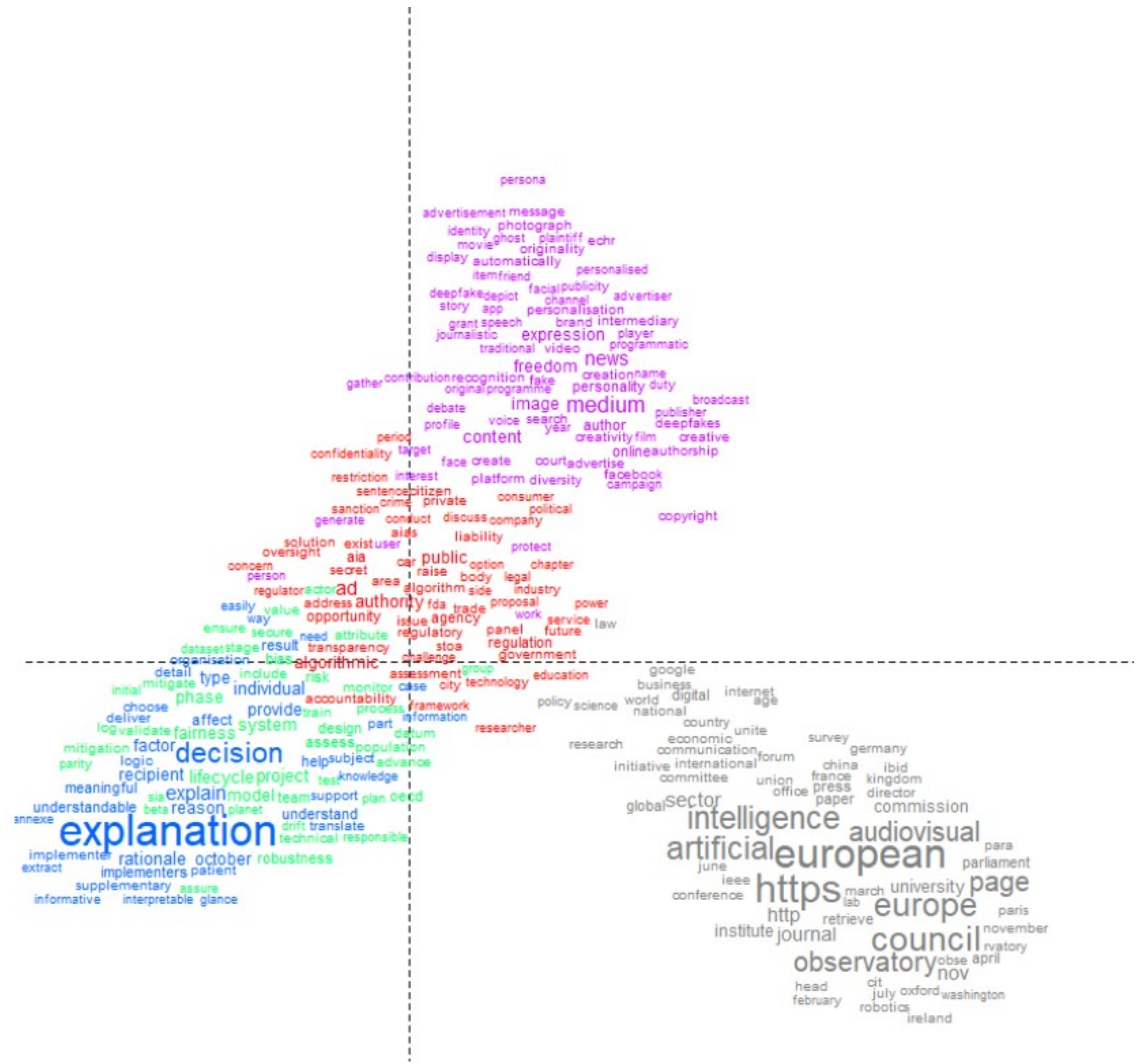- On **fairness**: no significant change

- On **AGI**: little scientific/technical, little regulation

- On **freedom/autonomy**: heavy focus on regulation and social justice

Gornet, M., Delarue, S., Boritchev, M., & Viard, T. (2024, June). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 127-140).

AI ethics through its manifestos

- A structured space, mostly concretised around 3 poles:
  - **Technical** discourses
  - **Regulatory** discourses
  - Business/**innovation** discourses
- Over time, a **structuration** of AI ethics into **AI regulation**
- Still a **lack of operationalisation**, except domain-specific
- Still a **lack of visibilisation**

Gornet, M., Delarue, S., Boritchev, M., & Viard, T. (2024, June). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 127-140).

AI ethics through its manifestos

- A structured space, mostly concretised around 3 poles:
  - **Technical** discourses
  - **Regulatory** discourses
  - Business/**innovation** discourses
- Over time, a **structuration** of AI ethics into **AI regulation**
- Still a **lack of operationalisation**, except domain-specific
- Still a **lack of visibilisation**

Gornet, M., Delarue, S., Boritchev, M., & Viard, T. (2024, June). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 127-140).

The structuration of AI regulation

AI Act works start (Dec. 2017)

The EU Commission nominates the High-Level Expert Group to draft expert guidelines on AI regulation

Jobin et al. (Dec. 2018)

Researchers denote the large number of charters on ethical AI, and their lack of operationalisation.

AI Act is voted (Aug. 2024)

The EU Council and EU Parliament vote on the final text

JTC 21 works start (2021)

The EU Commission nominates the High-Level Expert Group to draft expert guidelines on AI regulation
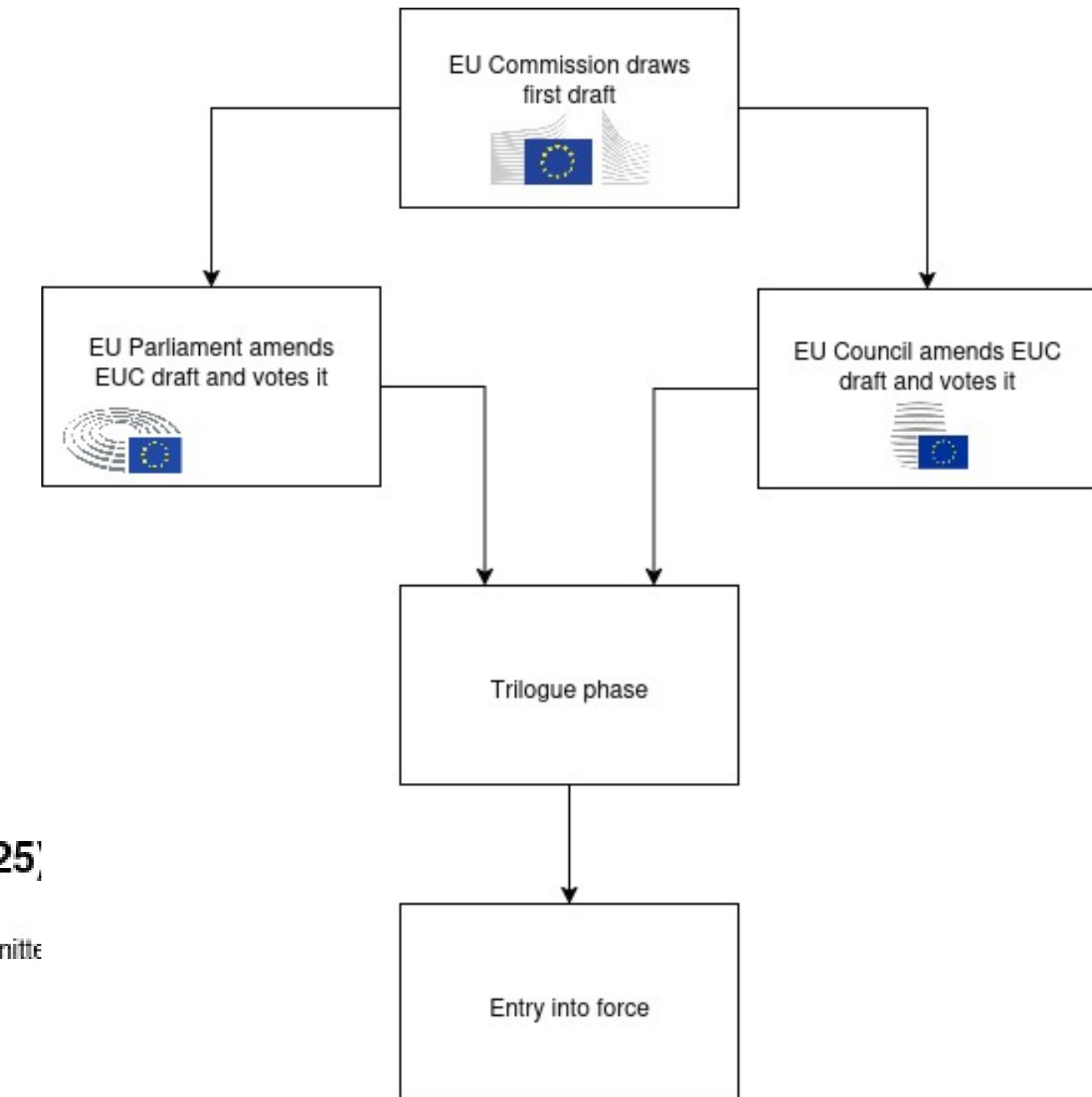
AI Board publishes CoP (July 2025)

AI Office publishes the work of the code of practice committee the implementation of Article 56

EU Commission publishes draft (Mar. 2021)

The EU Commission elaborates a first draft to be transmitted to the EU Council and Parliament

EU Commission draws first draft

EU Parliament amends EUC draft and votes it

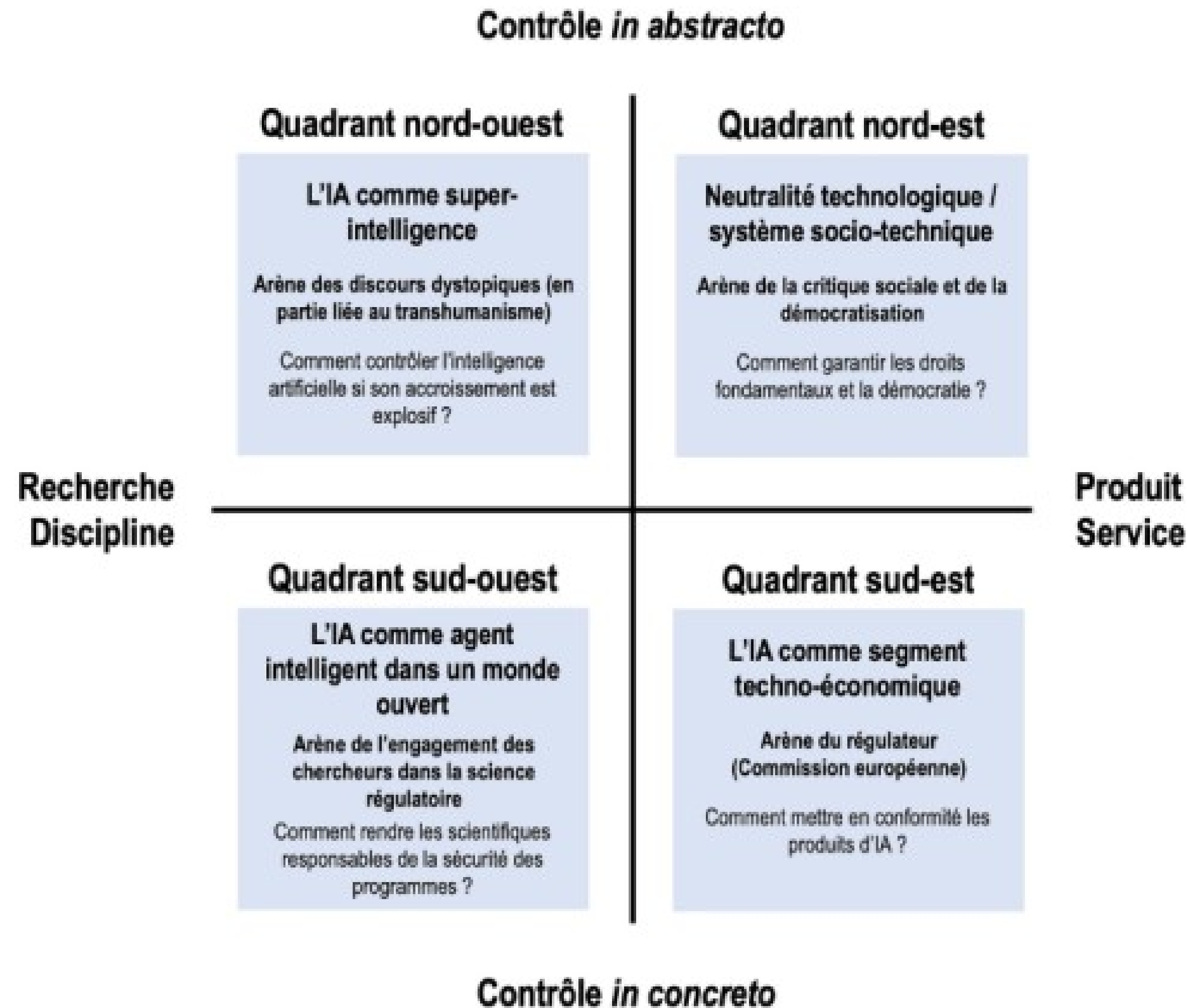EU Council amends EUC draft and votes it

Trilogue phase

Entry into force

A continuum from **charters to regulation and standards**

AI ethics and AI safety: different framings of risk

- AI safety: a recent arena, focusing more on **existential risks** (see later today)
- Spearheaded by **private sector AI leaders** (Musk, Altman), a few **high-profile AI academics** (Hinton, Bengio), and **transhumanist and hyperutilitarian philosophers** (Bostrom, Kurzweil)
- Heavily **North-America centered**
- Has garnered policymaker attention: AI Act code of pratice, Bletchley conference, AI Summit, AI safety institute, Center for AI safety…
- Key argument: AI is **comparable to nuclear technology**, and should be **documented/regulated** as such
  - A **technology-centered approach to regulation**, that ignores non-technical aspects
  - Ties to the AI as 'super intelligence' technoreligion

THIBAULT PRÉVOST

LES PROPHÈTES DE L'IA

Jacques Ellul, critique de la technique

LANGDON WINNER

LA BALEINE ET LE RÉACTEUR

À LA RECHERCHE DE LIMITES AU TEMPS DE LA HAUTE TECHNOLOGIE

ÉDITIONS LIBRE
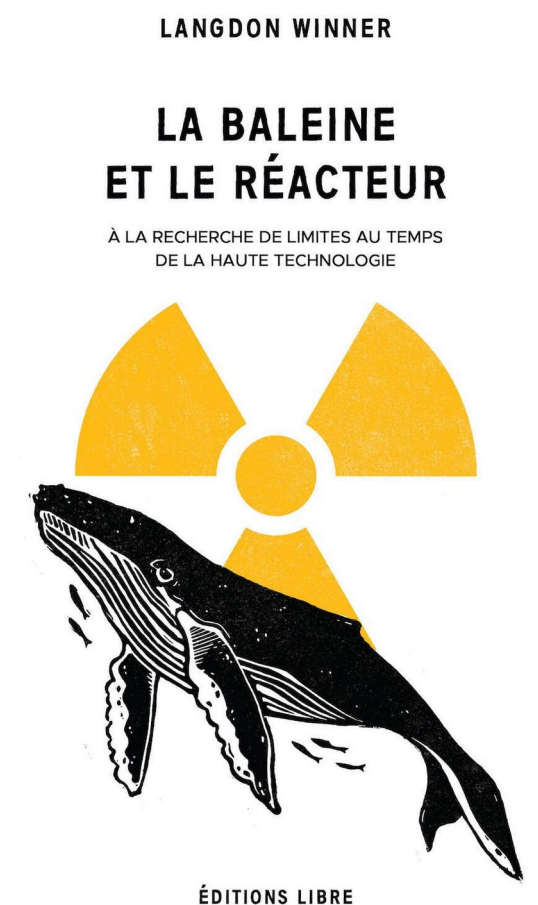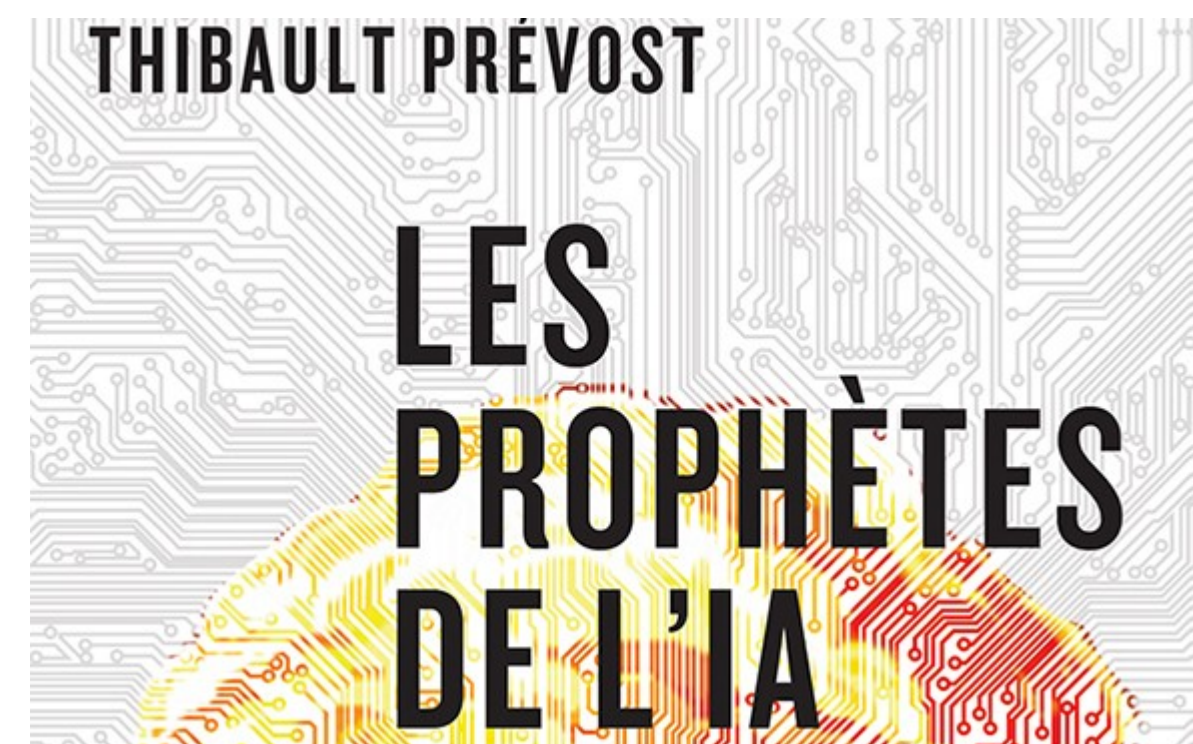
Categorising discourses in normative arenas

AI as super-intelligence

Focused on transhumanist perspectives and the fear of loss of control

Base assumption: the risk for an AI "takeover" is small but must be taken seriously

- A recurring theme since AI's inception: (Wiener, 1950), (Good, 1966), (Moravec, 1988), (Russell, 2019);

- Now outside of specialists' discourses, spearheaded by a few influent individuals (100);

- Has led to the creation of institutes (e.g. Future of Life Institute), research labs.

Main criticism: focusing on **hypothetical long-term harms** distracts from **addressing real short-term harms** (Bender et al., 2021)

Intelligent agents in an open world

The above abstract control of AI is seen as a lukewarm take. Focus on **short-term risks**.

The Asilomar meeting on long-term AI (2009) aims at gathering public scientific positions on AI, in response to "perception of emergency by non-experts"

- Initiative by AAAI, a leading research organization,

- Calls for regulation of "open-world" AI,

- Aims at allowing experts to identify risks and countermeasures

- AI is seen through is technical component, that can be certified

In 2009, machine learning was less prevalent; questions of social justice, environmental harm have taken space

One Hundred Year Study on AI, Partnership on AI

Tech neutrality and sociotechnical systems

Previous arenas focused on technical solutions. But AI is also a concrete "service", and its control must be holistic.

2 main axes of AI criticism:

- AI comforts power structure instead of challenging them (Weizenbaum, 1976), (Chamak, 2004)

- Machine learning presents systematic bias that are under-addressed (Buolamwini et Gebru, 2018)

This criticism is addressed through:

- audits, policy reports and social sciences studies,

- financing academia to provide technical solutions to ethical problems

There is a strong tension between **technical ethical solutions** and seeing AI as an indicator of **deeper political issues**.

AI through technical and economical lenses

- Goal: building **conformity** for AI products, especially on the **European market**.

- Large definitions create **judicial insecurity**; the response has been in creating adaptable texts.

- Von der Leyen's mandate is marked by digital regulation: GDPR, DSA, DMA, AI Act

    Leading to a *Brussels effect* on american and asian partners (Anu Bradford)

- Initial solution: a broad in-text definition with easily amendable appendices

- Search for an equilibrium between in- and out- text definitions (and associated conceptions)

- Conflicting definitions and conceptions of law: see for example the *right to explanations*

The regulation of AI: the EU's AI Act

- First works by the High Level Expert Group in 2021

- Final vote in May 2024

- Entry into force starting May 2025

- Main definitions: **"AI systems"** and **"high-risk AI systems"**

- Reliant on **standards**, **national supervisory authorities**, and now **codes of practice**

- Much left to be decided!

Further resources on EU's AI Act

- https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

- https://artificialintelligenceact.eu/

- Thomas Le Goff's AI Act game: https://www.canva.com/design/DAGF2FfogqE/QpjWW1ghhCr_GMZT_gLJ3A/view

A dense regulatory ecosystem
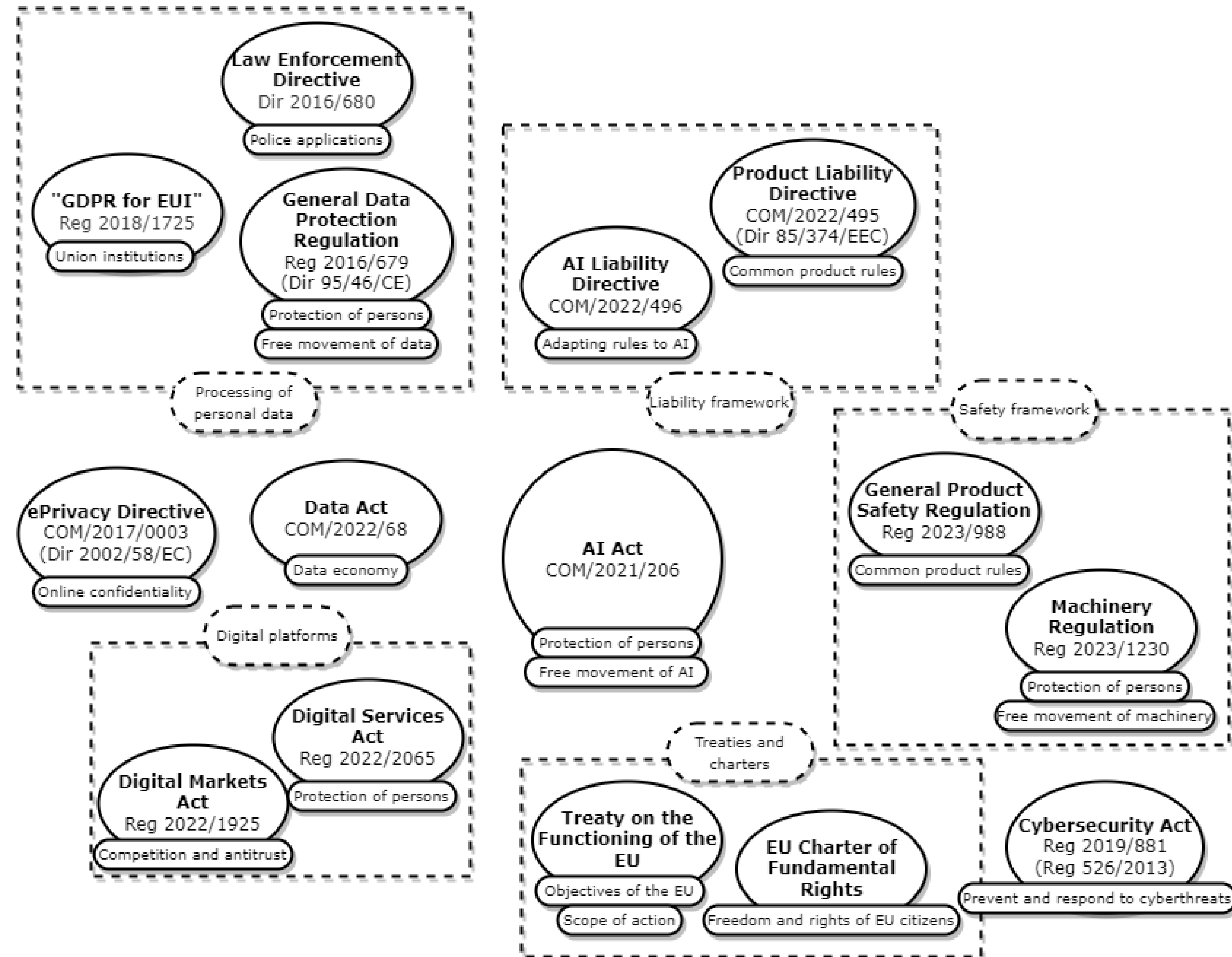
Figure by Mélanie Gornet, reproduced with authorisation

# The regulation of AI: standards
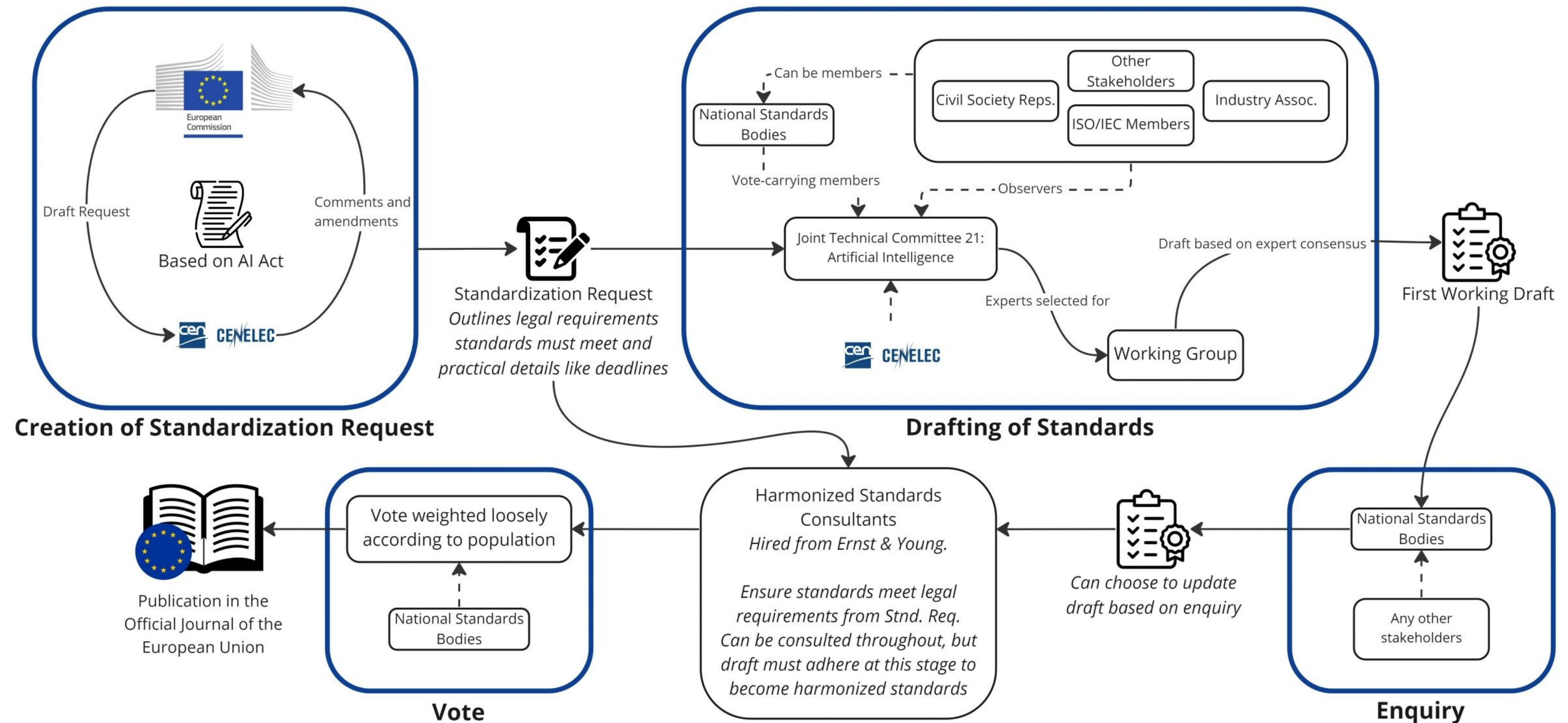
The Development Process for Harmonized Standards

Standards and AI

- Compliance = testing against **harmonized standards** (hENs)

- A common tool for product safety, associated with the **CE mark**

- **AI Act novelty:** certification of fundamental rights

- Difficult process: fundamental rights are heavily **context-dependent** and require **judicial determination**

- **Democratic processes** in standard organisations are put into question (CJEU, 5 march 2024)

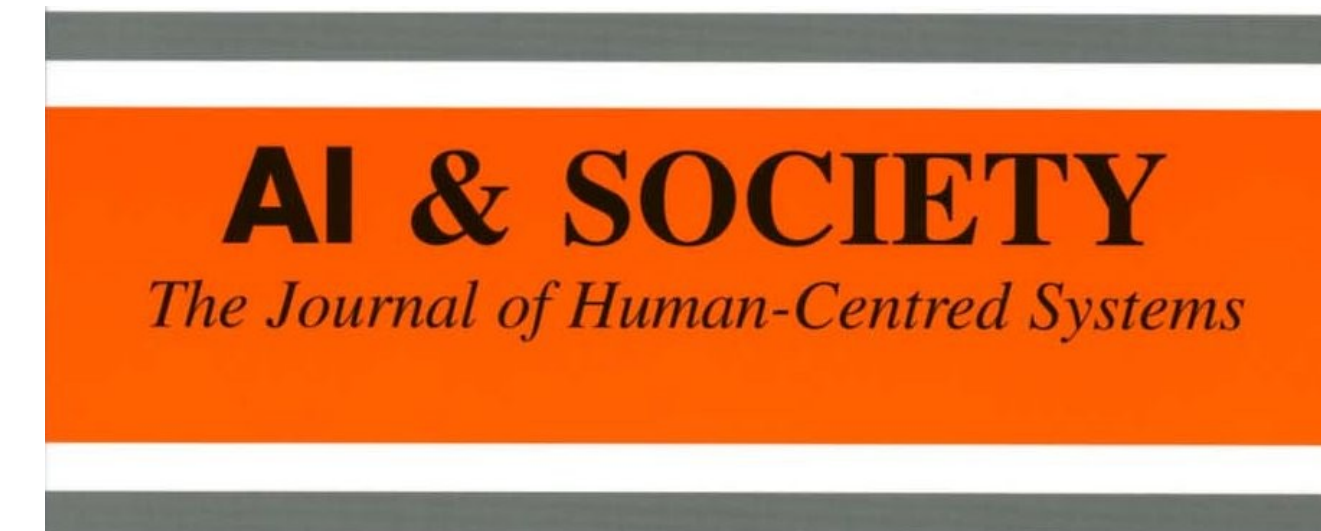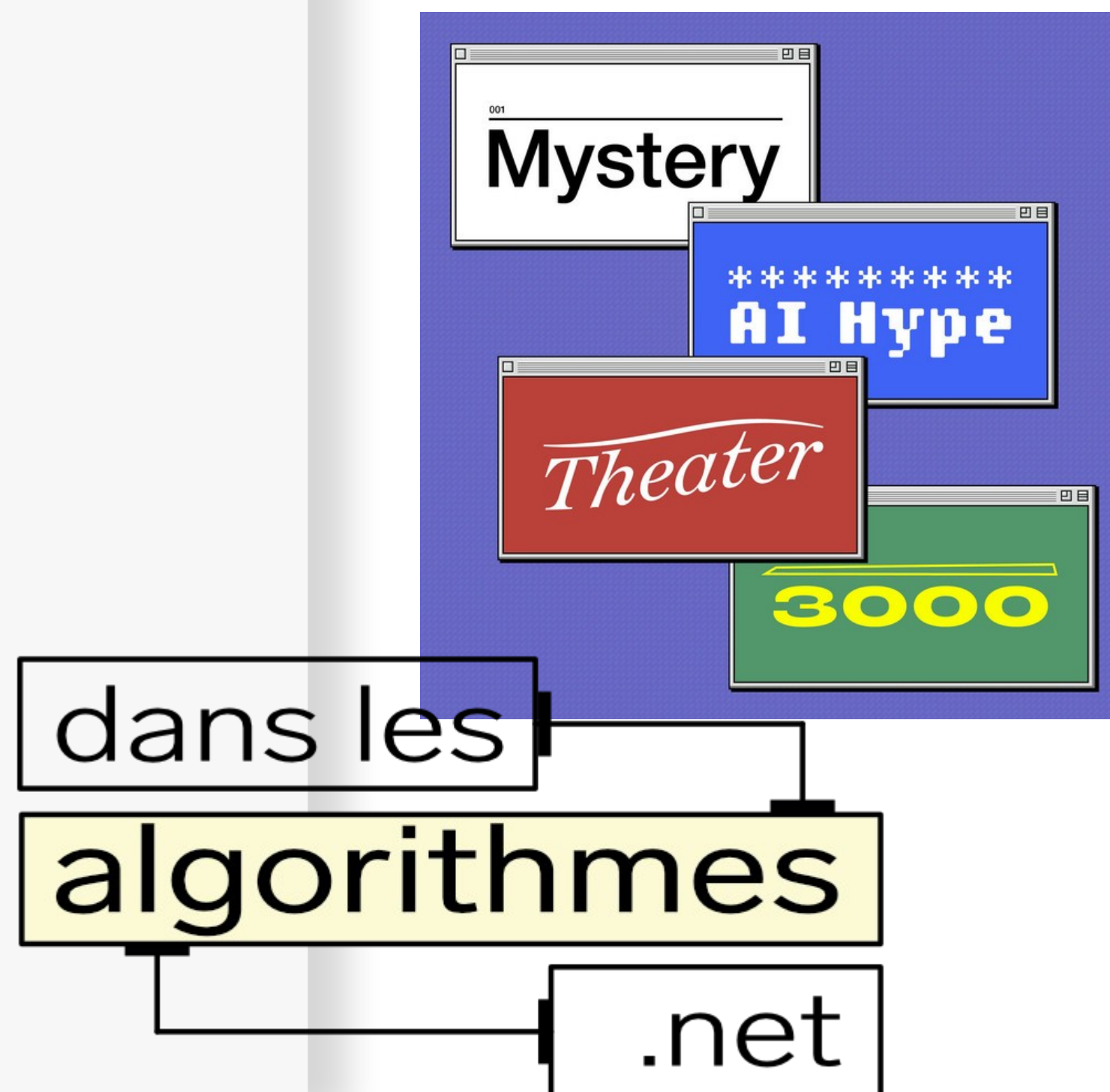# Back to the AI sociotechnical system
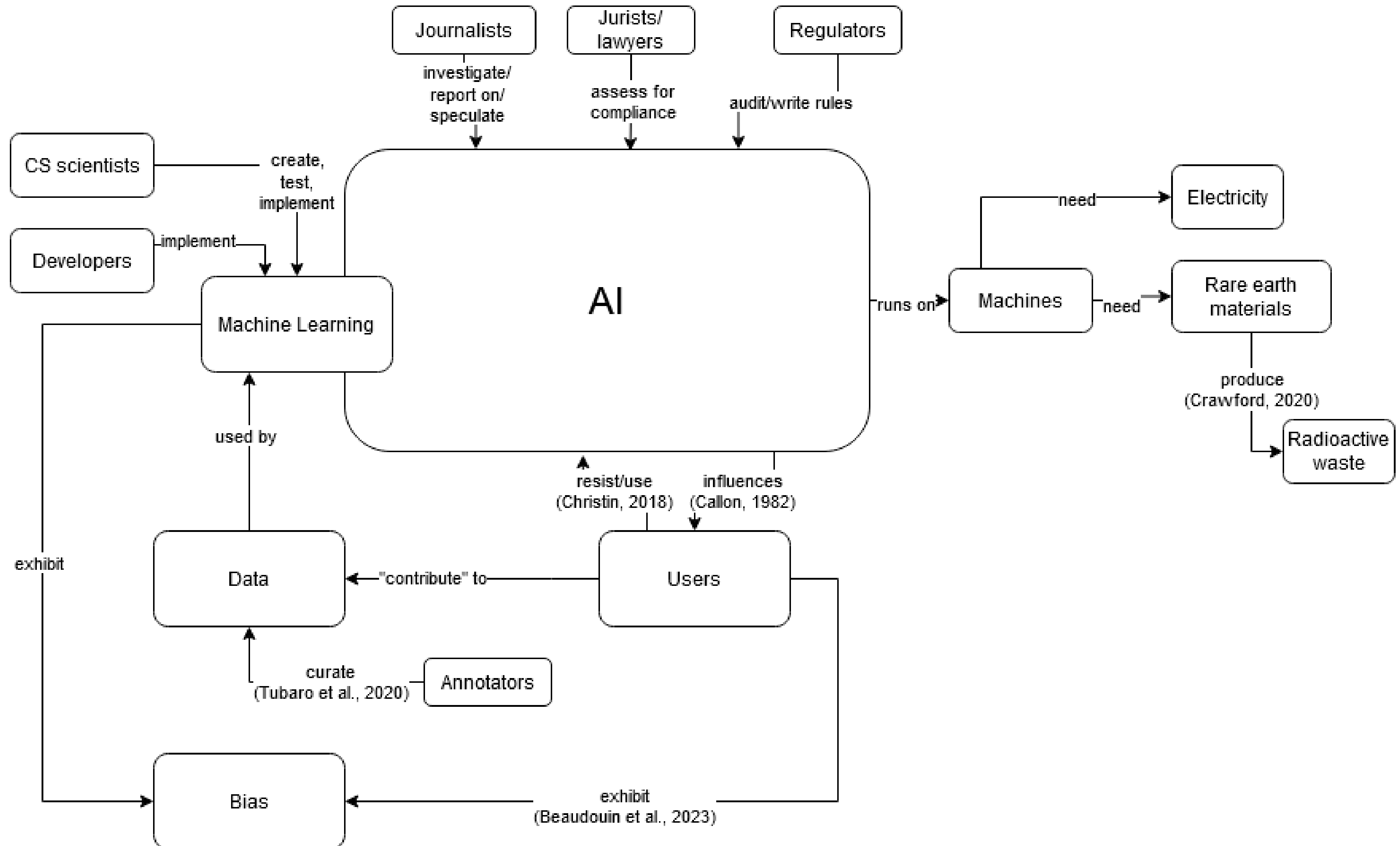
Wrapping up and concluding

What have we learned?

- Defining AI is difficult and shows a sequence of fights for scientific space

- AI has **material consequences**, both in terms of **labour** and **physical resources**

- AI should be studied broadly as a **sociotechnical system**

- AI is a **highly polysemic term**, with different definitions in technical, media, regulatory and scientific arenas

- It is not recent: a socio-history of AI shows **recurrent conflicts of definition**

- Defining AI and AI ethics is a **confrontational space**: the presence (or absence) of discourses is a (sub-)political fight

- We can organize this space in **arenas**, i.e. coherent groups of people with **tenets**

- **Framing, tracking controversies** are key tools in the setting of the course

Thank you!

Next lectures: the AI Act in deeper detail, the ethnography of (machine learning) algorithms, case study discussions, the materiality of AI

Form groups quickly! You will have the case studies by the end of the month.