# An end-to-end approach to ethical AI

## Socio-economic dimensions of the production and deployment of automated technologies

Antonio Casilli

# Why AI ethics?

- AI has been around for 70 years.
- But ethical AI guidelines started proliferating in the second half of the 2010s.
  * Loss of control to some ‘super-intelligence’?
  * Ethical challenges in autonomous technologies (weapon systems, cars...)

Russell S., Dewey D. & Tegmark M. 2015. Research priorities for robust and beneficial Artificial Intelligence. *AI Magazine*, 36(4): 105-114.

# Main debates

# AI guidelines are largely divergent

1. Transparency
2. Justice and fairness
3. Non-maleficence
4. Legal responsibility
5. Privacy

nature machine intelligence

PERSPECTIVE

https://doi.org/10.1038/s42256-019-0088-2

## The global landscape of AI ethics guidelines

Anna Jobin, Marcello Ienca and Effy Vayena*

**In the past five years, private companies, research institutions and public sector organizations have issued principles and guidelines for ethical artificial intelligence (AI). However, despite an apparent agreement that AI should be 'ethical', there is debate about both what constitutes 'ethical AI' and which ethical requirements, technical standards and best practices are needed for its realization. To investigate whether a global agreement on these questions is emerging, we mapped and analysed the current corpus of principles and guidelines on ethical AI. Our results reveal a global convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy), with substantive divergence in relation to how these principles are interpreted, why they are deemed important, what issue, domain or actors they pertain to, and how they should be implemented. Our findings highlight the importance of integrating guideline-development efforts with substantive ethical analysis and adequate implementation strategies.**

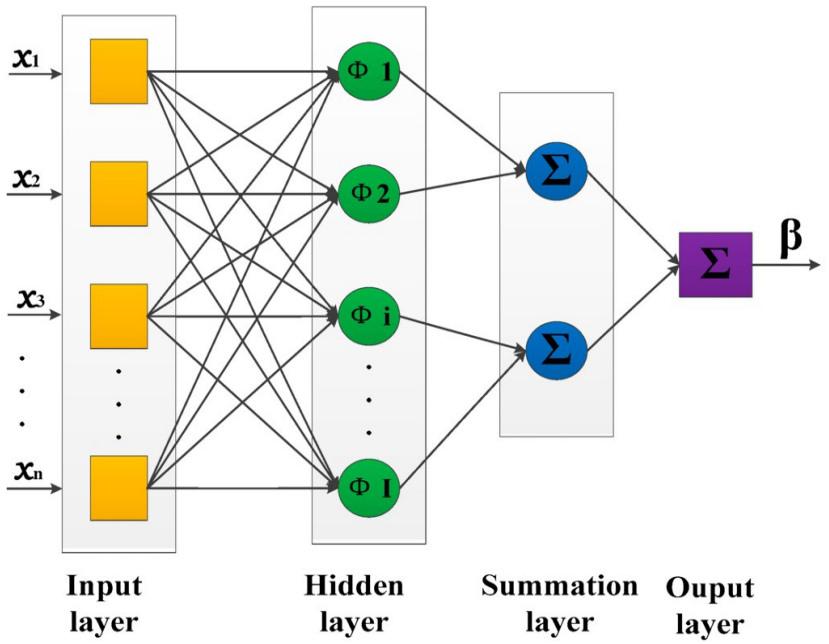
Artificial intelligence (AI), or the theory and development of computer systems able to perform tasks normally requiring human intelligence, is widely heralded as an ongoing "revolution" transforming science and society altogether[1,2]. While approaches to AI such as machine learning, deep learning and artificial neural networks are reshaping data processing and analysis[3], autonomous and semi-autonomous systems are being increasingly used in a variety of sectors including healthcare, transportation and the production chain[4]. In light of its powerful transformative force and profound impact across various societal domains, AI has sparked ample debate about the principles and values that should guide its development and use[5,6]. Fears that AI might jeopardize jobs for human workers[7], be misused by malevolent actors[8], elude accountability or inadvertently dis-

Reports and guidance documents for ethical AI are instances of what is termed non-legislative policy instruments or soft law[10]. Unlike so-called hard law—that is, legally binding regulations passed by the legislatures to define permitted or prohibited conduct—ethics guidelines are not legally binding but persuasive in nature. Such documents are aimed at assisting with—and have been observed to have significant practical influence on—decision-making in certain fields, comparable to that of legislative norms[11]. Indeed, the intense efforts of such a diverse set of stakeholders in issuing AI principles and policies is noteworthy, because they demonstrate not only the need for ethical guidance, but also the strong interest of these stakeholders to shape the ethics of AI in ways that meet their respective priorities[12,13]. Specifically, the private sector's involvement in the AI ethics arena has been called into question

Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

# AI guidelines are largely divergent

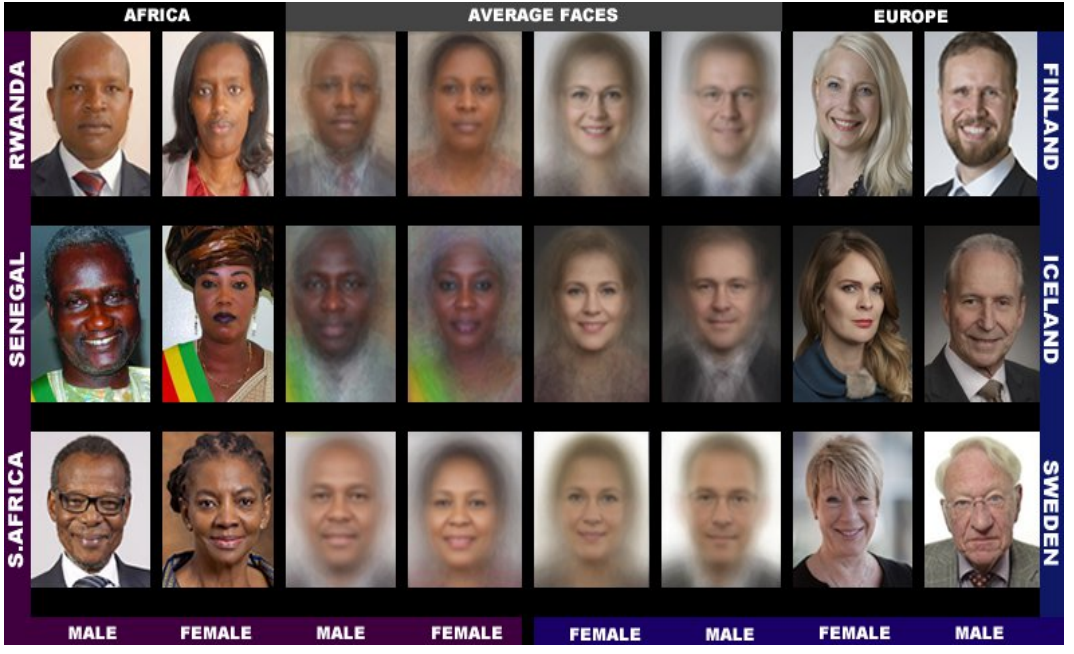
1. Transparency
2. Justice and fairness
3. Non-maleficence
4. Legal responsibility
5. Privacy




Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

# AI guidelines are largely divergent

1. Transparency
2. Justice and fairness
3. Non-maleficence
4. Legal responsibility
5. Privacy




Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

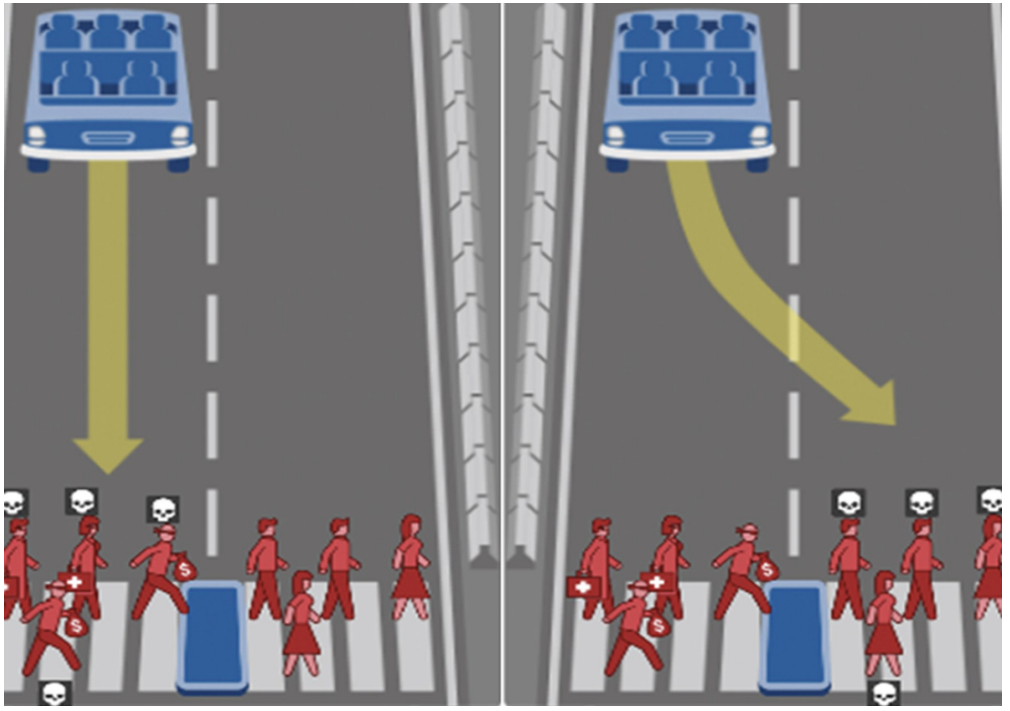
# AI guidelines are largely divergent

1. Transparency
2. Justice and fairness
3. Non-maleficence
4. Legal responsibility
5. Privacy

Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

# AI guidelines are largely divergent

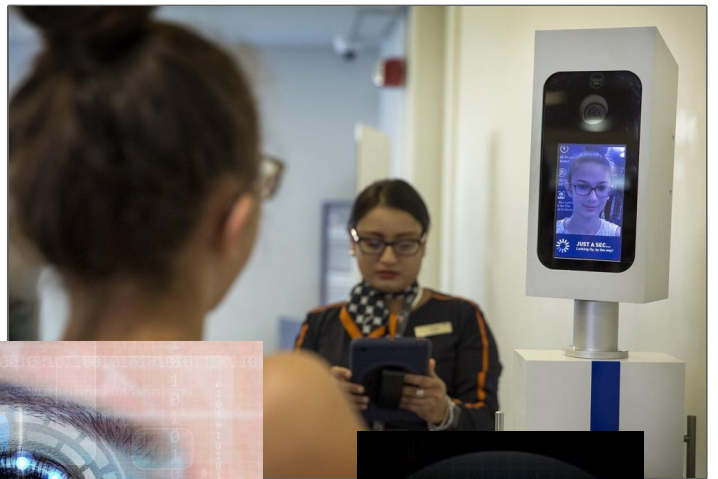
1. Transparency
2. Justice and fairness
3. Non-maleficence
4. Legal responsibility
5. Privacy

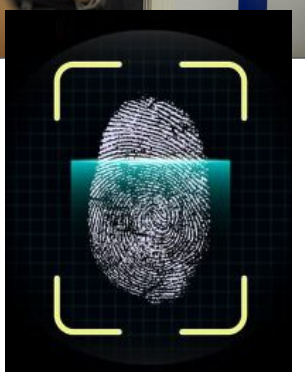


Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

# AI guidelines are largely divergent

1. Transparency
2. Justice and fairness
3. Non-maleficence
4. Legal responsibility
5. Privacy





Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

# AI guidelines are largely divergent

- Different interpretations of shared principles
- Remarkable absence of sustainability, trust, autonomy, dignity, solidarity...
- No links between principles and actionable requirements

Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

# Alignment with industrial interests

- Tech giants co-opt and neutralize critique: denigrate dissenting research + fund weakest critics
- To avoid more stringent regulation
- Focus on ethics to frame issues of power, dominance, inequalities, oppression...

The Steep Cost of Capture

Meredith Whittaker, New York University

This is a perilous moment. Private computational systems marketed as artificial intelligence (AI) are threading through our public life and institutions, concentrating industrial power, compounding marginalization, and quietly shaping access to resources and information.

In considering how to tackle this onslaught of industrial AI, we must first recognize that the "advances" in AI celebrated over the past decade were not due to fundamental scientific breakthroughs in AI techniques. They were and are primarily the product of significantly concentrated data and compute resources that reside in the hands of a few large tech corporations. Modern AI is fundamentally dependent on corporate resources and business practices, and our increasing reliance on such AI cedes inordinate power over our lives and institutions to a handful of tech firms. It also gives these firms significant influence over both the direction of AI development and the academic institutions wishing to research it. Meaning that tech firms are startlingly well positioned to shape what we do—and do not—know about AI and the business behind it, at the same time that their AI products are working to shape our lives and institutions.

Examining the history of the U.S. military's influence over scientific research during the Cold War, we see parallels to the tech industry's current influence over AI. This history also offers alarming examples of the way in which U.S. military dominance worked to shape academic knowledge production, and to punish those who dissented.

Today, the tech industry is facing mounting regulatory pressure, and is increasing its efforts to create tech-positive narratives and to silence and sideline critics in much the same way the U.S. military and its allies did in the past. Taken as a whole, we see that the tech industry's dominance in AI research and knowledge production puts critical researchers and advocates within, and beyond, academia in a treacherous

Insights

- Big tech's control over AI resources made universities and other institutions dependent on these companies, creating a web of conflicted relationships that threaten academic freedom and our ability to understand and regulate these corporate technologies.
- To ensure independent and rigorous research and advocacy capable of understanding and checking these technologies, and the companies behind them, we need to organize, within tech and within the university.



Whittaker,M. (2021). The steep cost of capture, IX Interactions, XXVIII(6): 50-55.

# Alignment with industrial interests

- Tech giants account for the majority of money spent on AI research.
- Authors with corporate ties from 43% to 79%.
- Publications by Alphabet, Microsoft increasing more than fivefold.

Figure 2: Corporate and Big Tech author affiliations.

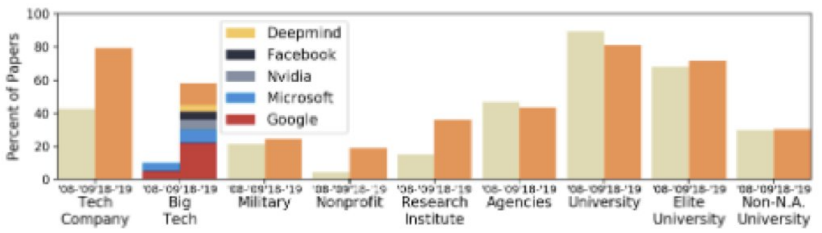
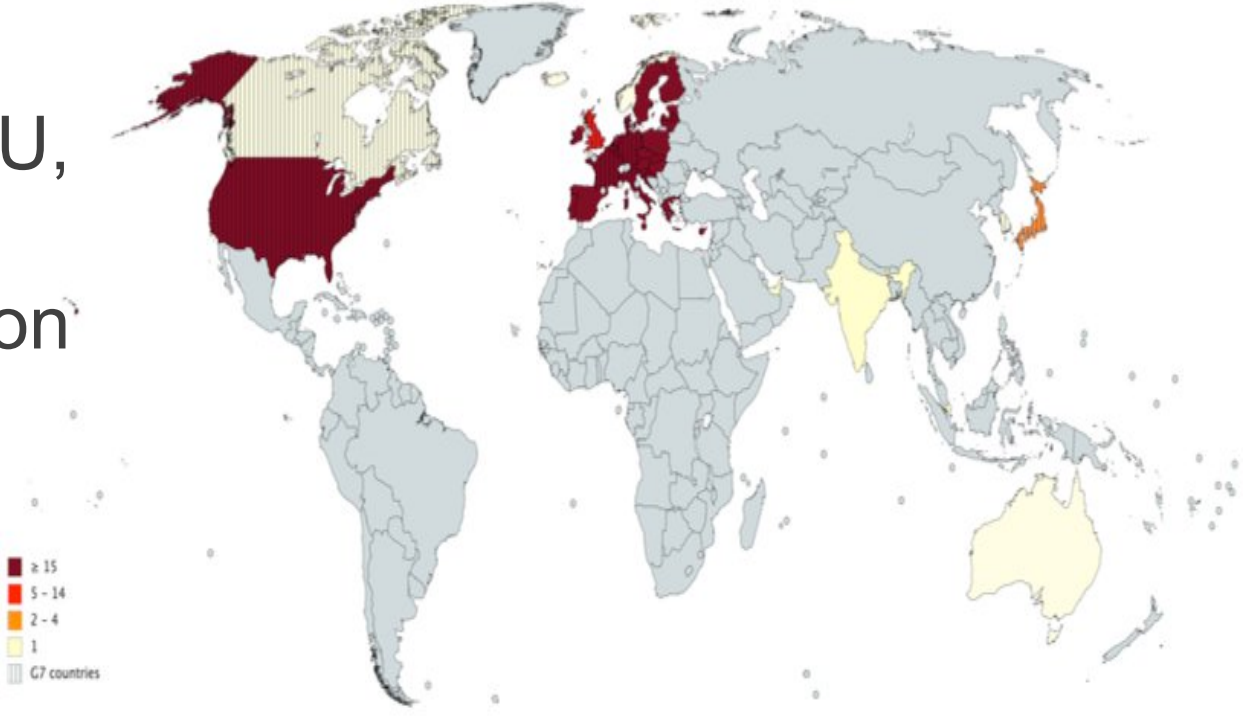



Figure 3: Corporate affiliations and funding ties. Non-N.A. Universities are those outside the U.S. and Canada.

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., Bao, M. (2021). The Values Encoded in Machine Learning Research, arXiv:2106.15590 [cs.LG].

# Global North initiatives

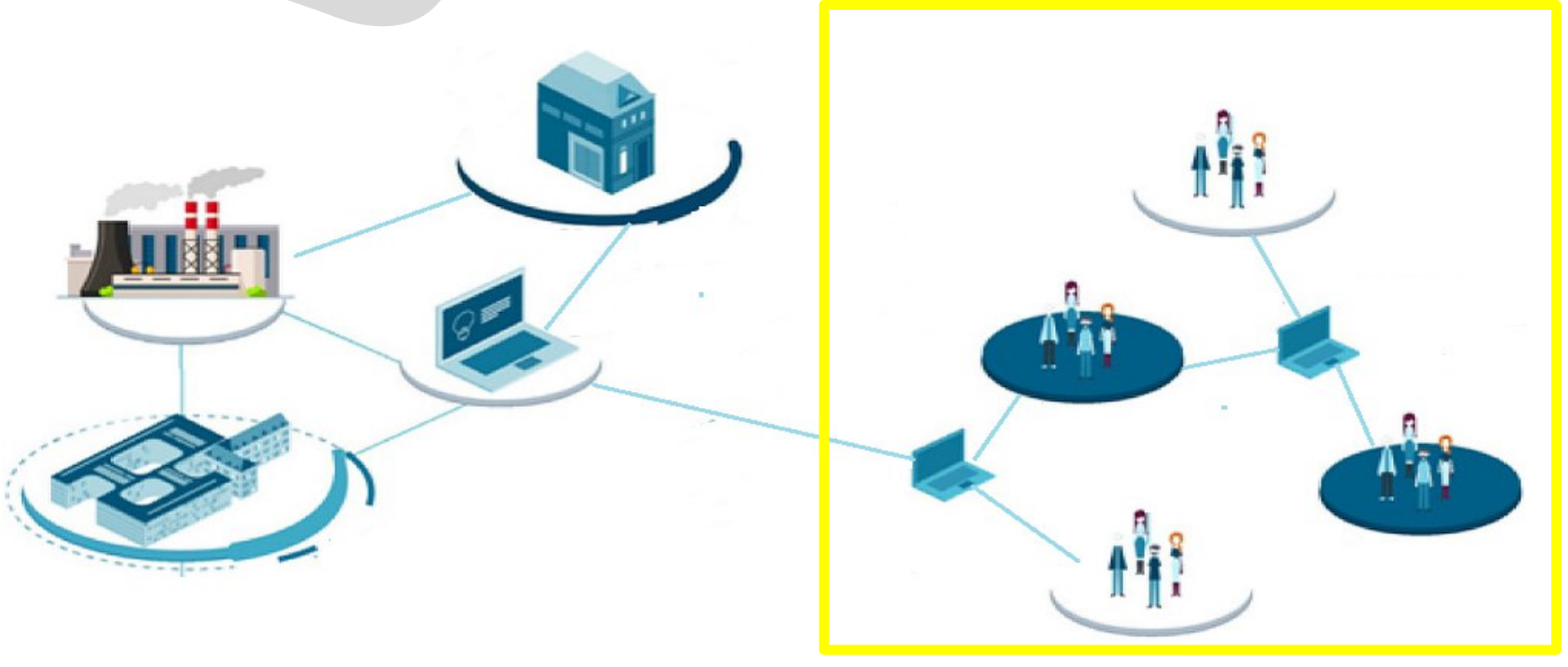
- Most guidelines released in USA, EU, UK and Japan
- Under-representation of the Global South
- May undermine pluralism and cultural awareness




Jobin A., Ienca M. & Vayena E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1: 389-399.

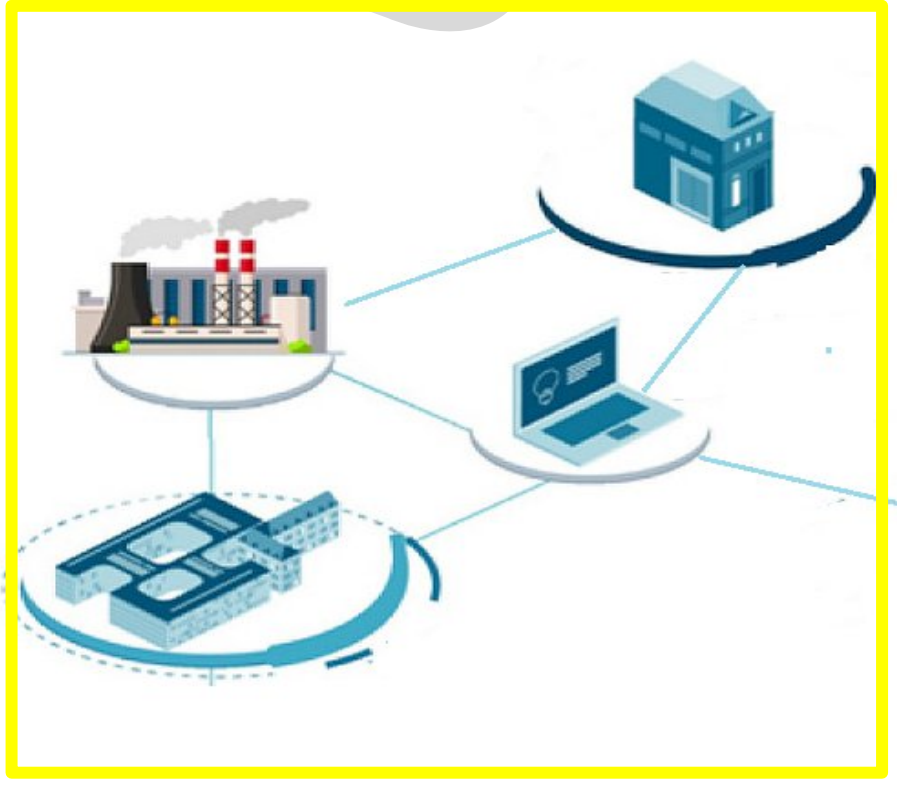
# Placing the human back in the loop

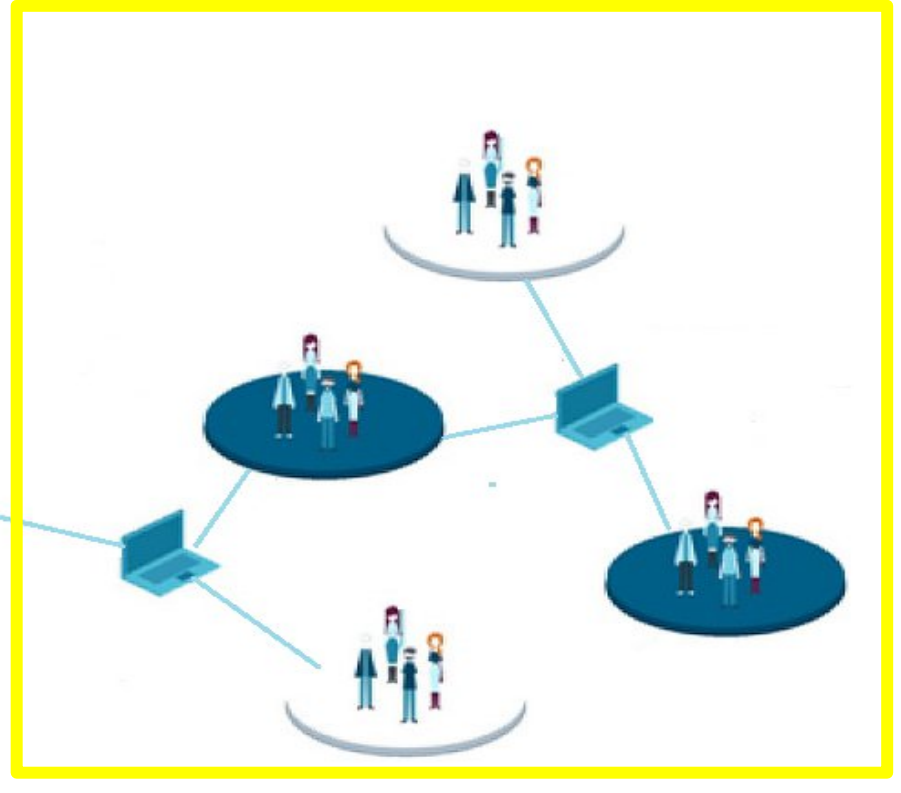
## Production phase vs. deployment phase



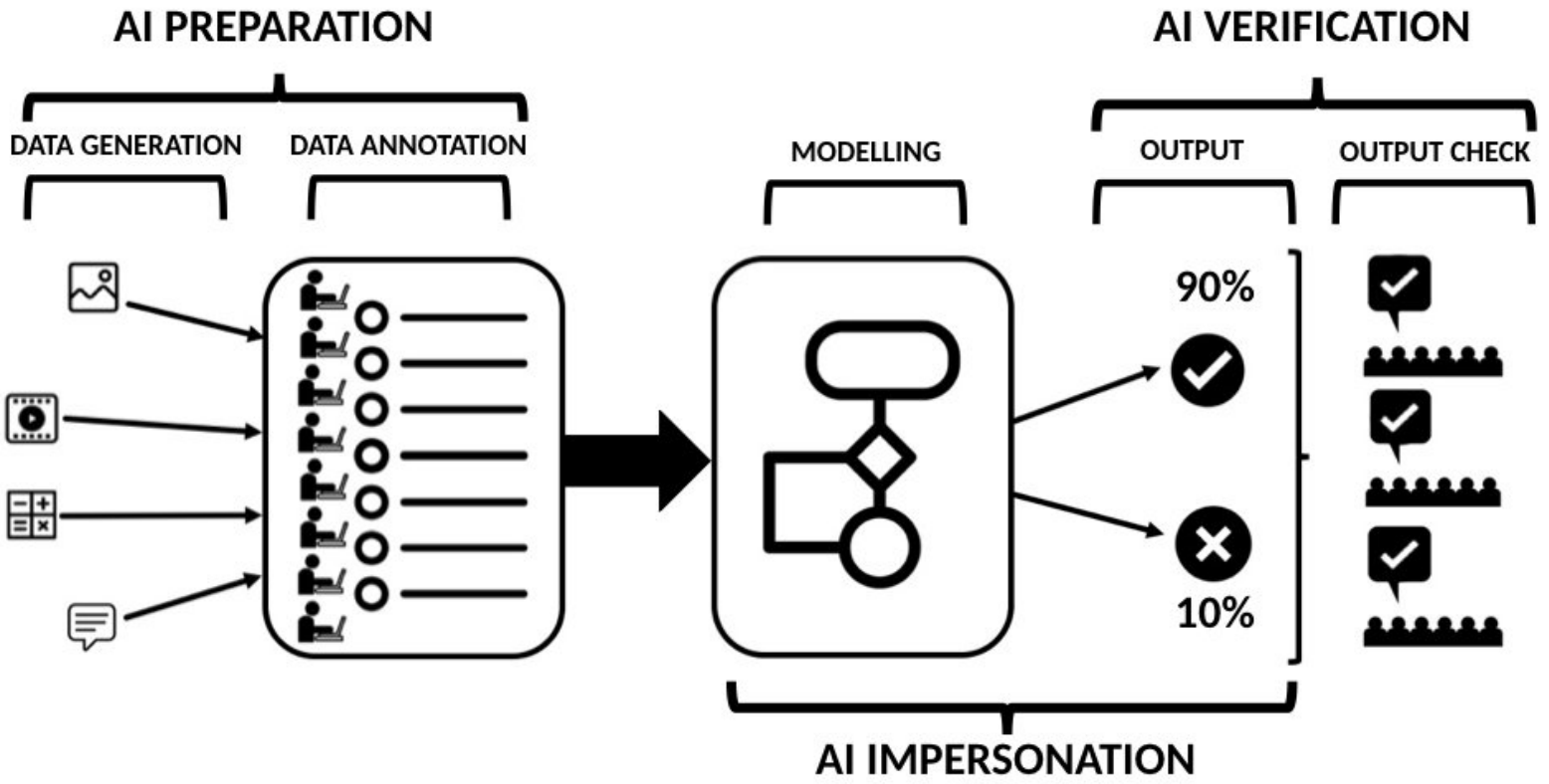
# Placing the human back in the loop

## Production phase vs. deployment phase



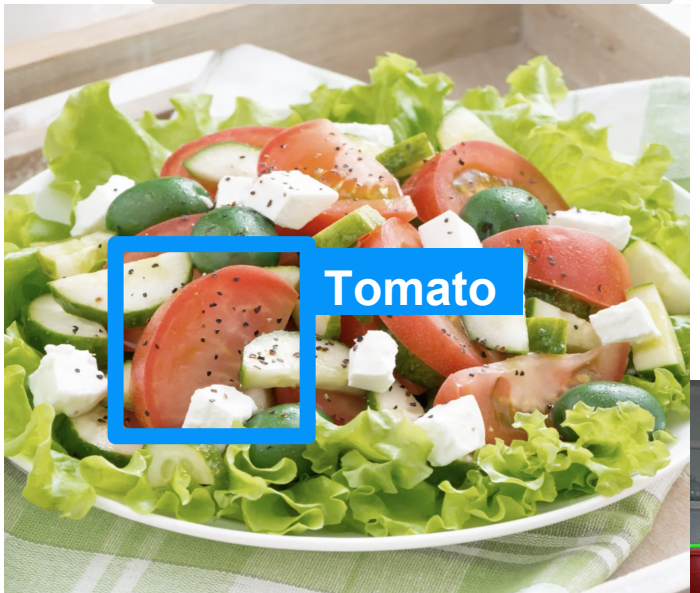


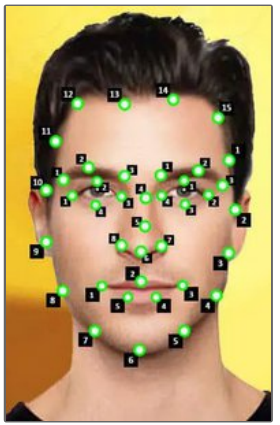
# Workers behind AI

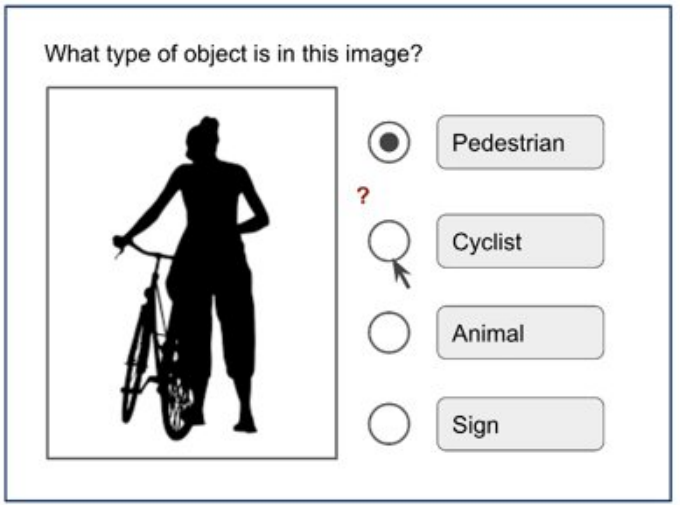


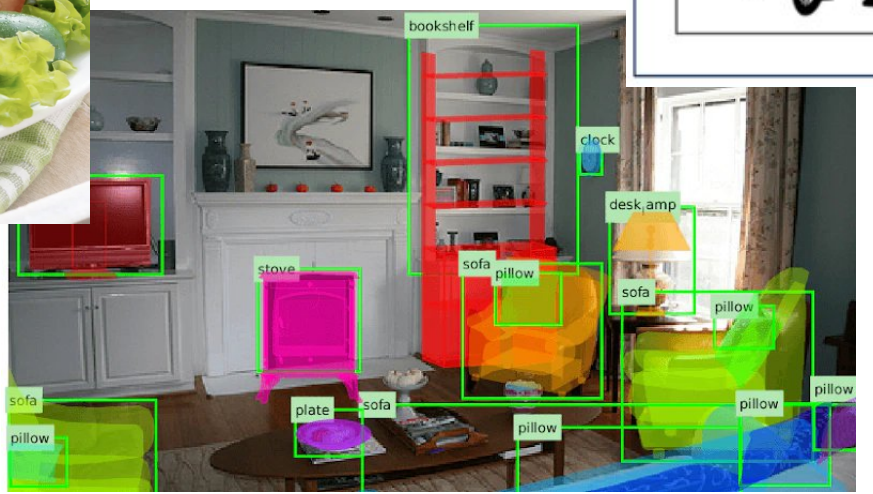

Tubaro P., Casilli A.A. & Coville M. 2020. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1).

# Data labor: training models



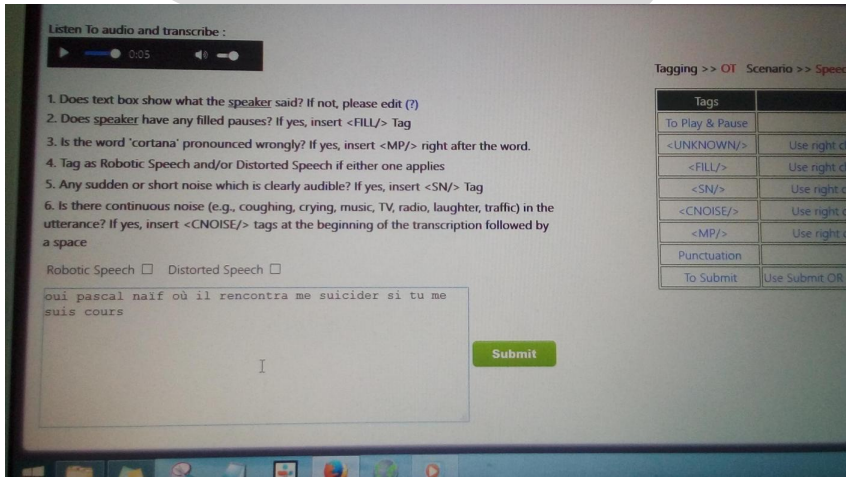


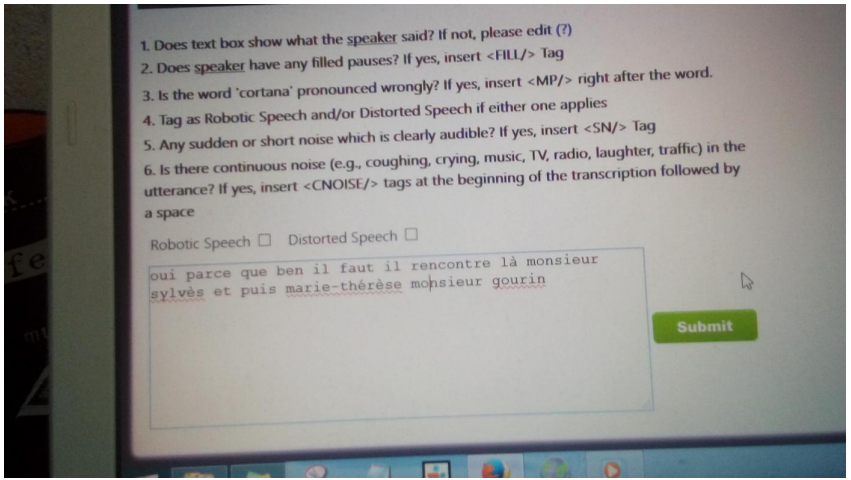




*'They give us a picture of a plate and say 'draw a square around a tomato', we don't know why, everyone knows what a tomato is, I hope...'*

T., FR, 45 yo, 2017

# Data labor: verifying outputs




*'I listened to the audio recording, then a text appeared on screen, showing what [the vocal assistant] understood and transcribed. My job was to check if it was accurate - if not, I had to correct the text'*

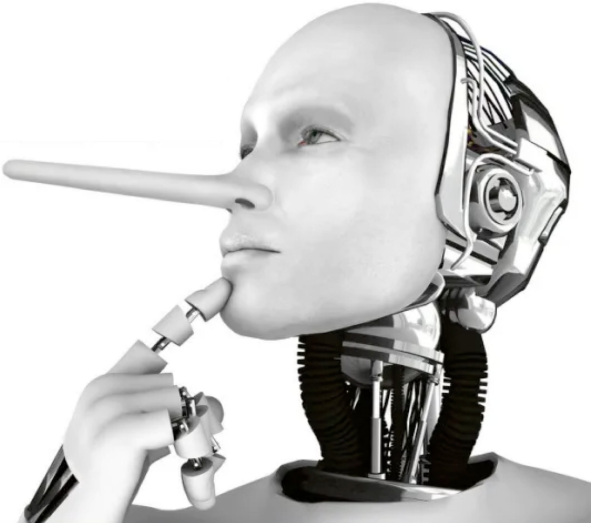
J., FR, 26 yo, 2019

Tubaro P., Casilli A.A. 2022 (in press). Human listeners and virtual assistants: Privacy and labor arbitrage in the production of smart technologies. In F. Ferrari & M. Graham (eds.), *Digital Work in the Planetary Market*. MIT Press, pp. 175-190.

1
2
3
4
5
6

# Data labor: impersonating AI

Technology Behind
**Google Duplex AI**



*'The overwhelming majority of B2B startups we know are human-based. But I understand them because for them, it's a bet on the future. They have to create the data for the appointments and then do machine learning and hope, one day, that the process will be automated'*

K., FR, 42 yo, 2018

*‘Madagascar is the leader in French AI.’*
*K., FR, 42 yo, 2018*

Source: Invisibles – ClickWorkers, France.tv, DW.

# Geographies of data production

HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

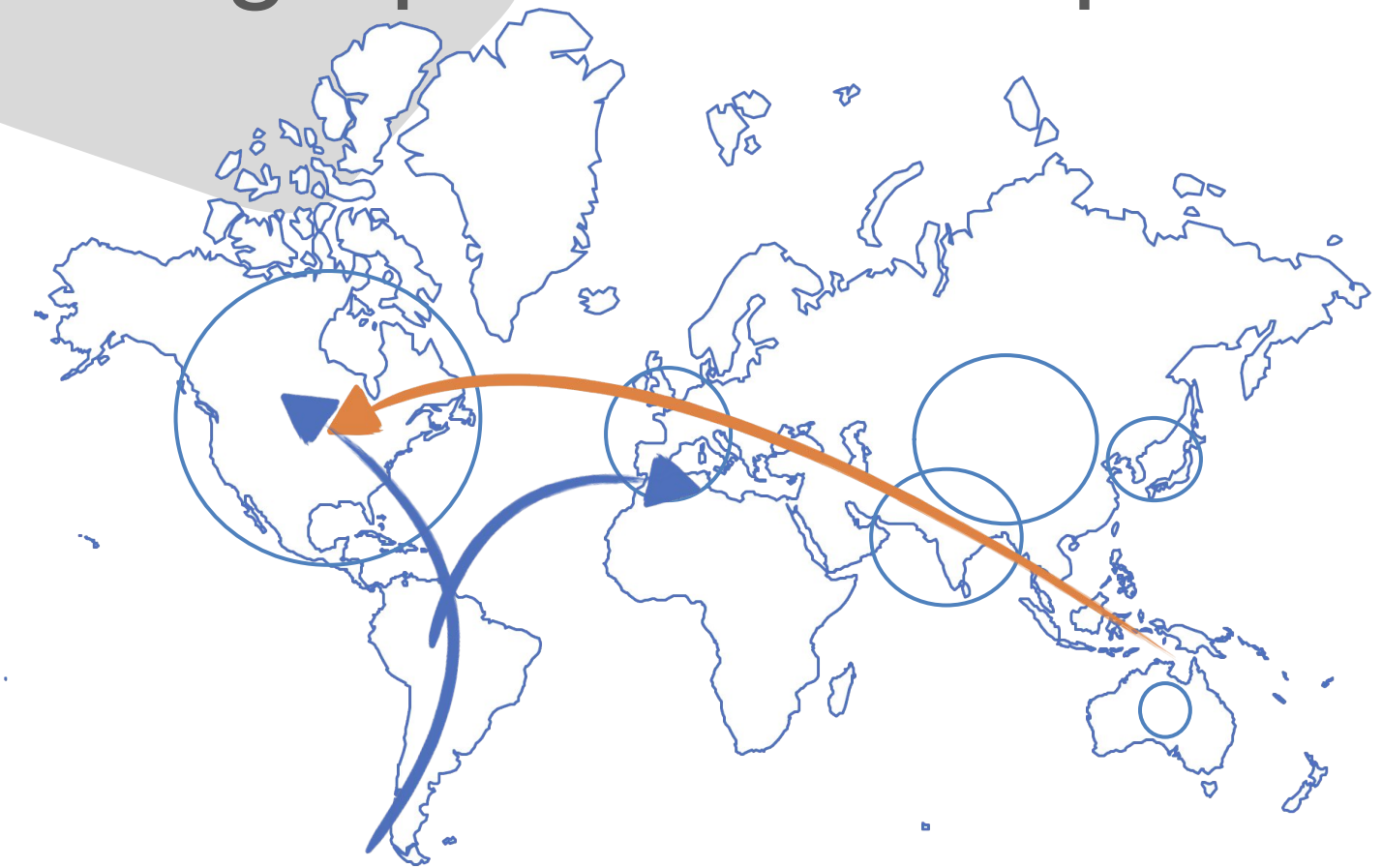
# Geographies of data production

HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

# Geographies of data production

HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

# Geographies of data production

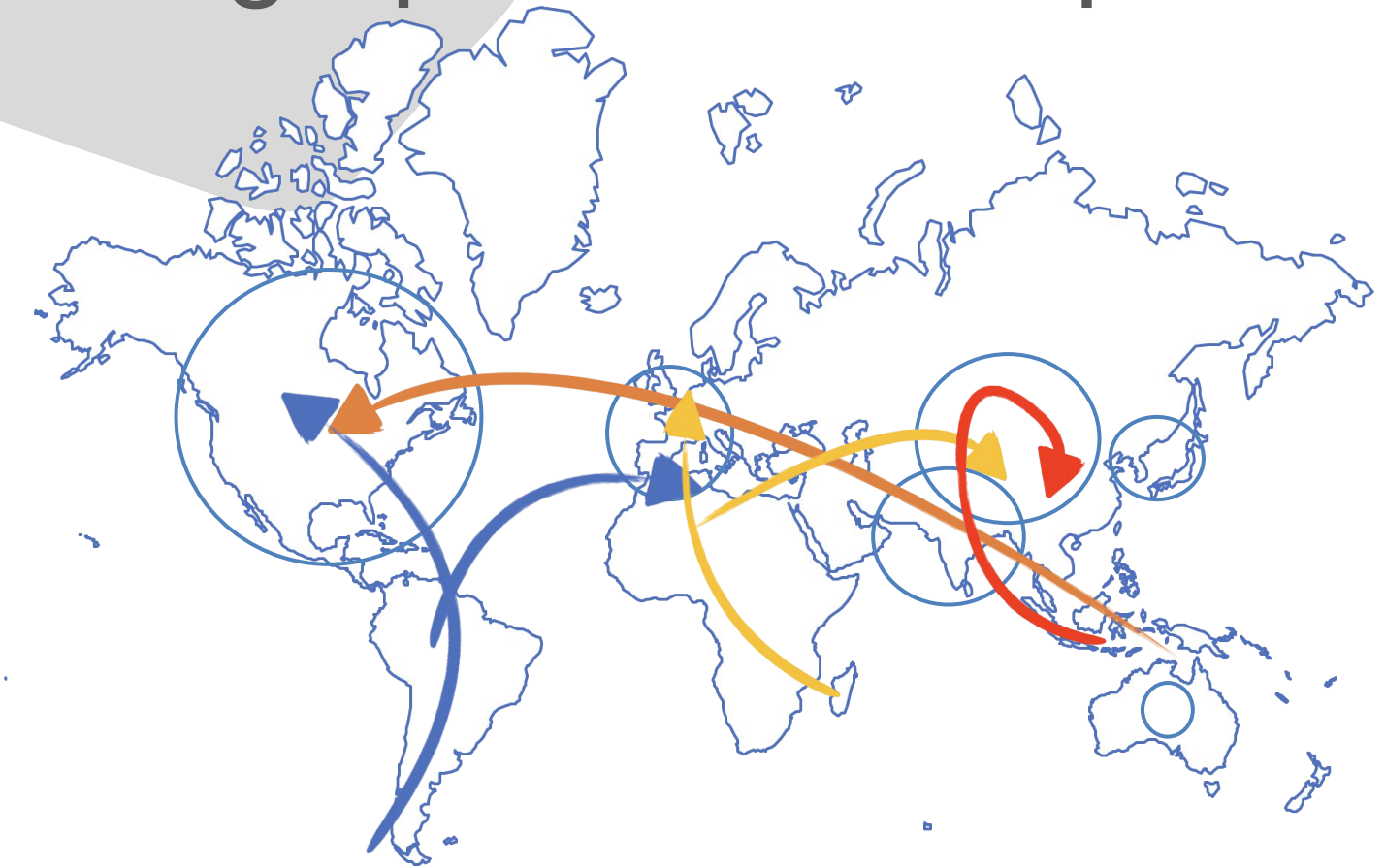


HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

# Geographies of data production

HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

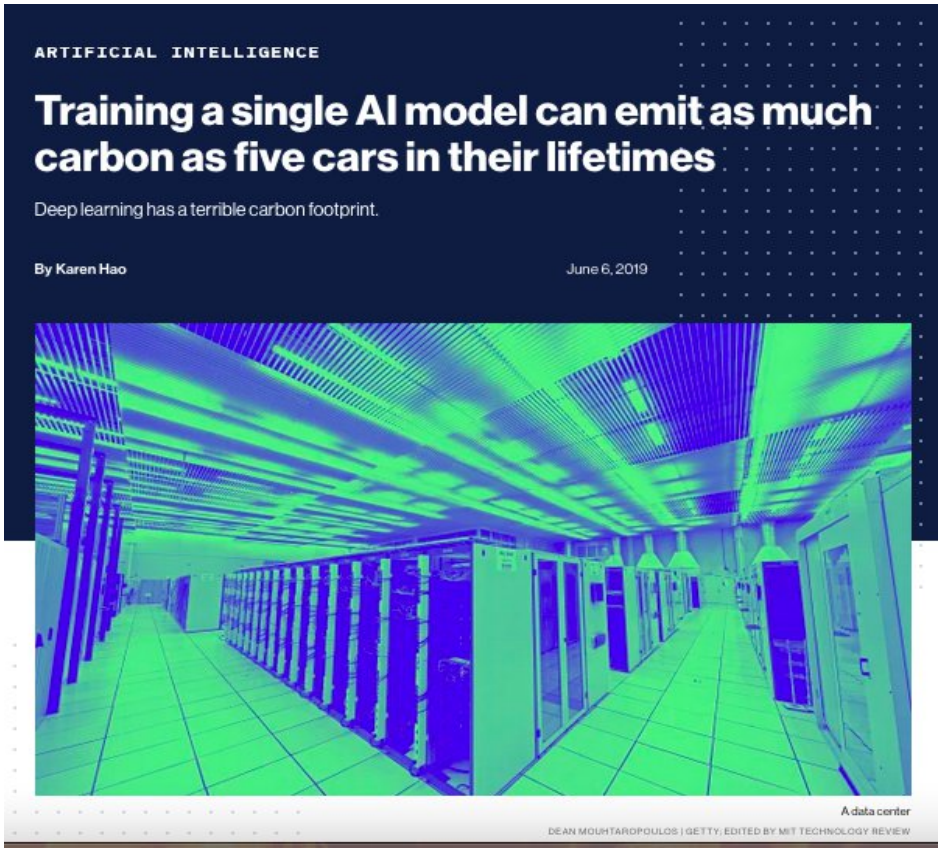
# Geographies of data production



HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

# Environment and natural resources

- Very high environmental costs of ML
- Usually solutions consist paradoxical use of AI to reduce carbon footprint and waste and deadweight assets of AI

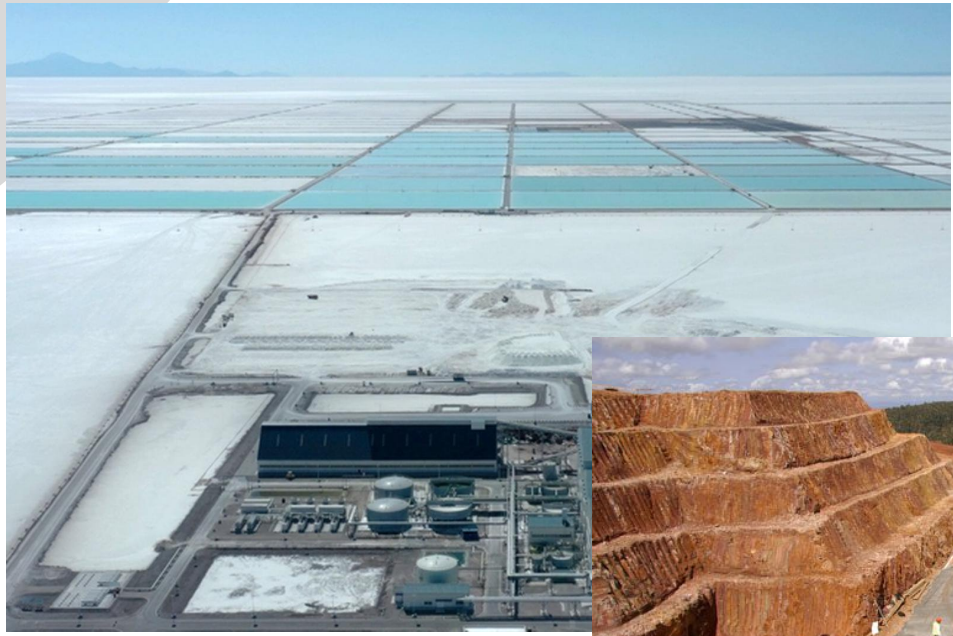



Crawford K. 2021. *Atlas of AI*. Yale University Press.

# Environment and natural resources

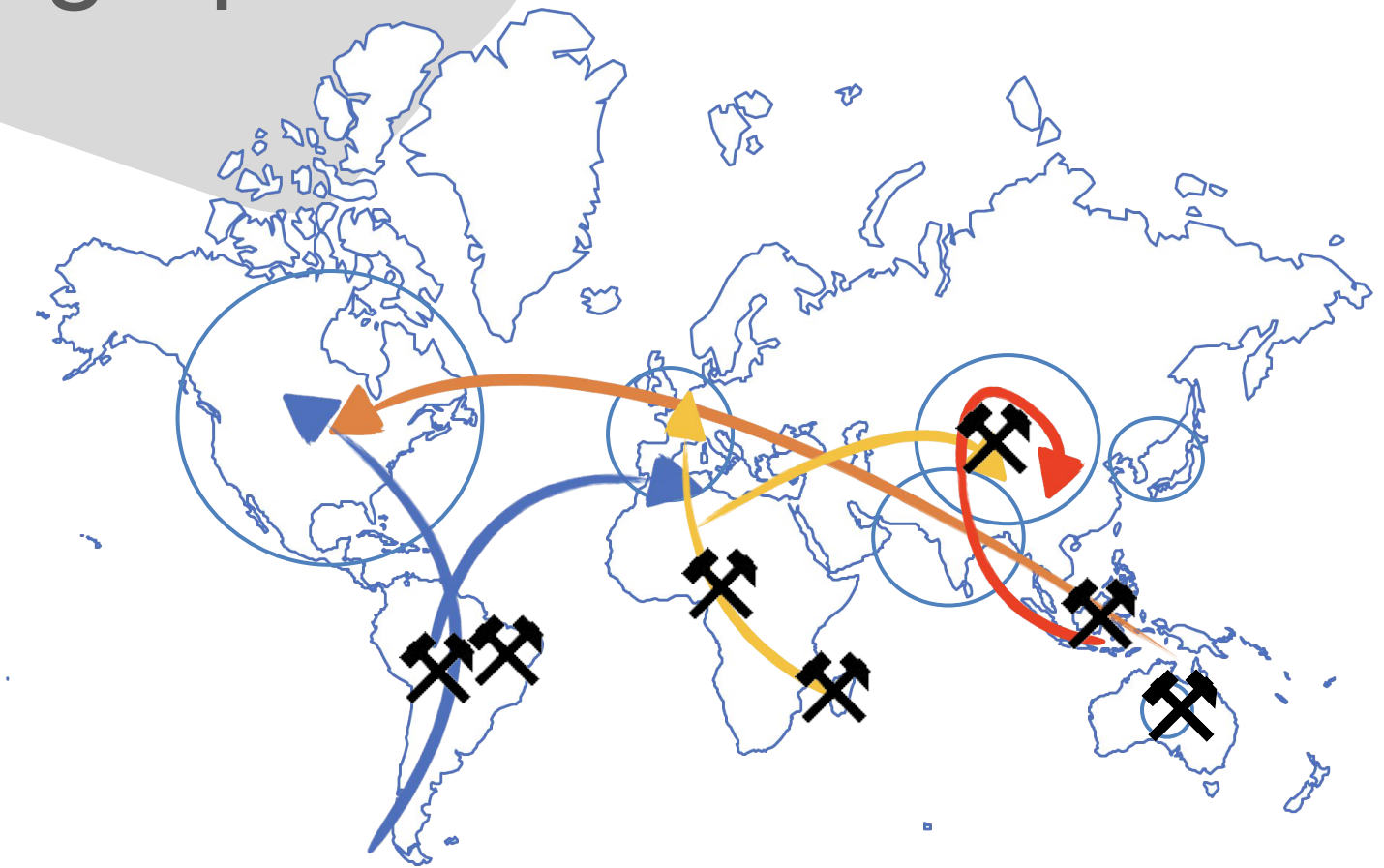
- Extractive industry
  - Cobalt
  - Nickel
  - Lithium
  - Rare earths
- Disrupting economies
- Conflict over land
- International transportation

  ⇨ Huge quantity of labor



Crawford K. 2021. *Atlas of AI*. Yale University Press.

# Geographies of resource extraction



HUSH "The Human Supply Chain Behind Smart Technologies" (2020-2024); TRIA "The Labor of Artificial Intelligence: Ethics and Governance of Automation" (2021-2024)

# Conclusions

# An end-to-end approach to AI?

- Consider AI ethics in the broad framework of its production systems – not only its (future) deployment
- Tech-only solutions insufficient
- Take into account natural and human/social environment
  - Consider the political economy of AI, embedded in global dependencies
  - Insure humane working conditions
  - Enforce stringent standards all along the AI supply chain

# Thank you


- antonio.casilli@telecom-paris.fr