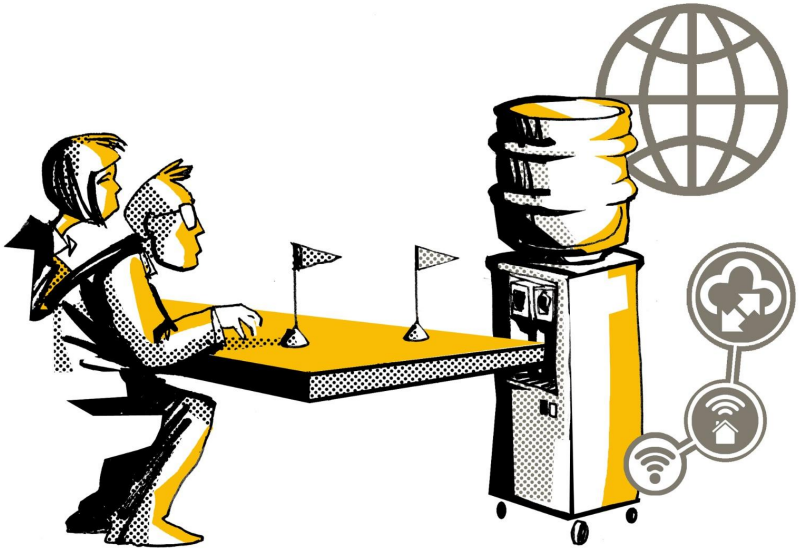
Biases about biases

What we often miss in discussions on AI inclusion

DiPLO Geneva Internet Platform

Context



- AI is here to stay. How do we leverage it for inclusive and sustainable development?

What do we talk about when we talk about inclusion?

- ❏ Access inclusion
- ❏ Data inclusion
- ❏ Labour inclusion
- ❏ Business inclusion
- ❏ Financial inclusion
- ❏ Gender inclusion
- ❏ Geographical inclusion
- ❏ Legal inclusion
- ❏ Knowledge inclusion
- ❏ Language inclusion
- ❏ Identity inclusion
- ❏ Youth inclusion
- ❏ Elderly inclusion
- ❏ Disability inclusion
- ❏ Policy inclusion
- ❏ …

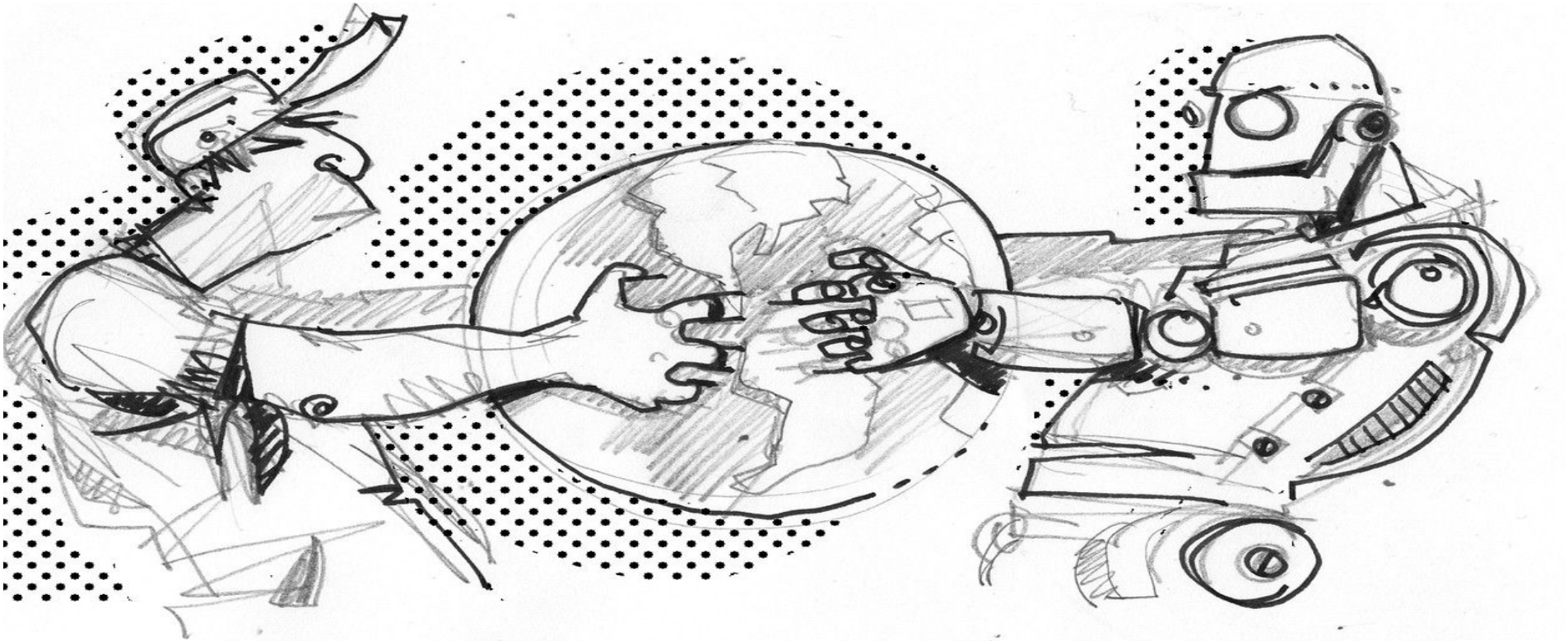
AI has the potential to help break down barriers and promote equity and inclusion, but it can also dangerously amplify existing divides.

Can AI contribute to a more inclusive and equitable world?

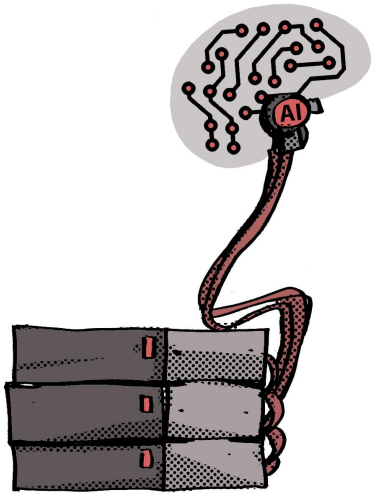
Yes, but only if...

- Transparency and accountability are upheld.
- Ethical considerations guide AI development and deployment.
- Bias is actively addressed.
- Diversity and inclusivity are prioritized.
- Access to AI technology is made equitable.
- Privacy and data protection are safeguarded.
- Fairness is embedded in AI algorithms and decision-making processes.
- Collaboration between policymakers, industry, and civil society is fostered to ensure equitable AI governance.
- Education and training programs promote digital literacy and AI skills among underserved communities.
- Responsible AI procurement practices are adopted by organizations to avoid supporting biased or discriminatory systems.
- Continuous monitoring and evaluation mechanisms are implemented to detect and correct biases and disparities in AI applications.
- Social impact assessments are conducted to anticipate and mitigate potential negative consequences of AI on marginalized groups.
- Regulatory frameworks are established to enforce ethical standards and prevent AI systems from perpetuating inequalities.
- Cross-border cooperation and knowledge sharing foster global collaboration in developing equitable AI policies and practices.

Biases at the core of the ‘battle’ for AI inclusion



The biases we often talk about



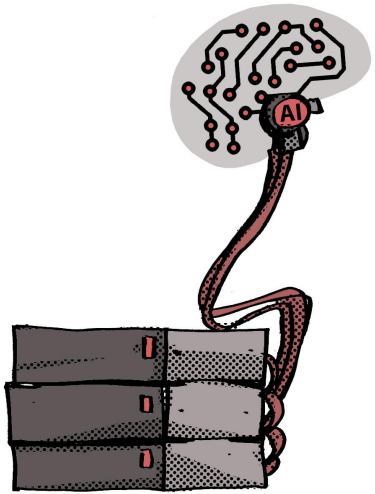
Machine/technical biases

- ❏ Data bias
- ❏ Labelling bias
- ❏ Relevance bias
- ❏ Algorithmic bias
- ❏ Interaction bias

Societal biases

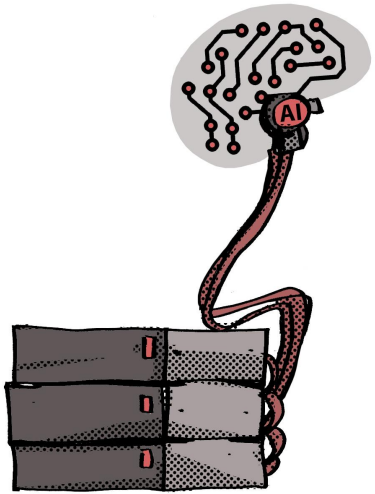
- ❏ Gender bias
- ❏ Racial bias
- ❏ Age bias
- ❏ …

The biases we need to talk more about



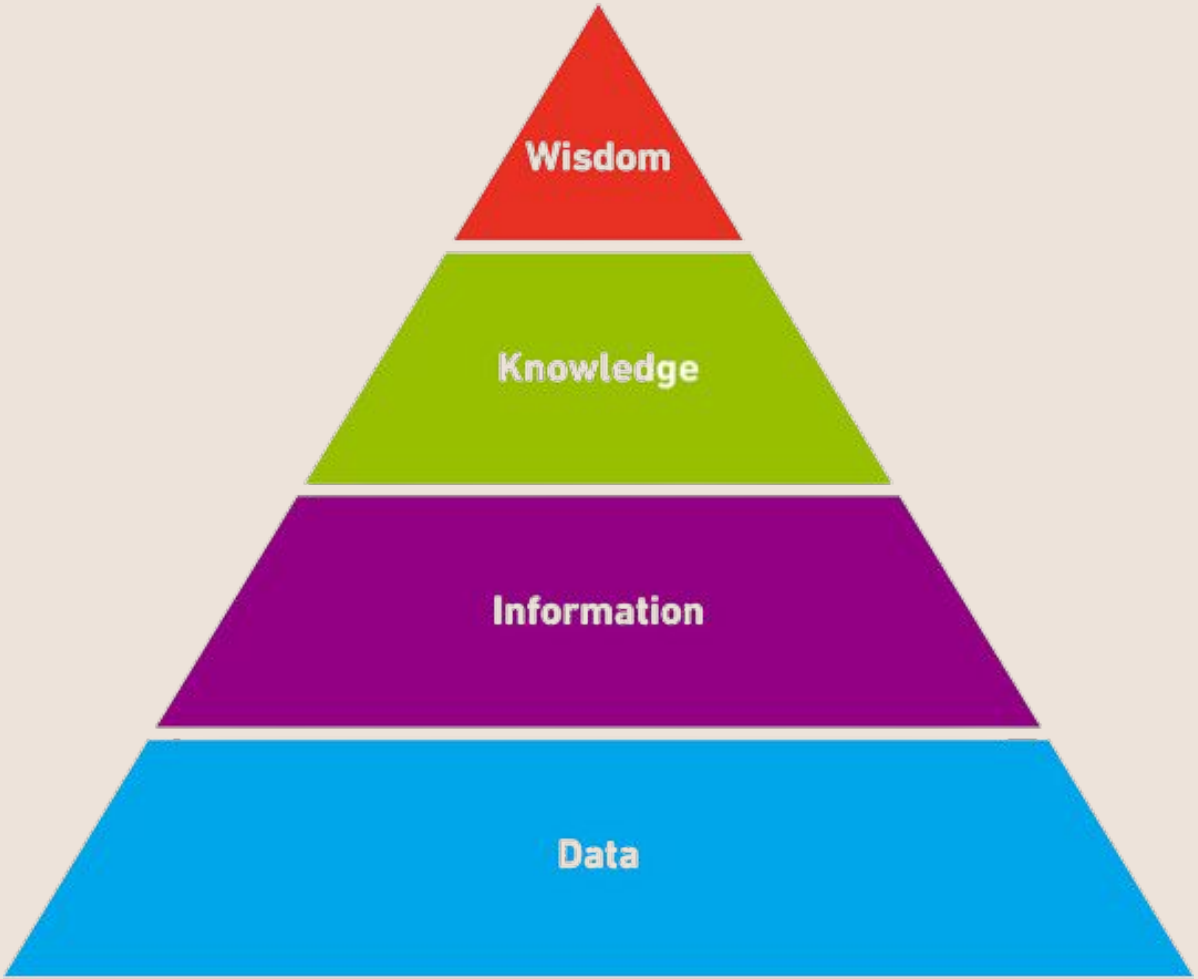
- Denial/minimisation bias
- Selective perception bias
- Bias blind spot
- Attribution bias
- False consensus bias
- Self-serving bias
- Availability bias
- Anchoring bias
- In-group bias
- Confirmation bias
- …

In simple terms

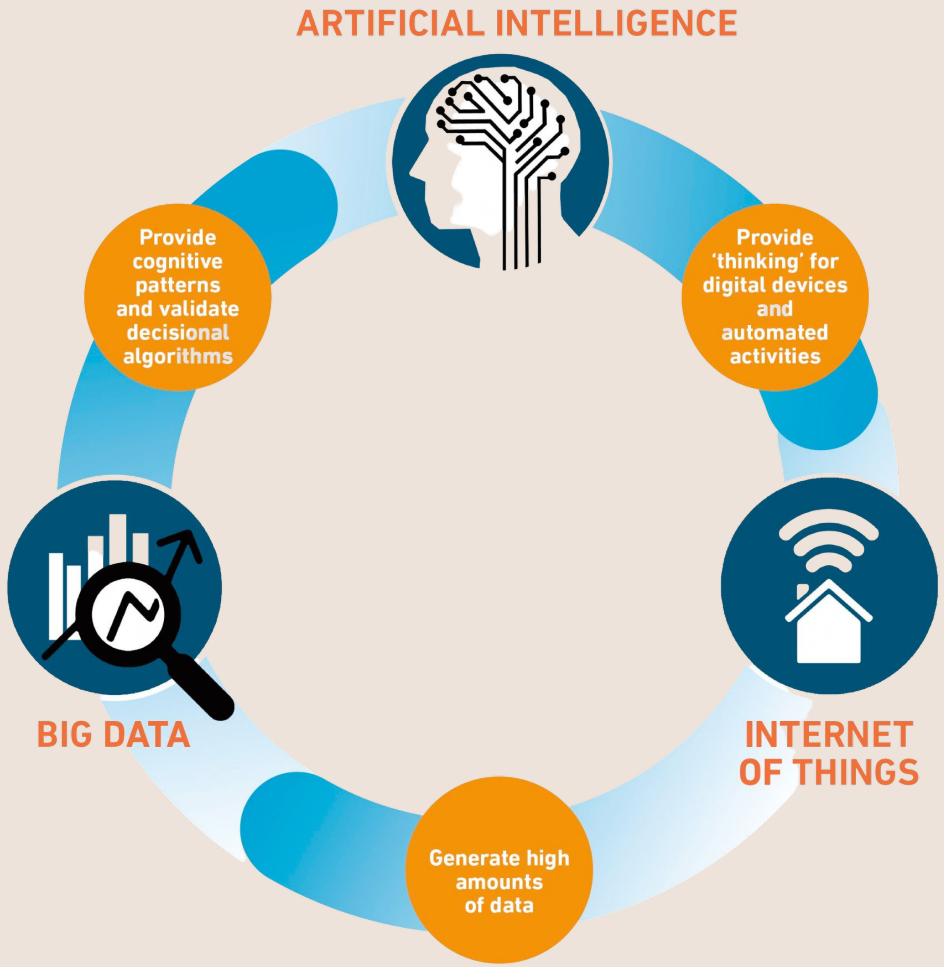


- AI is on the verge of reshaping economies and societies worldwide. But to what extent are AI solutions – tech and policy alike – sensitive to, and reflective of, local cultural and societal contexts?

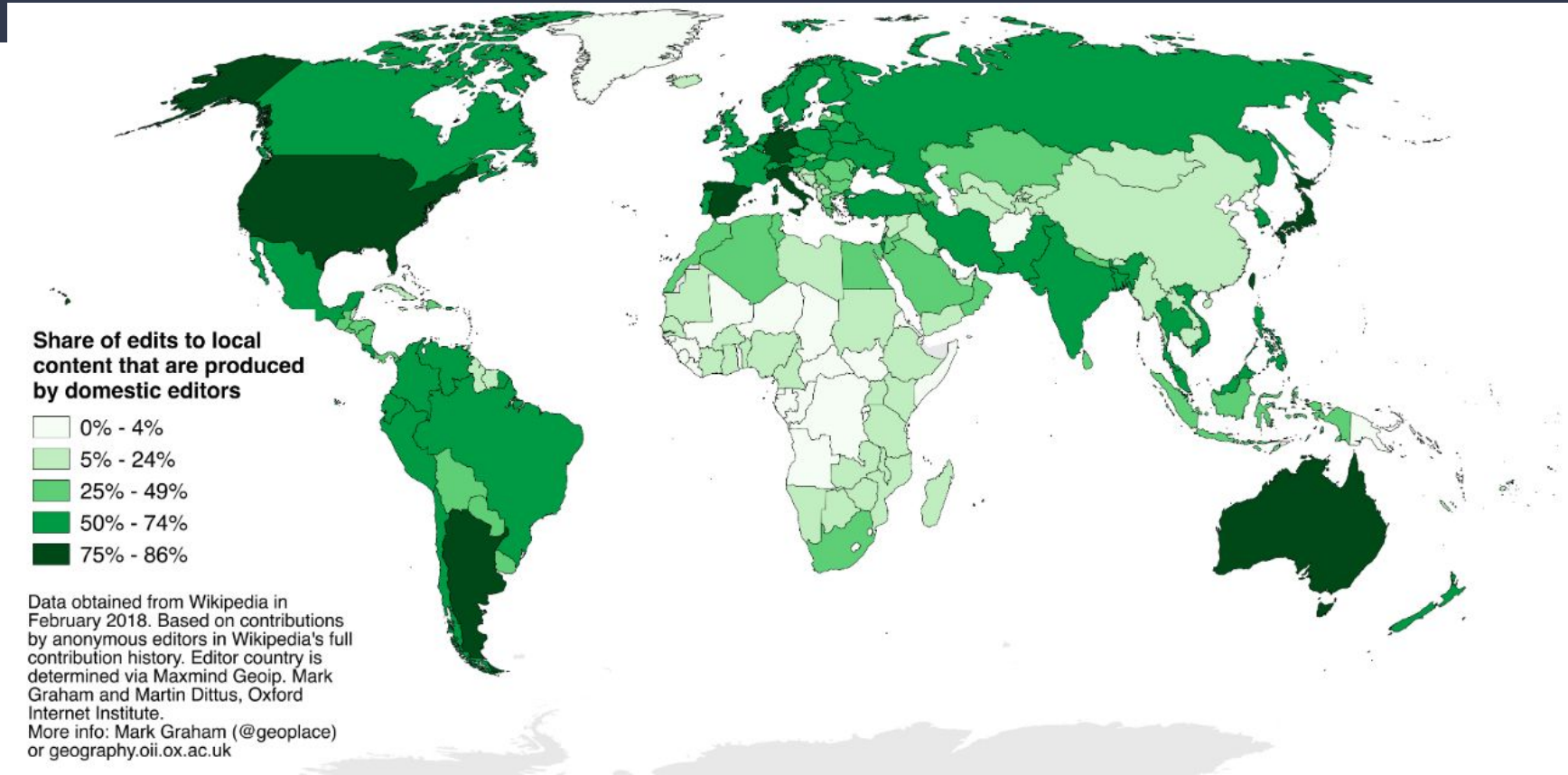
Data leads to knowledge and wisdom



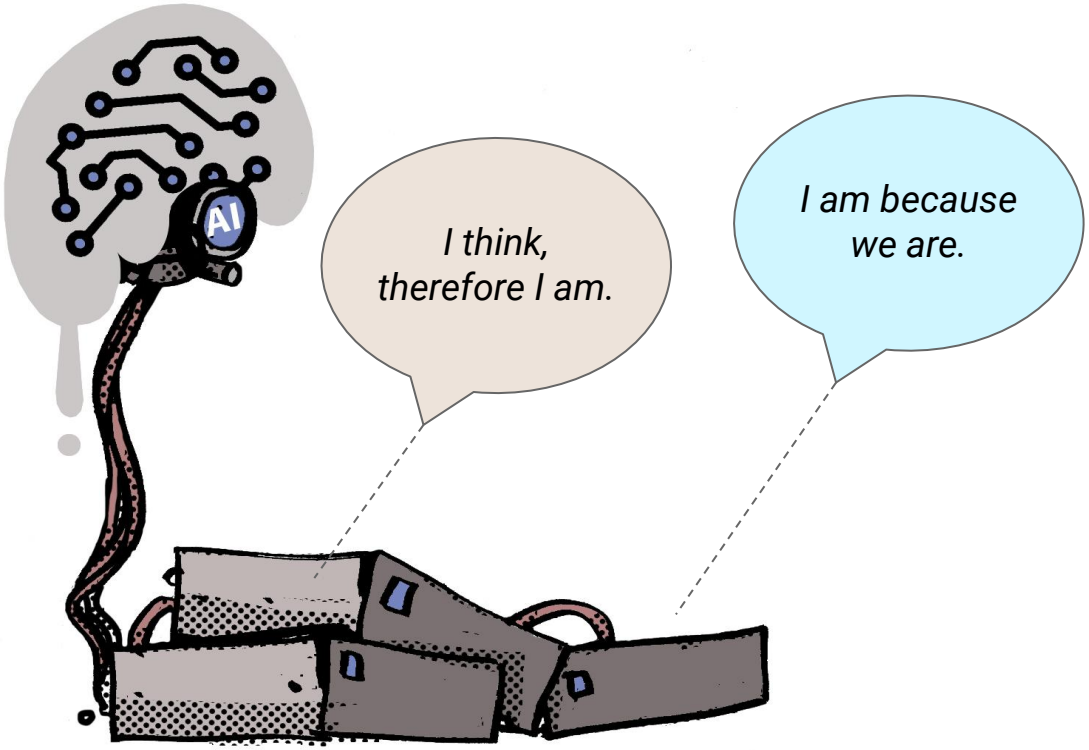
Data also powers tools, services, business models



But the tide doesn’t lift all boats equally

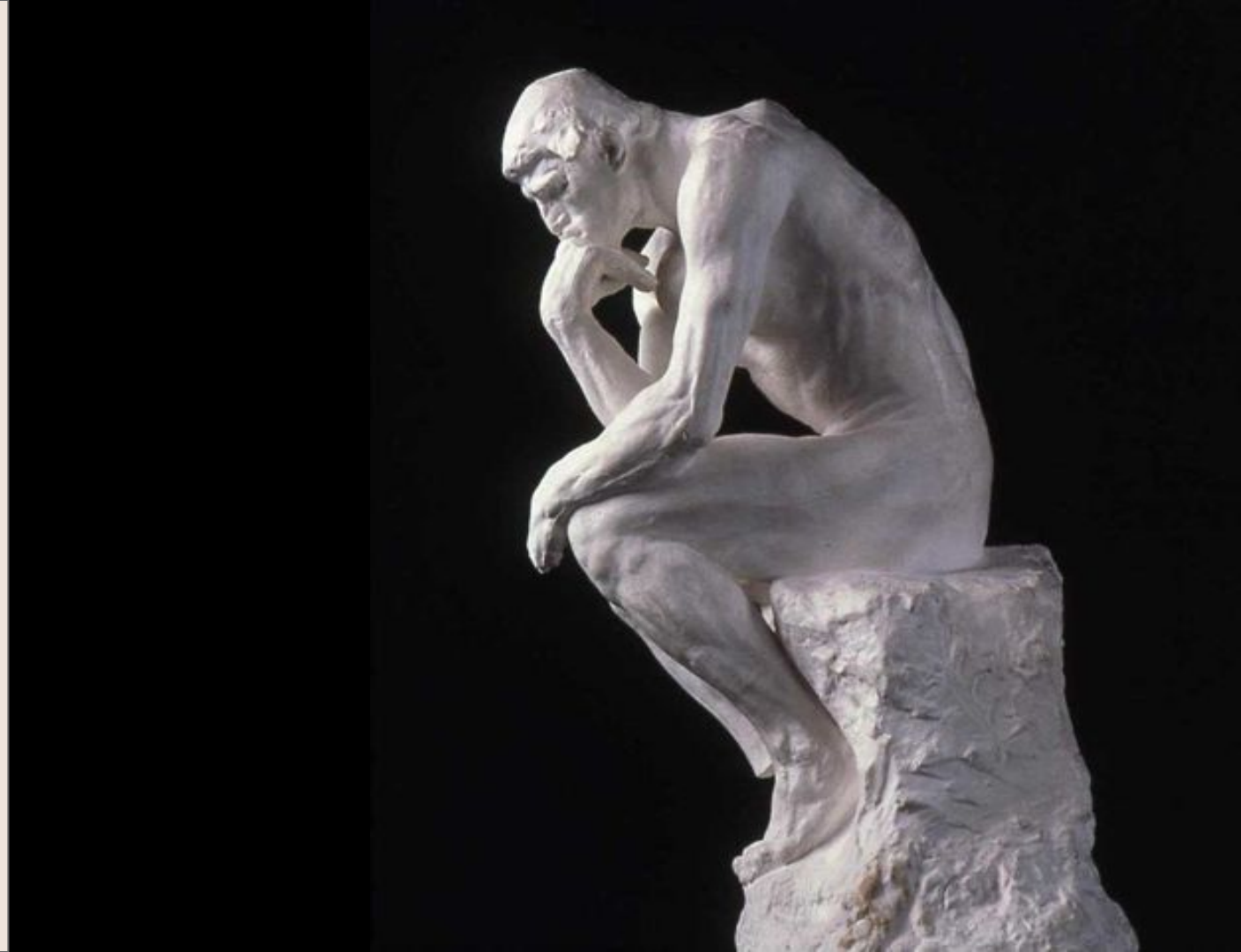


… not at all



What does this mean for the future of AI and humanity?

A time for reflection



And a new social contract?

‘I see three foundations for a renewed social contract fit for the twenty-first century:
(a) trust;
(b) inclusion, protection and participation; and
(c) measuring and valuing what matters to people and the planet.’

Fixing top-down AI

Make AI more representative of diverse communities, in a transparent way.

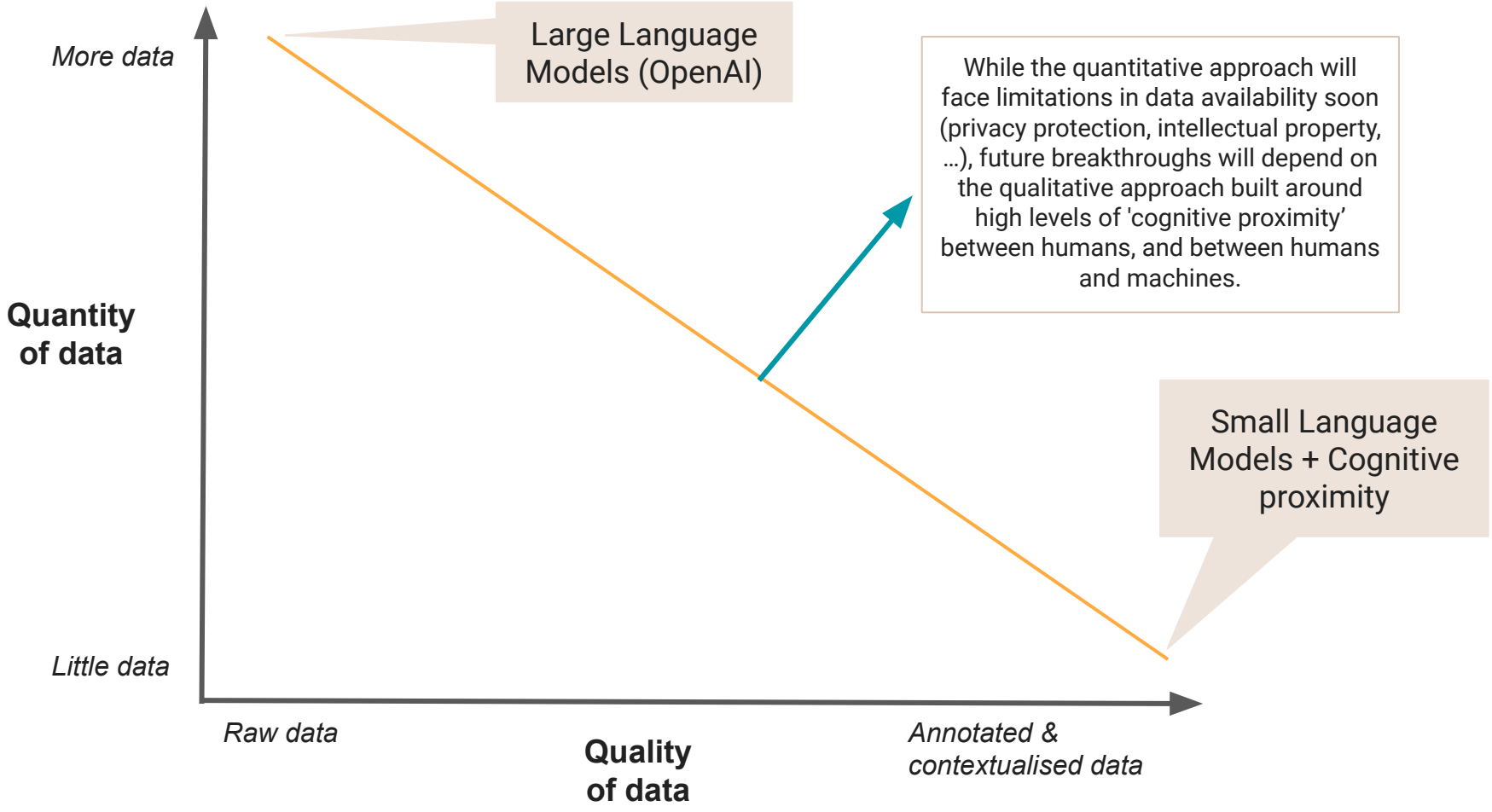
For instance, address intercultural bias and its biases by:

- ❏ Ensuring that data sets represent cultures and languages worldwide.
- ❏ Better codifying oral cultures and representing them in AI models.
- ❏ Giving higher weight in AI analysis to critical philosophical and religious texts that have a major impact on specific cultural contexts.

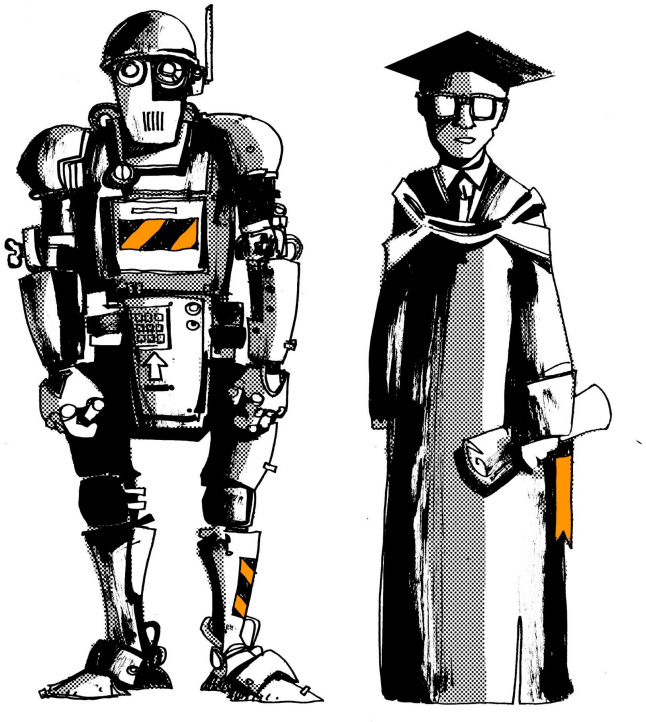
Encouraging bottom-up AI

Return AI to citizens and communities and address the risks of power centralisation.

- ❏ Reliance on small, high-quality data sets embedded in cultural, social, work contexts.
- ❏ Ownership – at the level of individuals and small groups – over the patterns derived from our data.


More data
Quantity of data
Little data
Large Language Models (OpenAI)
While the quantitative approach will face limitations in data availability soon (privacy protection, intellectual property, ...), future breakthroughs will depend on the qualitative approach built around high levels of 'cognitive proximity' between humans, and between humans and machines.
Small Language Models + Cognitive proximity
Raw data
Quality of data
Annotated & contextualised data

A right to be humanly imperfect?



Preserving and protecting the right to be humanly imperfect even when it is not economically optimal and efficient.

- ❏ Right to remain ‘natural’
- ❏ Right to disconnect
- ❏ Right to be anonymous
- ❏ Right to be employed over machines
- ❏ …

Diplo’s work on AI

AI policy and governance

- ❏ Digital Watch: Coverage of AI tech, policy and regulation
- ❏ Online course: AI technology and policy
- ❏ Debates: AI and diplomacy

AI tools

- ❏ Speech generator
- ❏ DiploGPT: AI drafting and summarising tool
- ❏ Bottom-up AI

Thank you!

diplomacy.edu
dig.watch
humainism.ai

sorinat@diplomacy.edu

DiPLO Geneva Internet Platform