CENTRE FOR ARTIFICIAL INTELLIGENCE RESEARCH

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Humanities 100. Since 1919

SACAIR
Southern African Conference for Artificial Intelligence Research

UNESCO

unesco

The UNESCO RECOMMENDATION on the ETHICS of AI

Joint ICTP-IAEA Advanced School/Workshop on Machine Learning in Citizen Science

2 March 2023

Emma Ruttkamp-Bloem

CONCENTRATION OF AI PATENTS

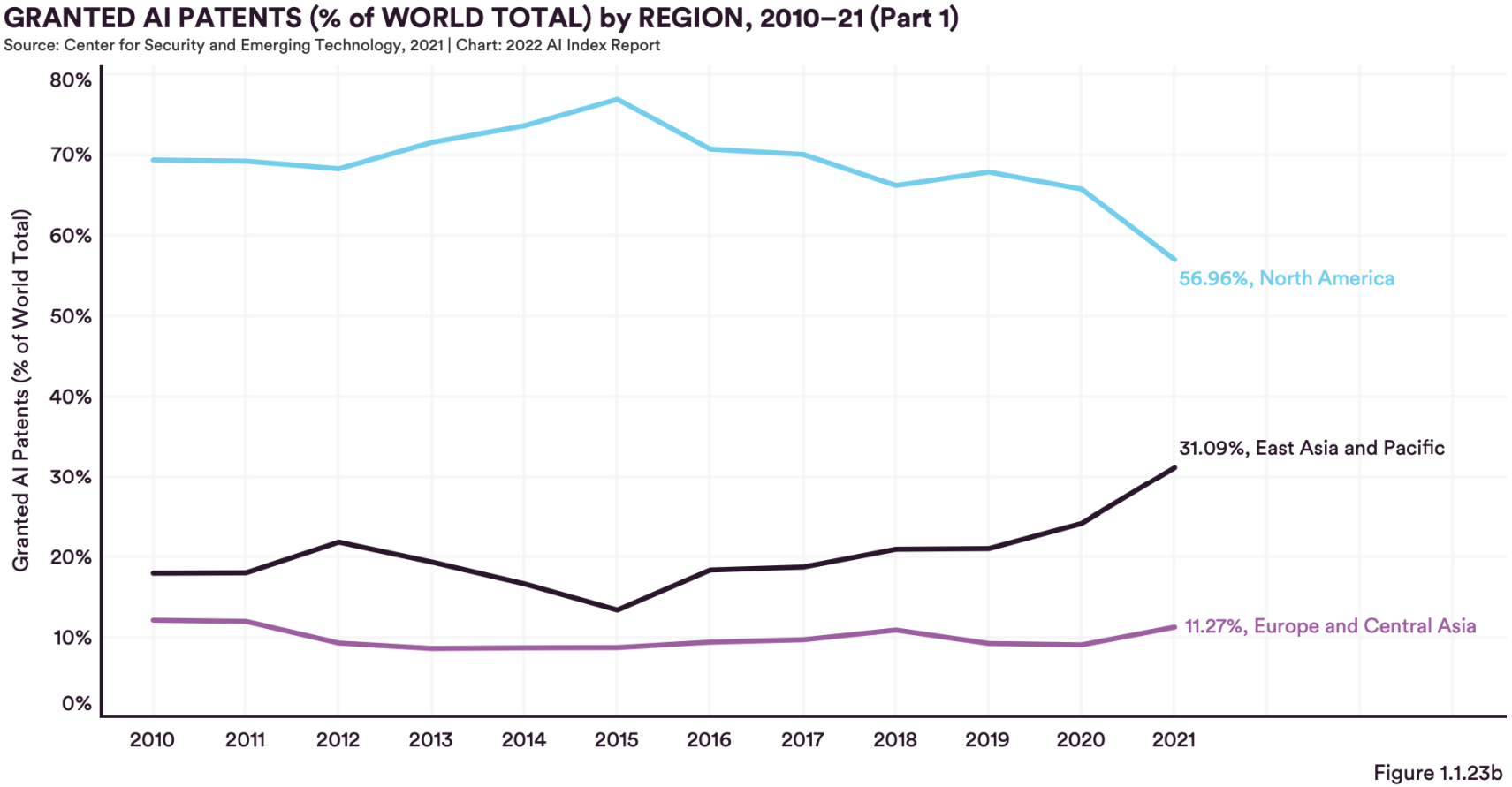


Figure 1.1.23b

AI research and development is **highly concentrated**. The large majority AI patents are filed in and granted to groups in **the U.S. and China**.

FAST FORWARD TOGETHER: AI HIRING

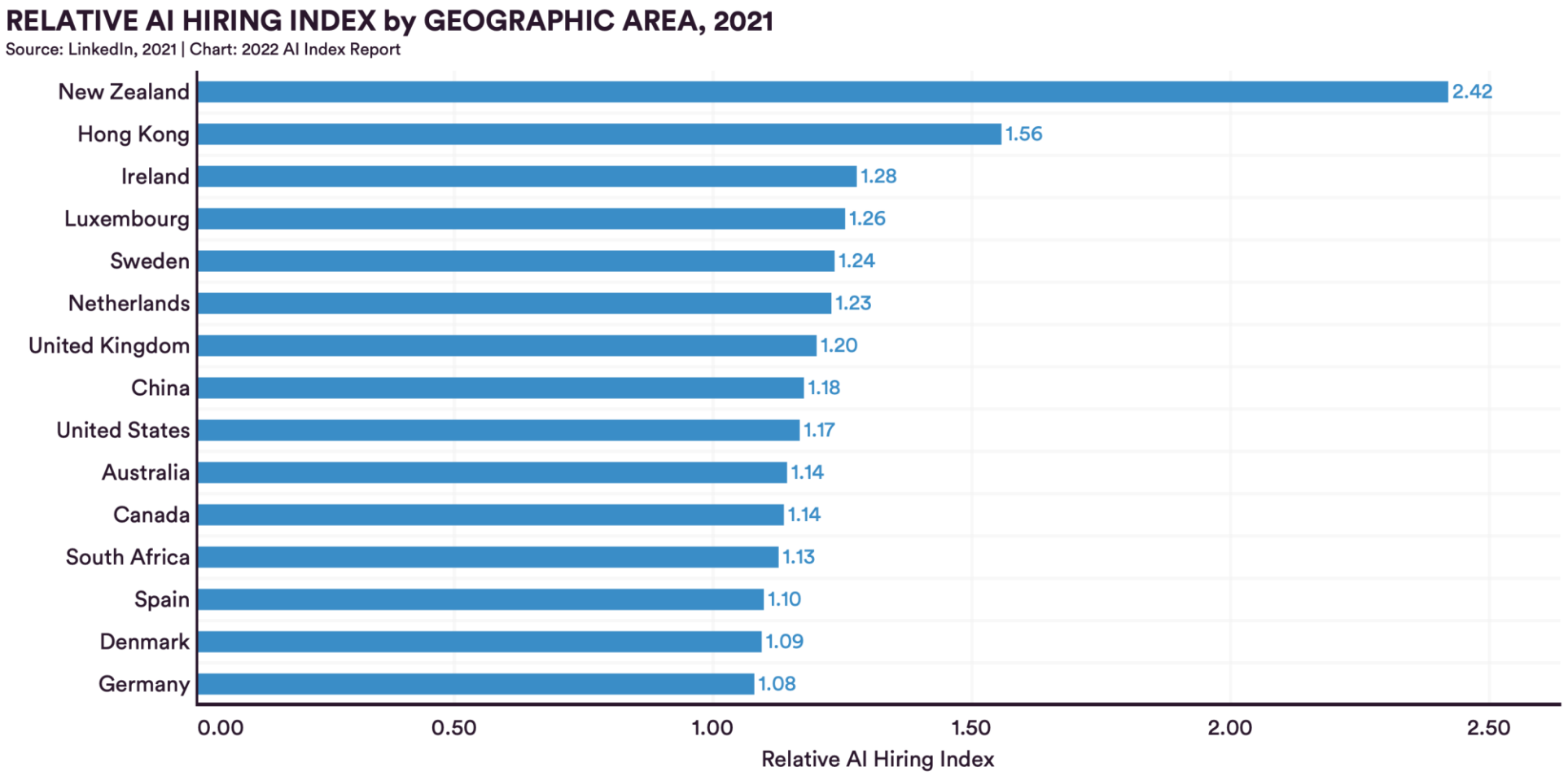
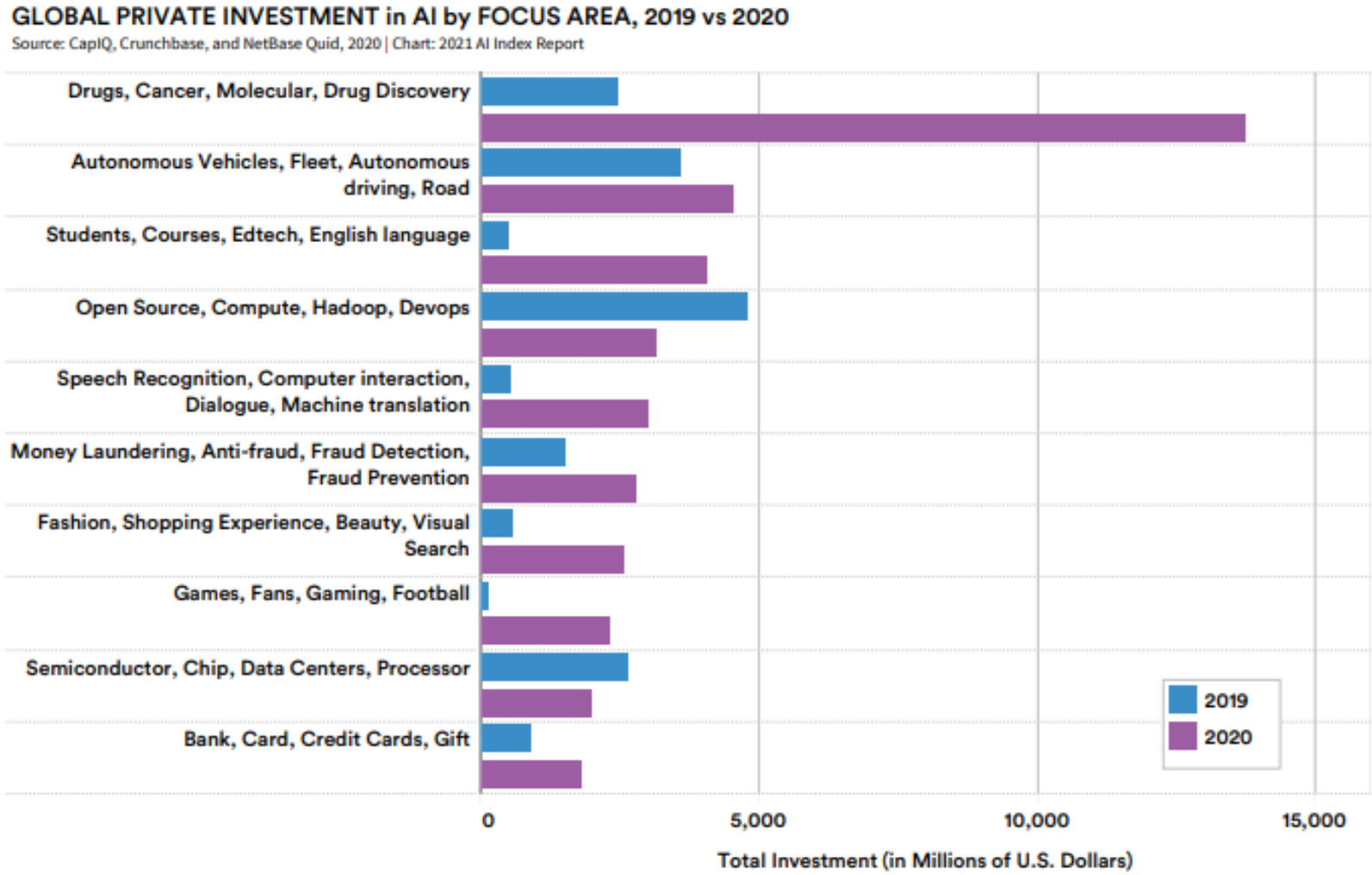


Figure 4.1.1

FAST FORWARD TOGETHER: PRIVATE INVESTMENT



Downsides of Self-Regulation

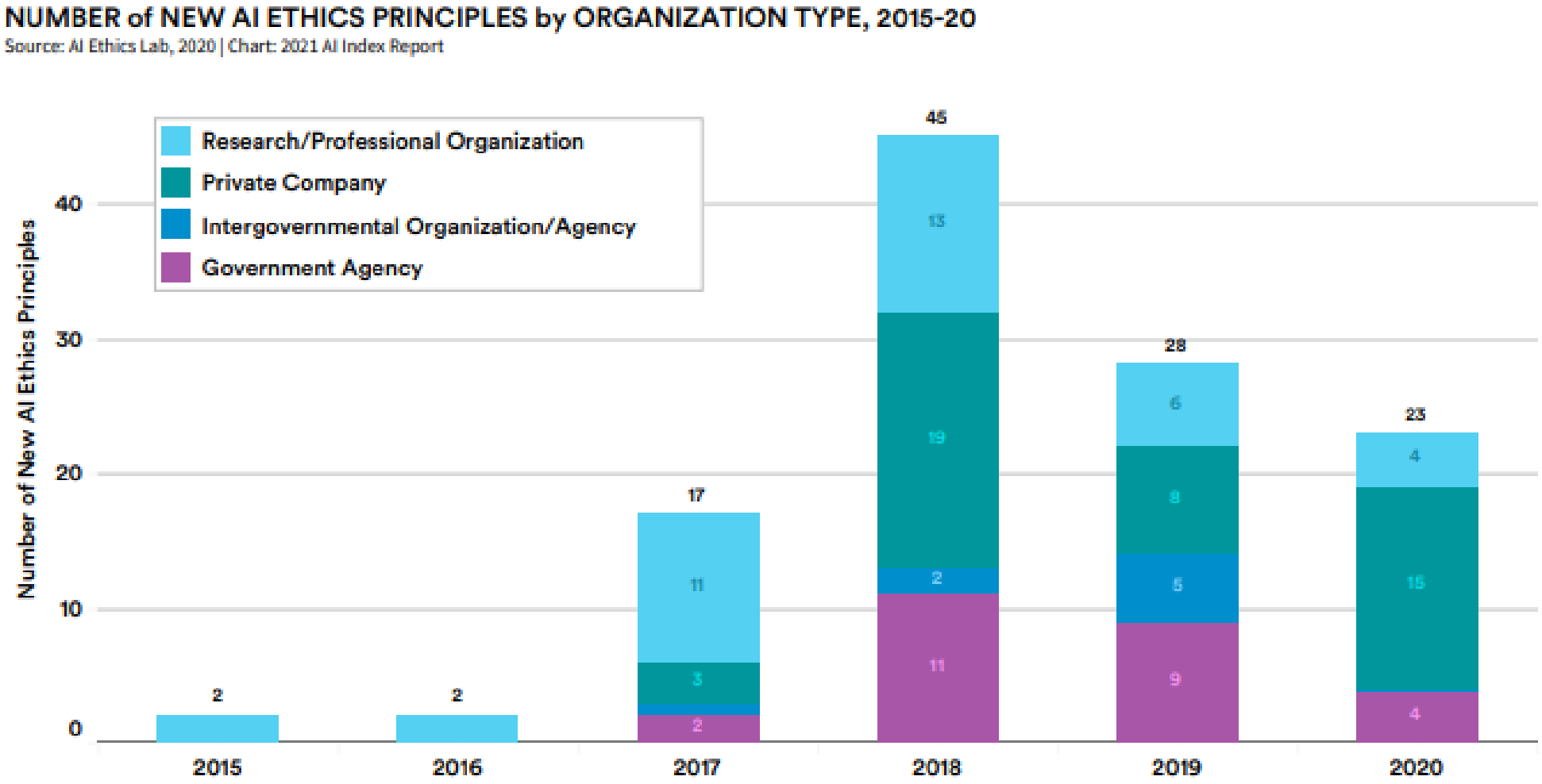
Human Rights Violations (ethics blue-washing)

Privacy

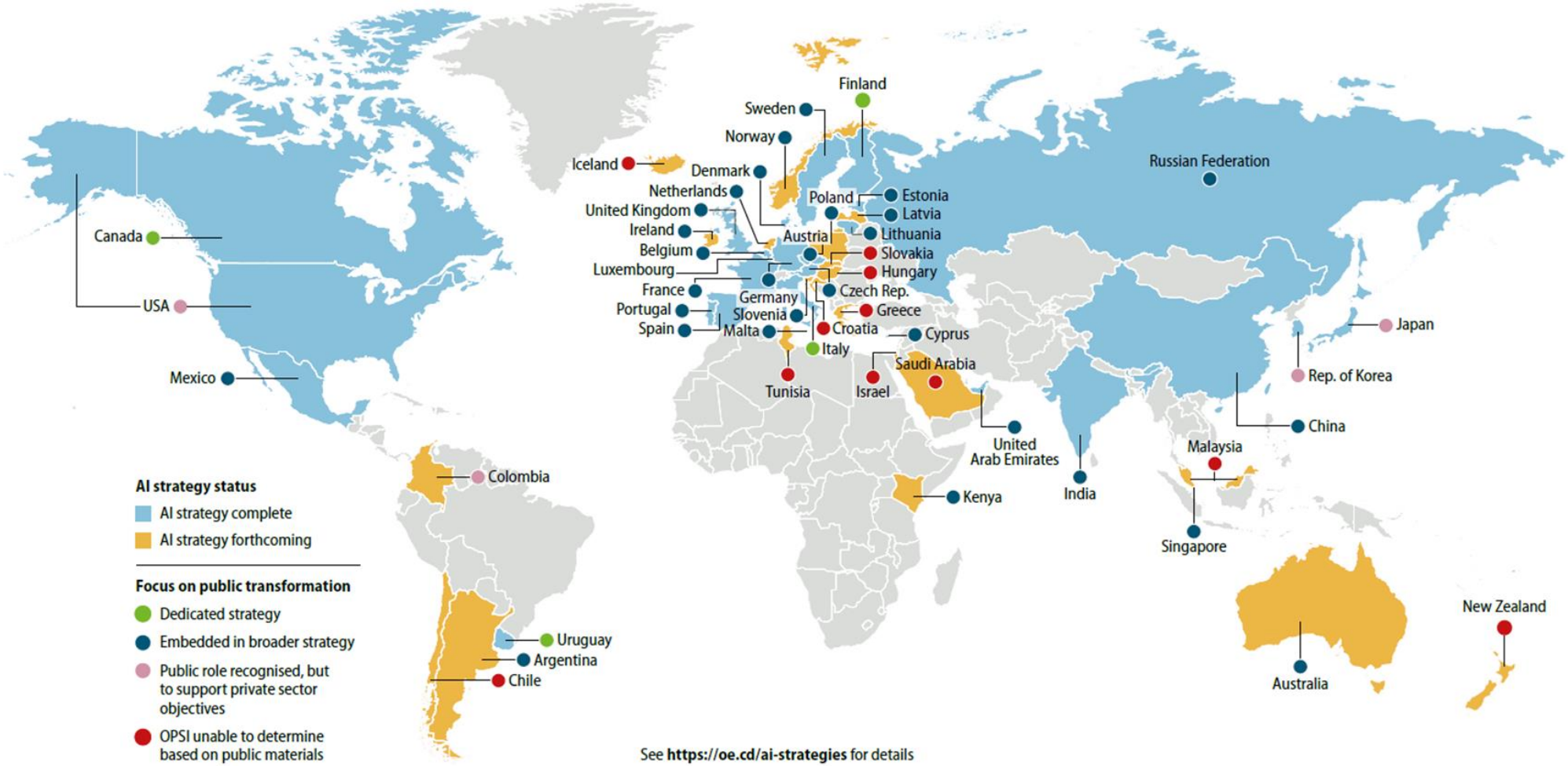
Diversity Inclusion Bias

Data Ownership

LACK OF A GLOBAL STANDARD AND BENCHMARKS



LACK OF A GLOBAL STANDARD AND BENCHMARKS



Global regulation remains necessary

AI technologies threaten harm to individuals in such deep layers of their lives that ultimately the harm will be to humanity as a whole

Contemporary 'risk society': Contemporray risks are qualitatively different from those of earlier historical periods

- Today's risks have global impact, with no geographic frontiers or limits, and are democratized, affecting all human beings, even future generations (Beck 1992)

Sustainable and ethical AI development requires international cooperation – ethics dumping, etc.

Widening of inequality will impact negatively on everyone and weaken adherence to regulation (e.g., the power of youth in Africa)

→ A recognition of the mutual human bondedness among all humans (Ubuntu, Mahayana Buddhism, Daosim) should underpin global AI ethics regulation precisely to ensure diversity

- Every human belongs to a greater whole, which thrives when all its constituent parts are enabled to thrive – thus every human should be *allowed* to thrive!

Global regulation is possible if there is:

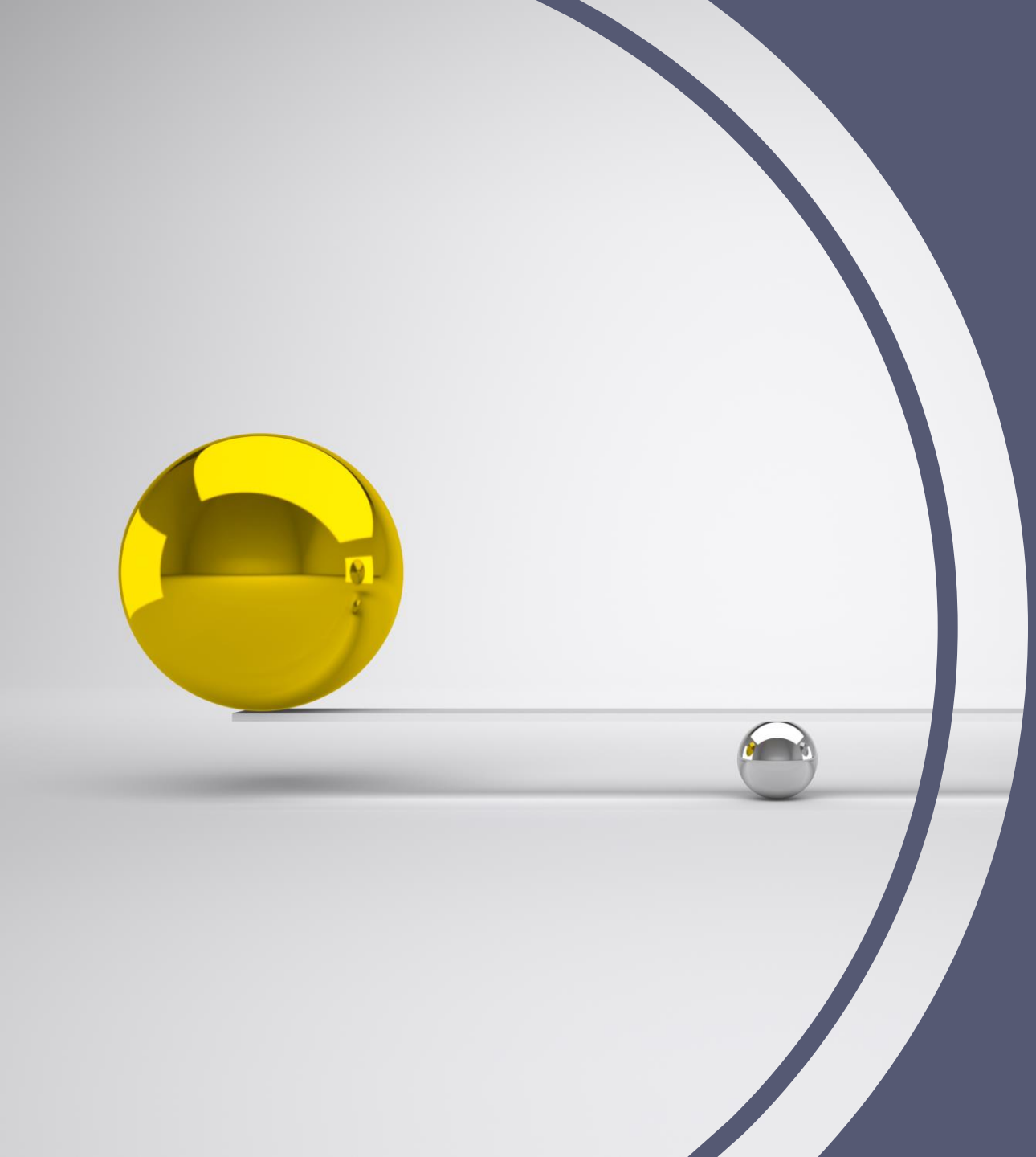
Epistemic justice – hear each other, recognise each other's interpretations of the world

> The meaning ascribed to values should be articulated & communicated in a context of openness to diverse engagement

Sensitivity to formulation – We need to reflect on formulation, and recognise culture as a calculus of values, an interpretation tool to measure inclusivity and predict adherence

> E.g., the right to privacy in an AI context remains core in all cultures, but how it is engaged with and expressed in a rights-based individualist culture may differ from how it is expressed in a duty-based collectivist culture

An active awareness of and support for fighting data colonialism (including data activism)



Values & Principles

Values

Respect, protection and promotion of human rights and fundamental freedoms and human dignity

Environment and ecosystem flourishing

Ensuring diversity and inclusiveness

Living in peaceful, just and interconnected societies

Principles

Proportionality & do no harm

Safety & security

Fairness & non-discrimination

Sustainability

Right to privacy & data protection

Human oversight & determination

Transparency & explainability

Responsibility & accountability

Multi-stakeholder & adaptive governance & collaboration

Awareness & literacy

Actionability

Concrete enough to make an impact

Firm and open enough to ensure adherence

Supple enough to have validity in the future

→ Readiness status & policy actions

"UNESCO recognizes that Member States will be at different stages of readiness to implement this Recommendation, in terms of scientific, technological, economic, educational, legal, regulatory, infrastructural, societal, cultural and other dimensions.

- It is noted that 'readiness' here is a dynamic status" (Policy Areas, UNESCO 2021)

UNESCO is developing a Readiness and an Ethical Impact Assessment tool

Policy Areas

Ethical impact assessment

Ethical governance and stewardship

Data policy

Development and international collaboration

Environment and ecosystems

Gender

Culture

Education and research

Communication and information

Economy and labour

Health and social well-being

Respect, protection and promotion of human rights and fundamental freedoms and human dignity (value)

- The inviolable and inherent dignity of every human constitutes the foundation for the universal, indivisible, inalienable, interdependent and interrelated system of human rights and fundamental freedoms …
- Human dignity relates to the recognition of the intrinsic and equal worth of each individual human being …
- Human rights and fundamental freedoms must be respected, protected and promoted throughout the life cycle of AI systems. Governments, private sector, civil society, international organizations, technical communities and academia must respect human rights instruments and frameworks in their interventions in the processes surrounding the life cycle of AI systems. New technologies need to provide new means to advocate, defend and exercise human rights and not to infringe them.

Ensuring diversity and inclusiveness (value)

- Respect, protection and promotion of diversity and inclusiveness should be ensured throughout the life cycle of AI systems, consistent with international law, including human rights law. This may be done by promoting active participation of all individuals or groups …

Living in peaceful, just and interconnected societies (value)

- The value of living in peaceful and just societies points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment …
- The notion of humans being interconnected is based on the knowledge that every human belongs to a greater whole, which thrives when all its constituent parts are enabled to thrive …

Basically, all other principles and all policy areas ☺

Preventing unwanted harms, security and safety risks, as well as unintended or problematic societal impacts

Proportionality and do no harm (principle)

- … none of the processes related to the AI system life cycle shall exceed what is necessary to achieve legitimate aims or objectives and should be appropriate to the context …

Safety and security (principle)

- Unwanted harms (safety risks), as well as vulnerabilities to attack (security risks) should be avoided and should be addressed, prevented and eliminated throughout the life cycle of AI systems to ensure human, environmental and ecosystem safety and security … [this is also about accuracy and robustness]

Right to privacy and data protection (principle)

- Privacy, a right essential to the protection of human dignity, human autonomy and human agency, must be respected, protected and promoted throughout the life cycle of AI systems. …
- Adequate data protection frameworks and governance mechanisms should be established in a multi-stakeholder approach at the national or international level, protected by judicial systems, and ensured throughout the life cycle of AI systems. …

Data policy & Culture & Communication and information & Health and social well-being (policy areas)

Awareness and literacy (principle)

- Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training led jointly by governments, intergovernmental organizations, civil society, academia, the media, community leaders and the private sector, and considering the existing linguistic, social and cultural diversity, to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence. … [civil society is the most powerful weapon against possible harmful AI systems]

Transparency and explainability (principle)

- While efforts need to be made to increase transparency and explainability of AI systems, including those with extra-territorial impact, throughout their life cycle to support democratic governance, the level of transparency and explainability should always be appropriate to the context and impact, as there may be a need to balance between transparency and explainability and other principles such as privacy, safety and security. …
- From a socio-technical lens, greater transparency contributes to more peaceful, just, democratic and inclusive societies. …

Ensuring diversity and inclusiveness, including gender equality

Fairness and non-discrimination (principle)

- AI actors should promote social justice and safeguard fairness and non-discrimination of any kind in compliance with international law. This implies an inclusive approach to ensuring that the benefits of AI technologies are available and accessible to all …
- Member States should work to promote inclusive access for all, including local communities, to AI systems with locally relevant content and services, and with respect for multilingualism and cultural diversity …
- AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems. Effective remedy should be available against discrimination and biased algorithmic determination. …

Gender (policy area)

- … Member States should ensure that the potential of AI systems to advance the achievement of gender equality is realized. They should ensure that these technologies do not exacerbate the already wide gender gaps existing in several fields in the analogue world, andinstead eliminate those gaps. …

Respect, protection and promotion of human rights and fundamental freedoms and human dignity & Living in peaceful, just and interconnected societies (vaues)

Economy and labour (policy area)

Achieving accountability and human oversight

Human oversight and determination (principle)

- Member States should ensure that it is always possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems, as well as in cases of remedy related to AI systems, to physical persons or to existing legal entities. …
- … the decision to cede control in limited contexts remains that of humans … As a rule, life and death decisions should not be ceded to AI systems.

Responsibility and accountability (principle)

- The ethical responsibility and liability for the decisions and actions based in any way on an AI system should always ultimately be attributable to AI actors corresponding to their role in the life cycle of the AI system. …

Ethical impact assessment & Ethical governance and stewardship (policy areas)

Contributing to sustainability, sustainable development and environmental protection

Sustainability (principle)

- The continuous assessment of the human, social, cultural, economic and environmental impact of AI technologies should … be carried out with full cognizance of the implications of AI technologies for sustainability as a set of constantly evolving goals across a range of dimensions, such as currently identified in the Sustainable Development Goals (SDGs) of the United Nations. …

Environment and ecosystems (policy area)

- … assess direct and indirect environmental impact throughout the AI system life cycle, including, but not limited to, its carbon footprint, energy consumption and the environmental impact of raw material extraction"
- … introduce incentives, when needed and appropriate, to ensure the development and adoption of rights-based and ethical AI-powered solutions for disaster risk resilience; the monitoring, protection and regeneration of the environment and ecosystems; and the preservation of the planet …

Environment and ecosystem flourishing (value)

- Environmental and ecosystem flourishing should be recognized, protected and promoted through the life cycle of AI systems …

Education and research (policy area)

Realising inclusive, multi-stakeholder approaches to AI development and governance

Multi-stakeholder and adaptive governance and collaboration (principle)

- International law and national sovereignty must be respected in the use of data. …
- Participation of different stakeholders throughout the AI system life cycle is necessary for inclusive approaches to AI governance, enabling the benefits to be shared by all, and to contribute to sustainable development. …
- Measures should be adopted to take into account shifts in technologies, the emergence of new groups of stakeholders, and to allow for meaningful participation by marginalized groups, communities and individuals …

Awareness and literacy (principle)

- Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training led jointly by governments, intergovernmental organizations, civil society, academia, the media, community leaders and the private sector, and considering the existing linguistic, social and cultural diversity, to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence. …

Development and international collaboration (policy area)

Implementation Strategy

1. Elaboration of the capacity-building and monitoring tools:
    - Readiness Assessment Methodology
    - Ethical Impact Assessment
2. AI Experts Without Borders: a roster of experts to be managed by UNESCO for deployment in beneficiary countries for targeted capacity-building interventions
3. Women4EthicalAI: to spearhead the implementation and deployment of the Recommendation from gender perspective
4. Observatory of AI Ethics: to gather state of the art analysis around the world on ethical AI
5. Global Forum on Ethics of AI: as an annual high-level flagship event

Conclusion

There is a long road ahead of us

We should

1. Share best practice, help each other – and UNESCO is willing to do that

2. Apply the value of interconnectedness as the golden threat to weave together a global AI ethics

3, Embrace difference and apply culture as the needle not the scissors – to calculate and differentiate and distill values in manners that include every person in ways that are

4. Openly act against data colonialism and any form of exclusion from AI and AI ethics discourses

5. This should all happen in non-negotionably open, epistemic just, and transparent discourses

Thank you