“Artificial intelligence” and social sustainability – does ethics of artificial intelligence as a global challenge also need global answers?

EBAPE Graduate Seminar

Johann Čas

25. January 2023

Co-funded by the European Union

This project has received funding from the European Union's Horizon 2020 research and innovation programme H2020-MSCA-RISE-2019 under grant agreement No [873119]
("Promoting Research on Digitalisation in Emerging Powers and Europe towards Sustainable Development")

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

Overview

- The Institute of Technology Assessment (ITA)
- PRODIGEES - Promoting Research on Digitalisation in Emerging Powers and Europe Towards Sustainable Development
- Introduction to my project on Digitalisation and Sustainability - Poverty and Inequality
- AI and Ethics
- Impacts on SDG 1 and 10
- Conclusions and suggestions for solutions

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

The Institute of Technology Assessment (ITA)

- Founded in 1988
- Approx. 25 staff
- Interdisciplinary team: both humanities, social sciences and natural sciences and engineering
- One of 25 research institutes of the Austrian Academy of Sciences (ÖAW), Austria's largest non-university research and science institution with its statutory mission to "promote science in every way".
- Financed by Academy and by third parties on a project basis (European Union, research funds, government, parliament, …)
- Member of NTA, EPTA, globalTA

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

ITA's double mission: academic and policy-oriented

"The ITA studies the impact of new technologies on the environment, economy and society. The results of its scientific work support policy-makers, administration and the public with regard to issues of technology policy.

The ITA carries out interdisciplinary technology studies with three main aims:

- to understand the complex interplay between technology and society from multiple perspectives,
- to concomitantly analyze technology development, and
- to contribute to socially responsible technology policy by advising policy-makers and society."

Clients of ITA

- Executive/Administration
  - Austrian Council for Research and Technology Development
  - Federal Ministries of Innovation and Technology, Research, Environment, Health, …
  - EU Commission (in particular via FP)
- Legislative
  - Austrian Parliament
  - European Parliament (STOA Panel)
- Social partners
  - Federal Chamber of Labour
  - Economic and Social Council
- Research funds (national/international)
- OECD and other international organisations
- Internal funds for basic research

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

ITA's current four research areas

1. Digital technology, democracy and society
2. Emerging technosciences, values and uncertainty
3. Innovation, the environment and sustainable futures
4. Reflexive studies: methods, concepts and frameworks of TA

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

PRODIGEES

Promoting Research on Digitalisation in Emerging Powers and Europe Towards Sustainable Development

January 1, 2020 - ~~December 31, 2023~~ June 30, 2025

PRO DIG EES

This project has received funding from the European Union's Horizon 2020 research and innovation programme H2020-MSCA-RISE-2019 under grant agreement No [873119]
("Promoting Research on Digitalisation in Emerging Powers and Europe towards Sustainable Development")

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

- PRODIGEES is an exploration of trends and possibilities for digitalisation in development processes.

- PRODIGEES is a transnational knowledge cooperation and exchange between Global North and Global South partners.

- PRODIGEES is a project which addresses the United Nations' 2030 Agenda for Sustainable Development, as well as the economic framework announced in the European Green Deal.

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ITA INSTITUTE OF TECHNOLOGY ASSESSMENT

IDOS German Institute of Development and Sustainability

Istituto Affari Internazionali (IAI), Italy

Universität Hamburg (UHAM), Germany

Centre for Strategic and International Studies (CSIS), Indonesia

Stellenbosch University (SU), South Africa

Instituto Mora (MORA), Mexico

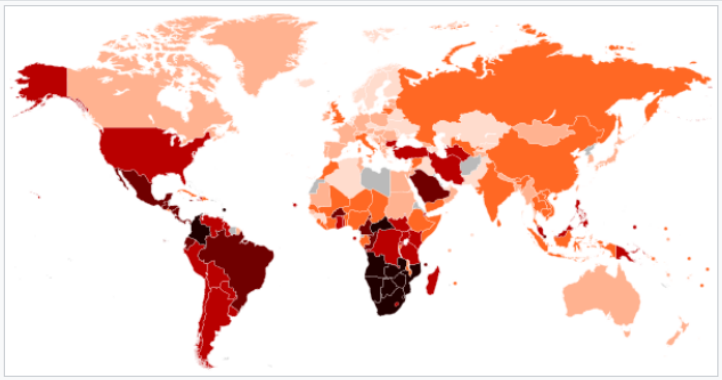
Fundação Getúlio Vargas (FGV), Brazil

Research and Information System for Developing Countries (RIS), India

Luiss University, Italy

Motivation

- Rising poverty and inequality are among the most critical developments threatening social sustainability and political stability
- Gini indices:
  - Brazil: 48,9 (2020) (pI ca. 56?)
  - European Union: 30,1 (2021)
  - Austria: 26,7 (2021) (pI 49)
- Inequality is growing at an increasing pace



https://en.wikipedia.org/wiki/Gini_coefficient

> Income disparities risk increasing polarization and resentment within societies

Focus on "Artificial Intelligence" and SDGs 1 and 10 – No poverty and Reduced Inequality

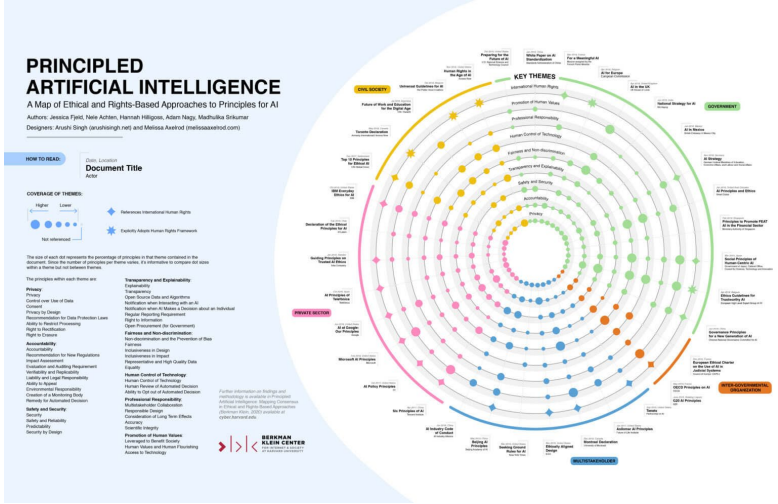
- Apparent mismatch between (attributed) potential and attention given
- Artificial intelligence (AI) is an area of strategic importance and a key driver of social and economic development (https://ai-watch.ec.europa.eu/topics_en)
- But hardly directly addressed in specific activities, e.g. the AI for Good Initiative of the International Telecommunications Union (https://aiforgood.itu.int/)
- In general, SDGs 3, 6, 13, 7, 8, and 4 having been assessed the most often among the 17 SDGs

Research questions

- To what extent are aspects of SDGs 1 and 10 addressed in debates on ethics and AI at European, global and national levels in Brazil?
- To what extent are they included in current regulatory projects?
- What direct or indirect effects can be expected from these efforts?
- Can differences in the importance of these goals be observed depending on the level (European Union, global, national)?
- Which dimension of inequality are addressed?
- How realistic are ambitions to contribute to social sustainability through AI?
- What regulatory changes and measures are considered appropriate or necessary to achieve such goals?

“Artificial Intelligence” and Ethics

- Huge number of initiatives
- They can be summarised into a limited number of basic principles (transparency; justice and fairness; non-maleficence; responsibility; privacy; …)



- Many of the guidelines focus on aspects that can be implemented technically
- Many of them overlook or disregard social aspects

AI Ethics – European Union I
Ethics Guidelines for Trustworthy AI (2019)

- Produced by the High-Level Expert Group on Artificial Intelligence
  - Predominantly members from academia and industry
- Seven key requirements for Trustworthy AI:
  - Human agency and oversight; Technical robustness and safety; Privacy and Data governance; Transparency; Diversity, non-discrimination and fairness; Societal and environmental well-being; Accountability
- Four ethics principles:
  - Respect for human autonomy; Prevention of harm; Fairness; Explicability
- 15 pages Assessment List, but only a few, unspecified lines on Societal Wellbeing

EU II: White Paper on Artificial Intelligence
A European approach to excellence and trust (2020)

- Promoting the development and deployment of AI, based on European values,
- but also preparing the turn to a risk based regulation
- Stating that the use of AI systems can have a significant role in achieving the Sustainable Development Goals,
- but not addressing at all in which way this should be realised
- Final version quite different from leaked draft versions.

EU III: Proposal for an Artificial Intelligence Act (2021)

- Draft Act considered to be passed in 2023 and set into force by 2025
- Risk-based instead of value based
- Differentiating between uses of AI that create an unacceptable risk, a high risk, and low or minimal risk
- Prohibited artificial intelligence practices, e.g.
  - practices that have a significant potential to manipulate persons through subliminal techniques
  - AI-based social scoring for general purposes done by public authorities
  - use of 'real time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement is also prohibited unless certain limited exceptions apply

EU IV: Proposal for an Artificial Intelligence Act (2021)

- High risk AI systems
  - Mainly AI systems intended to be used as safety component of products and explicitly listed stand-alone AI systems with mainly fundamental rights implications
  - High risk AI systems must fulfill a number of specific requirements
- Limited risk
  - Transparency obligations
- Low or minimal risk AI systems
  - No obligations
- No relation with sustainability anymore (apart from mentioning environmental sustainability)

Global Initiatives I: UNESCO Recommendation on the ethics of artificial intelligence (2021)

- In line with United Nations Sustainable Development Goals
- Detailed recommendations for 11 Policy Areas
- Policy Area 1: Ethical Impact Assessment
  - Member States should also be able to assess the socioeconomic impact of AI systems on poverty and ensure that the gap between people living in wealth and poverty, as well as the digital divide among and within countries, are not increased with the massive adoption of AI technologies at present and in the future.
- Policy Area 10: Economy and Labour
  - Focus on labour markets
- Policy Area 11: Health and Social Well-being
  - Focus on health

Global Initiatives II: OECD – Recommendation of the Council on Artificial Intelligence (2019)

- Provides the first intergovernmental standard for AI policies
- Reference to related activities (UNESCO, EU, …)
- Five principles
- i) Inclusive growth, sustainable development and well-being
  - Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as …, reducing economic, social, gender and other inequalities, …, thus invigorating inclusive growth, sustainable development and well-being.
- ii) human-centred values and fairness; iii) transparency and explainability; iv) robustness, security and safety; and v) accountability
- Five policy recommendations

Global Initiatives III: IEEE Code of Ethics (2020)

IEEE Code of Ethics

We, the members of the IEEE, in recognition of the importance of our technologies in affecting

I. To uphold the highest standards of integrity, responsible behavior, and ethical conduct in professional activities.

1. to hold paramount, the safety, health, and welfare of the public, to strive to comply with ethical design and sustainable development practices, to protect the privacy of others, and to disclose promptly factors that might endanger the public or the environment;

2. to improve the understanding by individuals and society of the capabilities and societal implications of conventional and emerging technologies, including intelligent systems;

3. to avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do

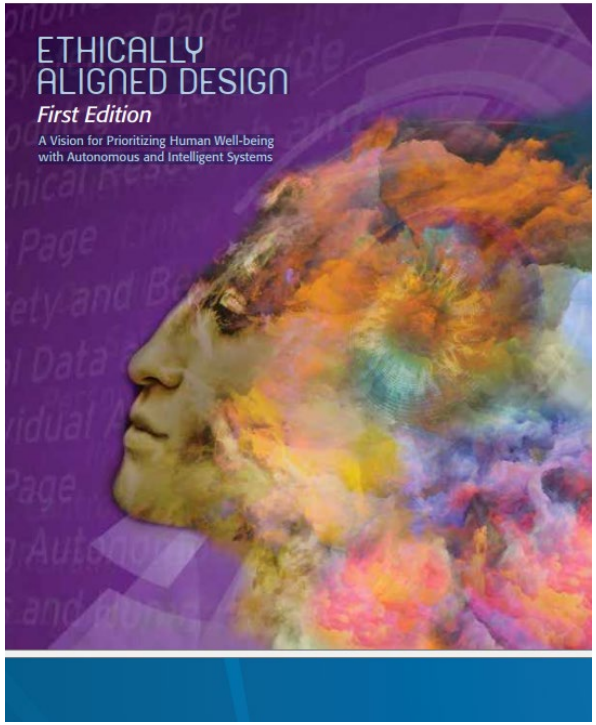
II.

7.

8.

9.

Global Initiatives III cont.: IEEE Ethically Aligned Design Report (2019)



https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf

Extracts of Brazilian Artificial Intelligence Bill, n. 21/2020 (based on Non-official Translation)

- **Art. 3.** The application of artificial intelligence in Brazil is aimed at scientific and technological development, as well as:
- I – incentivizing sustainable and inclusive economic development and societal welfare;
- **Art. 5.** The following are the principles for the development and application of artificial intelligence in Brazil:
- I – beneficial purpose: artificial intelligence systems shall seek beneficial results for humanity;
- **Art. 6.** When regulating the application of artificial intelligence, the government must observe the following guidelines:
- III – risk-based management: … risks, always evaluated in reference to: a) the potential social and economic benefits offered by that artificial intelligence system;

Impacts I – What can be expected?

- Different prospects for the areas of **poverty** and inequality
- Without intervention, both positive and negative effects are conceivable for the reduction of the poverty problem
- Mitigating poverty
  - Higher productivity, higher wages, economic growth, ...
  - Relative increase in income due to cheaper services and products, …
  - Indirect effects via other SDGs - more efficient agriculture, better health care, more educational opportunities, …
- Increasing poverty
  - Automation and job losses, higher unemployment and pressure on wage levels, rising inequalities

Impacts II – Rising inequalities

- Without intervention, negative effects are to be expected for SDG 10 (Reduce inequality within and among countries)
- Global race to become the most competitive AI region between China, Europe and the USA
- Repeated on EU and national levels (inequality within countries is higher than inequality between countries)
- However with very different starting conditions and prerequisites
  - Economic power
  - Required technical infrastructure
  - Educational system and skilled workforce
  - Political system
  - …

Conclusions and Solutions I – Some ideas and suggestions

- AI technology is not a panacea or an end in itself, but a powerful tool
- To serve human interests it must be designed, implemented and used accordingly
- Otherwise, AI systems reflect and reinforce existing and past social relations and imbalances rather than redressing them
- Aiming at making good use instead of becoming the best in global competition
    - Example financial versus labour markets
- Establish instruments to assess the social/societal impacts
- Establish instruments between engineering and social sciences

Conclusions and Solutions II – Some ideas and suggestions

Issue: Current roadmaps for development and deployment of A/IS are not aligned with or guided by their impact in the most important challenges of humanity, defined in the seventeen United Nations Sustainable Development Goals (SDGs), which collectively aspire to create a more equal world of prosperity, peace, planet protection, and human dignity for all people.[4]

IEEE (2019), Ethically Aligned Design, p. 142

- Particularly true for the Draft AI Act of the European Union
- The Draft AI Act of the European Union will not repeat the success story of the GDPR
- The Draft AI Bill appears to better reflect the UN SDGs
- Brazil is very rich in nature, people and resources – an appropriate AI policy can help to make better use of them

Thank you for your attention!

Contact:
Johann Čas
A-1010 Vienna, Bäckerstraße 13
Austria
T: +43 (1) 51581 6581
F: +43 (1) 51581 6570
jcas@oeaw.ac.at
oeaw.ac.at/en/ita

Current affiliation and contact:
Fundação Getulio Vargas (FGV), EBAPE
Edifício Roberto Campos –
Rua Jornalista Orlando Dantas, 30
Botafogo, Rio de Janeiro - RJ, 22231-010
Telefone: +55 21 972 831 042
johann.cas@fgv.br