AI Governance and Ethics:

Industry Standards as vehicle to address socio-technical challenges from AI

Ansgar Koene,

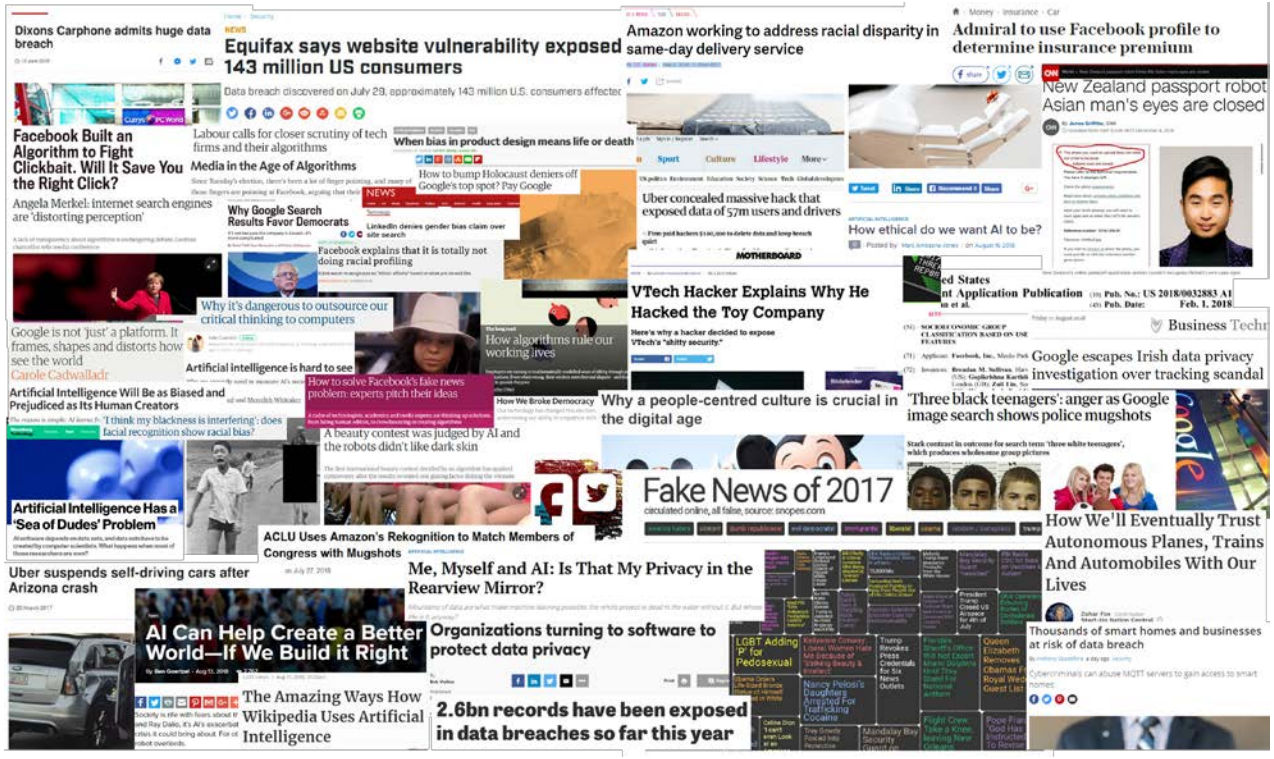
IEEE P7003 Standard for Algorithmic Bias Considerations

Undesired impact

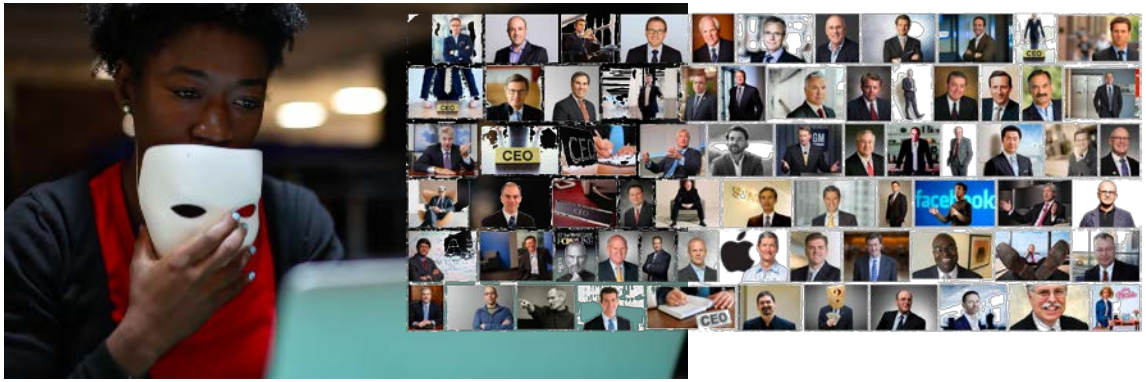
The need for ethics, legal, social, economic intervention

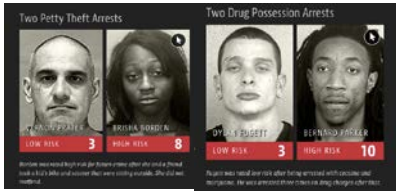
IEEE

Algorithms in the news

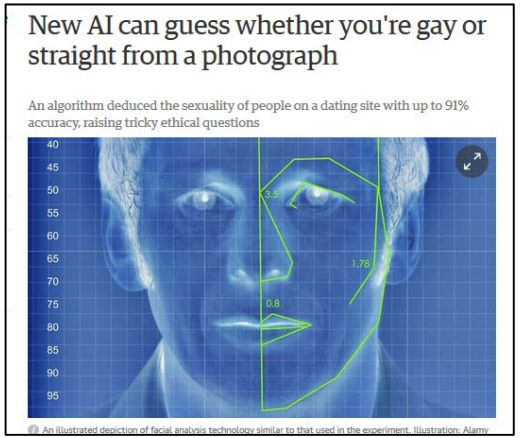


Algorithmic Discrimination





Machine Bias: There's software used across the country to predict future criminals. *Propublica*



Algorithmic systems are socio-technical

- Algorithmic systems do not exist in a vacuum
- They are built, deployed and used:
  - by people,
  - within organizations,
  - within a social, political, legal and cultural context.

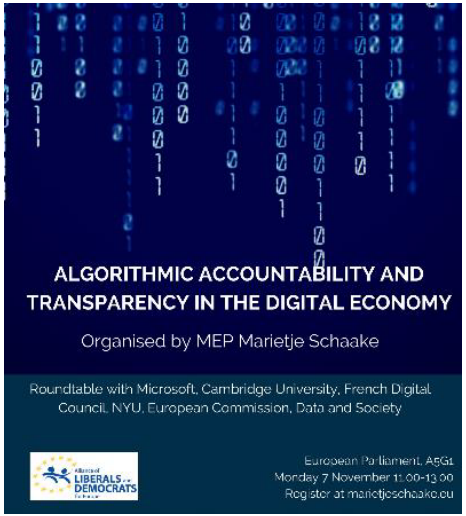
- The outcomes of algorithmic decisions can have significant impacts on real, and possibly vulnerable, people.

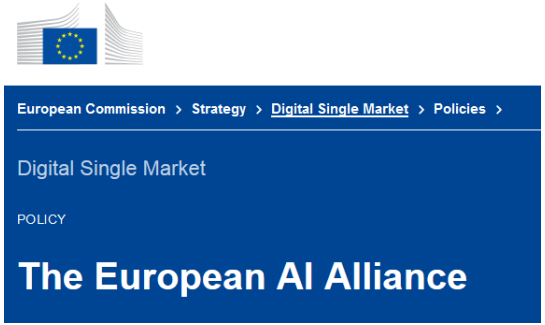
Governance frameworks for algorithmic systems

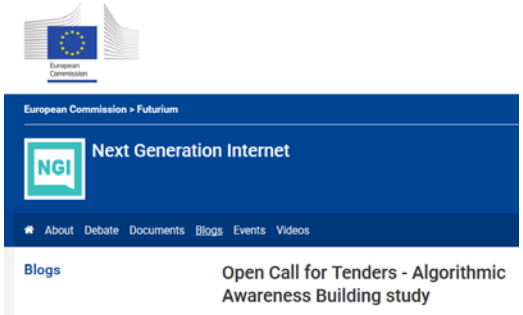
Public and Private sector responses to Undesired impacts of AI

IEEE

EU response (in addition to GDPR)







EU Parliament Science and Technology Options Assessment (STOA) panel request for study on “Algorithmic Opportunities and Accountability”

IEEE

Governance options

Table II Selected market solutions and governance measures by categories of risk

Risks	*Market solutions* *Demand side*	*Supply side*	*Companies: self-organization*	*Branches: self-regulation*	*Co-regulation*	*State intervention*
Manipulation		×	×	×		×
Bias	×	×				
Censorship	×	×	×			×
Violation of privacy rights	×	×	×	×	×	×
Social Discrimination	×		×			×
Violation of property rights		×	×	×		×
Abuse of market power			×			×
Effects on cognitive capabilities						
Heteronomy						

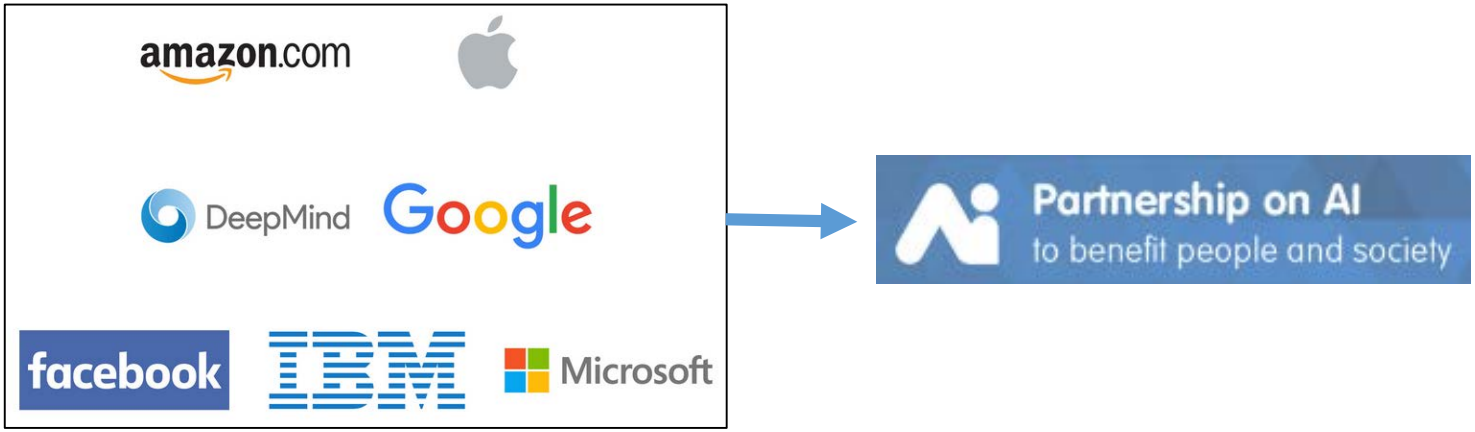
Florian Saurwein, Natascha Just, Michael Latzer, (2015) "Governance of algorithms: options and limitations", info, Vol. 17 Issue: 6, pp.35-49, doi: 10.1108/info-05-2015-0025

Demand side Market solutions

Supply side Market solutions

Companies self-organization (aka CSR)

Branches self-regulations



Branches self-regulations: Industry standards

- British Standards Institute (BSI) – BS 8611 *Ethics design and application of robots*

- IEEE P70xx Ethics of Autonomous and Intelligent Systems standards

- ISO/IEC JTC1 SC42 (launched at start of 2018)
  - *Artificial Intelligence Concepts and Terminology*
  - *Framework for Artificial Intelligence Systems Using Machine Learning*
    - *SG 2 on Trustworthiness*
      - *transparency, verifiability, explainability, controllability, etc.*
      - *robustness, resiliency, reliability, accuracy, safety, security, privacy, etc.*

- The EU standards bodies CEN and CENELEC officially created a Focus Group on AI, in support of ISO/IEC SC42 (December 2018).

- Jan 2018 China published “Artificial Intelligence Standardization White Paper.”

Co-regulation

Privacy Shield
Framework

IEEE

State Intervention

Oversight by regulatory organisations

- An FDA for algorithms – Andrew Tutt (2016)
- An FAA for algorithms – paraphrasing Alan Winfield

(Chair of IEEE P7001 Standard for Algorithm Transparency)

Setting legal requirements

Industry Standards

IEEE P70xx Standards, developed as part of the Global Initiative for Ethics of Autonomous and Intelligent Systems

IEEE

IEEE P70xx Standards Projects

IEEE P7000: Model Process for Addressing Ethical Concerns During System Design

IEEE P7001: Transparency of Autonomous Systems

IEEE P7002: Data Privacy Process

IEEE P7003: Algorithmic Bias Considerations

IEEE P7004: Child and Student Data Governance

IEEE P7005: Employer Data Governance

IEEE P7006: Personal Data AI Agent Working Group

IEEE P7007: Ontological Standard for Ethically Driven Robotics and Automation Systems

IEEE P7008: Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

IEEE P7009: Fail-Safe Design of Autonomous and Semi-Autonomous Systems

IEEE P7010: Wellbeing Metrics Standard for Ethical AI and Autonomous Systems

IEEE P7011: Process of Identifying and Rating the Trustworthiness of News Sources

IEEE P7012: Standard for Machines Readable Personal Privacy Terms

IEEE P7013: Inclusion and Application Standards for Automated Facial Analysis Technology


IEEE P7003™,
Standard for Algorithmic
Bias Considerations
Working Group
IEEE Computer Society/Software & Systems
Engineering Standards Committee (C/S2ESC)

P7003 - Algorithmic Bias Considerations

- All non-trivial* decisions are biased
- We seek to minimize bias that is:
  - Unintended
  - Unjustified
  - Unacceptable
- as defined by the context where the system is used.

*Non-trivial means the decision space has more than one possible outcome and the choice is not uniformly random.

IEEE

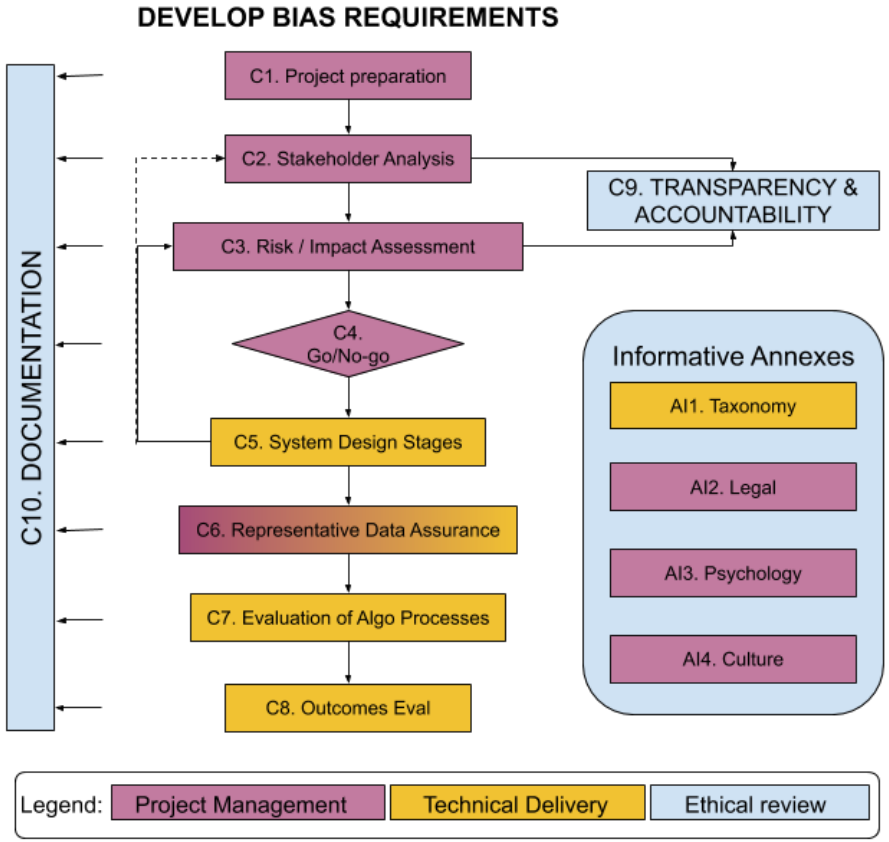
Causes of algorithmic bias

- Insufficient understanding of the context of use.
- Failure to rigorously map decision criteria.
- Failure to have explicit justifications for the chosen criteria.

Key question when developing or deploying an algorithmic system

- Who will be affected?
- What are the decision/optimization criteria?
- How are these criteria justified?
- Are these justifications acceptable in the context where the system is used?

IEEE P7003 general structure



5Rights Universal Standards for childhood and adolescence

A collaboration between 5Rights and IEEE-SA

Why these Standards are needed

- That presence of children in the digital environment must be anticipated
  - by design
  - by default
- Technology must be provided in children in a way that upholds their rights and meets their needs.
  - Not simply about content
  - The ways it uses their data
  - The behaviours it encourages
  - The responsibility it takes for the impact of its services.

The approach

Look systemically for the drivers and inhibitors in technological systems, to create a good environment for child

- How does the system impact on, or promote the autonomy of a child?
- What effect might a child's engagement have on their health or well being?
- Have you considered both their physical and emotional wellbeing?
- What processes are in place to inform the child of the likely impact of using your service?

From the big '**is it fair**' and '**does it uphold the rights of the child**' to the entirely granular about where on the screen a button might be better placed.

A suite of Standards/guidance - industry connections group

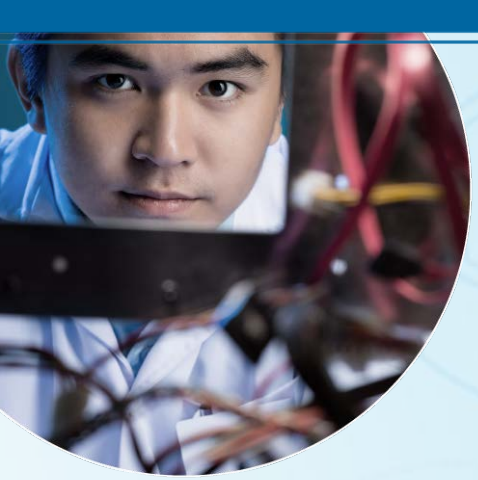
- Age Appropriate Contract – that is to determine what terms and conditions, or community rules should offer when the end user is a child.
- Standards that cover:
  - security of IoT,
  - Child Online Protection Issues,
  - Privacy differentials,
  - context capacity authentication,
  - guidance for duty of care,
  - appropriate governance structures,
  - reporting standards,
  - flagging systems,
  - data minimization standards,
  - best practice geolocation
  - etc.

Interplay between Standards and Legislation

- 5rights Founder and Chair, Baroness Beeban Kidron is the architect of ground-breaking new Age Appropriate Design Code, an enhanced GDPR for children under 18.

- 5Rights as an organization, believes that standards should both anticipate and be an alternative to legislation.

- Creating standards from a trusted source allows all businesses – small and big – access to the thoughtful and ethical digital services for children.

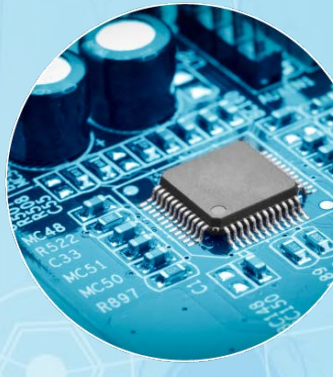


Thank you!

ansgar.koene@Nottingham.ac.uk

IEEE P7003 Standard for Algorithmic Bias Considerations project site:
http://sites.ieee.org/sagroups-7003/

https://5rightsfoundation.com/



IEEE