Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

What is the Internet Doing to Me:

Ethics on the Internet – AI and Data Governance

Dave Lewis, dave.lewis@scss.tcd.ie
Thanks to: Wessel Reijers, Arturo Calvo, Killian Levacher

Student Online Teaching Advice Notice

The materials and content presented within this session are intended solely for use in a context of teaching and learning at Trinity.

Any session recorded for subsequent review is made available solely for the purpose of enhancing student learning.

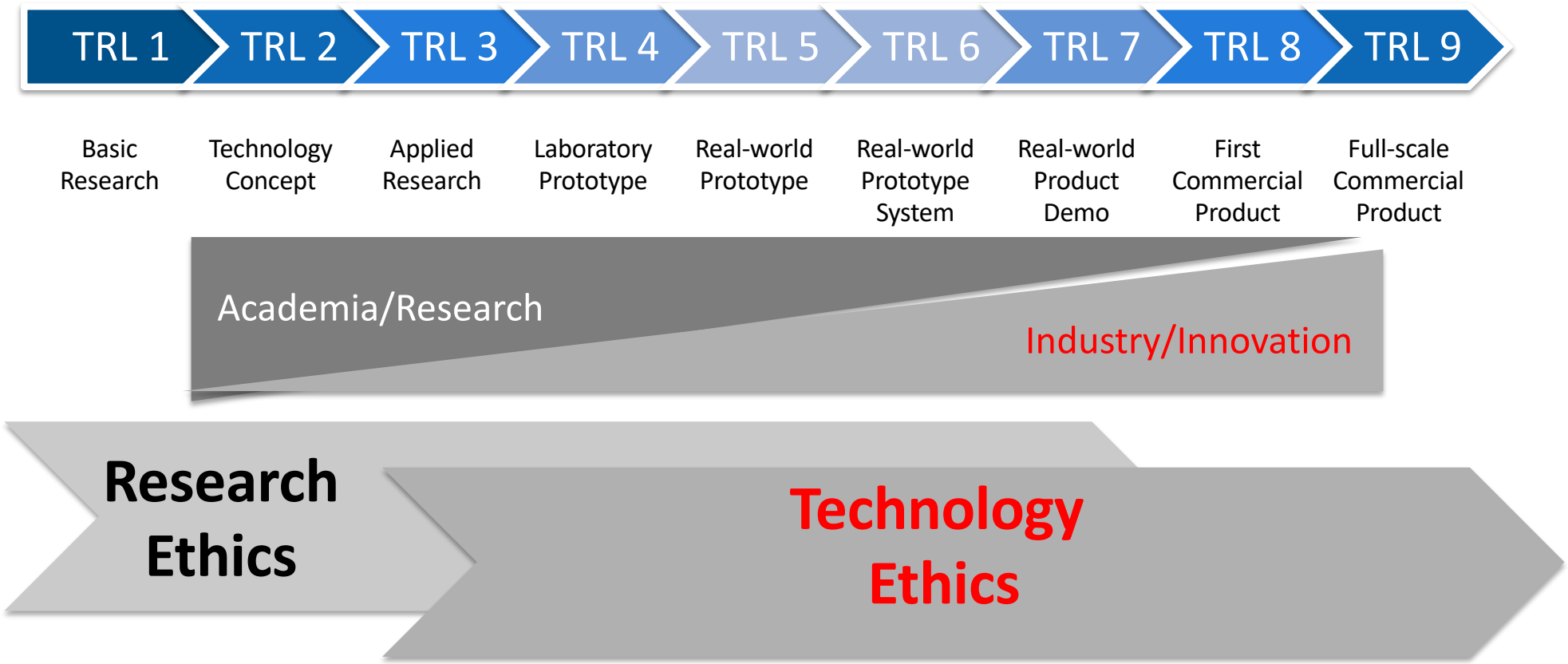
Students should not edit or modify the recording in any way, nor disseminate it for use outside of a context of teaching and learning at Trinity.

Please be mindful of your physical environment and conscious of what may be captured by the device camera and microphone during videoconferencing calls.

Recorded materials will be handled in compliance with Trinity's statutory duties under the Universities Act, 1997 and in accordance with the University's policies and procedures.

Further information on data protection and best practice when using videoconferencing software is available at https://www.tcd.ie/info_compliance/data-protection/.

Ethics in Technology: Research vs. Innovation



Why Should Tech Innovators be Concerned with Ethics?

- Because new technologies have a **profound impact** on the way **we live**, on the **relationships we have**, on the **societal & political processes we engage in**.

- For tech innovators?
  - First: because it is good for the image of your business (instrumental goal)
  - Second: because it actually improves the service you provide! (substantive goal)
  - Third: because it is the *good* thing to do, it contributes to your idea of a better society and being a good person (normative goal)

Dominant Views in Technology Ethics

- The neutrality thesis: technologies are *instruments* that we can use to attain our own goals.
  - "People kill people"
- The determinism thesis: technologies *dictate* everything we do, they determine who we are.
  - "Guns kill people"
- The co-shaping thesis: technologies and humans together "construct" our social world.
  - "Gun-men kill people"

Technology Impact: Example

Software Malfeasance: Example

Unintended Impacts - Example: Gender in Google Translate

- Some languages, like Turkish, don't has gender specific pronouns
- Google translate has to guess the gender when translating in English
- Statements allocating gender to role reveal gender bias
- What is the source of this?
- Is it a problem?

Sample Google Translate output:

he is a soldier
she's a teacher
he is a doctor
she is a nurse

https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/

Power of Big Data: Example: Cambridge Analytica

- Academic research into Psychographics (U. Cambridge) revealed the link between psychological profiles and Facebook profiles
- Correlated major psychological types to elements in the social graph: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism
- Cambridge Analytica applied psychographics to help target political ads in 2016 US elections….

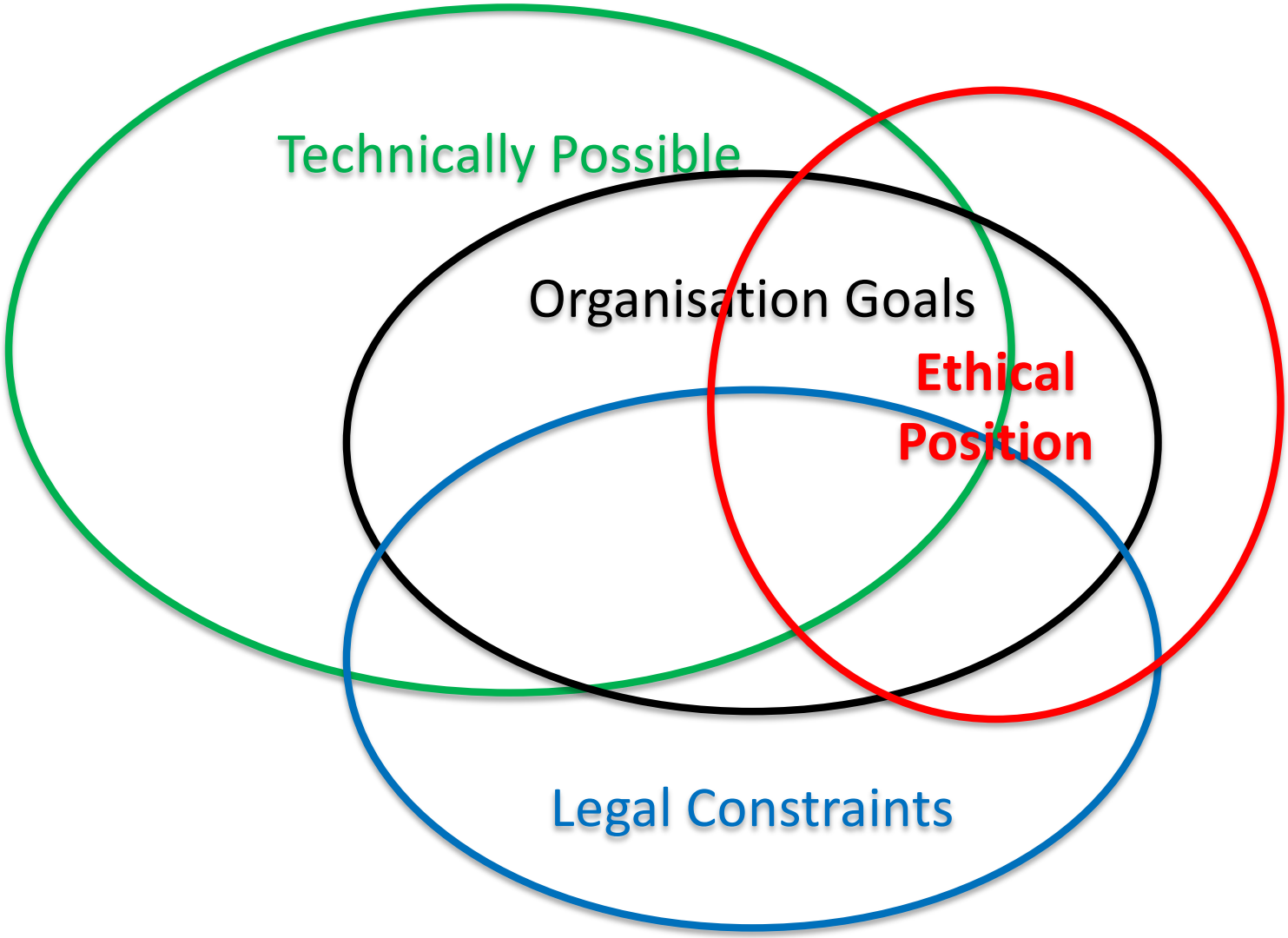
https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Algorithmic Power on Behaviour & Worldview

- “Race to the Bottom … of the Brain Stem” Tristian Harris
- 70% of YouTube views are based on algorithmic recommendations
- Business model maximises video views to maximise ad views
- Outrage/fear/anger the most reliable reactions that drive us to keep watching
- -> Recommender algorithm inevitably drive us to content that builds outrage to keep us watching

- Evidence to US Congress: https://www.youtube.com/watch?v=WQMuxNiYoz4
- Agenda: https://humanetech.com/wp-content/uploads/2019/06/Technology-is-Downgrading-Humanity-Let%E2%80%99s-Reverse-That-Trend-Now-1.pdf

Ethics in a Technology Development



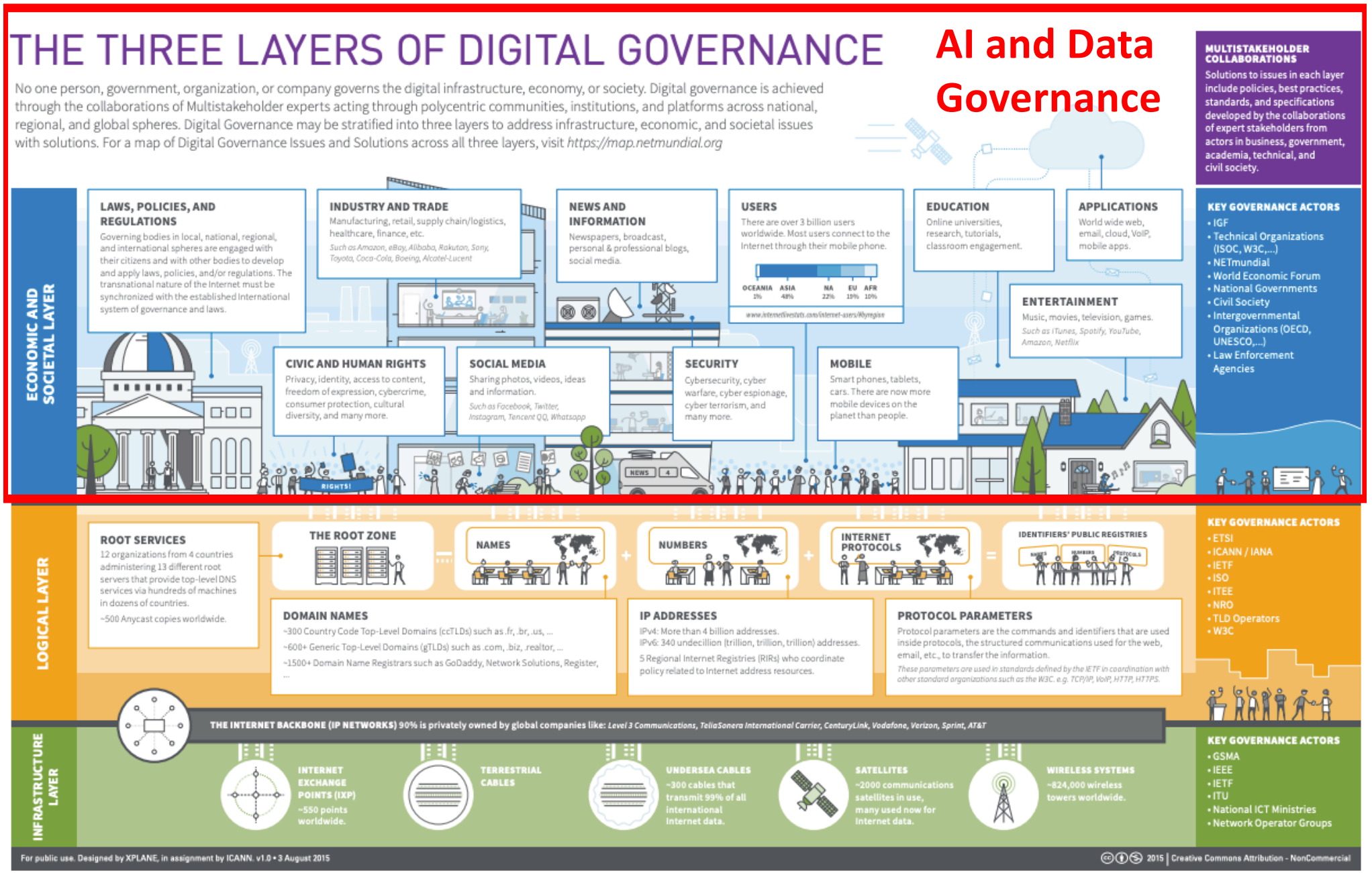
[IBM]

Ethical Risks of AI on the Internet: Algorithmic Selection of digital content

- Manipulation of individuals or groups
- Diminishing variety that creates biased views and distortion of reality
- Constraints on communication and freedom of expression
- Threats to privacy and data protection rights
- Social discrimination
- Violation of intellectual property rights
- Impact on the human brain and cognitive capacity
- Algorithmic power over human behavior and development

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. Handbook on the Economics of the Internet, (October 2014), pp 395–425. Retrieved from https://doi.org/10.4337/9780857939852.00028

Governance over the Internet



Big Data and AI

Big Data are extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

- Examples: location traces; social media posts/likes/comments; digital content in the form of text, audio, video; geospatial data; sensor data

Artificial Intelligence (AI) is a family of computational techniques that aim to mimic human capabilities such a learning and problem solving

Machine Learning is an increasing successful form of AI using mathematical models trained on Big Data rather than explicit coding instructions

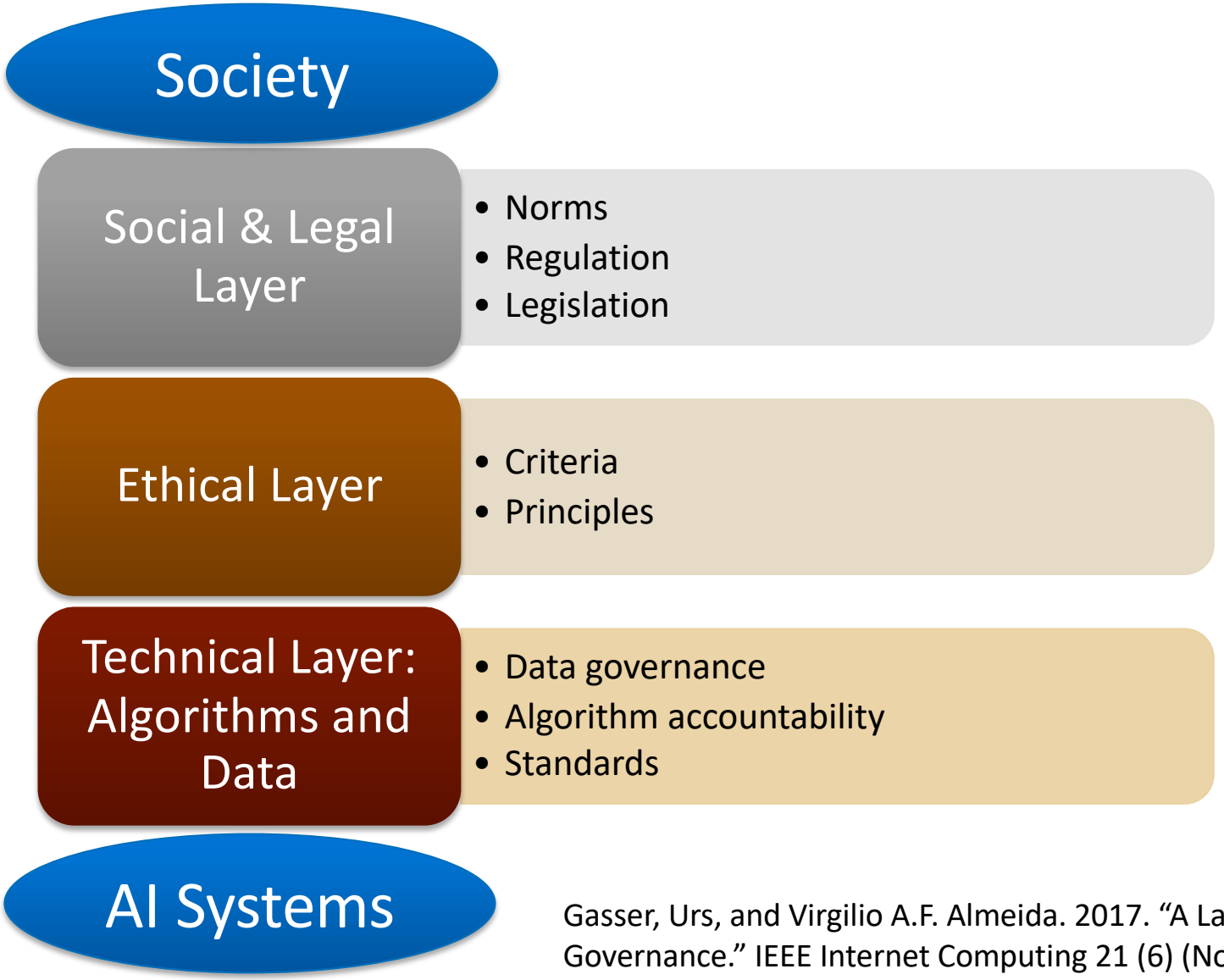
- Example applications: media recommender systems, speech recognition, face recognition, natural language processing, machine translation, search, predictive data analytics

Why AI and Data Governance?

- Companies harvest and utilise personal data on a **massive scale**
- Growing concerns about the **collection, linking, use and leakage** of personal data from **mobile devices, bio-sensors, cameras, GPS trackers and social media**.
- Machine Learning deliver new levels of **insights and predictions** about an individual's behaviour and also feeds increasingly **personalised AI-driven interactive digital experiences - Ads to Alexa**
- Individuals and groups **struggle to understand** the impact of personal information processing
- Companies, especially SMEs, often lack the knowledge and expertise needed to address these **complex legal and ethical issues**.

.

AI Governance: Layered Model

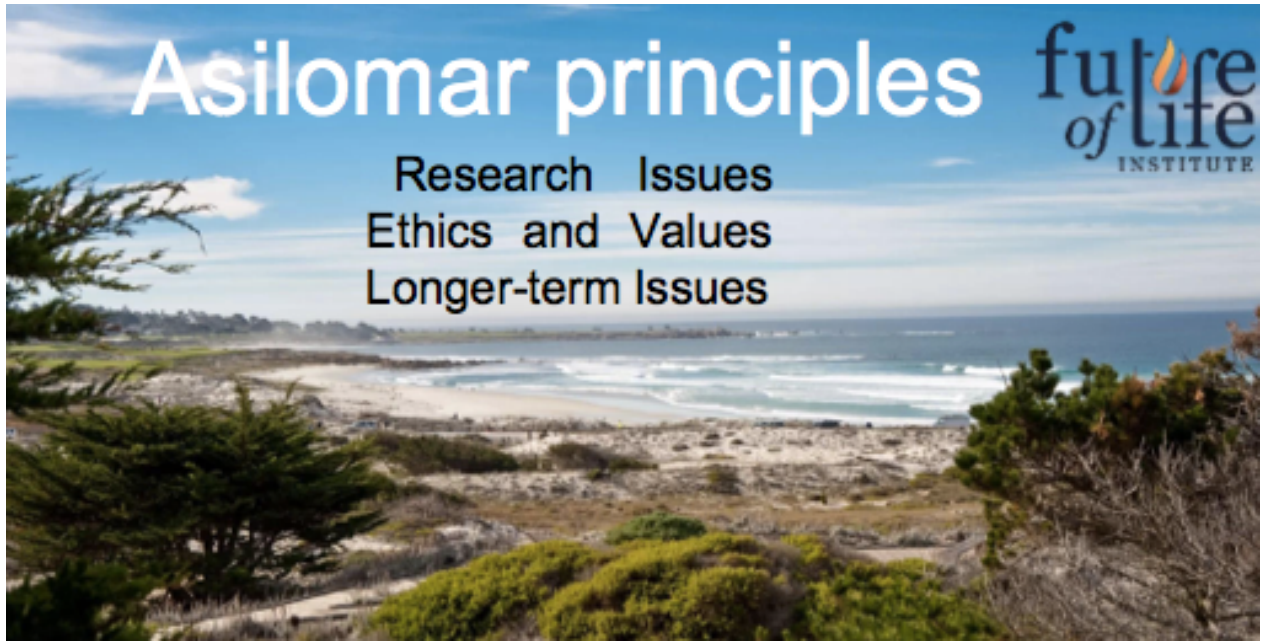


Gasser, Urs, and Virgilio A.F. Almeida. 2017. "A Layered Model for AI Governance." IEEE Internet Computing 21 (6) (November): 58–62.

Asilomar Principles

Ethical AI Principles

- Safety
- Failure Transparency
- Judicial Transparency
- Responsibility
- Value Alignment
- Human Values
- Personal Privacy
- Liberty and Privacy
- Shared Benefit
- Share Prosperity
- Human Control
- Non-subversion
- AI Arms Race



https://futureoflife.org/ai-principles/

EU Ethics Guidelines for Trustworthy AI - 2019

Ethical Principles mapped from EU Charter of Fundamental Right

International AI Policy Differentiator for EU

Ethical AI, alongside Lawful AI and Robust AI

Requirements

- Human Agency and Oversight
- Technical Robustness and Safety
- Privacy and Data Governance
- Transparency
- Diversity, Non-Discrimination and Fairness
- Societal and Environmental Well Being
- Accountability

https://ec.europa.eu/futurium/en/ai-alliance-consultation

EU Ethics Guidelines for Trustworthy AI

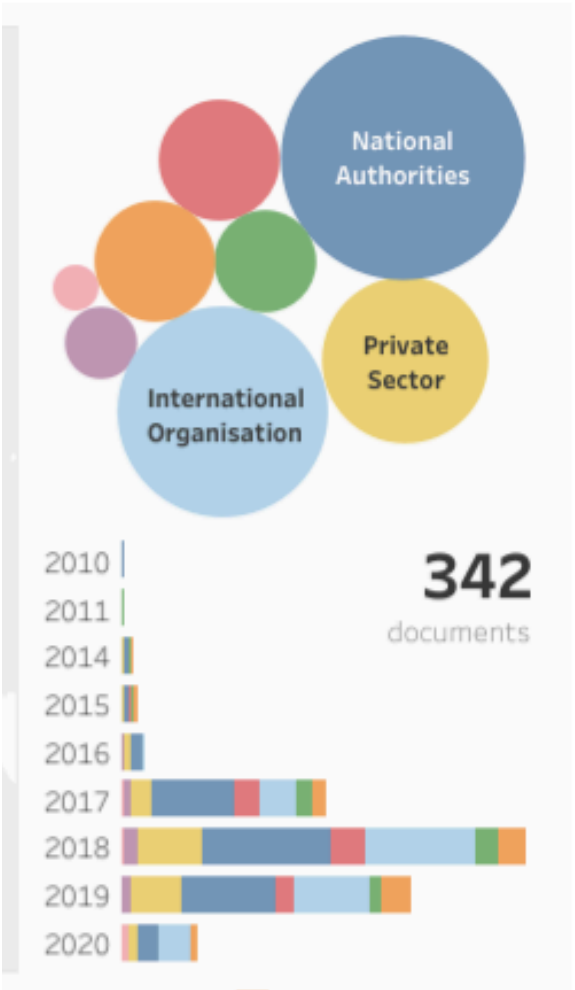
Risk Mitigation Methods

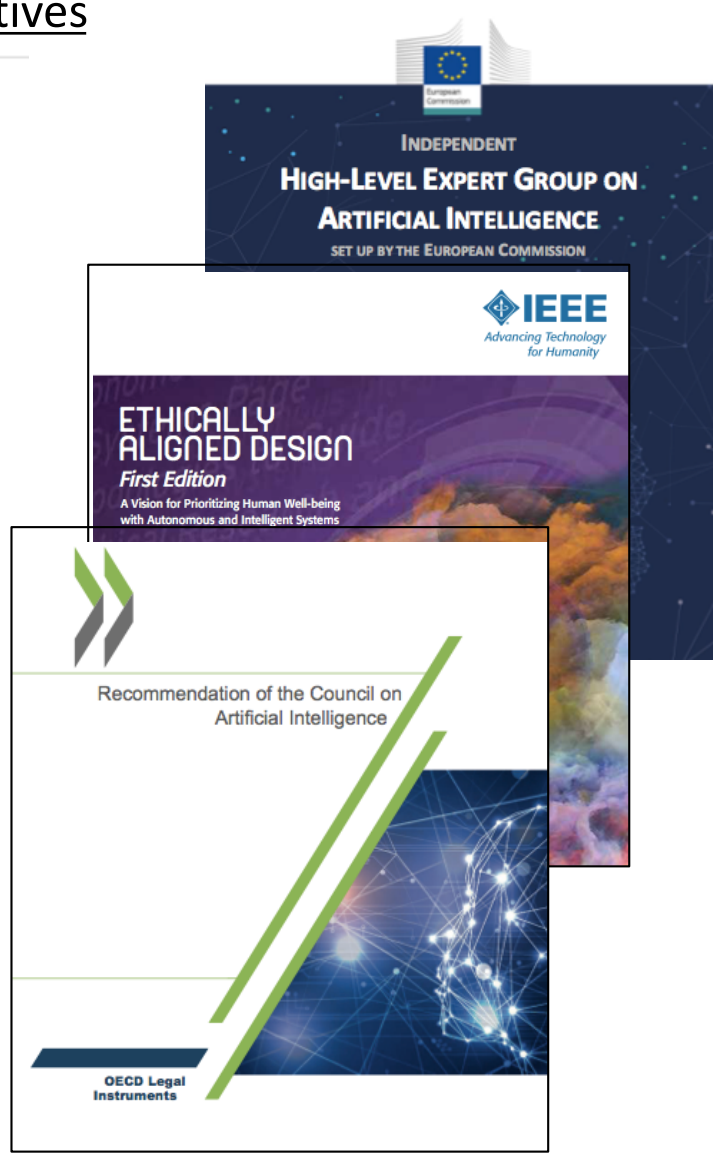
- Technical:
  - Architecture,
  - Ethics/privacy-by-design,
  - Explanation,
  - Testing/validation,
  - QoS Indicators
- Non Technical:
  - Regulation
  - Code of Conduct
  - Standardisation
  - Certification
  - Accountability via Governance Frameworks
  - Education & Awareness
  - Stakeholder Participation
  - Diverse Design Teams

https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Competing/Converging Sets of Principles

https://www.coe.int/en/web/artificial-intelligence/national-initiatives



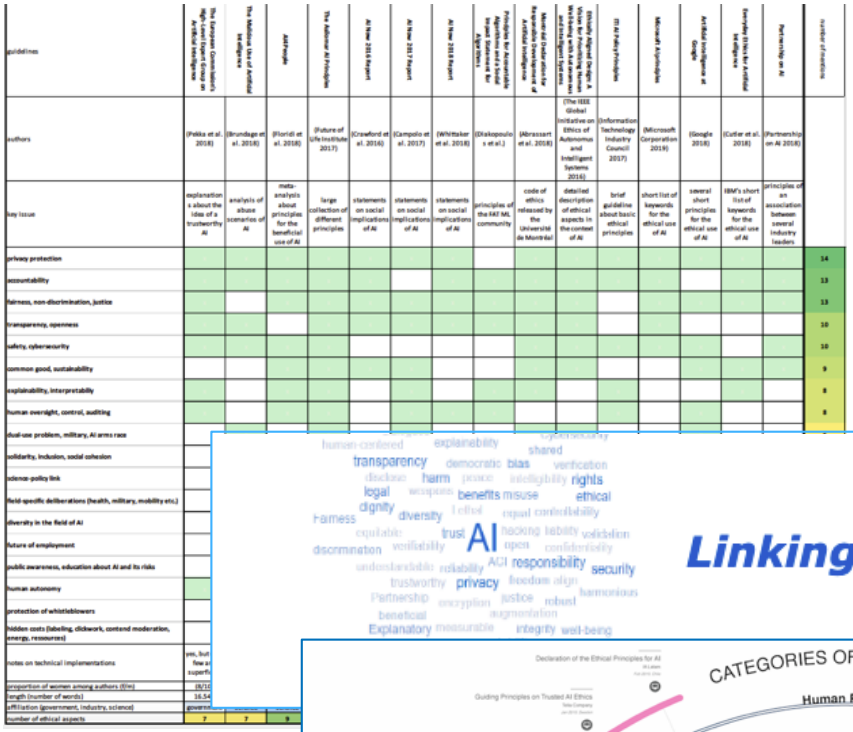


Consensus on principles of

- Transparency
- Justice
- Non-maleficence
- Responsibility
- Privacy

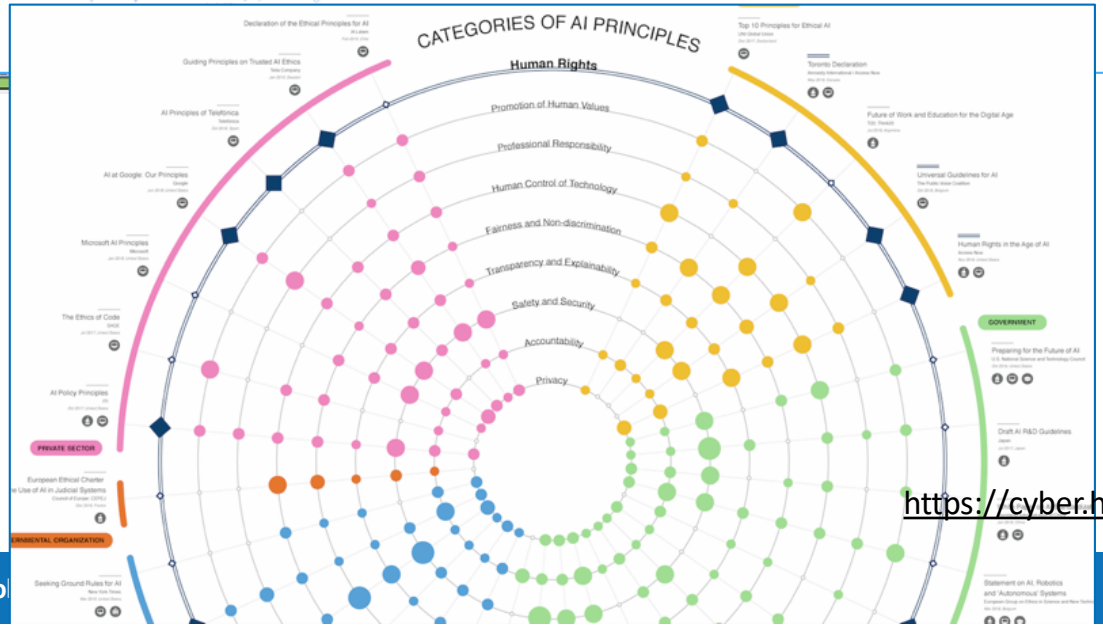
Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. Nat Mach Intell 1, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2

Growing body of policy analysis



The Ethics of AI Ethics -- An Evaluation of Guidelines,
Thilo Hagendorff, Feb 2019,
https://arxiv.org/abs/1903.03425

http://www.linking-ai-principles.org/



BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

https://cyber.harvard.edu/publication/2020/principled-ai

Trustworthy AI - Policy Gaps

- Government use
- Automation and labour
- Autonomous Weapons
- Environmental impact
- Collective vs. individual harms and rights
- “Ethical” vs. Societal – “ethicswashing”
- Over-focussed on technically solvable areas, e.g. Bias vs. Fairness

AI Governance: Challenges

- **Definition**: Difficult to reach stable consensus on what defines AI
- **Discreetness**: Growing access to AI skills and computing power, it can be developed out of sight
- **Diffuseness**: AI used in a diffuse set of locations and jurisdictions
- **Discreteness**: Impact of an AI component only apparent when assembled into a system
- **Opacity:** Modern machine learning yields results without clear explanations
- **Forseeability**: AI-driven autonomous system can behave in unforseeable ways – 'liability gap'
- **Control**: AI can work in ways/speeds out of control of those responsible for them

Scherer, M.U. Regulating Artificial Intelligence System, Harvard Journal of Law and Technology, 29(2) 2016

Headwinds to International Consensus on AI Governance

- **Pacing:** AI tech and applications develop faster that societies ability to regulate it
- **Securitisation**: International competition as AI perceived as a strategic economic/military resource
- **Innovation:** Perceived impediment to AI-based innovation and its economic and social benefits
- **Asymmetry:** Power of AI concentrated in a few digital platforms that benefit from massive network effects

BUT

- GDPR provides new rights with growing uptake
- Governments now starting to consider AI Regulation

https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en