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

What is the Internet Doing to Me:

Ethics on the Internet
Emerging Practice with AI and Data

Dave Lewis, delewis@tcd.ie,
Delaram Golpayegani, sgolpays@tcd.ie
Thanks to: Wessel Reijers, Arturo Calvo, Killian Levacher, Harsh Pandit

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.

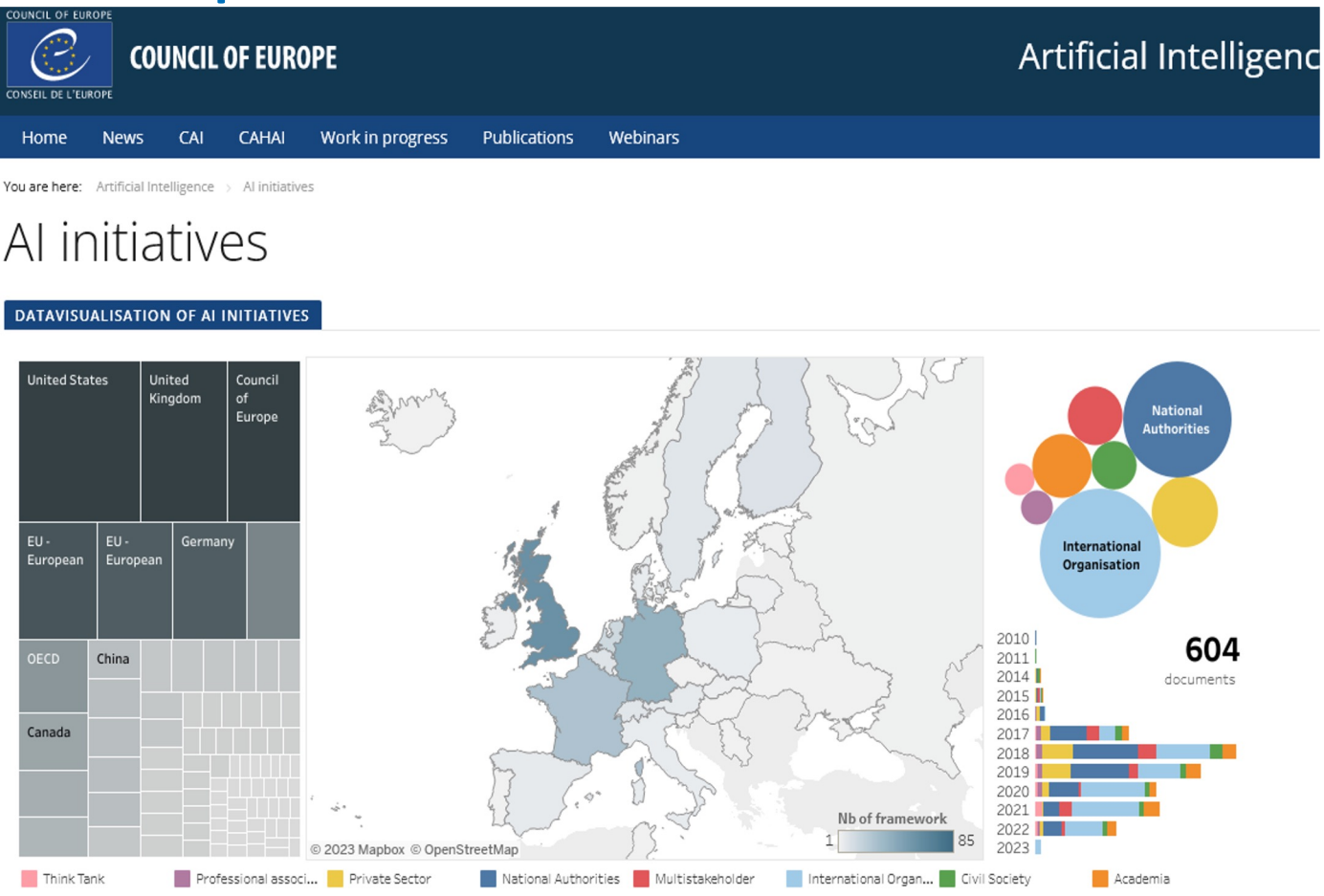
Students should not edit or modify the recording in any way, nor disseminate it for use outside of a context of teaching and learning at Trinity.

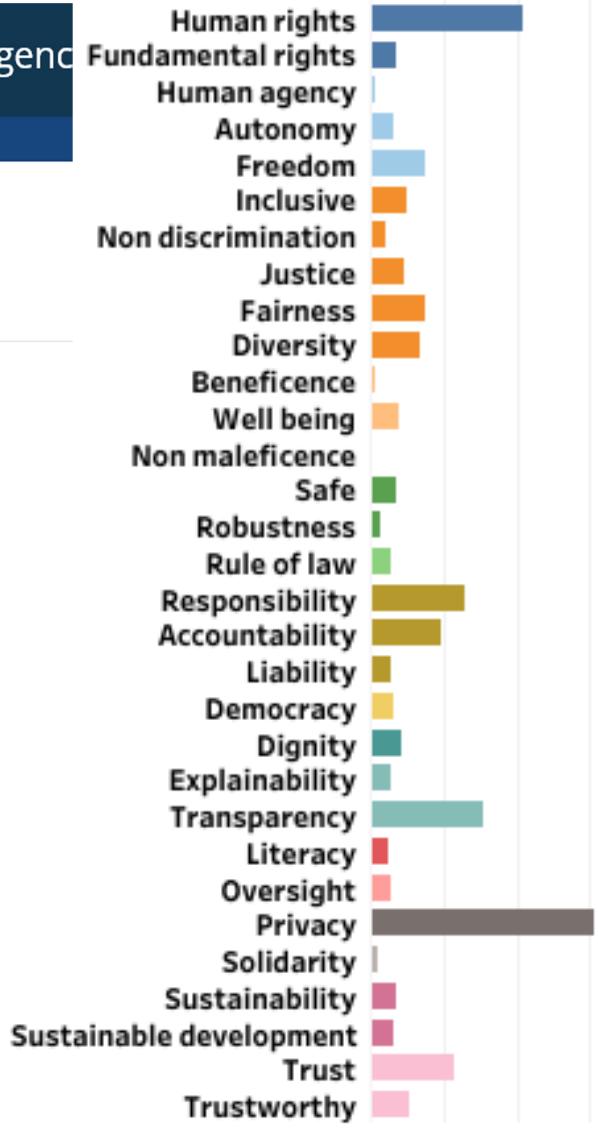
Please be mindful of your physical environment and conscious of what may be captured by the device camera and microphone during videoconferencing calls.

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/.

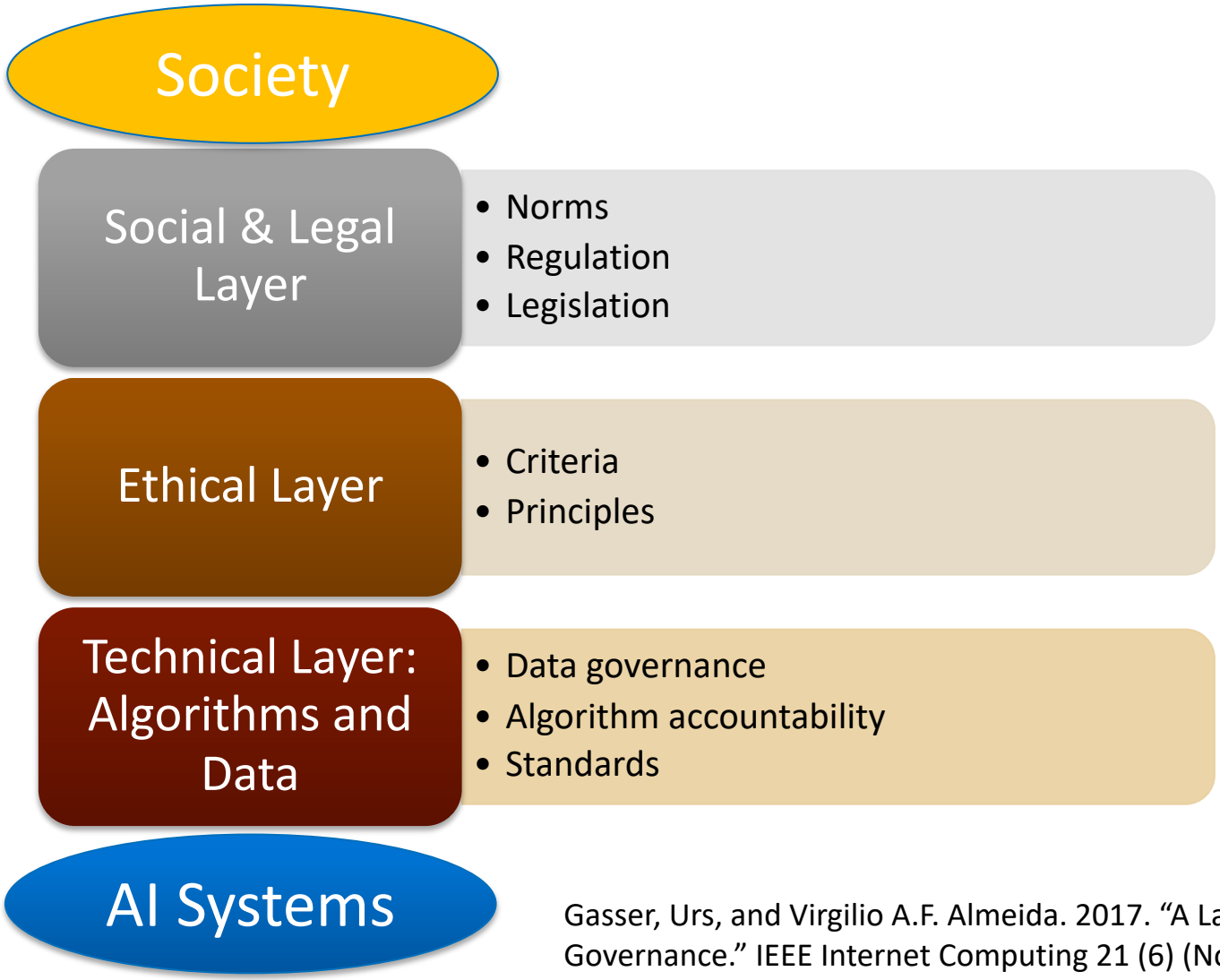
Flourishing of Trustworthy/ Ethical/ Responsible AI initiatives





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

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.

Overview - Trustworthy AI and Data Governance: complexities moving from principles to practice:

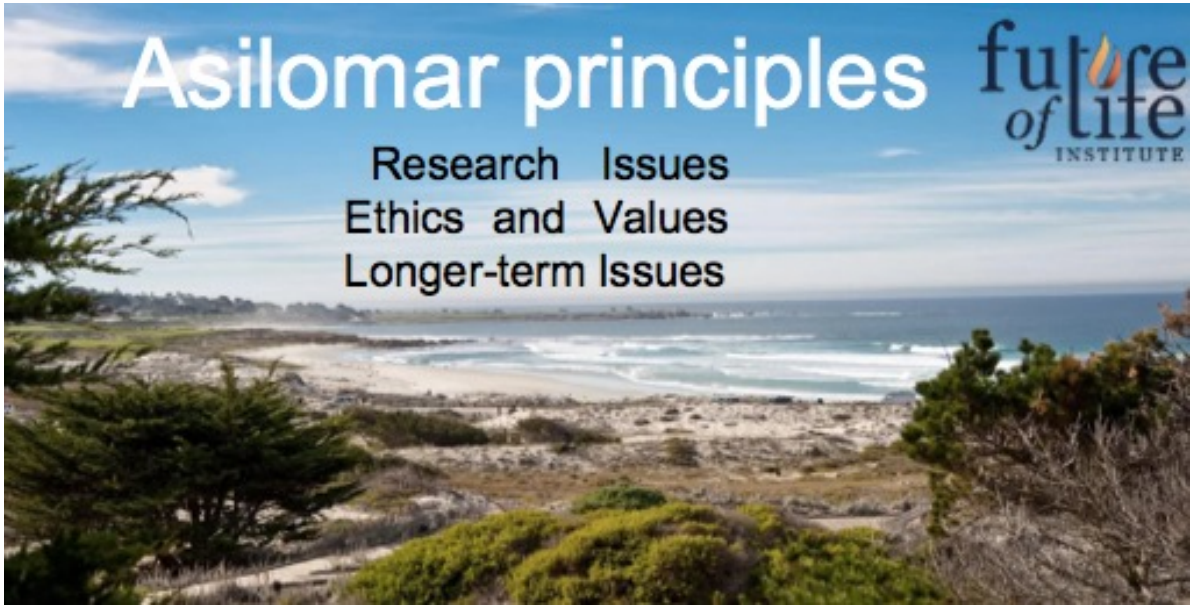
- **Significant initiatives:**
  - Organization level policies: tech platforms, public sector, sectoral/vertical, researchers
  - Emerging Standards: ISO/IEC JTC1 SC42, CEN/CLC JTC 21, IEEE P7000, National Bodies
  - Public Policy: EU HLEG, OECD, UNESCO
  - Emerging Legislation: EU AI Act, US National AI Initiative Act

- Systemic Concerns
  - Workable guidelines
  - Regulatory load vs. benefits
  - Value-chain complexities & liability
  - Who wields oversight authority
  - Stakeholder engagement
  - Ethics Washing
- Divergent pressures
  - Disciplinary: ‘problemists’ vs ‘solutionists’
  - Sectorial
  - Technological
  - Jurisdictional

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/

Examples of Ethical 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/Principles

- 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

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

Challenges in Governing Ethical AI and Data technology

- **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 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

Approaches to AI/Data Governance

- Self Governance by Organisations
    - Internal tech ethics boards – current examples lack transparency
- Self Regulation by Industries
    - Example: Partnership for AI: https://www.partnershiponai.org/
    - Lack of transparency and enforcement
- Government Regulation – Co Regulation – e.g. EU AI Act
    - Oversight shared between state and value-chain actors
    - Labelling of AI projects akin to energy efficiency

      Ethics much more complicated than energy consumption
    - Supplier declaration of conformance for data sets or trained models
    - External certification of declaration processes
- Machine Ethics
    - Stuart Russell: Human Compatible: AI and the Problem of Control
    - AI asks people for input on ethical issues, leanrs from that but if uncertain switches itself off

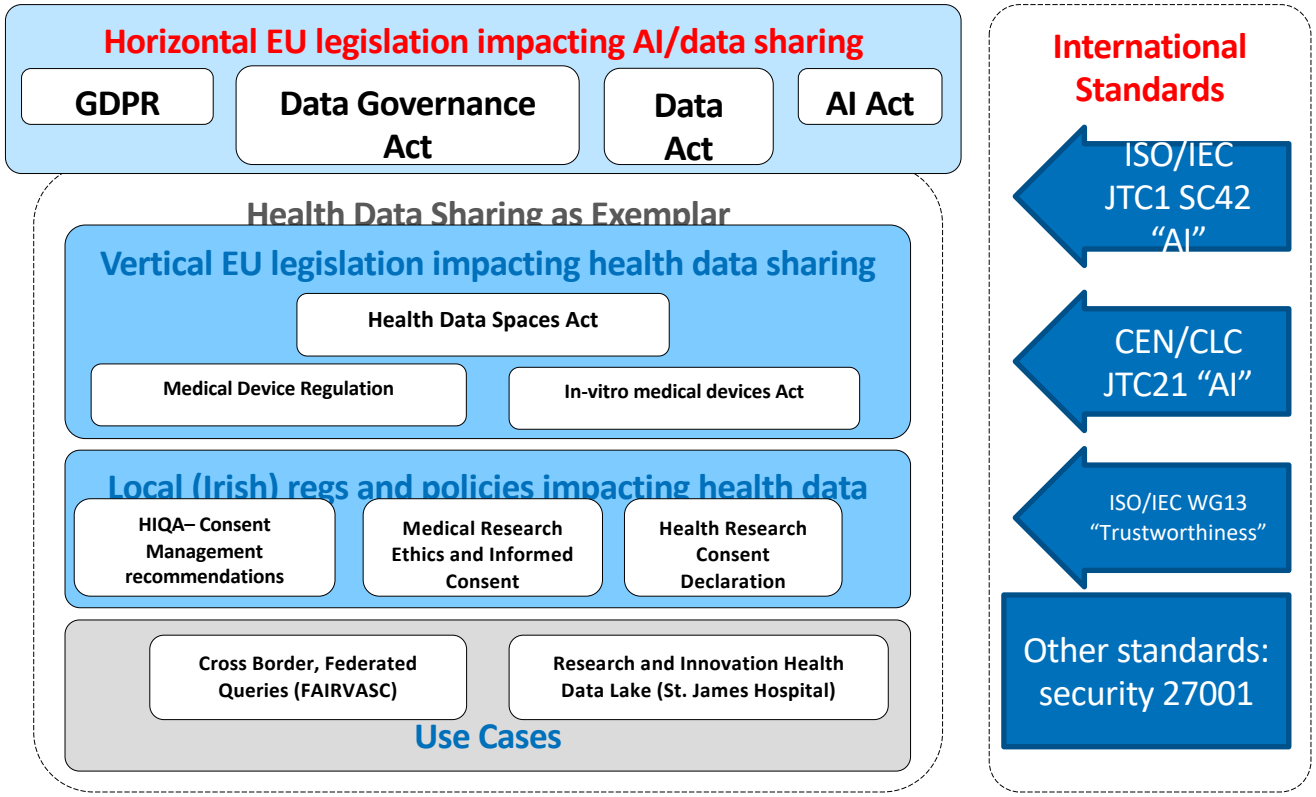
Recap: High Variety of Ethical Frameworks for AI in Research & Innovation

Your Role?	Accountable to Whom?	What must you demonstrate?	Possible Sanction?
Planning Publicly-funded Research	Institution Ethics Committee	Clear research goals, minimise risks to human participants, data protection	Research prevented or delayed
Processing Personal Data (GDPR)	Org's Data protection Office, Data Protection Authorities, Courts	Data Protection: risk assessment, by design, security measures, minimisation, storage limits, subject rights	Public determinations, Legal sanctions, large fines, compensation to subjects
Publishing Research	Peer reviewers/editors	Consideration of professional standards and discipline-specific concerns, e.g. ACM, ACL	Feedback to improve, required corrections,
Seeking Funding	Funding body & its reviewers	Ethics & Data Management, specific measures in high risk areas: clinical, human subject, animals, nuclear, biomed and AI.	Funding denied/delayed on ethics/GDPR grounds
Deploying Product or Service (future AI Act)	External- or Self-Certification Authorities, AND Market Surveillance Authorities, the Courts	GDPR compliance, AI Quality Management System, Impact assessments & Risk Management, technical documentation	Denied product certification, products removed from market, large fines, commercial liability

Roll out of Digital Regulation in the EU

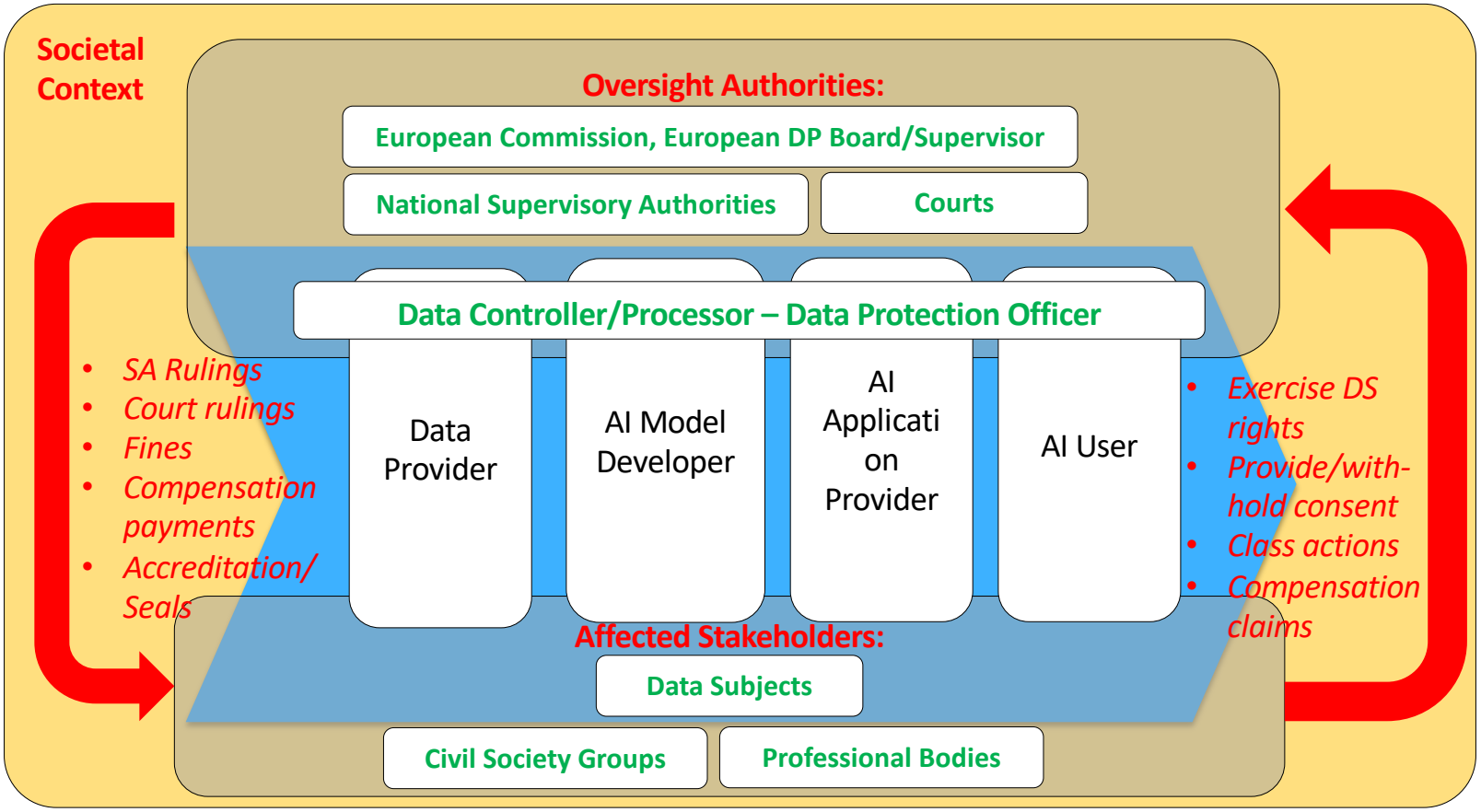
Law	Stage	Area
GDPR	MAY-2018	personal data
Digital Services Act	NOV-2022	services
Digital Markets Act	MAY-2023	market
Data Governance Act	SEP-2023	data markets
AI Act	voting	technology
ePrivacy Reg	draft	communication
Data Act	proposed	data
Health Data Space	proposed	health data
Interoperability Act	proposed	data

Rapidly Growing Regulatory Complexity: e.g. for health data sharing



- **GDPR was fairly 'self-contained'**
- **Raft of new digital EU legislation emerging:**
  - AI Act,
  - AI Liability Directive
  - Data Governance Act
  - Data Act

Recap: Trust building signals and affordances of GDPR

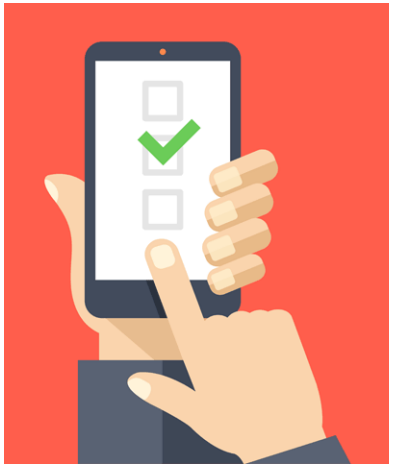


Does GDPR help in Governing AI/Data

- Informed consent, data minimization and storage limitations help
  - Data Subject Rights and Supervisory Authority offer some Transparency
  - Fines, SA judgements and compensation offers some Accountability
- Right to Portability: intended to give ‘market’ power to data subjects, but where to port to?
- Pseudo anonymization: if personal information _can_ be extracted from data sets - it is subject to GDPR
  - As AI gets better at re-identification, more data sets are subject to GDPR
  - Attempts to balance with statistical techniques, e.g. differential privacy
  - Profiling in GDPR: inference of new data that “evaluates personal aspects”
- Right to Explanation and automated decision making:
  - Human explanation and intervention in automated decisions
  - Explainability of machine learning can make this a challenge

Limitations of GDPR

GDPR's **Notice and Consent** affordance model limited by asymmetry in: Knowledge, Expertise & Time



Informed consent

Specific purpose, Data minimisation, Storage limitation

Sharing transparency

Rights to access, erase, correct, object, port

Cons

Pros

Vague but reassuring purposes "improve service"

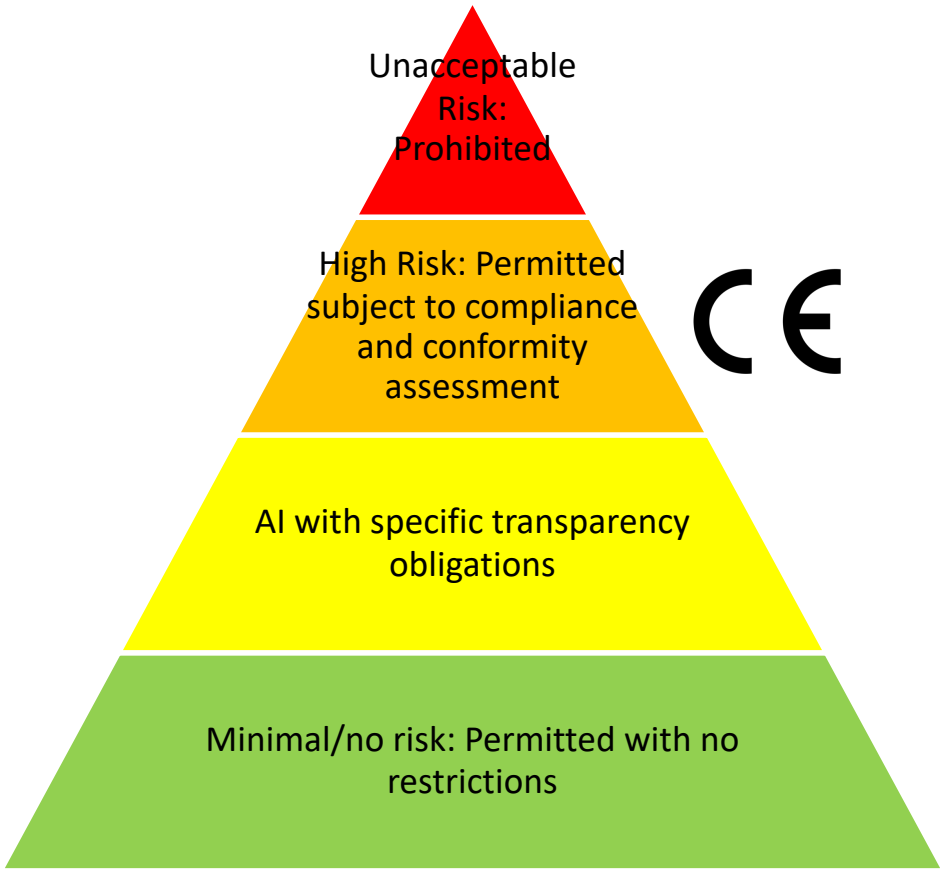
Dark Patterns for consent

"Legitimate Interest"

Poor enforcement

Data drives ever-more enticing and addictive apps

Proposed Framework for EU AI Act



Aims to:

- ensure AI systems are **safe** and respect **fundamental rights**
- Provide legal certainty for innovation, promote public trust and support single market

A **Risk-based** approach to regulating AI

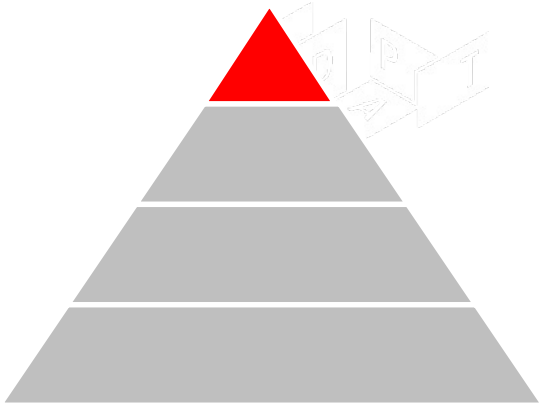
Large penalties: 30M or 6% of global turn-over

https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

Can Standards protect Fundamental Rights, deliver Ethical AI?

https://www.citizensinformation.ie/en/government-in-ireland/european-government/eu-law/charter-of-fundamental-rights

Prohibited AI Systems



Subliminal manipulation resulting in physical/ psychological harm

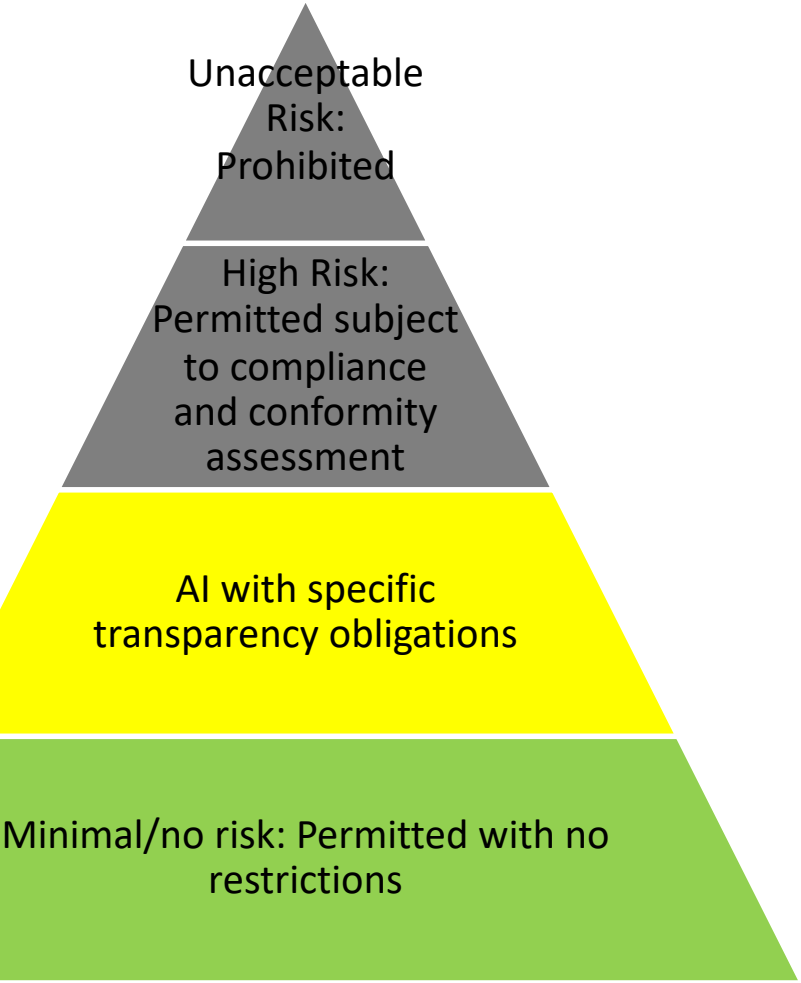
Exploitation of children or mentally disabled persons resulting in physical/psychological harm

General purpose social scoring

Remote biometric identification for law enforcement purposes in publicly accessible spaces (with exceptions)

https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

Some (most?) AI systems won't be high risk



New transparency obligations for certain AI systems (Art. 52)

- **Notify humans** that they are **interacting with an AI system** unless this is evident
- Notify humans that emotional recognition or biometric categorisation systems are applied to them
- Apply **label to deep fakes** (unless necessary for the exercise of a fundamental right or freedom or for reasons of public interests)

Possible voluntary codes of conduct for AI with specific transparency requirements (Art. 69)

- No mandatory obligations
- Commission and Board to encourage drawing up of codes of conduct intended to foster the **voluntary application of requirements to low-risk AI systems**

https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

High-risk Artificial Intelligence Systems (Title III, Chapter 1 & Annexes II and III)

1 **SAFETY COMPONENTS OF REGULATED PRODUCTS**

(e.g. medical devices, machinery) which are subject to third-party assessment under the relevant sectorial legislation

2 **CERTAIN (STAND-ALONE) AI SYSTEMS IN THE FOLLOWING DOMAINS**

- ✓ Biometric identification and categorisation of natural persons
- ✓ Management and operation of critical infrastructure
- ✓ Education and vocational training
- ✓ Employment and workers management, access to self-employment
- ✓ Access to and enjoyment of essential private services and public services and benefits
- ✓ Law enforcement
- ✓ Migration, asylum and border control management
- ✓ Administration of justice and democratic processes

AI Act: Transparency and Accountability

Requirements for high-risk AI systems (Title III, Chapter 2)

Establish and implement **risk management system** & in light of the **intended purpose** of the AI system

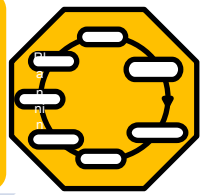
- Use high-quality **training, validation and testing data** (relevant, representative etc.)
- Draw up **technical documentation** & set up **logging capabilities** (traceability & auditability)
- Ensure appropriate degree of **transparency** and provide users with **information** on capabilities and limitations of the system & how to use it
- Ensure **human oversight** (measures built into the system and/or to be implemented by users)
- Ensure **robustness**, **accuracy** and **cybersecurity**

AI Quality Management System

Requirements on High Risk AI Systems

Determine if classified as High Risk

Ensure Design and Development are in compliance with Reg: risk management of data quality, transparency, tracebility, auditability, robustness, accuracy, cybersecurity, human oversight



AI Quality Management System

Third Party or Internal Conformity assessment procedure

Affix CE mark

CE

Put on Market/into Service

Can the AI Act deliver Ethical AI? …… Not without Standards

- Reminder: overall goal to protect health, safety and fundamental rights
- Existing regulation is referenced that has well established risk and quality models for **health and safety**
- **No direct guidance on how to protect fundamental rights – Act references 'harmonized standards'**
- **Harmonized standards are international standards approved through consensus of National Standards Bodies, e.g. National Standards Authority of Ireland and approved by European Commission**
- **CEN/CENELEC is the consensus forming body for standards in Europe**
  - **Joint Technical Committee 21 on AI established in 2021**
- **ISO/IEC JTC1 is the global consensus forming body for ICT standards**
  - **Subcommittee 42 established in 2017 to develop AI standards**

"standardisation is arguably where the real rulemaking in the Draft AI Act will occur"

Demystifying the Draft EU Artificial Intelligence Act, M.Veale, F.Z.Borgesius Computer Law Review International (2021), 22(4) 97-112, https://doi.org/10.48550/arXiv.2107.03721

Can International Standard Guide Ethical AI?

- SC42 follow established model of identifying specific consideration (for AI) within existing standards
  - Management System, Risk Management, Quality Management
  - Organisation and data governance
- AI-specific standards identify types of technical metrics that can be used:
  - Bias
  - Testing of Neural Networks
- Who develops these standards?:

Trustworthy AI Standards: Some Key Concepts

- **trustworthiness:** ability to meet stakeholder's expectations in a verifiable way [JTC1 AG]
- **stakeholder:** any individual, group, or organization that can affect, be affected by, or perceive itself to be affected by a decision or activity [ISO/IEC 38500:2015]
- **accountable:** answerable for actions, decisions, and performance [ISO 31000:2018]
- **risk**: effect of uncertainty on objectives [ISO 31000:2018]
- **control**: measure that maintains and/or modifies *risk* [SOURCE ISO 31000:2018]
- **bias:** favouritism towards some things, people, or groups over others

Harmonised Standards: ISO/IEC SC 42 AI Some key projects

JWG 1 [Governance of AI] - completed

- ISO/IEC 38507:2022 — Governance of IT — Governance implications of the use of artificial intelligence by organizations

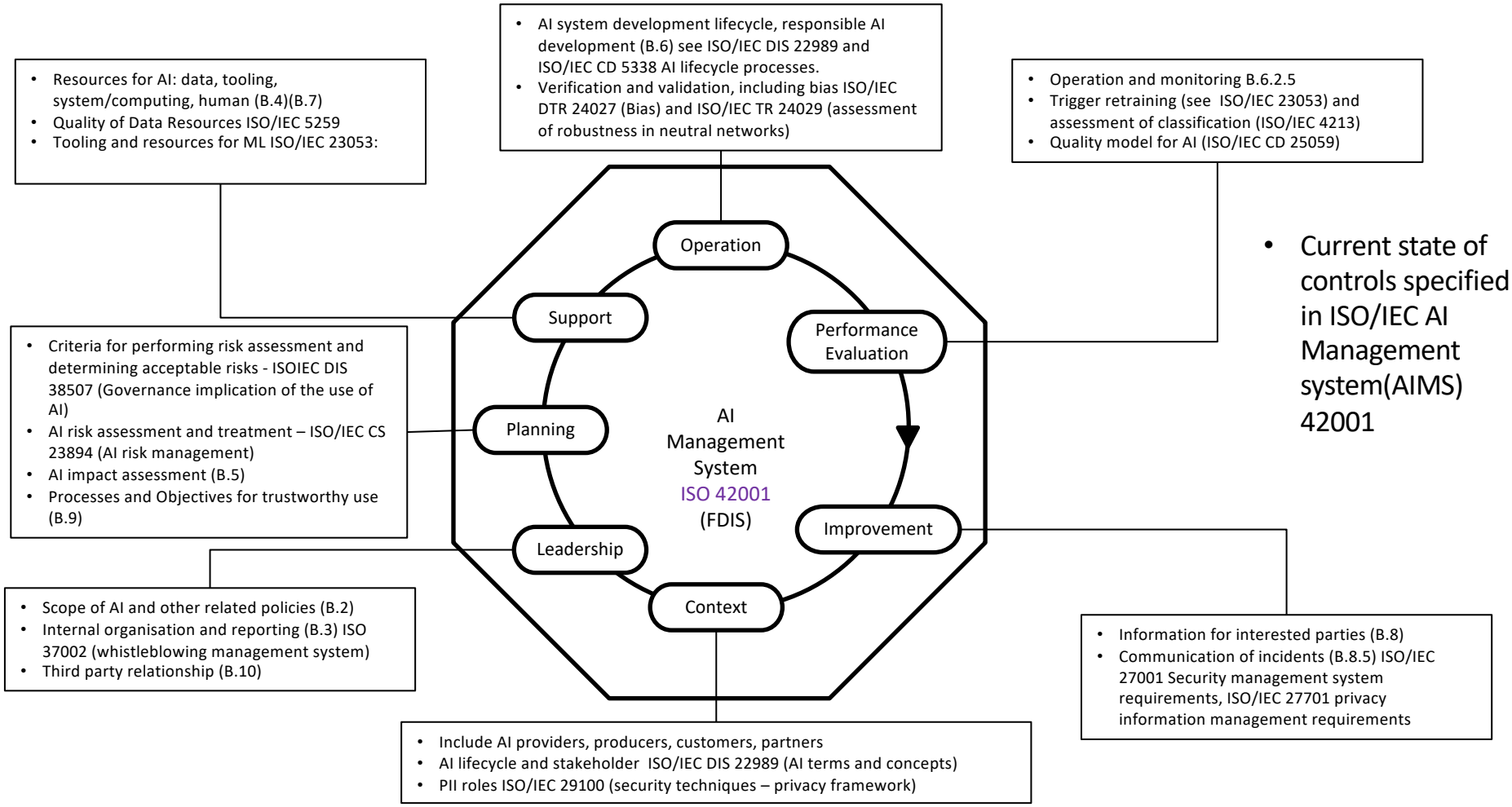
WG 1 [Foundational]

- ISO/IEC 22989:2022 Artificial intelligence concepts and terminology
- ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
- ISO/IEC DIS 42001 Artificial intelligence — Management system

WG 3 [Trustworthiness] 300+ experts

- ISO/IEC TR 24028:2020 Overview of trustworthiness in artificial intelligence
- ISO/IEC TR 24027:2021 Bias in AI systems and AI aided decision making
- ISO/IEC TR 24029-1:2021 Assessment of the robustness of neural networks — Part 1: Overview
- ISO/IEC TR 24368:2022 — Overview of ethical and societal concerns
- ISO/IEC 23894:2023 — Risk Management
- https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0

Expanding concept and requirement modelling to AIMS Control and their links to other SC42 standards



- Current state of controls specified in ISO/IEC AI Management system(AIMS) 42001

Can International Standard Guide Ethical AI?

- Standards should not and will not resolve consensus on disputed concepts which often frame ethical issues

- Example: should ‘fairness’ in allocating education or healthcare resources be based on:
  1. Sameness/equality?
  2. Deservedness/meritocracy?
  3. Need?

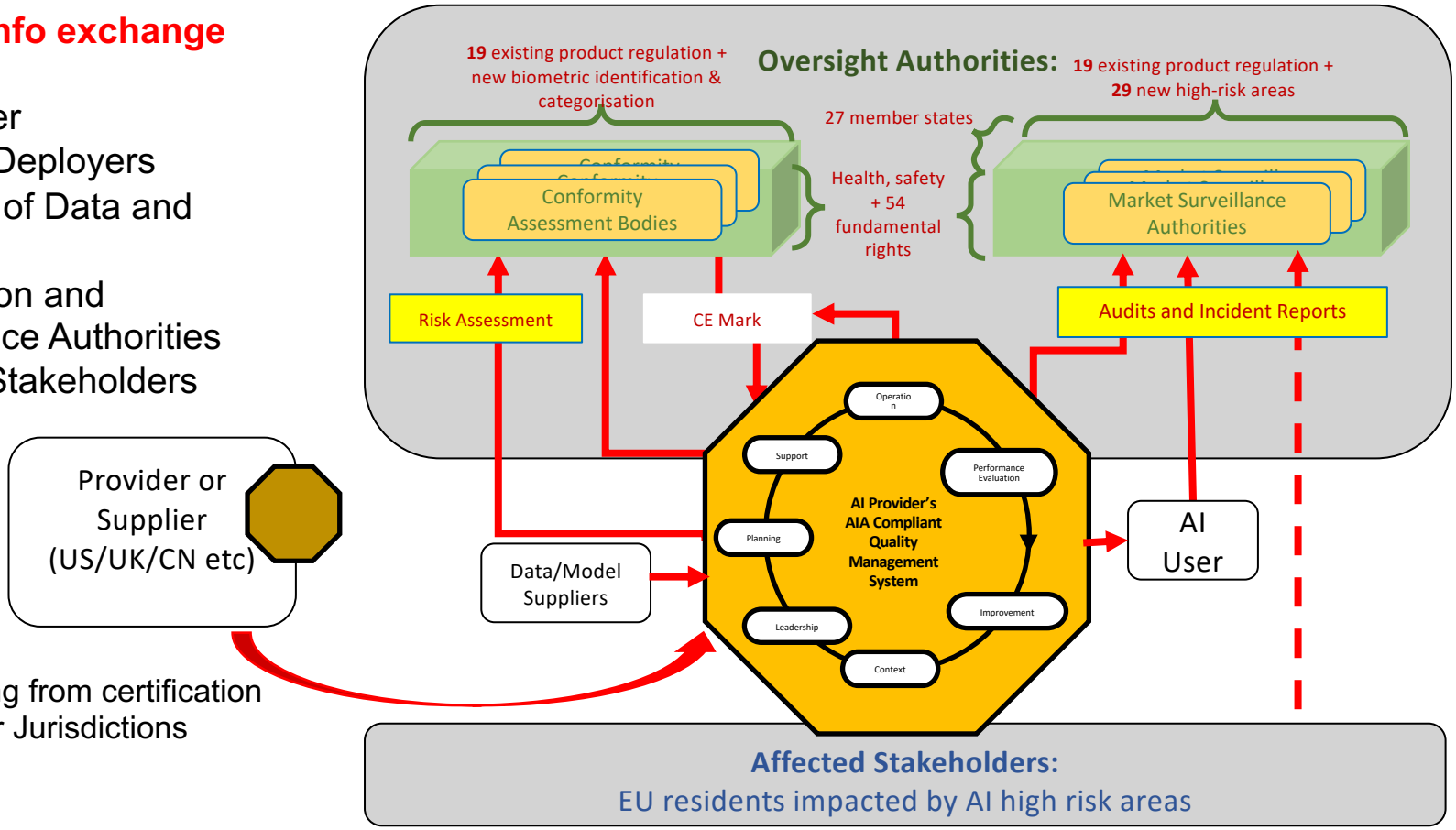
- Such societal-level disputes must be resolved through political processes, not by technical experts employed by large companies
- Standards may be able. To provide ‘knobs and levers’, e.g. definition of tests for bias
- It is a societal responsibility to define acceptable levels of risk
  - E.g. risk of mis-recognizing speech from those with less common accents
  - Same for education, ambulance dispatch, asylum?

31

Information Requirements for the AI Act

Complex info exchange between

- AI Provider
- AI Users/Deployers
- Suppliers of Data and Models
- Certification and Surveillance Authorities
- Affected Stakeholders



- Mapping from certification in other Jurisdictions

Information Required for Determining High-Risk AI

Engaging Content
Engaging People

By analysis of Annex III of AI Act, we identified **5 core concepts** for determining high-risk applications of AI

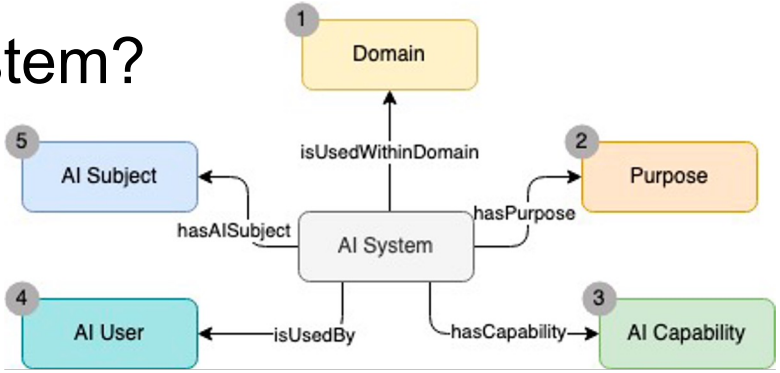
(1) In which **Domain** is the AI system used?

(2) What is the **Purpose** of the AI system?

(3) What is the **Capability** of the AI system?

(4) Who is the **User** of the AI system?

(5) Who is the **AI Subject**?



https://w3id.org/airo

Example of Identification of High-Risk AI Using the 5 Concepts

(1) In which **Domain** Education is the AI system used?

(2) What is the **Purpose** Assigning People to Educational Institutes of the AI system?

(3) What is the **Capability** Named Entity Recognition of the AI system?

(4) Who is the **User** Department of Education of the AI system?

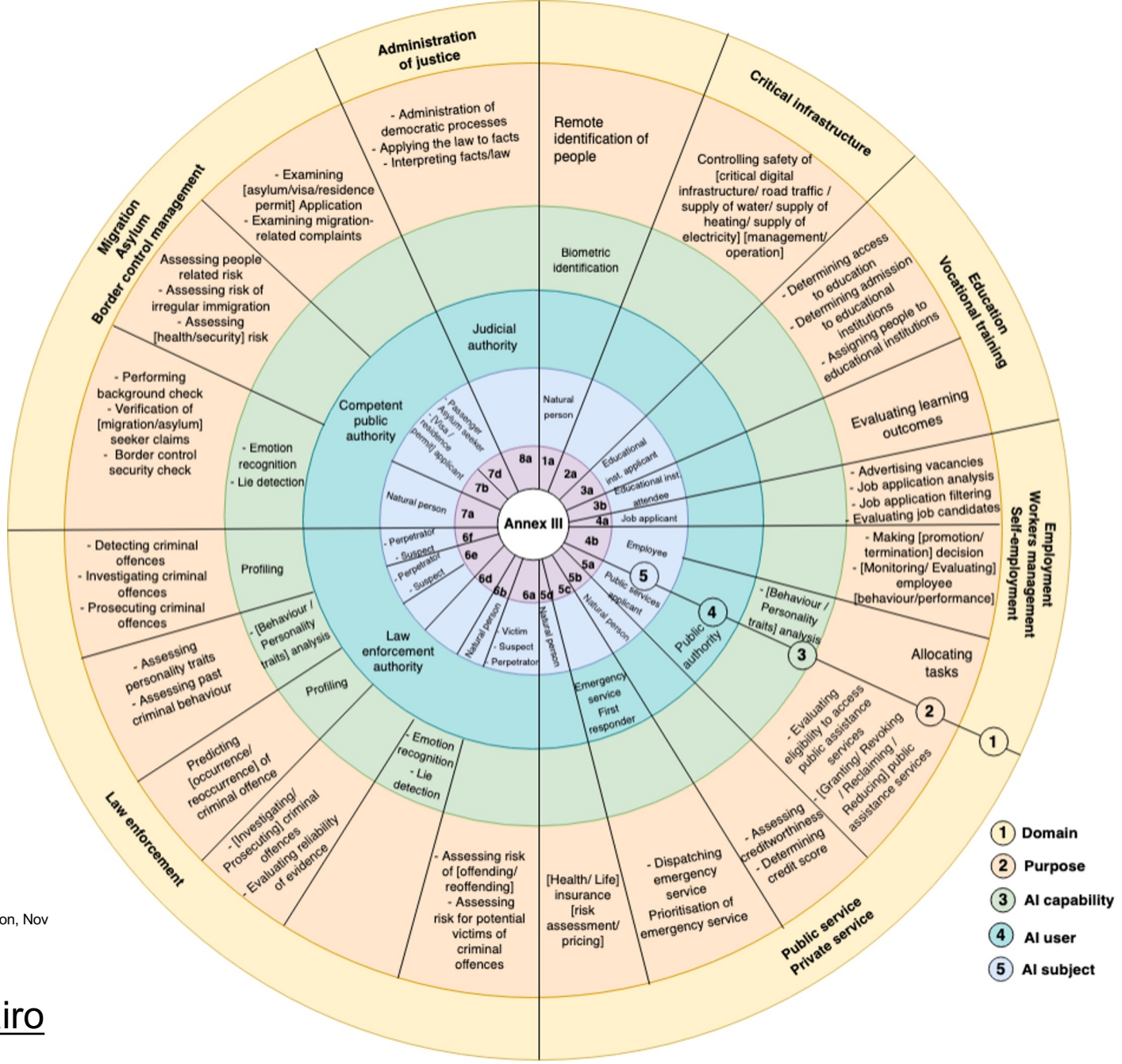
(5) Who is the **AI Subject**? Applicants

⬇

The AI system is highly likely to be **High-Risk** according to Annex III, 3a

34

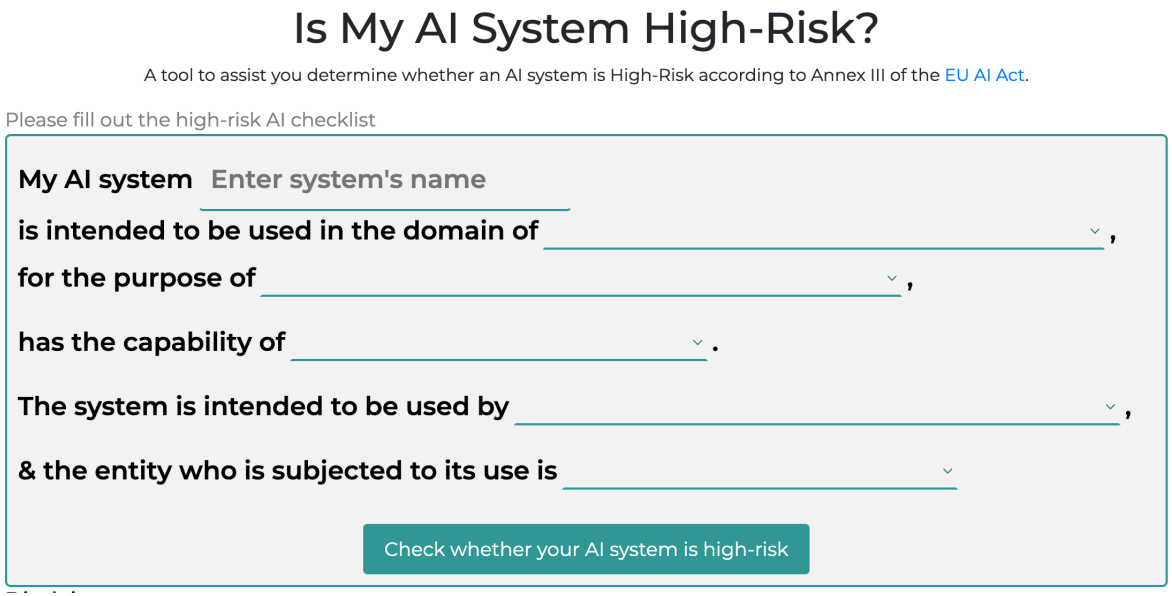
Risks Models for High-Risk Domains



Based on the Council's common position, Nov 2022

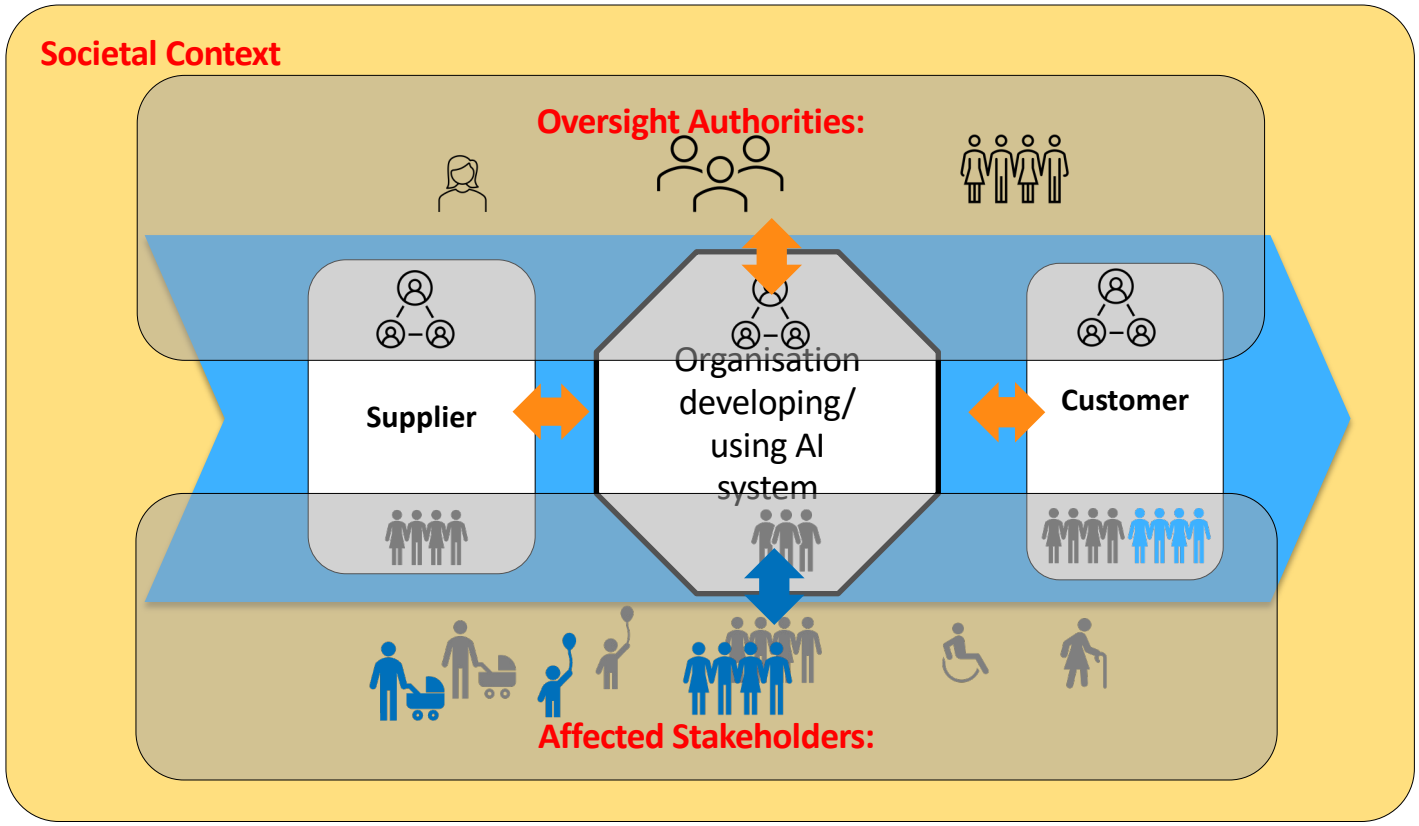
https://w3id.org/airo

Tool for Determining High-Risk AI



https://regtech.adaptcentre.ie/highrisk

Identifying Stakeholders in AI/Data Value Chains
Social Responsibility Perspective



- Labour Practices (workers)
- The Environment (future generations)
- Fair Operating Procedures (suppliers, customers, regulators)
- Consumer Issues (consumers)
- Community Involvement and Development (local communities)
- Human Rights (everyone)

Based on ISO 26000

SC42: Social Responsibility for AI

- Ethical and Societal Issues:
  - ISO need international consensus BUT avoids importing specific value-sets
  - Needs principles, which ones?
- ISO already has non-ICT specific principles: ISO 26000 – Social Responsibility
- Identification and engagement with stakeholder is key

Principles	Core Subjects (stakeholders)
• Accountability • Transparency • Ethical behavior • Respect for stakeholder interests • Respect for the rule of law • Respect for international norms of behaviour • Respect for human rights	• Organizational Governance Mitigations (governance board, managers, shareholders) • Human Rights (everyone) • Labour Practices (workers) • The Environment (future generations) • Fair Operating Procedures (suppliers, customers, regulators) • Consumer Issues (consumers) • Community Involvement and Development (local communities)

Human Rights issues for Social Responsibility

Risks	Mitigations
• **Legal**, from impacts in equality, privacy, access to justice • **Reputational**, from impacts to dignity, physical and mental integrity • **Complicity** in partner violations of rights • **Conflicts between stakeholder**, e.g. investors vs consumers, suppliers vs local communities • **To civil & political rights**: e.g. fake news social media bots, deep fake video impacting elections, filter bubbles, censorship • **To economic, social, cultural rights**: education, healthcare, wellbeing • **To just and favourable work**: casualised and deskilling labour of gig and click workers	• *Due diligence*: Human rights policy , Fundamental Rights Impact Assessment • *Avoid* value chain partners that may commit violations • Establish *grievance and redress mechanism*: transparent, accessible, external scrutiny, AI explanations • *Monitor for discrimination* towards vulnerable groups in AI decision making, e.g. insurance, justice, recruiting • *Education and access* for all groups to benefits of AI • Ensure worker *freedom of association* and collective bargaining

Labour Practice issues for Social Responsibility

Risks	Mitigations
• Legal arising from **discrimination** in AI assisted recruiting • To reputation: **worker dissatisfaction**, e.g. to intrusive monitoring • AI automation leading to **labour displacement** • **Deskilling** of work, e.g. translators correct machine translations • To worker **physical and mental health**	• *Recognition* of secure employment & decent working conditions • Engage in *social dialo*gue with worker and affective professional and community representative • Employee *retraining* • *Health and safety practices*, e.g. for robot coworkers, offensive content moderators • Protect *personal data of employees* • Seek *assurance* of good labour practices in value chain partners

40

Environment issues for Social Responsibility

Risks	Mitigations
• Increased **carbon emission** due to AI training and service operation • Resource usage and **pollution** from AI-driven product creation and disposal, e.g. sensors, batteries	• Monitor and plan reduction of non sustainable energy and resource use over whole product lifecycle • Make AI services available for environmental monitoring and analysis

Fair Operating Procedure issues for Social Responsibility

Risks	Mitigations
• Use of AI in **corrupt or anti-competitive practices**, e.g. finance, investment, procurement • Use of AI to **undermine the public political process**, e.g. through deep fakes, targeted manipulation or misinformation online • Violation of **intellectual property** rights	• Ensure *transparency and other safeguards* against abuse of power or complicity, e.g. protecting whistleblowers • Promote responsible behaviour in *value chain partners*, e.g. through requesting ethical impact assessment from AI partners • Identify, respect and fairly compensate *right holders*, e.g. annotators, translators, content providers

Consumer Issues in Social Responsbility

Risks	Mitigations
• Conveying **deceptive, misleading, fraudulent or unfair** information to consumers • Endangering **consumer health and safety**, e.g. mental health, self image • Incentivising **unsustainable consumption** • Misuse of **personal data** • Biased access to **essential services**	• Clearly *identify promoted content* and its sponsors • Monitor and benchmark *safety performance* and correct problems promptly • Consumer '*nutrition labels*', e.g. performance and failure envelope, energy usage • Clear and accessible *complaint and redress* mechanisms • Compliance with *privacy regulations*, e.g. transparency on data held or shared and its use • Consumer *education and awareness* raising, regardless of their capabilities or accessibility needs • *No discrimination or censorship* in access to services and information

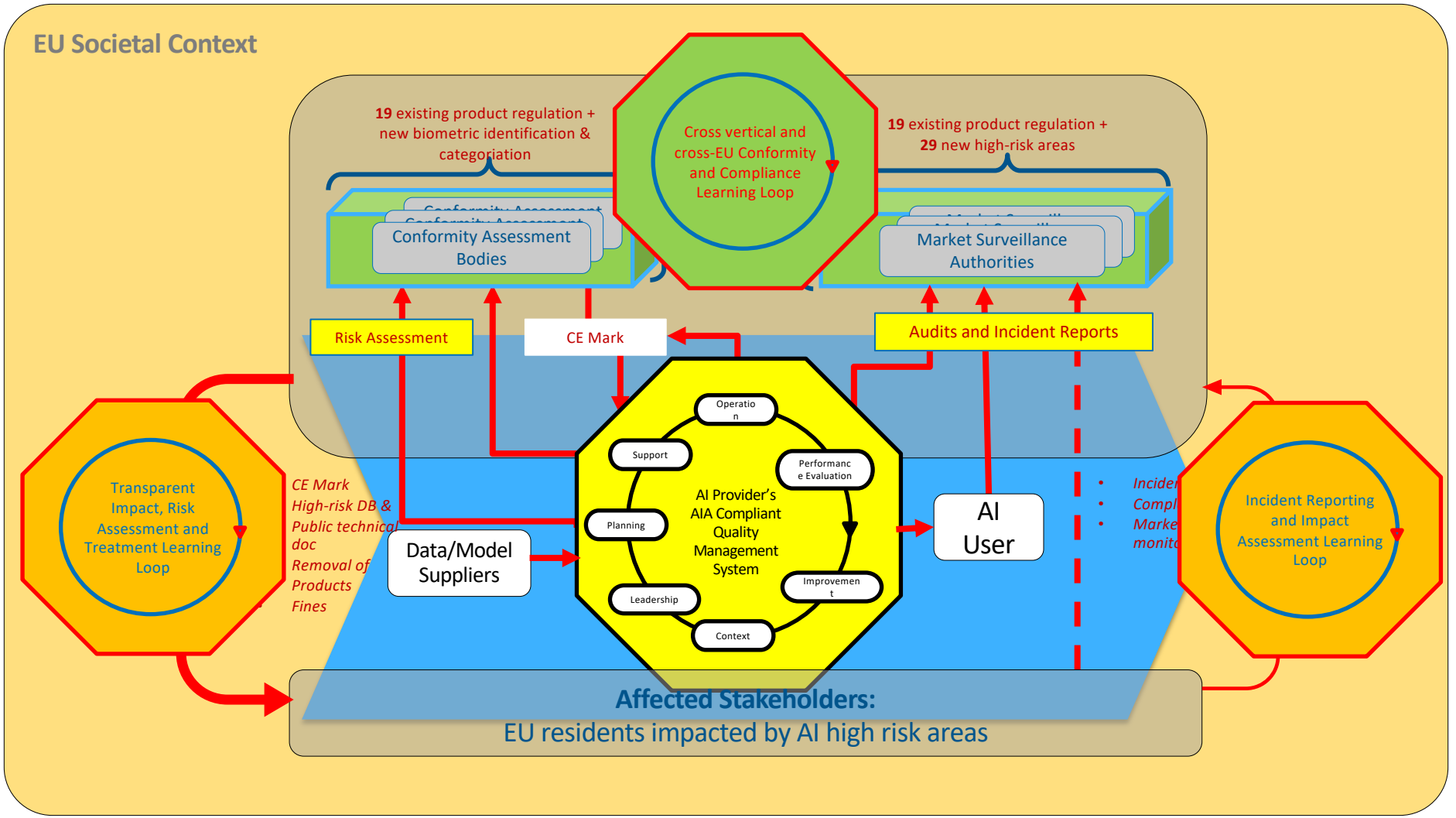
43

Community Involvement and Development

issues in Social Responsibility

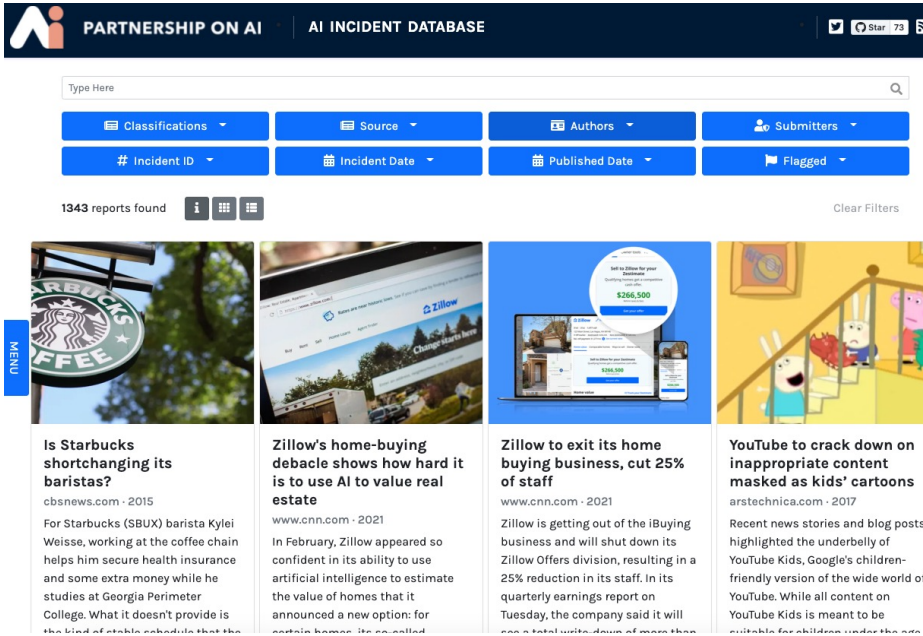
Risks	Mitigations
• Difficulty **identifying communities** suffering negative impact of AI use (including non-users), e.g. social networks, road users, medical patients • Negative impact on **local health, employment and wellbeing**, e.g. deskilling, child development, culture wars • **Concentration of AI's wealth and income creation** away from local communities • Perpetuating **local dependence on philanthropic** activities	• *Consult* with early and widely communities, especially where vulnerable • Be *transparent* on engagement with local authorities • Promote *education and preservation* of local cultures • Support *employment creation and skills development* in impacted communities and along value chain • Direct AI to *solve local* social and environmental issues • Enhance *local scientific and technological* development and entrepreneurship • Promote *economic diversification*, support local suppliers and employment

Trust Building through Multistakeholder Learning for the AI Act



Public Awareness: collecting and collating incident reports

- Move from anecdotal to evidence based assessment of risk
- Incident reporting repositories emerging – based on news stories
  - https://www.aiaaic.org/
  - https://incidentdatabase.ai/
- How to classify and collate incidents from different sources?
- Can AI Act market surveillance incidents contribute to public debate?



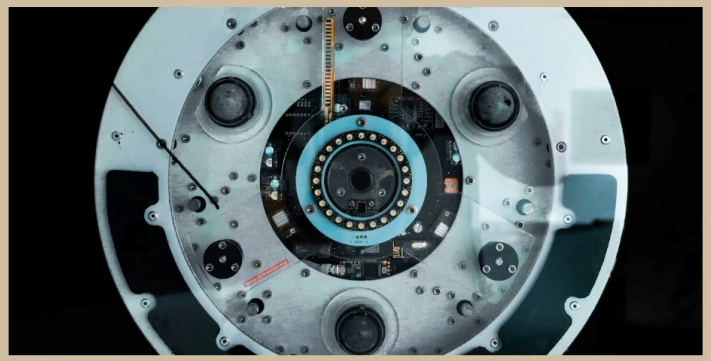
AI Incident Reports

https://www.aiaaic.org/

AIAAIC

- Home
- Projects
- AIAAIC Repository
- Resources
- Get involved
- About AIAAIC

AIAAIC is an independent, non-partisan, public interest initiative that examines and makes the case for real AI, algorithmic, and automation transparency and openness. More



Get to grips with the risks and harms of AI, algorithms, and automation

AIAAIC's independent, free, open library identifies and assesses incidents and controversies driven by and relating to AI, algorithms, and automation.

- About the AIAAIC Repository
- How the AIAAIC Repository is managed
- Classifications and definitions

Latest entries

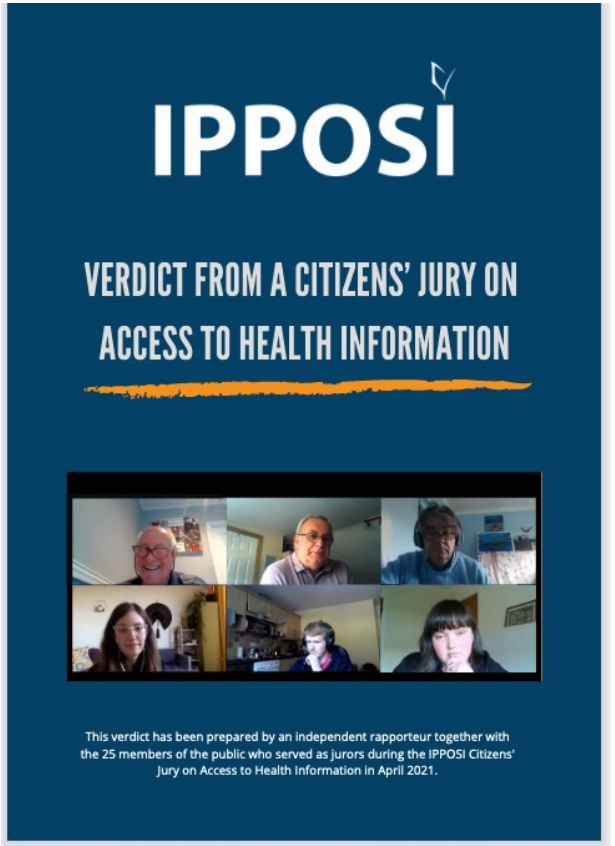
- Inaccurate auto translation denies Pashto-speaking refugee asylum
- Amazon uses AI to generate 'Fallout' series promo art
- Carmel school students attack authorities with deepfakes
- AI image generators accept 85% of election manipulation prompts
- Adobe sells AI-generated Israel-Hamas war image

Updated

- RealPage/YieldStar automated rent-setting
- Books3 AI training dataset
- Sistema de Reconocimiento Facial de Prófugos
- Replika 'encouraged' Queen Elizabeth II assassination

Public Consultation: Mini-Publics to Deliberate Ethical AI Issues

- Recent ‘Citizen Jury’ on Health Information
- National AI Strategy promises Youth Citizen Assembly on AI and an AI Ambassador
- Relevant Structure for AI act?
  - Initiated at a national level?
  - Initiated at a Sectorial level, e.g. education, public services, finance, labour?
- GDPR contained provision for public consultation – but has not materialised



Summary of Trustworthy AI Governance

- Regulation of AI is rapidly approaching
- The EU's proposal relies heavily on standards but has little opportunity for citizen's voice on ethical issues
- Areas that may raise the citizen's voice:
  - Standards for social responsibility
  - Publicly collecting and collating incident reports
  - Integrating public consultation on emerging ethical issues
- Don't forget the data, its governance, and people's control over that.

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

Data and Trustworthy AI: EU Data Governance Act

Role of **Data** in Trustworthy AI and Ethics

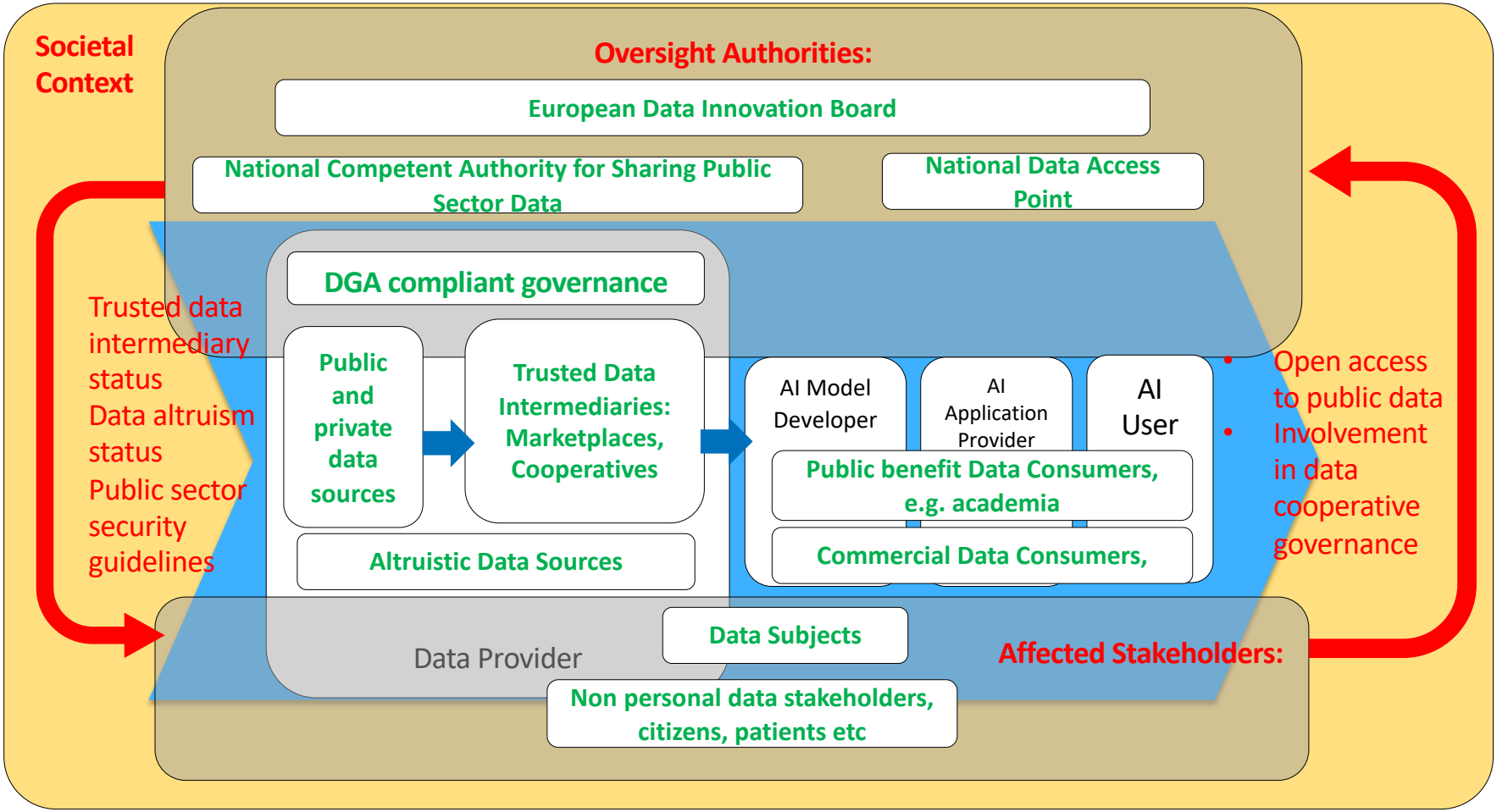
- Problem: for Data, **Possession** is 9.9/10^{ths} of the Law
- Power of AI-driven digital engagement (and then it potential power over us) grows with the volume (and quality) of its training data
- **Controlling the Flow of Data** is the Key to Governing AI

Data Governance

- Regulation can improve Transparency and Accountability of data handling in organisation, but can it do enough to maintain legitimacy?
- Regulations are highly technical and suffer pacing problem – how can Democratic oversight and control be exercised?
- Concentration of power over data, how can this be decentralised?

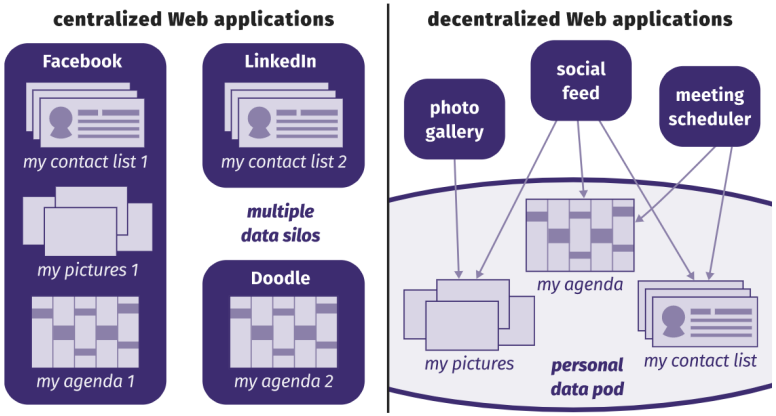
EU Data Governance Act

- Governance measures to improve trust and confidence in data sharing
- Reduce barriers to sharing personal and non-personal data, improve reuse



Personal Online DataStores

- Edge Platforms emerging for **maintaining possession** of Personal Data:
  - Inrupt.com (based on solid.org)
  - hubofallthings.com
- Software to keep your data in a store you control - “personal data pod”
- Companies that want data have to agree to **your** T&Cs rather than you agreeing to theirs



Pros:

- Can consistently enforce your preferences for sharing data
- Can more readily rescind/renegotiate access
- Pool understanding of requests for your data
- No centralized data store to attract hackers

Cons:

- Requires some effort to monitor and control own data and set T&Cs
- Trustworthiness and legal basis of “Pod” provider – EU Data Governance Act - certified trusted intermediary

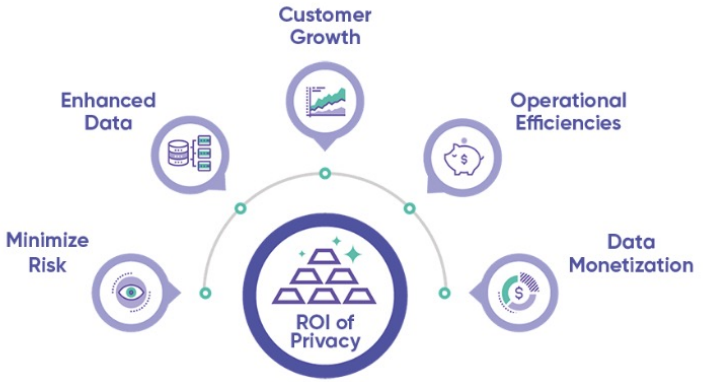
Could new patterns of Data Stewardship Help?

Organizations could already transfer governance responsibility to more representative groups:

- Data Unions – Data as Labour – https://blog.singularitynet.io
- Data Trusts
- Data Co-ops

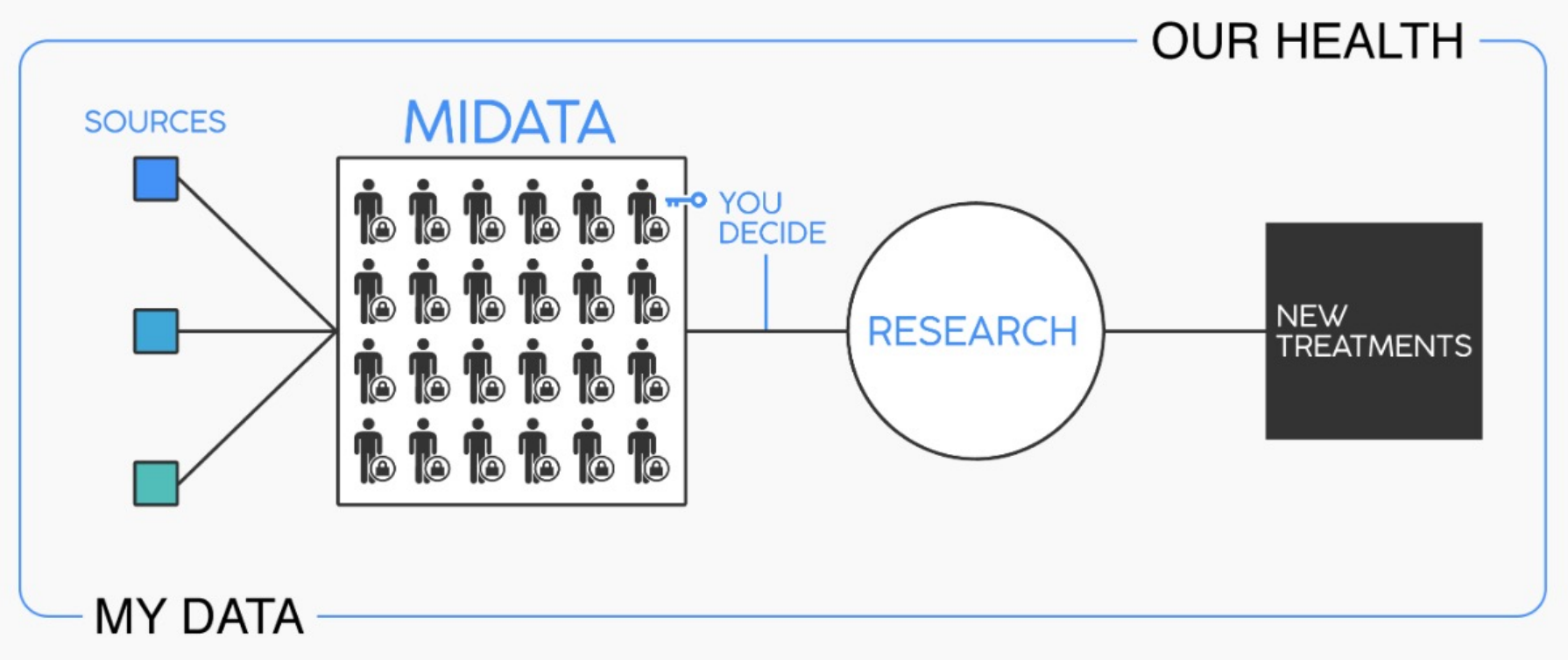
Example: Data Trusts

- Truata.com: Anonymised Data Analytics Services



- Offers large clients secure, anonymised analytics services of their own data
- Outsources data protection risks without loss of benefits from data analytics
- Part of business model is a Data Trust - constituted separately to the business/profit driven part of the company
- Data Trust gives clients (and their customers) confidence that the rules can't change for business reasons
- Other examples: https://theodi.org/article/odi-data-trusts-report/

Example: MIDATA Medical Data Coop



- Control of Data from hospitals and medical studies handed to MIDATA
- Operates as a **cooperative** in the interest of its members medical data subjects
- Management appointed and operated under **democratic principles**

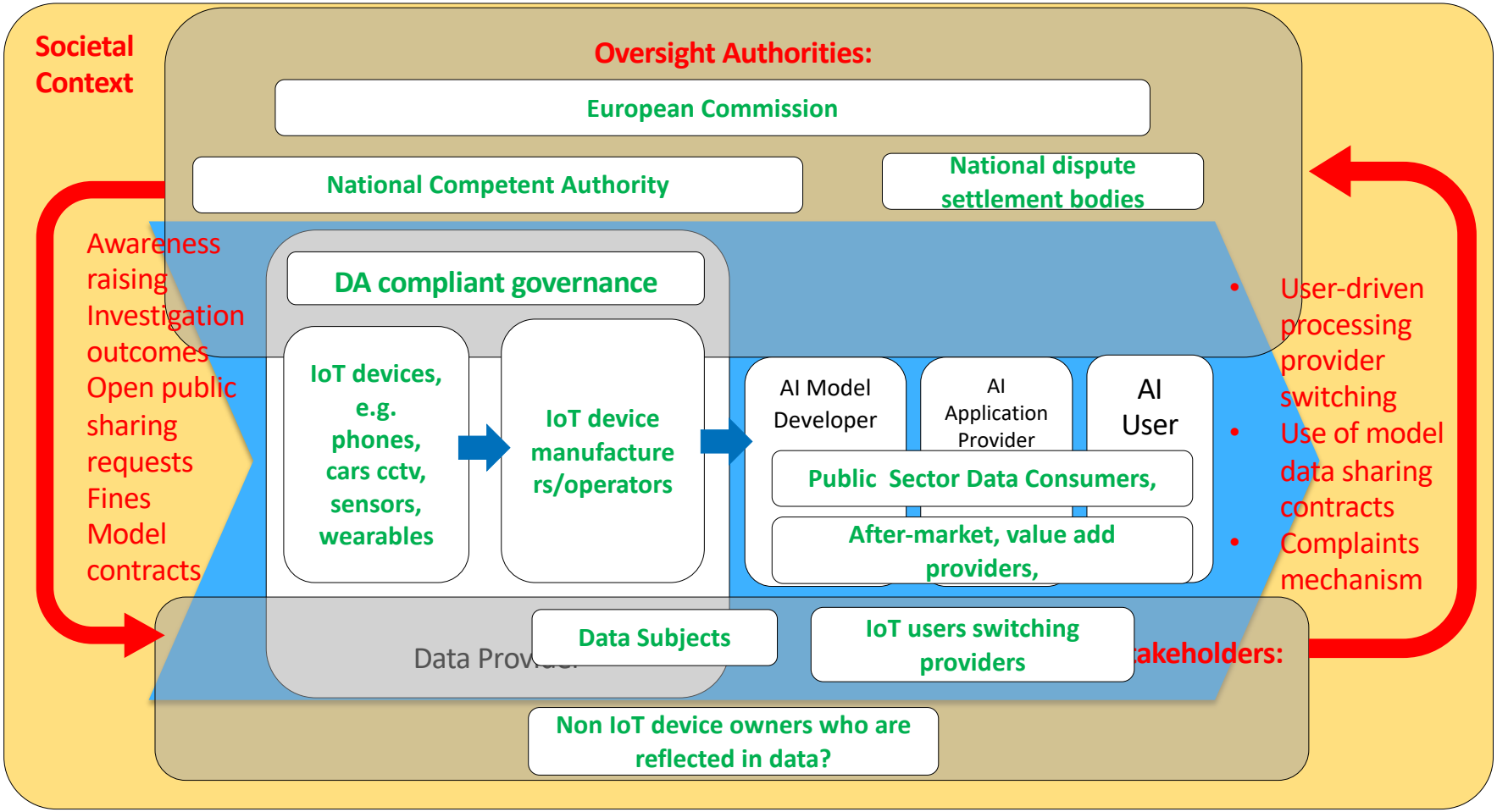
https://www.midata.coop

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

Other EU Digital Legislation

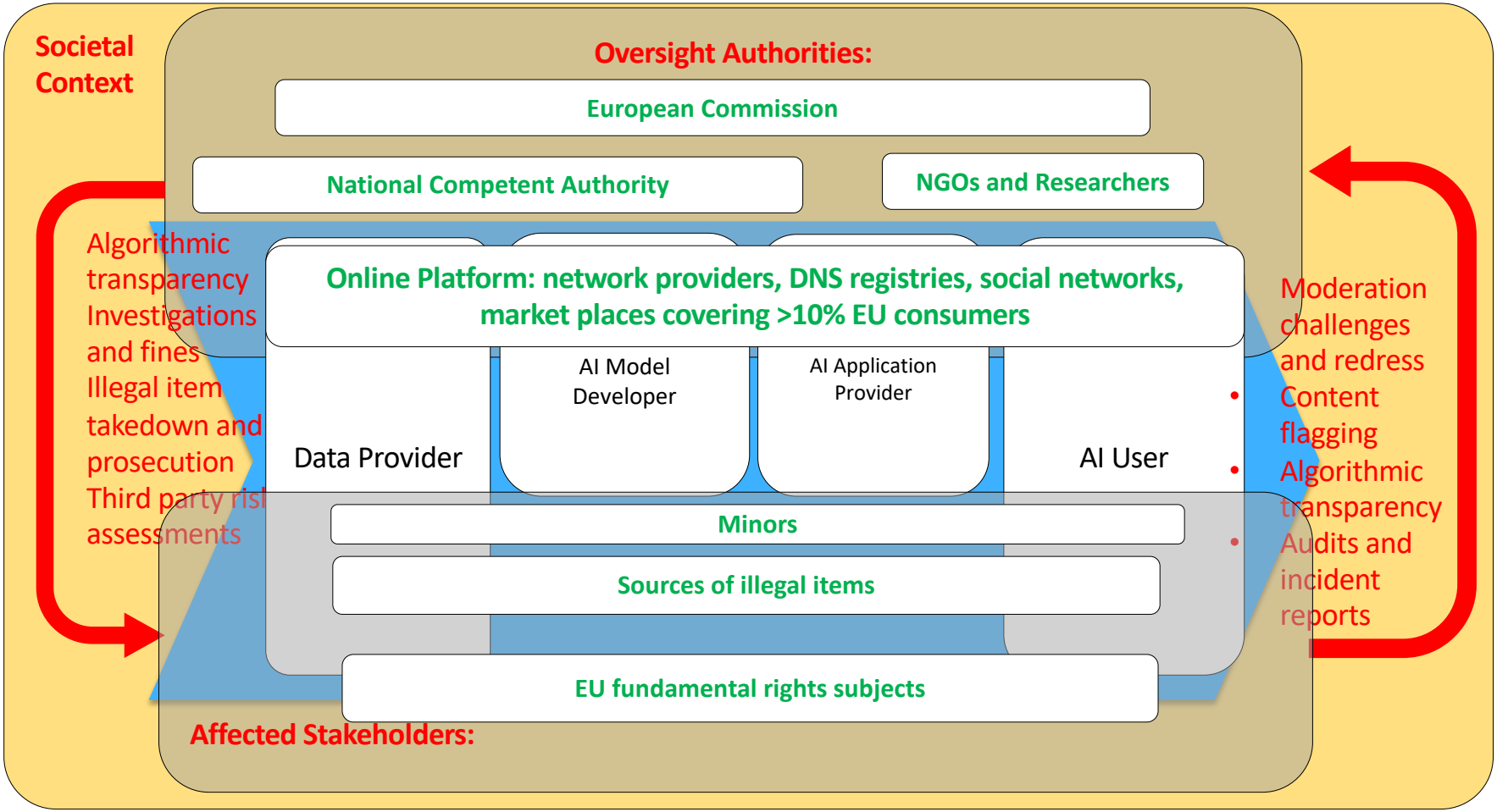
EU Data Act

- Rules for more open data sharing from IoT devices
- Access for users, fairer terms for SMEs and public sector – improve competitiveness in markets



EU Digital Services Act

- Safe and Accountable online environment
- Illegal items, content moderation, risk assessment, protection of minors



EU Digital Market Act

- Applies to Gatekeepers, e.g. who control online market places, operating systems, cloud services, search engines
- Addresses unfair practices to improve digital market access and competition
- Rules on unfair access terms, promotion of own products/services over others

Conclusions

- Critical thinking about the Governance of AI is still in its infancy
- Understanding risks for stakeholder is key - Assignment
- Future approaches may need:
  - Transparency and stakeholder engagement on risk assessment
  - Collection and sharing incident reports
  - More decentralised approaches to data governance
- Regulation is expanding, but need to critically observe design and enforcement

References

- Principled Artificial Intelligence: Mapping Consensus in Ethical and Right-based Approaches to Principles for AI, Jessica Fjeld, Adamn Nagy, Berkman Klein Centre, Jan 15, 2020, https://cyber.harvard.edu/publication/2020/principled-ai
- AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Floridi, L., Cowls, J., Beltrametti, M. et al. Minds & Machines (2018) 28: 689. https://doi.org/10.1007/s11023-018-9482-5
- Gasser, Urs, and Virgilio A.F. Almeida. 2017. “A Layered Model for AI Governance.” IEEE Internet Computing 21 (6) (November): 58–62.
- Hagendorff, T., (2019) "The Ethics of AI Ethics – An Evaluation of Guidelines", https://arxiv.org/pdf/1903.03425.pdf
- Adamson, G., Havens, J. C., & Chatila, R. (2019). "Designing a Value-Driven Future for Ethical Autonomous and Intelligent Systems". Proceedings of the IEEE, 107(3), 518–525. https://doi.org/10.1109/JPROC.2018.2884923
- Leenders, G. (2019). "The Regulation of Artificial Intelligence — A Case Study of the Partnership on AI". Medium, April, Retrieved from: https://becominghuman.ai/the-regulation-of-artificial-intelligence-a-case-study-of-the-partnership-on-ai-c1c22526c19f
- EU High Level Expert Group on AI, (2019). "Ethics Guidelines for Trustworthy AI", April 2019, Retrieved from https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines
- "Mapping regulatory proposals for artificial intelligence in Europe", Access Now, Nov 2018. Retrieved from https://www.accessnow.org/mapping-regulatory-proposals-AI-in-EU
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé, H., Crawford, K. (2018) "Data sheets for Datasets", in Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning, Stockholm, Sweden, 2018, available at: https://arxiv.org/abs/1803.09010
- Buchanan, B., Miller, T. (2017) "Machine Learning for Policy Makers - What it is and Why it matters" Belfer Centre, available from: https://www.belfercenter.org/sites/default/files/files/publication/MachineLearningforPolicymakers.pdf
- Joshua A. Kroll , Joanna Huey , Solon Barocas , Edward W. Felten , Joel R. Reidenberg , David G. Robinson & Harlan Yu, "Accountable Algorithms", 165 U. Pa. L. Rev. 633 (2017). Available at: https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3
- Calo, Ryan, "Artificial Intelligence Policy: A Primer and Roadmap" (August 8, 2017). Available at SSRN: https://ssrn.com/abstract=3015350 or http://dx.doi.org/10.2139/ssrn.3015350
- Future of Life Institute. (2017). "Asilomar AI Principles". Future of Life Institute. Retrieved from https://futureoflife.org/ai-principles/
- Scherer, Matthew U., "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (May 30, 2015). Harvard Journal of Law & Technology, Vol. 29, No. 2, Spring 2016. Available at SSRN: https://ssrn.com/abstract=2609777 or http://dx.doi.org/10.2139/ssrn.2609777
- Saurwein, Florian and Just, Natascha and Latzer, Michael, "Governance of Algorithms: Options and Limitations" (July 14, 2015). info, Vol. 17 No. 6, pp. 35-49, 2015. Available at SSRN: https://ssrn.com/abstract=2710400

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

Thank You