AI and privacy risk governance

Sakthi Thangavelu
Principal Consultant – Data Privacy & AI Ethics
Ethically.in

By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.

Eliezer Yudkowsky

Privacy professional's imperative (and the agenda for today)

Get broader perspective of AI systems

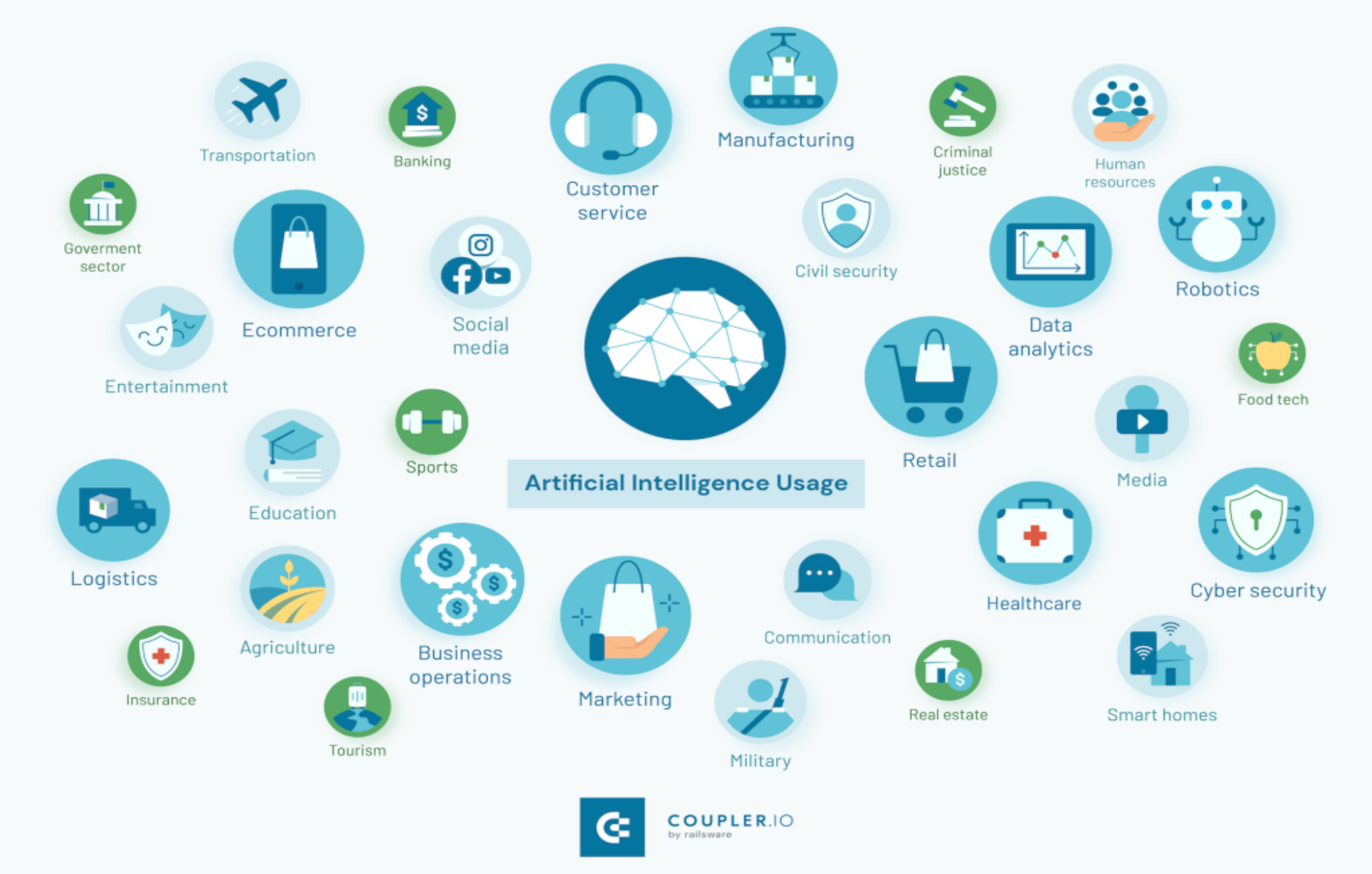
Understanding of AI risks

Understand AI governance structure

Catch up AI regulatory developments

Enhance the role of privacy function in tackling AI risks

AI is everywhere



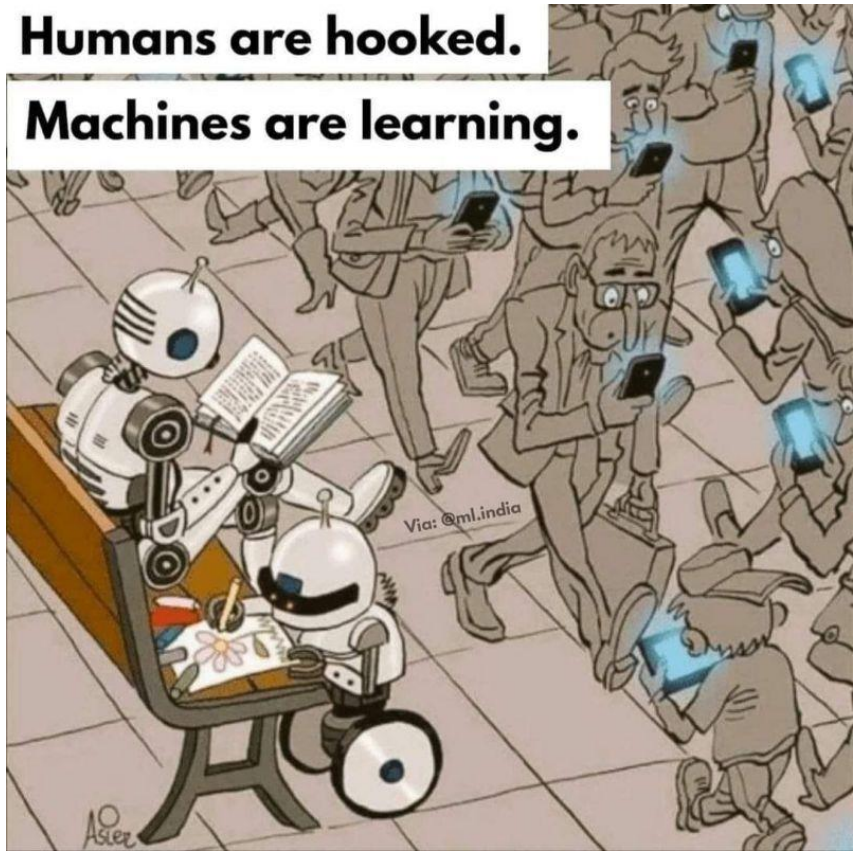
“Personal Data” in AIML systems

Artificial Intelligence is a collective term for computer systems that can sense their environment, think, learn, and take action in response to what they are sensing and their objectives.

FORMS OF USE TODAY

Automated intelligence (RPA, Intelligent automation)	Resume filtering
Assisted intelligence (aids decision making)	Chatbot
Autonomous intelligence (makes decisions)	Self driving cars

Machine learning types



Machines that learn from data

Underpinning many advances in artificial intelligence is machine learning. There are three forms of machine learning:

Reinforcement learning

The system will take action in a particular environment and assess whether those actions help to achieve its goals.

Those actions that lead to the best outcomes will be prioritised and thus the machine learns how best to achieve its goals.

Supervised learning

The system is given a block of training data containing both inputs and desired outputs.

It uses that information to "learn" how to complete the relevant task.

The system will then be tested against a separate block of testing data to confirm that it is generating the correct outputs.

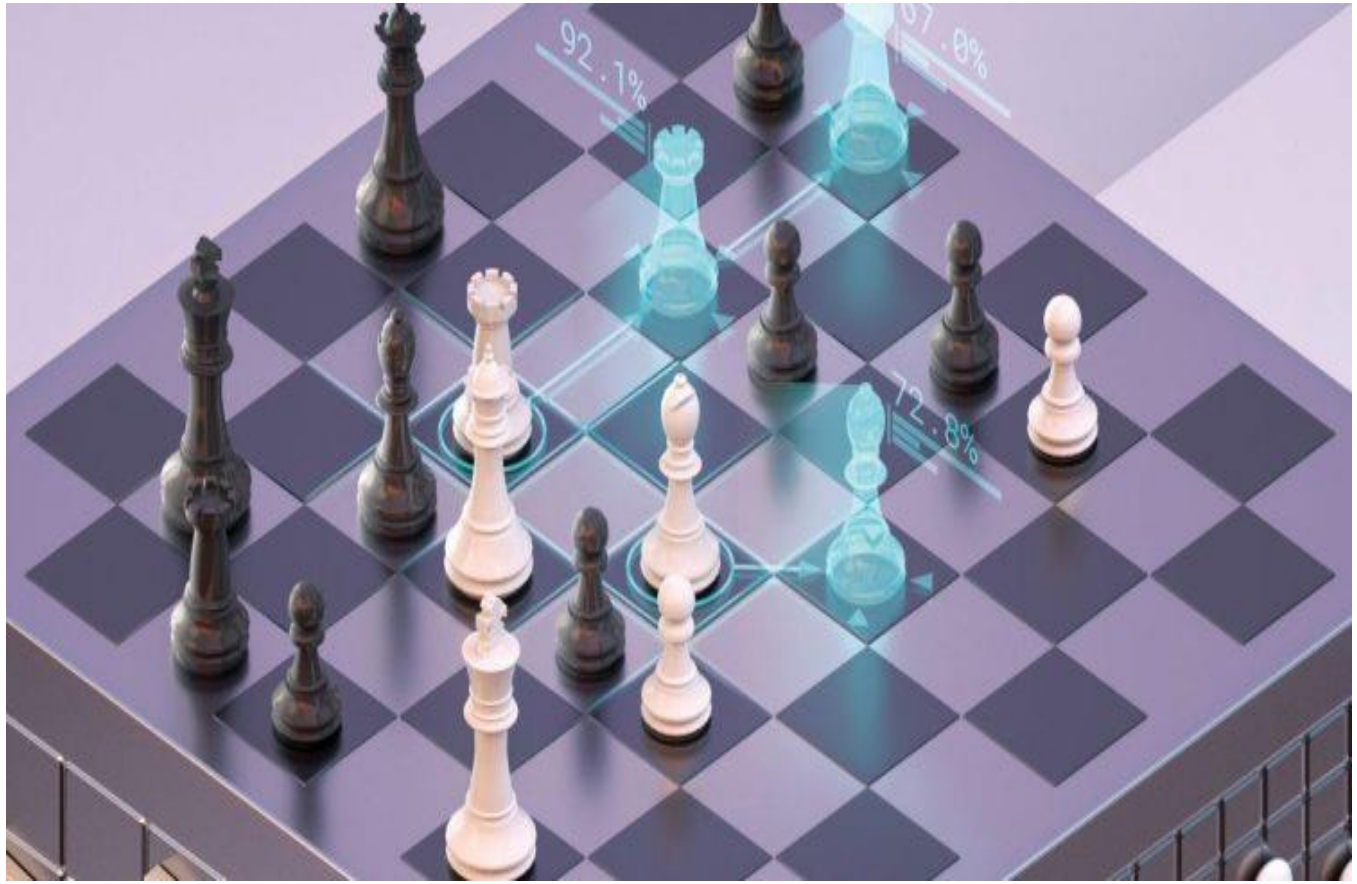
Unsupervised learning

With unsupervised learning, the system is given a block of data that has not been labelled (e.g. classified or categorised).

Since the data is not labelled it is not possible to ensure specific outcomes but it may still be possible to analyse the data to spot clusters or groupings.

Credit: Linklaters

A case of reinforcement learning



AlphaZero, the gameplaying system created by DeepMind, which was tasked with becoming a champion chess player.

It started with details of the rules of chess but no information about chess strategy, such as what constituted a good position or move.

To learn, it played itself around a billion times, using the data from those games for reinforcement learning – i.e. to identify what constitutes a good game state and strategy.

The Challenge

Our job is to support and ensure trust in AI systems through proactive identification and mitigation of privacy risks, but…

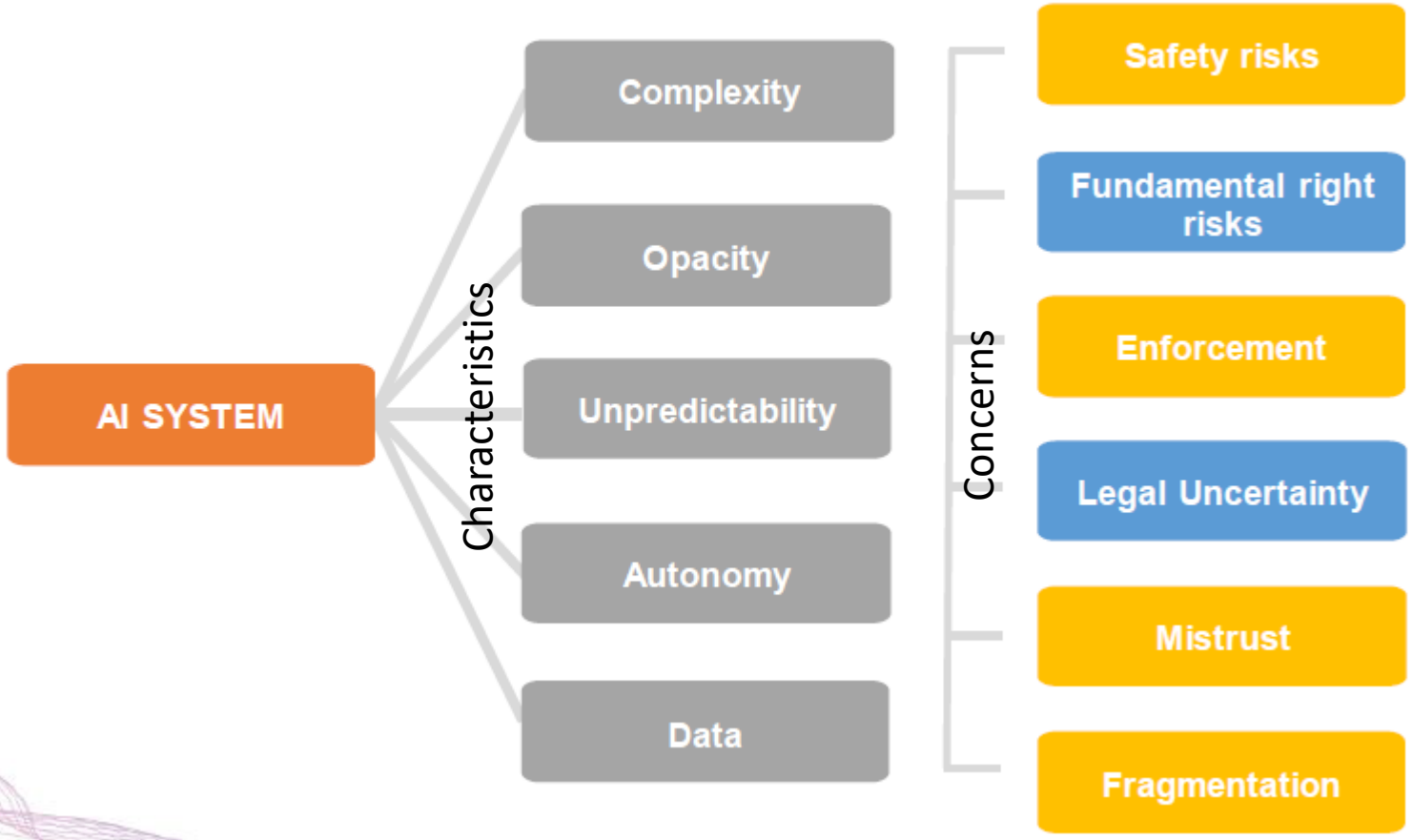
How do we ensure transparency in AI?

How do we ensure lawfulness in AI?

What do we need to know about data accuracy and statistical accuracy?

How do we ensure no bias or discrimination to individuals injected through AI?

Why do we regulate AI use cases?



AI systems not just create privacy risks, but there are concerns across every sector

Employment laws
Competition laws
Intellectual property laws
.....

Credit: European Commission

AI systems from GDPR lens

- AI is not new, GDPR is compatible with AI systems in privacy risk governance.
- Lawfulness, Fairness, Transparency, Data subject rights, Accuracy, Security … all applies to AI systems too.
- PIAs, LIAs, DPIAs – do a real check on AI systems.
- Note: Privacy risks may arise even without involvement of personal data

  E.g. A wrongly written rule or a model makes decisions leading to discrimination of individuals

The GDPR contains controls on the use of automated decision making, i.e.:
" a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

Guidance from regulators suggests that this will include a range of different activities, such as deciding on loan applications or changing credit card limits.

Automated decision making is only permitted in the following situations:

- Human involvement
- Consent
- Performance of a contract
- Authorized by law

“Personal Data” in AI/ML systems

- Structured data(database, files)
- Unstructured data (text, audio, video, images)

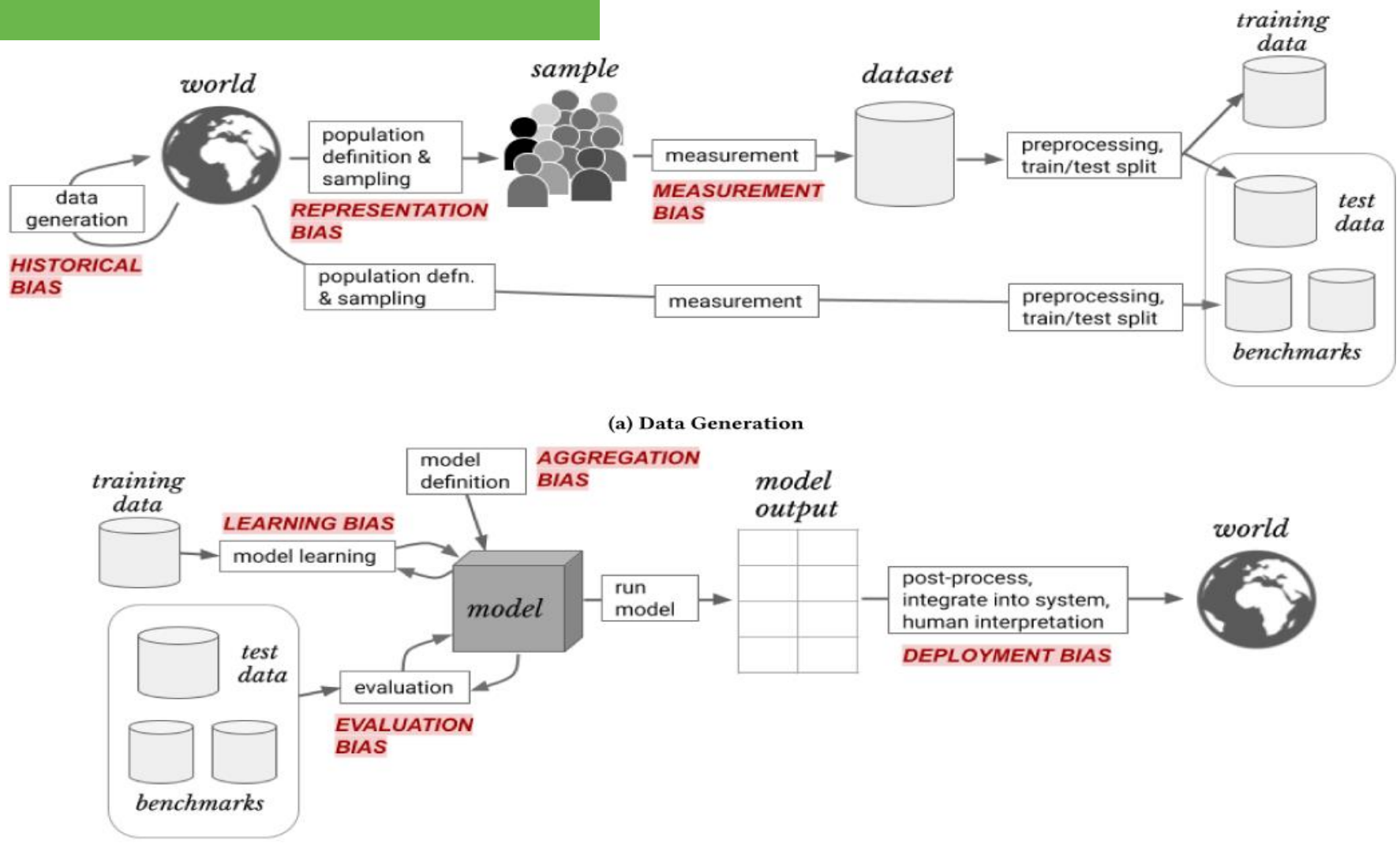
AI Life cycle

Training data to develop models

Output data
Output reveals identity of individuals

Input data
Real data fed into AI systems

AI Bias – an introduction



https://dl.acm.org/doi/fullHtml/10.1145/3465416.3483305

AI regulatory developments

Laws & regulation

- EU AI Act (draft)
- US Algorithmic accountability act (proposed)
- Sectoral laws (e.g. employment, healthcare, education etc)

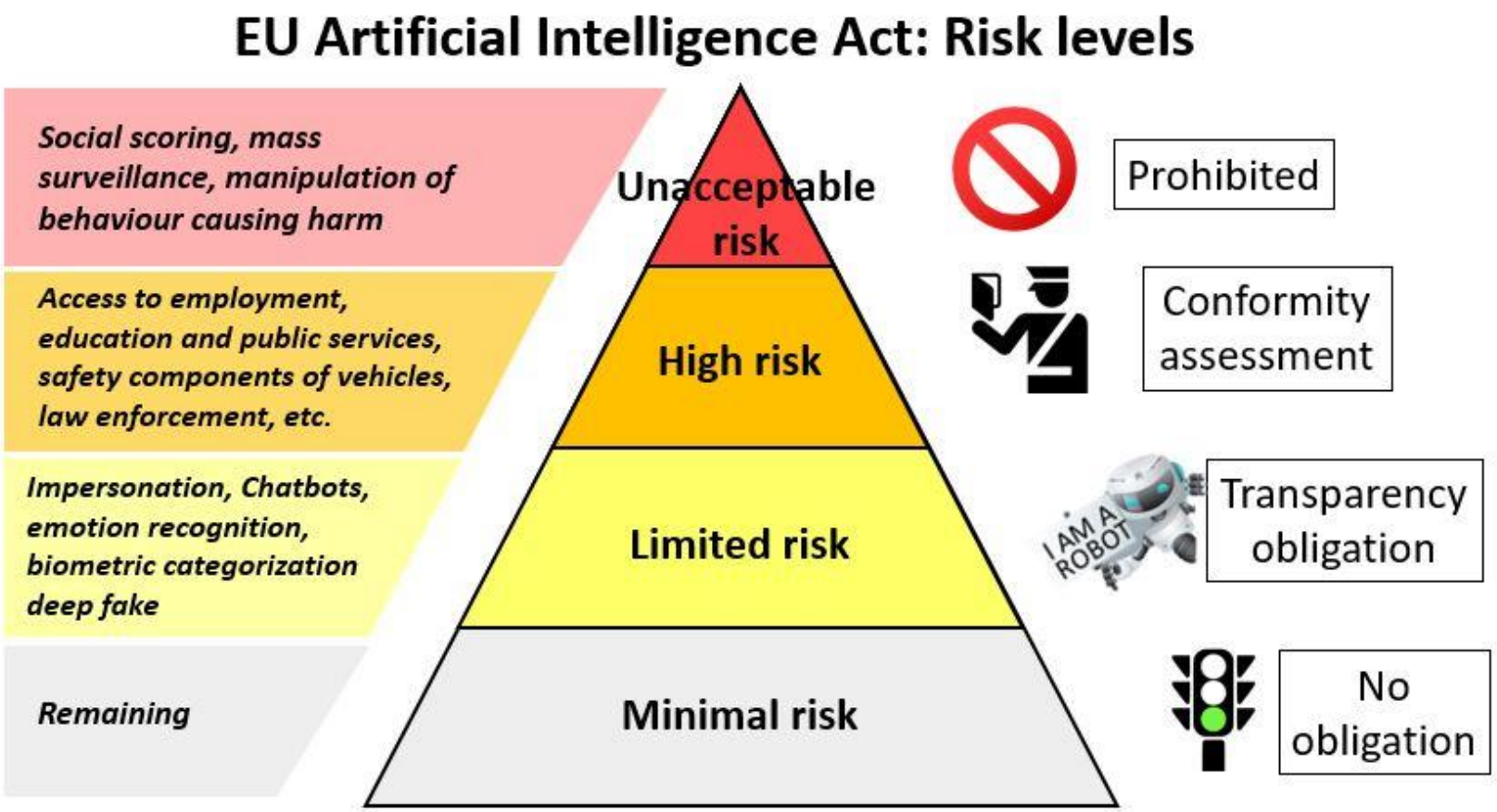
Guidance from authorities

- ICO guidance
- AEPD guidance
- CNIL guidance
- FTC guidance

Self regulation frameworks

- Multilateral & national
  - OECD AI principles
  - UK Ethics framework
  - UNESCO Trustworthy AI
- Private sector
  - Google, IBM, Microsoft – Trustworthy AI, Responsible AI principles
- Technical standards
  - NIST, ISO/IEC 23053:2022

Risk levels @ EU AI Act (draft)



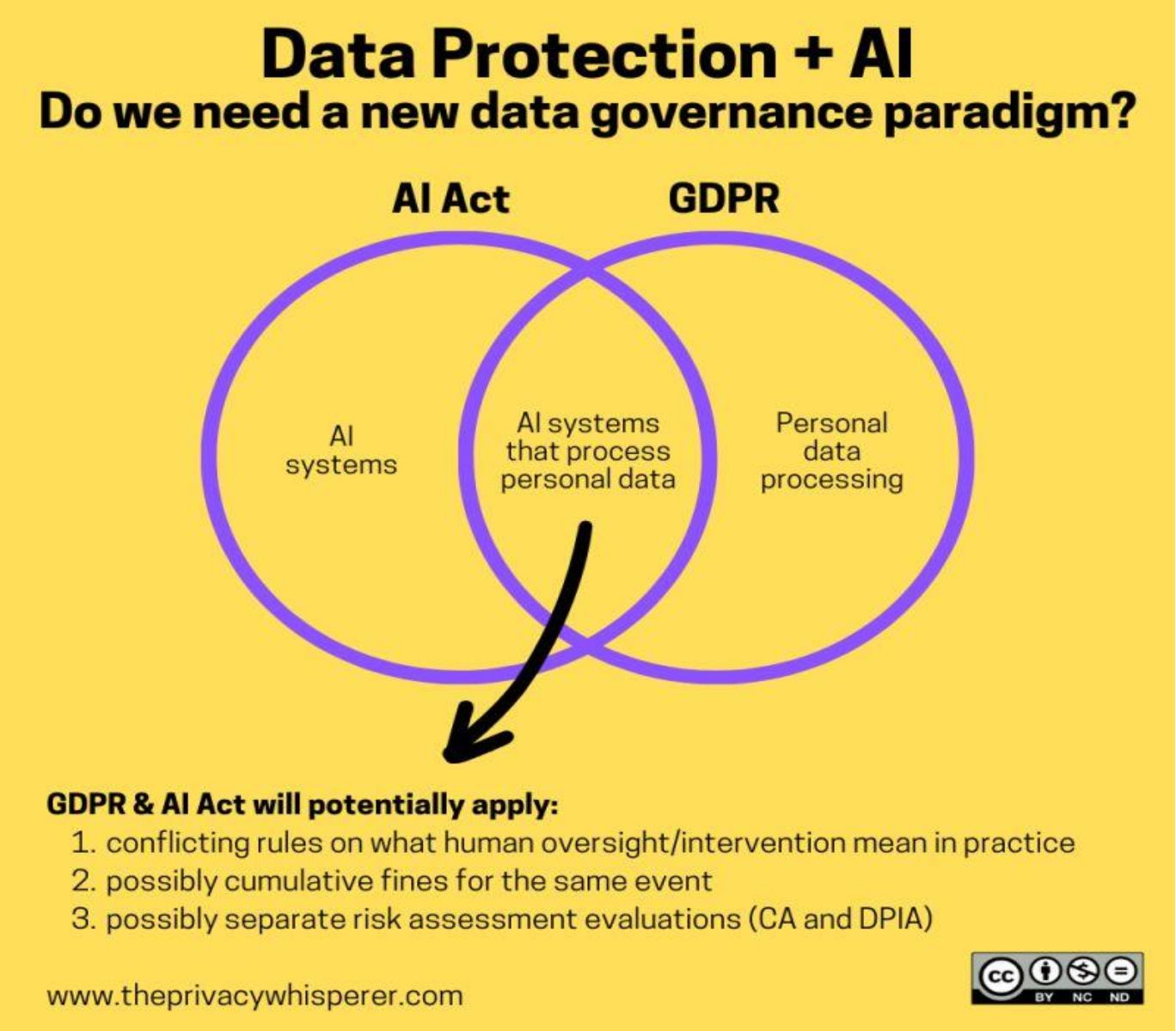
https://www.telefonica.com/en/communication-room/blog/a-fit-for-purpose-and-borderless-european-artificial-intelligence-regulation/

AI Risks and regulations

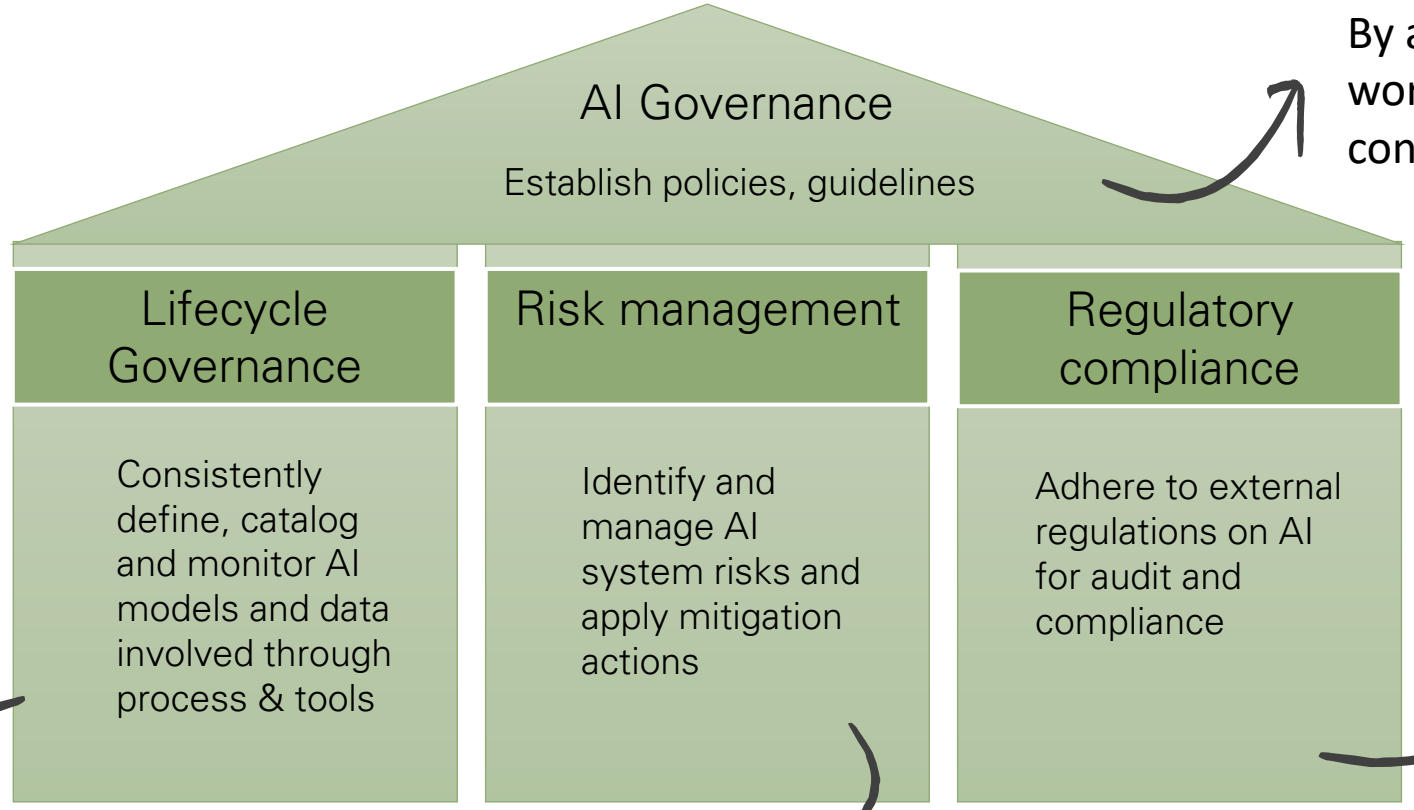
EU AI Act (draft) and GDPR are separate regulations

While Privacy office oversees GDPR compliance, which function in the org is responsible for AI regulatory compliance?

Who in the enterprise is working on AI governance?



3 Pillars of AI risk governance



By a task force or working group or consultants

By engineering teams involving Data scientists, AIML Engineers

By Legal, compliance functions

By Risk functions (security, privacy, business risks)

Global alignment on Trustworthy AI principles

UNESCO AI Principles	OECD AI Principles	EU AI Act Draft	Ethical principles of AI Singapore PDPC	NIST AI RMF	Ethical principles of AI Hong Kong PCPD
Safety and security	Robustness, security and safety	Robustness, Validation Safety, Security Cyber Security Resilient		Secure Resilient Safe	Reliability, Robustness and Security
Fairness and non-discrimination	Human-centered values and fairness	Non-discrimination	Fairness Inclusivity	Fair with harmful bias managed	Fairness
Sustainability	Inclusive growth, sustainable development and well-being				
Right to Privacy, and Data Protection		Privacy preserving measures	Human rights alignment	Privacy enhanced	Data Privacy
Human oversight and determination			Human Centricity and Well-being		Human Oversight
Transparency and explainability	Transparency and explainability	Transparency Explainable Interpretability	Auditability Explainability	Explainable Interpretable Transparent	Transparency Interpretability
Responsibility and accountability	Accountability	Accountability	Accountability	Accountable	Accountability
Awareness and literacy					
Multi-stakeholder and adaptive governance and collaboration			Progressiveness		
		Accuracy	Accuracy	Valid and reliable (accuracy and robustness)	
Proportionality and Do No Harm					Beneficial AI

Resource recommendation – OECD.AI

OECD.AI Policy Observatory

OECD.org | Going Digital Toolkit | EN

Blog | Experts | AI Principles | Policy areas | Trends & data | Tools & metrics | Countries | About

Policies, data and analysis for trustworthy artificial intelligence

Webcast

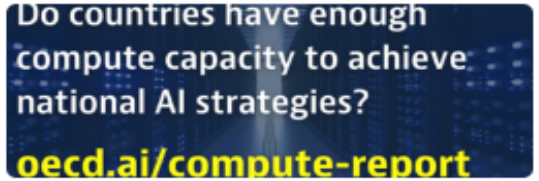
Expert forum on AI foresight and generative AI

On 19 Arpil, the OECD held two workshops on AI foresight and generative AI.

Contribute

Contribute to our Catalogue Tools & Metrics for Trustworthy AI

Do you know a tool or metric to help make AI trustworthy? Promote it on the Catalogue of Tools and Metrics for Trustworthy AI. You can also give feedback on one you have used.



Report

A blueprint for building national compute capacity

Countries need data and targeted plans for national AI compute capacity.

Priority projects

- Programme: AI in Work, Innovation, Productivity and Skills
- AI compute capacity
- Tools for trustworthy AI
- National AI policies
- Classification of AI Systems

Resource recommendation – UK ICO Toolkit

ICO - AI and Data Protection Risk Toolkit Free download @ https://ico.org.uk

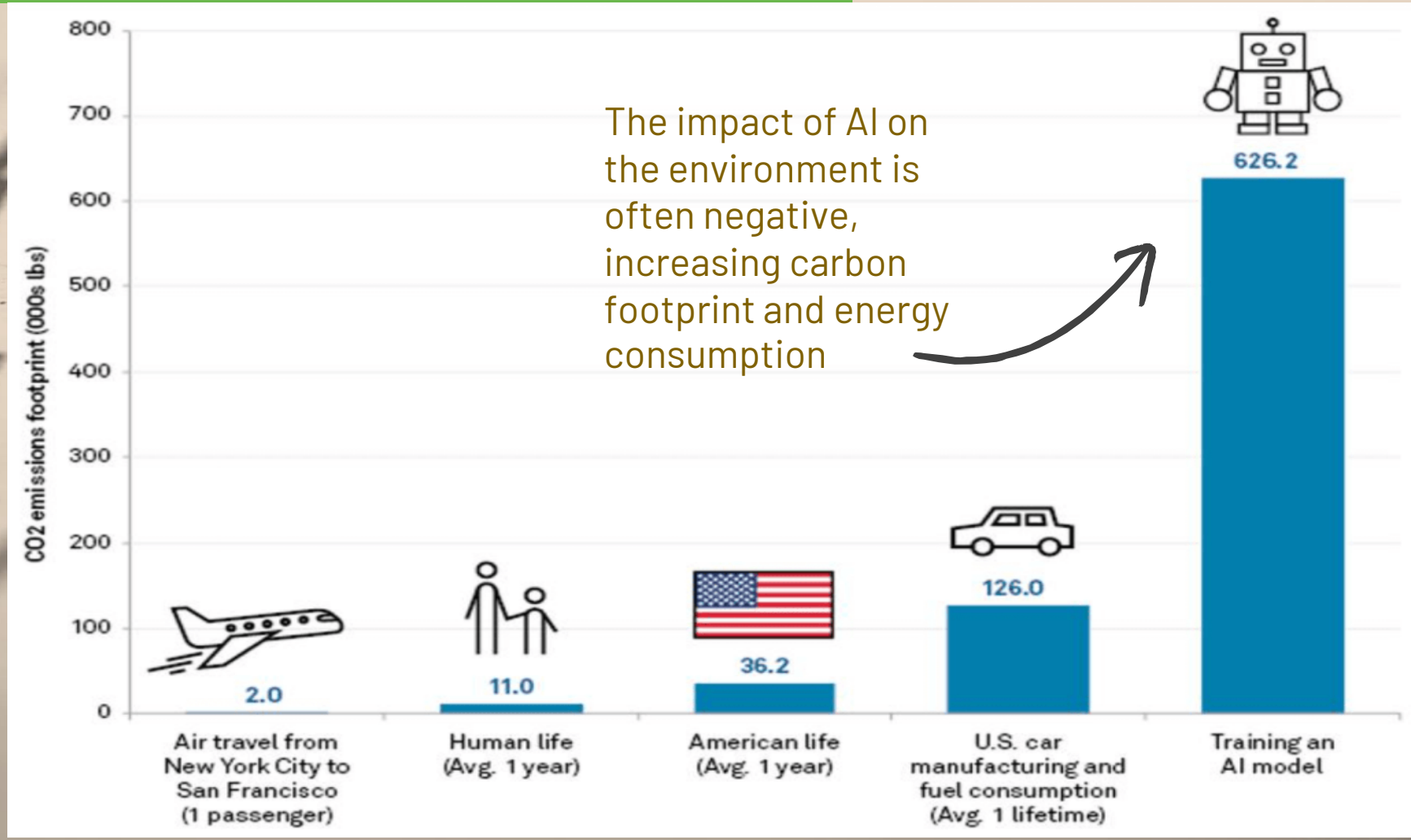
By AI lifecycle stages

Risk statements catalogue

Controls catalogue

AI Lifecycle Stage	ID	Risk area UK GDPR Reference	Data Protection Risk Statement	Risk Assessment Summary	Inherent Risk Rating	Control	Contr
Business requirements and design		Accountability Articles 5(2), 35 and 36 and Recitals 74-77, 84, 89-92, 94 and 95.	The misidentification of risks to individual rights and freedoms caused by not carrying out a risk assessment. As a consequence, an organisation cannot put in place appropriate technical and organisational measures to prevent harms occurring to individuals.			Conduct a data protection impact assessment (DPIA)	To identify risk appropriate te organisational them.

Sustainable AI



Datacenters are running on 100% renewable energy

Carbon footprint of Training models – will it beat the efficiency of human brain which just runs at 20 watts?

Source: Forbes (based on 2019 data compiled by The University of Massachusetts Amherst)

In summary

1. Enhance your AI knowledge to effectively interact with stakeholders
2. Understand the Data protection authorities guidance on AI, Follow the terms Responsible AI, Trustworthy AI, Ethical AI
3. Understand what your AI tech teams, Data science teams do on data governance
4. Partner with your Enterprise risk management teams and establish standards on AI risks

   Understand if/how IT teams do cataloging of AI systems
5. Expand the PIA process to support your company AI policy and DPA guidance


iapp
Thank You!