




# Take a seat we'll be starting shortly





**Dr. Maria Moloney**  
Senior Researcher  
and Consultant  
PrivacyEngine

**WEBINAR**

**AI and Privacy:  
Navigating Data  
Protection for DPOs  
in the Age of AI**

Friday 8th March 2024  
12:00 GMT



**Nollag Conneely**  
Head of Consultancy  
PrivacyEngine

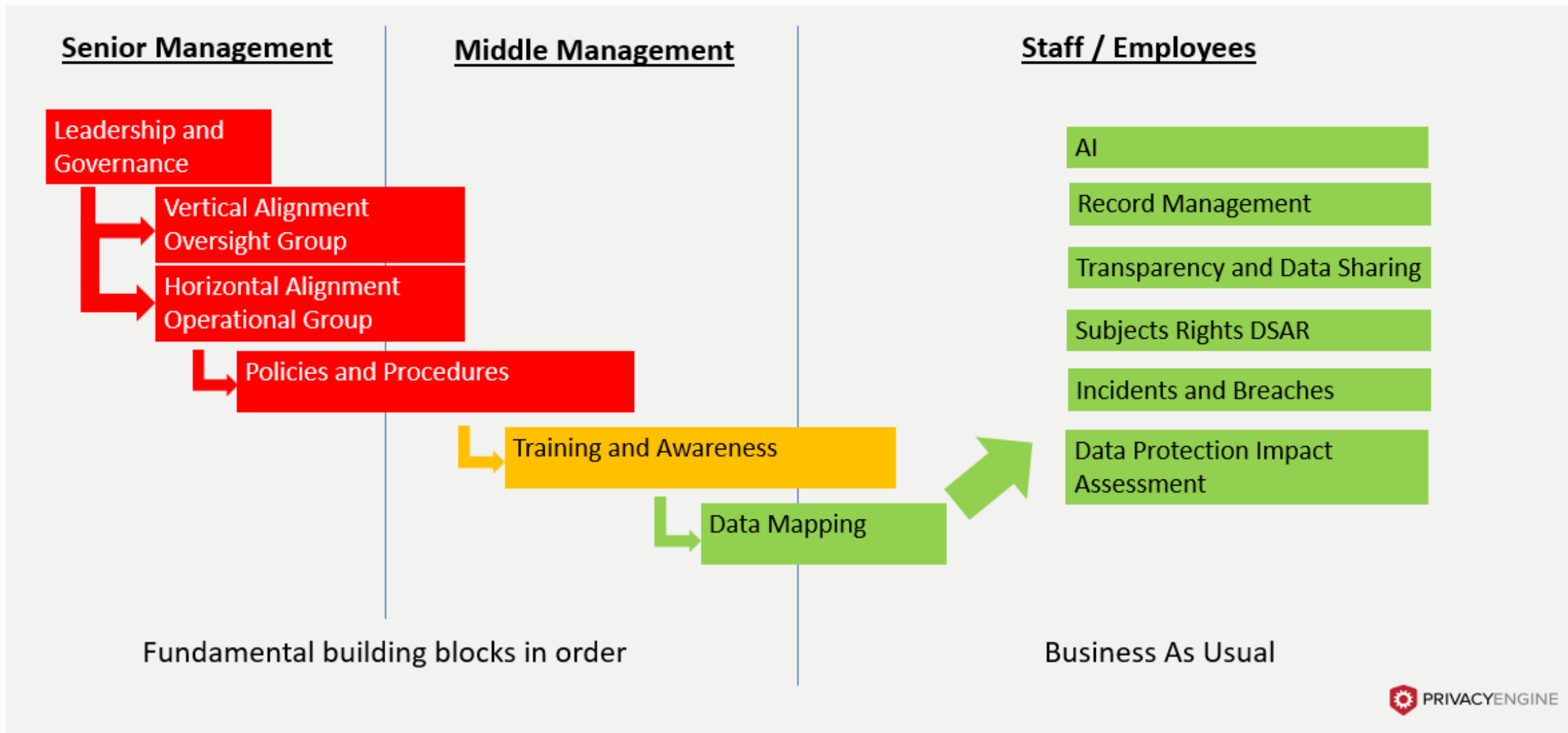
# Introduction

## ‘AI Exceptionalism’

*Treating AI as so different and special that we fail to see how the privacy problems with AI are the same as existing privacy problems, just enhanced. AI represents a future for privacy that has been anticipated for a long time.*

*Professor Daniel J Solove*

# Framework



# Agenda points

- What is AI
  - Definitions
  - Developer/Deployer
  - Risk levels
  - Scraping
- AI Act
- Stages of AI
- Case Studies and Red Flags

# What is an AI System?

- The EU AI Act defines an AI system as:
- *"a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."* [EU AI Act, Article 3(1)]
- *The OECD:*
- *An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems in their levels of autonomy and adaptiveness after deployment*

# What is General Purpose AI?

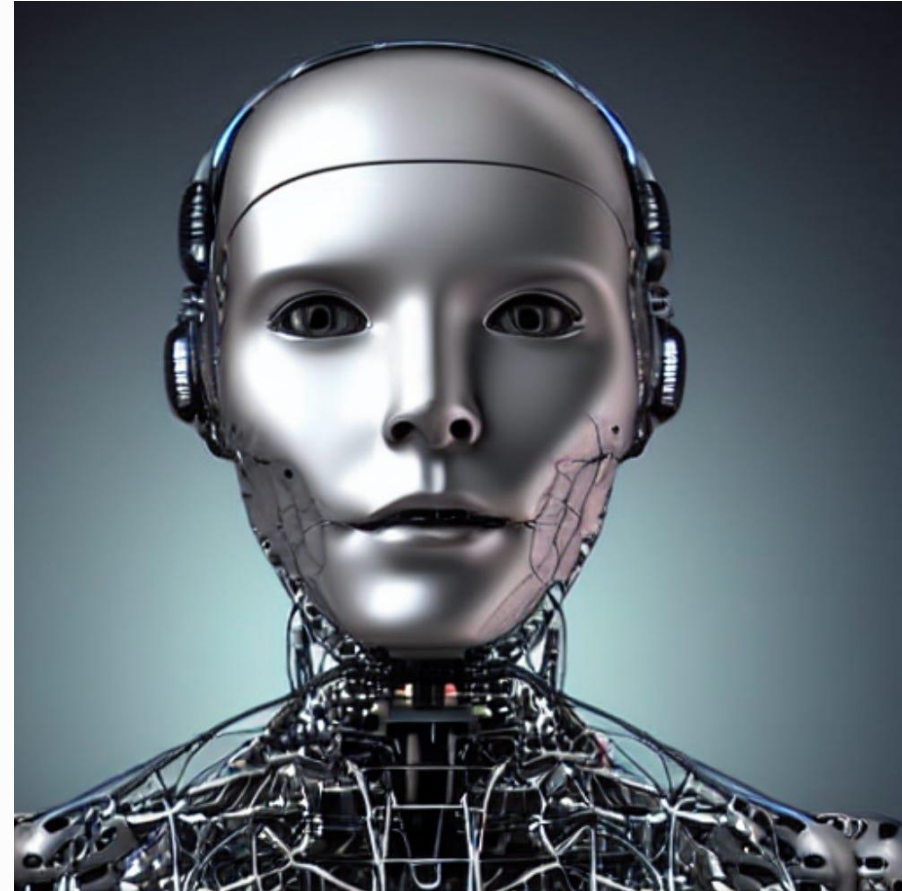
- A General Purpose AI (GPAI) system is defined as an AI system that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model was originally released on the market.
- These AI systems can be integrated into a variety of downstream systems or applications. GPAI models are also often referred to as Foundation Models. They have purposefully been designed to perform a wide range of tasks and to easily adapt to new situations.
- They are trained on very broad sets of unlabelled data and can be used for many different tasks without much fine-tuning.

# GPAI Model Protections: Two-Tier Risk Model

- **Tier 1:** several uniformed obligations for all GPAI models.
- **Tier 2:** an additional set of obligations for GPAI models with systemic risks.

A GPAI model falls in the category entailing systemic risks if it has high impact capabilities evaluated on benchmarks or per decision by the AI Office.

The Commission has the authority to adopt delegated acts to amend the thresholds and to supplement benchmarks and indicators in response to evolving technological developments, such as advancements in algorithms or improved hardware efficiency.



# Developer and Deployer – what's the difference?

- An AI Developer is typically the entity that creates and develops the AI system with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge.
- An AI Deployer is the entity that uses the AI system in the context of professional activity. They may use APIs to embed AI products within their own products or may simply use AI systems as internal tools.



# Developer & Deployer – High-Risk AI Systems

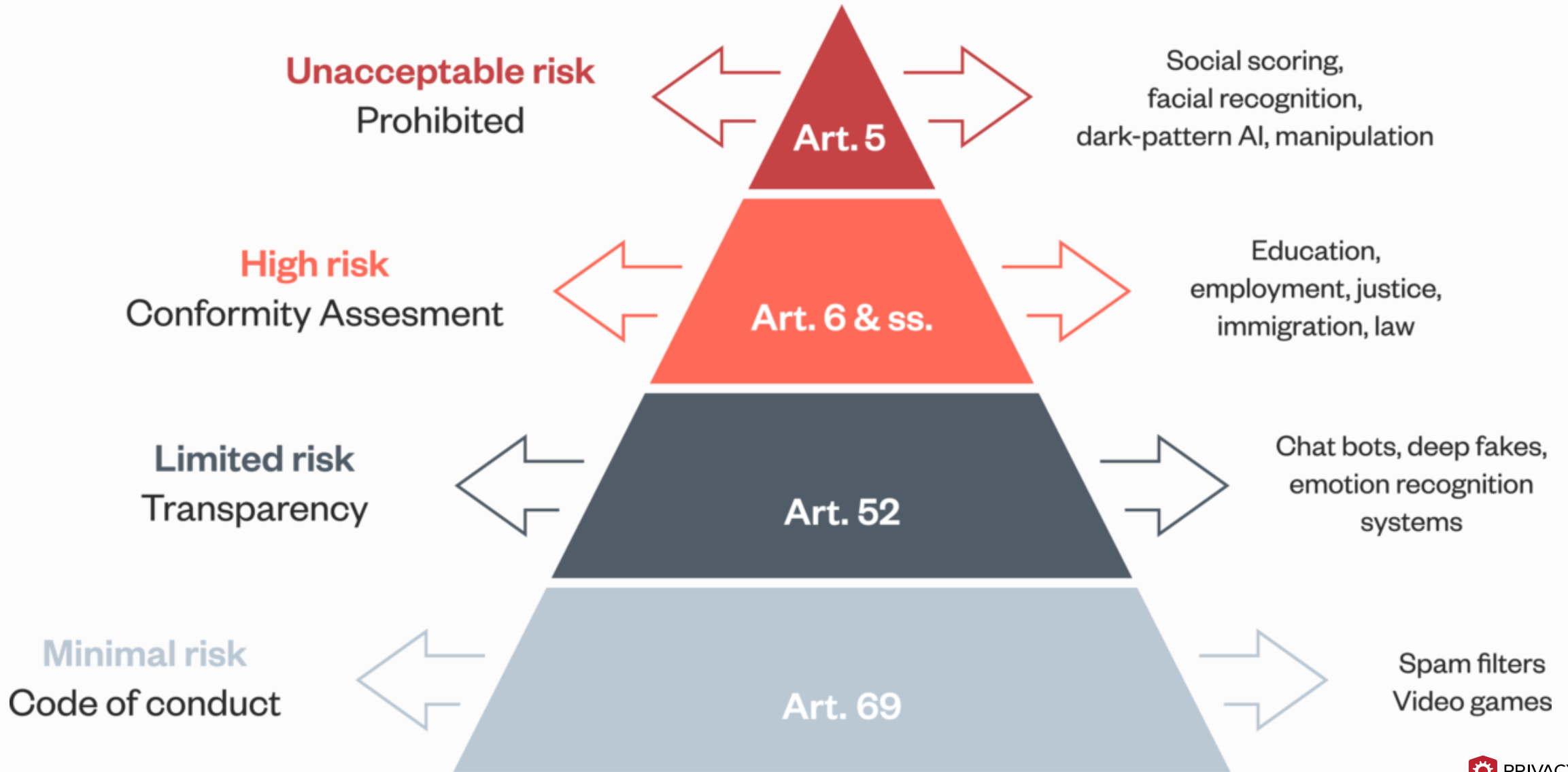
- **Obligations shared with developers and deployers of high-risk systems:**
  - **Risk assessment and mitigation:** Conduct a **comprehensive risk assessment** to identify and address potential risks associated with the AI system (Articles 29 & 30).
  - **Data governance:** Ensure **high-quality, accurate, and representative data** is used for development, training, and operation (Article 26). This includes measures to **mitigate bias and discrimination** (Article 22).
  - **Human oversight:** Implement **appropriate human oversight mechanisms** to ensure safe and responsible use (Article 29).
  - **Technical documentation:** Create and maintain **detailed technical documentation** outlining the system's design, development, and operation for regulatory scrutiny (Article 31).
  - **Record-keeping:** Maintain records of relevant information for a specific period (Article 44).
- **Additional obligations for developers:**
  - **Conformity assessment:** Before placing the system on the market, undergo a **conformity assessment procedure** to demonstrate compliance with the AI Act's requirements (Article 20).

# Developer & Deployer – Systemic-Risk AI Systems

- **Additional obligations for providers (developers):**

- **Model evaluation:** Conduct **model evaluations** including **adversarial testing** to assess and mitigate potential EU-level systemic risks (e.g., manipulation, societal discrimination) (details not yet finalized).
- **Cybersecurity:** Ensure an **adequate level of cybersecurity protection** against cyber threats and vulnerabilities (details not yet finalized).

# The EU AI Act Risk Levels



# Poll 1

- Q. What level of risk applies to your organisation?
- A. Prohibited (Social scoring, Profiling (SC) Manipulation, Facial Recognition)
  - B. High (Education, Employment, Justice, Immigration, Law, Insurance, Banking, Health)
  - C. Limited (Chat bots, Deep Fakes, Emotion Recognition)

# AI Training Data is Often Scraped off the Web

ICO consultation series on generative AI and data protection:

**Chapter one:** *The lawful basis for web scraping to train generative AI models.* Our first chapter covers the lawful basis for training generative AI models on web-scraped data and was open until 1 March 2024.

**Chapter two:** *Purpose limitation in the generative AI lifecycle.* Our second chapter covers how purpose limitation should be applied at different stages in the generative AI lifecycle and is open until 12 April 2024.

# Lawful basis for Scraped Web Data

**Chapter one:** *The lawful basis for web scraping to train generative AI models.*

Legitimate Interest - (Article 6(1)(f) of the GDPR):

1. the purpose of the processing is legitimate; however, if you don't know what your model is going to be used for, how can you ensure its downstream use will respect data protection and people's rights and freedoms?
2. the processing is necessary for that purpose; the ICO's understanding is that currently, most generative AI training is only possible using the volume of data obtained through large-scale scraping.
3. the individual's interests do not override the interest being pursued. Collecting data through web-scraping is an 'invisible processing' activity, where people are not aware their personal data is being processed in this way. This means people may lose control over how and what organisations process their personal data or become unable to exercise the information rights granted by EU data protection law. **Invisible processing and AI related processing** are both seen as **high-risk activities** that require a **DPIA**.

# Four Stages of AI

1. Purpose and Design
2. Data Collection
3. Training and Education
4. Review and Monitoring





# Poll 2, 3 & 4

- Does your organisation have an AI Policy?
  - Yes
  - No
  - Don't know
- Does your organisation currently use AI?
  - Yes
  - No
  - Don't know
- Do people in your organisation use ChatGPT?
  - Yes
  - No
  - Don't know

# Stages 1 Purpose and Design

- Governance
  - Critical as AI is context dependent
    - Problem formulation stage
    - Decision space
    - Construct Space
  - Senior Process owner
  - Proportionate to your use of AI
  - AI Policies and Procedures POLL
- Are you a Controller/Joint Controller/Processor
  - Questions
    - the source and nature of the data used to train an AI model
    - the target output of the model
    - What ML algorithms will be used to create models from the data (eg regression models, decision trees, random forests, neural networks)
    - What are the trade-off between false positives and false negatives

# Stages 1 Purpose and Design

- Subject rights
  - Model management system for tracking
  - Cookie settings/Opt out
  - Data Portability
- DPIA/Risk Assessment
  - AI Act Article 29
  - GDPR Article 35
  - Proportionate, necessity, collection, volume/variety, relationship, outcomes
- Function creep
  - Clear purpose, data flows for AI, privacy statement

# Stages 1 Purpose and Design

- Bias test
  - Training data POLL
  - Input data
  - Discrimination
  - GDPR Recital 71 ‘the controller should use appropriate mathematical or statistical procedures’
- Data Minimisation
  - Anonymisation
- Third Party Due Diligence
  - Assess all quality claims during procurement
  - Request testing reports
  - Training data reports

# Stages 2 Data Collection

## Poll 5

- Do you understand the role of training data in AI models?
  - Yes
  - No

# Stages 2 Data Collection



# Stages 2 Data Collection

- Ensure Purpose Limitation and Data Minimisation
- Accuracy
  - GDPR Accuracy
    - Article 5 Principles
- Training V Real World
  - Locker entry case study
  - Training data V's Live data
  - Labelling data
  - Overfitting/Underfitting of data
    - Appropriate representation
- Engage with individual or representatives

# Stages 3 Training and Testing

- Who
  - Internal
  - External
- Trade Offs
  - Minimisation V's Statistical accuracy
  - Statistical accuracy V's Discrimination
  - Transparency V's Security & IP
- Will models be continuously tested and updated:
  - How often
  - Using what kinds of data
  - How ongoing performance will be assessed
  - Must be part of due diligence
  - DPA
- Biases
  - Statistical, Societal, Evaluation & Emergent
  - Inference of Special Category or High Risk data
- AI Statistical Accuracy
  - Accuracy of the system



# Stages 4 Review and Monitor

- Proportionate
  - Article 22
  - No 'rubber stamping'
- System Accuracy
  - DP by design
  - Controller or Processor?
  - DPA
    - Data Distribution patterns
- AI data is not factual
  - Statistically informed guesses
  - Through DP by Design you need mitigate this
    - Labelling
    - Data flows
    - ROPA
- DPIA
  - Live document
  - Function/Scope creep, deployment bias
- EDPB Guidance
  - Log human interventions
  - Not in a routine fashion
  - 'meaningful' influence on the decision, including the 'authority and competence' to go against the recommendation
  - 'weigh-up' and 'interpret' the recommendation, consider all available input data

# The Dutch Child Benefit Case

- The Dutch childcare benefit scandal, also known as the "benefits affair," was a major political controversy in the Netherlands from 2019 to 2021.
- **What Happened:**
- The Dutch Tax and Customs Administration (Belastingdienst) wrongly accused thousands of parents of fraudulently claiming childcare benefits.
- Between 2005 and 2019, an estimated 26,000 families were targeted.
- The accusations were based on an algorithm that used factors like "foreign-sounding names" and "dual nationality" to flag potential fraud.

# The Dutch Child Benefit Case

## **The Problem:**

This algorithm led to racial profiling and discriminatory practices. Many innocent families, particularly those from immigrant backgrounds, were subjected to intense scrutiny and forced to repay benefits they had rightfully received.

Some families faced financial hardship and emotional distress due to the accusations. The entire process violated fundamental principles of fairness and due process.

## **Outcome:**

The scandal sparked public outrage and a parliamentary inquiry. Investigations revealed significant flaws in the tax authority's practices and highlighted institutional bias.

In January 2021, the Dutch government resigned in response to the scandal. Compensation programs were established to help affected families.

## Case C-634/21 of the European Court of Justice (ECJ)

- Case C-634/21 of the European Court of Justice involved proceedings between a citizen and Land Hessen, represented by the Data Protection and Freedom of Information Commissioner for Hesse (the ‘HBDI’), regarding the protection of personal data.
- The case revolved around SCHUFA Holding AG (‘SCHUFA’), a private company that provided a credit institution with a score for the citizen in question. This score served as the basis for the refusal to grant the credit for which the citizen had applied\*.
- The citizen requested SCHUFA to erase the entry concerning her and to grant her access to the corresponding data. However, SCHUFA only informed her of the relevant score and the principles underlying the calculation method for the score, without informing her of the specific data included in that calculation or of the relevance accorded to them in that context.

## Case C-634/21 of the European Court of Justice (ECJ)

- SCHUFA argued that the calculation method is a trade secret\*.
- In his Opinion, Advocate General Priit Pikamäe stated that the GDPR establishes a ‘right’ for the person concerned not to be subject to a decision based solely on automated processing, including profiling\*.
- The conditions for that right are satisfied because the procedure at issue constitutes ‘profiling’, the decision produces legal effects concerning the person concerned or similarly significantly affects him or her\*.

## Case C-604/22 of the European Court of Justice (ECJ)

- IAB Europe: Transparency and Consent Strings (TCS)
- Recital 26
  - Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person
- Recital 30
  - Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols..... when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them
- Ruling:
  - Court decided that a code generated to record the end-user's consent choice and document the GDPR's transparency obligation – the so-called TC String – is personal data
  - IAB Europe has, reasonable means allowing it to identify a particular natural person from a TC String

# Red Flags

- You represent and warrant that your privacy policy shall not conflict with the XXXXX Privacy Policy, and that in the event of such conflict, the terms of our Privacy Policy shall control
- No assurances that defects will be corrected, and service is free from inaccuracies.
- The Service is not directed to individuals under the age of 13. In the event that we discover that a child under the age of 13 has provided personally identifiable information to us, we will make efforts to delete the child's information if required by the Children's Online Privacy Protection Act.
- Regardless of where our servers are located, your personal data may be processed by us in the United States, where data protection and privacy regulations may or may not be to the same level of protection as in other parts of the world.
- **The good cases**
  - Microsoft Co-Pilot

# Deliverables

- Privacy Engine Framework including AI
- AI Gap Analysis
- Risk Assessment under AI Act Article 29 & 30
- DPIA under GDPR Article 35

[nollag.conneely@privacyengine.io](mailto:nollag.conneely@privacyengine.io)

[maria.moloney@privacyengine.io](mailto:maria.moloney@privacyengine.io)

<https://www.privacyengine.io/scheduleconsultation/>



# Thank You!

