

# AI & Data Protection Series

## Data privacy considerations in AI model training

**May 2025**

Alex Brown, Partner, Digital Business

Emilie Danglades-Perez, Managing Associate, Data Protection

# Our speakers



**Alex Brown**  
Partner  
London



**Emilie Danglades-Perez**  
Managing Associate  
Paris

# Data privacy considerations in AI model training

## Today's topics

Focus on AI developers + data privacy considerations associated with AI model development and training

- Lawful basis for the data collection and processing
- Fairness under the GDPR
- Transparency
- Data minimisation
- Accountability requirements
- Deployer considerations.

# Lawful basis

Why is personal data processing integral to AI model development?

- Effectiveness and accuracy
- Diverse data representation
- Contextual understanding
- Enhancing model robustness and safety
- Feedback and iterative improvement
- Real-world application



# Lawful basis

How to choose the appropriate legal basis?

- **Lawfulness**: obligation to rely on one of the legal bases outlined in Article 6 of the GDPR.
- **Legitimate interests** as an appropriate legal basis in most cases for AI development

**What are the consequences of lack of a valid legal basis when training AI models?**

Data protection authorities have imposed administrative fines ranging from €5 to €30 million

# Lawful basis

## Legitimate interest

1. Purpose Test: Identifying the legitimate interest
2. Necessity Test: Necessity of the processing
3. Balancing Test
4. Documentation and accountability

# Special category data

## Article 9 GDPR

- Explicit consent
- Substantial public interest
- Research and statistics
- Legal claims

## Special category data

- Limited conditions under which SCD can be processed:
  - Consent
  - Substantial public interest
  - Research and statistics
- EU case law (*GC and Others v CNIL*) and guidance – no processing of SCD until an individual rights request / verification of processing SCD

## Fairness

*“personal data is not processed by unfair methods, or by deception, or in a way that is ‘unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject’. Considering the complexity of the technologies involved, information on the processing of personal data within AI models should therefore be provided in an accessible, understandable and user-friendly way”*

- European Data Protection Board (EDPB)



# Transparency

## Article 12-14 GDPR

- Provision of information
- Timing and method of information provision
- Transparency in training processes
- Challenges in AI transparency

### What happens in case of absence of transparency?

- Investigations from data protection authorities in the EU/UK
- Temporary ban of the AI tool in one or several EU/UK countries
- Administrative fine

# Data minimisation

## Article 5 GDPR

### 1. Key aspects of data minimisation

- Data collection must have clear, specific purposes
- Use data that is adequate, relevant, and limited to the model's purpose

### 2. Techniques for data minimisation

- **Feature selection:** Use only relevant data features
- **Perturbation:** Add noise to reduce re-identification risk
- **Synthetic data:** Use data that mimics real data without personal information

### 3. Challenges in AI Models

# Accountability

## Article 5.2. GDPR

- Understanding accountability
- Conducting data protection impact assessments (DPIAs)
- Documentation and governance
- Ongoing monitoring and review
- Transparency and communication

# Deployers

## Key considerations

1. Due diligence on AI model development and training
  - Data sources and basis for processing
  - Data minimisation methodology
  - Bias and fairness measures
2. Understanding the model's functionality and limitations
  - Explainability
  - Testing and validation
3. Risk management and accountability

# Useful resources

CtrlTransfer



Digital Regulation Tracker



AI Literacy Programme



**Next session > 12 June:** AI and GDPR purpose limitation requirements