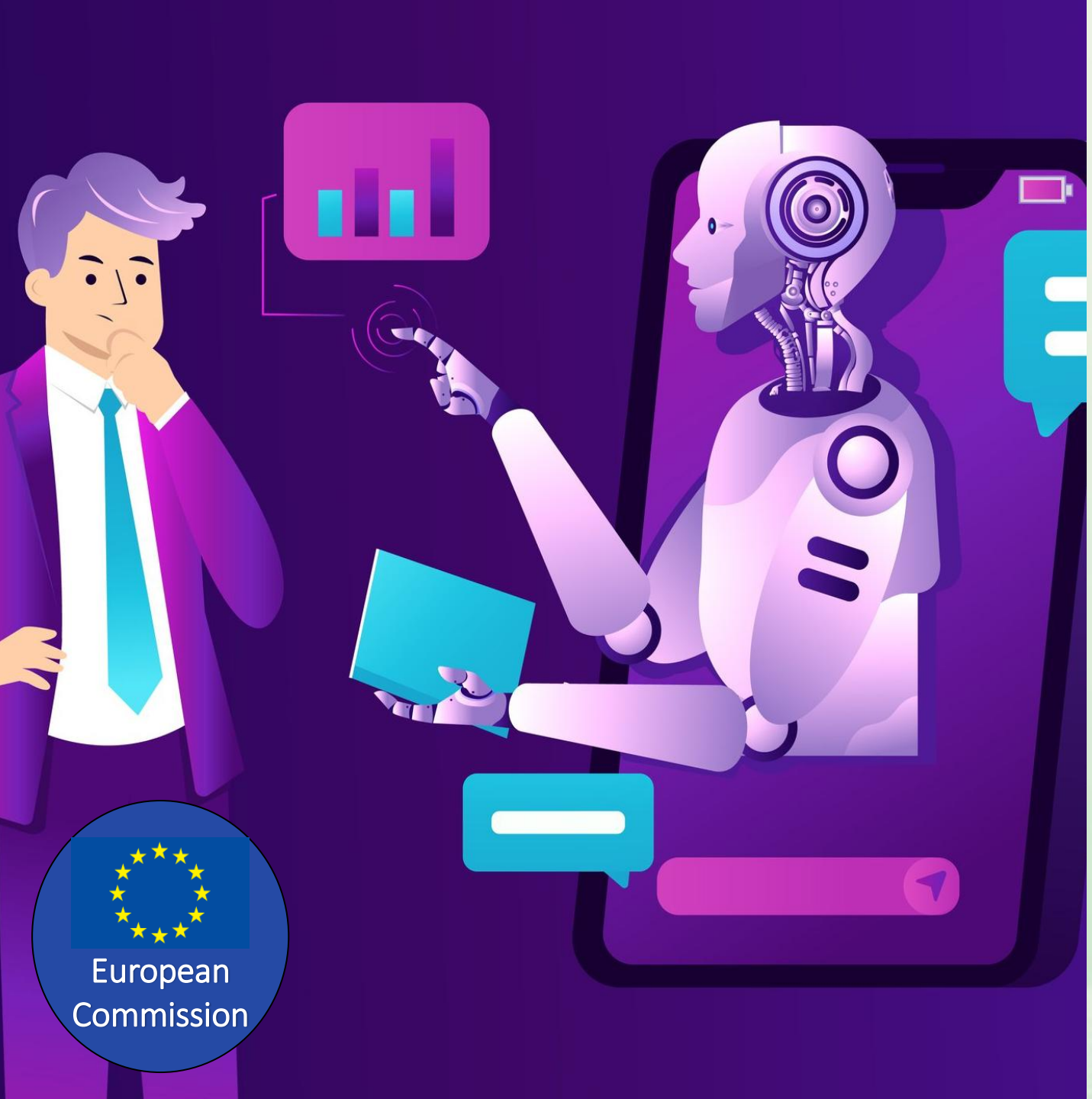


AI & Partners

Amsterdam - London - Singapore

Announcement: Prohibited AI systems must be removed from EU market

Prohibited AI practices are banned from 2nd February 2025 onwards.

What are prohibited AI practices?

See **Slides 4 – 11** for details.

Contents

AI & Partners
Amsterdam - London - Singapore

Overview

AI & Partners
Amsterdam - London - Singapore

What is the EU AI Act Handbook?

The EU AI Act Handbook is a comprehensive guide that outlines the legislative and other provisions made under the EU AI Act.

It is designed to ensure the safe and ethical development, deployment, and use of AI systems within the European Union. The Handbook provides detailed explanations of the Act's requirements, including prohibited AI practices, high-risk AI systems, and governance structures.

Content

Prohibited AI Practices: This section details AI practices that are strictly prohibited due to their potential to cause significant harm. Examples include subliminal techniques, exploitation of vulnerabilities, social scoring, predictive policing, emotion recognition, and real-time remote biometric identification.

Manipulative or Deceptive Techniques

AI & Partners
Amsterdam - London - Singapore

Description (Including Legislative Reference)

- **Legislative Reference**: Article 5(1)(a) of the EU AI Act
- **Description**: AI systems that deploy subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques. These techniques are designed to materially distort the behaviour of a person or group, impairing their ability to make informed decisions.

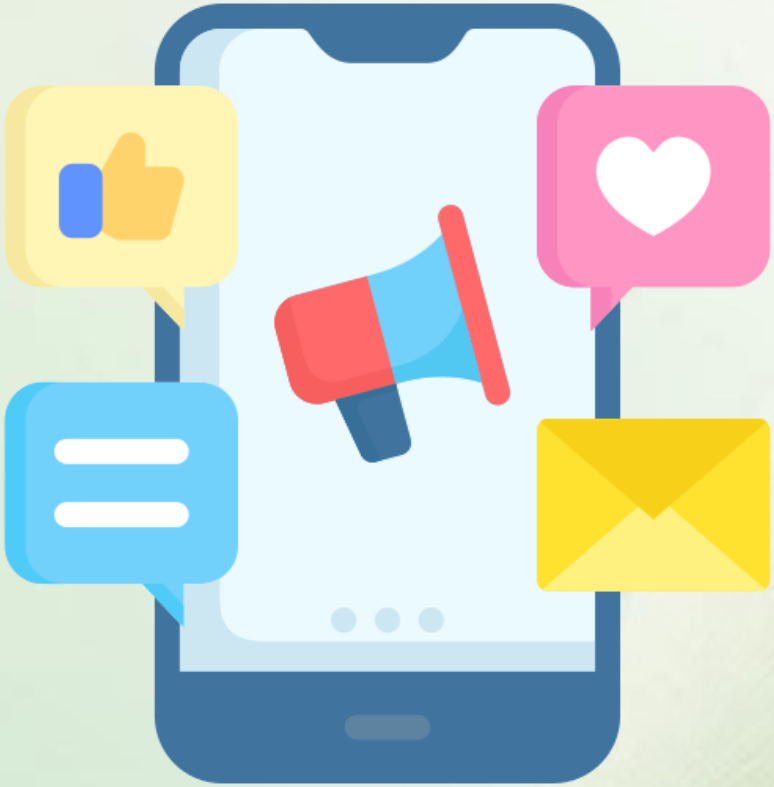
Factors to be Taken into Account

- **Nature of Techniques**: The use of audio, image, or video stimuli that are beyond human perception or other manipulative techniques that subvert or impair autonomy and decision-making.
- **Vulnerable Groups**: Special consideration for groups vulnerable due to age, disability, or specific social/economic situations.
- **Degree of Control**: The extent to which the AI system can control the stimuli presented to individuals, potentially through advanced interfaces like virtual reality or machine-brain interfaces.

Real-World Examples

- **Example 1**: An AI-driven advertising platform that uses imperceptible audio cues to influence consumer purchasing decisions without their conscious awareness.
- **Example 2**: A virtual reality application that subtly manipulates user emotions and decisions through controlled visual and auditory stimuli, leading to significant psychological impacts.
- **Example 3**: An AI system in social media that uses hidden algorithms to nudge users towards specific behaviours or opinions, potentially causing financial or psychological harm.

Advertising



Exploitation of Vulnerabilities

AI & Partners
Amsterdam - London - Singapore

Description (Including Legislative Reference)

- **Legislative Reference**: Article 5(1)(b) of the EU AI Act
- **Description**: AI systems that exploit vulnerabilities of specific groups due to age, disability, or social/economic situations. These systems are designed to materially distort behaviour, leading to significant harm.

Factors to be Taken into Account

- **Vulnerable Groups**: Special consideration for groups vulnerable due to age, disability, or specific social/economic situations, such as extreme poverty or minority status.
- **Degree of Exploitation**: The extent to which the AI system can manipulate or exploit the vulnerabilities of these groups, impairing their ability to make informed decisions.
- **Potential Harm**: The likelihood and severity of harm caused by the AI system, including physical, psychological, and financial impacts.
- **Regulatory Compliance**: Ensuring that the AI system does not violate existing laws and regulations designed to protect vulnerable populations.

Real-World Examples

- **Example 1**: An AI-driven lending platform that offers high-interest loans to individuals in extreme poverty, exploiting their financial desperation and lack of alternatives.
- **Example 2**: A targeted advertising system that manipulates elderly users into purchasing unnecessary and expensive products by exploiting their cognitive decline.
- **Example 3**: An AI-based recruitment tool that discriminates against candidates from specific socio-economic backgrounds, reducing their employment opportunities and perpetuating inequality.

Recruitment

Social Scoring

AI & Partners
Amsterdam - London - Singapore

Description (Including Legislative Reference)

- **Legislative Reference**: Article 5(1)(c) of the EU AI Act
- **Description**: AI systems used for evaluating or classifying individuals based on social behaviour or personal characteristics over time. These systems assign social scores that lead to detrimental or unfavourable treatment in contexts unrelated to the original data collection.

Factors to be Taken into Account

- **Nature of Data**: The types of data points used for social scoring, including social behaviour, personal characteristics, and the contexts in which the data was collected.
- **Unrelated Contexts**: The extent to which the social scores are applied in contexts unrelated to the original data collection, leading to unjustified or disproportionate treatment.
- **Impact on Individuals**: The potential for social scoring to result in discriminatory outcomes, exclusion, or unfavourable treatment of individuals or groups, violating their right to dignity and non-discrimination.

Real-World Examples

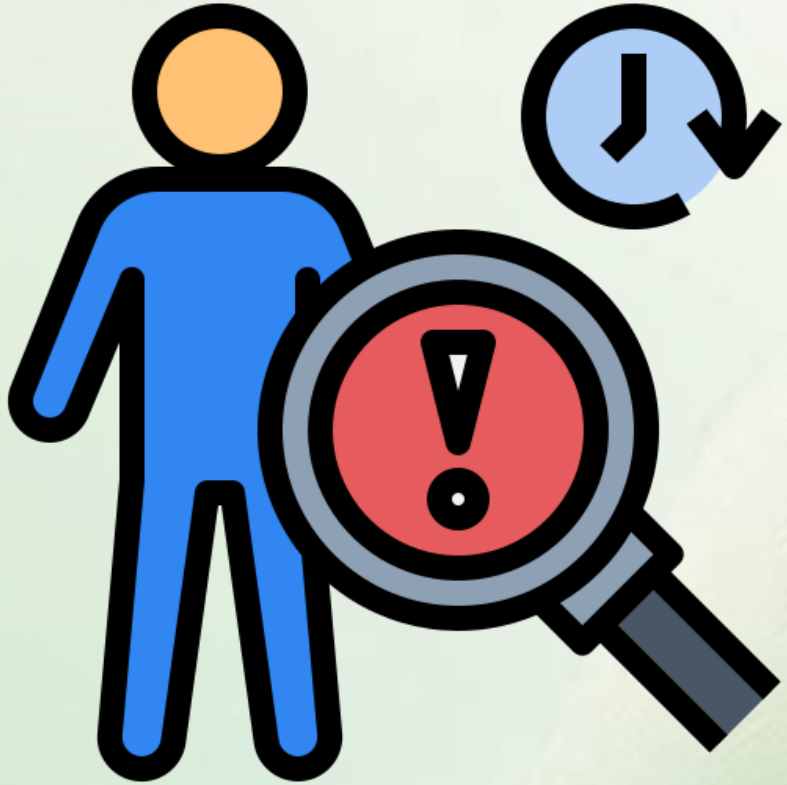
- **Example 1**: A social media platform using AI to assign social scores based on user interactions and posts, which are then used by employers to make hiring decisions, leading to potential discrimination.
- **Example 2**: A financial institution using AI to evaluate customers' social behaviour and personal characteristics to determine creditworthiness, resulting in higher interest rates or denial of services for certain individuals.
- **Example 3**: A government using AI to assign social scores to citizens based on their online activities and public behaviour, which affects their access to public services and benefits.

Social Score

Facial Recognition, Biometric Data Exploitation

AI & Partners
Amsterdam - London - Singapore

Predictive Policing



Description (Including Legislative Reference)

- **Legislative Reference**: Article 5(1)(d) of the EU AI Act
- **Description**: AI systems used for making risk assessments to predict the likelihood of individuals committing criminal offenses. These systems are based solely on profiling or assessing personality traits and characteristics.

Factors to be Taken into Account

- **Nature of Profiling**: The extent to which the AI system relies on profiling or assessing personality traits and characteristics to make predictions.
- **Potential for Discrimination**: The likelihood that the AI system will lead to discriminatory outcomes, particularly against marginalized or vulnerable groups.
- **Accuracy and Reliability**: The accuracy and reliability of the AI system in making predictions, including the quality of the data used for training and validation.

Real-World Examples

- **Example 1**: An AI-driven tool used by law enforcement to predict future criminal behaviour based on an individual's past criminal record and socio-economic background, leading to increased surveillance and potential bias.
- **Example 2**: A predictive policing system that assesses the likelihood of reoffending based on personality traits and demographic data, resulting in harsher sentencing or parole decisions for certain individuals.
- **Example 3**: An AI system used to identify potential suspects in a neighbourhood based on historical crime data and personal characteristics, leading to disproportionate targeting of specific communities.

Emotion Recognition in Work or Education

AI & Partners
Amsterdam - London - Singapore

Emotion Recognition

Description (Including Legislative Reference)

- **Legislative Reference**: Article 5(1)(e)-(g) of the EU AI Act
- **Description**: AI systems that create or expand facial recognition databases through untargeted scraping of images or infer emotions in workplaces and educational institutions. These systems categorize individuals based on biometric data, deducing sensitive attributes like race, or sexual orientation.

Factors to be Taken into Account

- **Nature of Data Collection**: The method of data collection, particularly untargeted scraping of images from the internet or CCTV footage, which can infringe on privacy rights.
- **Context of Use**: The specific contexts in which these AI systems are used, such as workplaces and educational institutions, where there is a significant power imbalance and potential for misuse.
- **Potential for Discrimination**: The likelihood that the AI system will lead to discriminatory outcomes, particularly against marginalized or vulnerable groups, by deducing sensitive attributes.

Real-World Examples

- **Example 1**: An AI system used in a workplace to monitor employees' emotions through facial recognition, potentially leading to biased performance evaluations and discriminatory treatment.
- **Example 2**: A school implementing an AI system to infer students' emotions during classes, which could result in unfair disciplinary actions based on misinterpreted emotional states.
- **Example 3**: A social media platform using AI to scrape images from users' profiles to expand its facial recognition database, categorizing users based on inferred sensitive attributes like political opinions or sexual orientation, leading to targeted advertising or content moderation.

Real-Time Remote Biometric Identification

AI & Partners
Amsterdam - London - Singapore

Biometric Identification

Description (Including Legislative Reference)

- **Legislative Reference**: Article 5(1)(h) of the EU AI Act
- **Description**: Use of real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes. These systems capture, compare, and identify biometric data instantaneously or near-instantaneously to identify individuals without their active involvement.

Factors to be Taken into Account

- **Nature of the Situation**: The seriousness, probability, and scale of the harm that would be caused if the system were not used. This includes situations like searching for victims of abduction, preventing imminent threats, or identifying suspects of serious crimes.
- **Impact on Rights and Freedoms**: The consequences of using the system for the rights and freedoms of all persons concerned, particularly the seriousness, probability, and scale of those consequences.
- **Proportionality and Necessity**: The use must be strictly necessary and proportionate to achieve the specific objectives, with limitations on the period of time, geographic scope, and personal scope.

Real-World Examples

- **Example 1**: Law enforcement using real-time remote biometric identification to locate a missing child in a crowded public event, ensuring rapid identification and rescue.
- **Example 2**: Deployment of real-time biometric systems at an airport to prevent an imminent terrorist threat, identifying suspects based on live video feeds.
- **Example 3**: Police using real-time biometric identification during a public protest to identify individuals with outstanding warrants for serious crimes, ensuring public safety while respecting legal constraints.

AI & Partners
Amsterdam - London - Singapore

Contact Us

AI & Partners

Amsterdam - London - Singapore

E-mail

contact@ai-and-partners.com

Website

https://www.ai-and-partners.com/

Need help to remove prohibited AI systems?

Book a call for more information: