



Engaging Content  
Engaging People

A World  
Leading SFI  
Research  
Centre



# To Be High-Risk, or Not To Be —Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards

Delaram Golpayegani\*, Harshvardhan Pandit\*\*, Dave Lewis\*

sgolpays@tcd.ie, harshvardhan.pandit@dcu.ie, delewis@tcd.ie

\* ADPAT Centre, Trinity College Dublin, Dublin, Ireland

\*\*ADAPT Centre, Dublin City University, Dublin, Ireland

Presentation for ACM Conference on Fairness, Accountability, and Transparency (FAccT 2023)



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



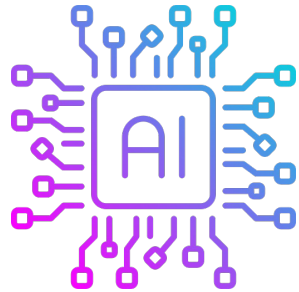
**Protect**

Slides



# Overview of the AI Act

## Scope



Placing on the market, putting into service, and use of AI systems within the EU

## Key Legal Roles Subjected to Obligations

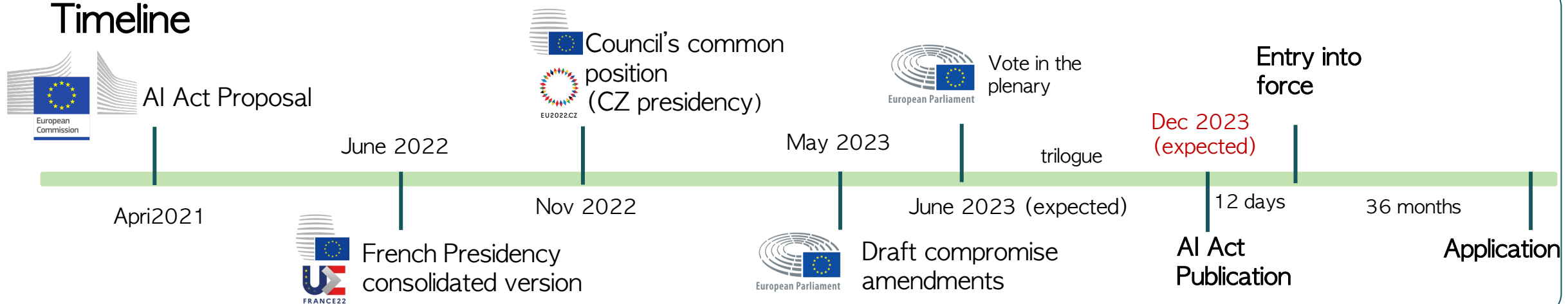
### AI Provider

An entity who develops an AI system or that has an AI system developed and places that system on the market or puts it into service ... (Art. 3(2))

### AI User

An entity under whose authority the system is used (Art. 3(4))  
Not the same as “end-uses”

## Timeline



In this presentation, the Council's common position is adopted.

# AI Act's Risk Hierarchy

## Unacceptable Risk

Prohibited to be placed on the market, put into service, or used

## High Risk

Title III, Chapter 2: Requirements for High-Risk AI Systems  
Title III, Chapter 3: Obligations of Providers and Users of High-Risk AI Systems and Other Parties

## Limited Risk

Title IV: Transparency Obligations for Providers and Users of Certain AI Systems

## Minimal Risk

Title IX: Codes of Conduct

Harmful risk to three areas:



Health



Safety



Fundamental rights

# What is Classified as High-Risk AI?

Under the AI Act, an AI system is high-risk if it is:

## Article 6

1

A **product** covered by the Union harmonisation legislation (**Annex II**)  
&  
is required to undergo a **third-party conformity assessment**

OR

2

Used as a **safety component** of a product covered by **Annex II** legislation  
&  
is required to undergo a **third-party conformity assessment**

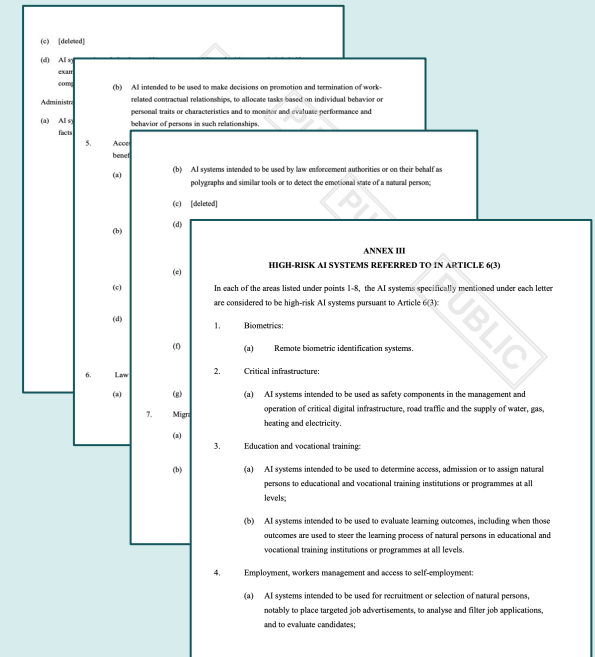
OR

3

An **Application** referred to in **Annex III**

## Annex III

High-risk uses are listed under 8 areas

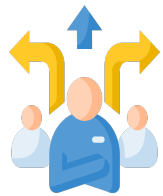




**What information** is needed to make a decision about whether an application of AI is high-risk?



**When** should the decision be re-visited?

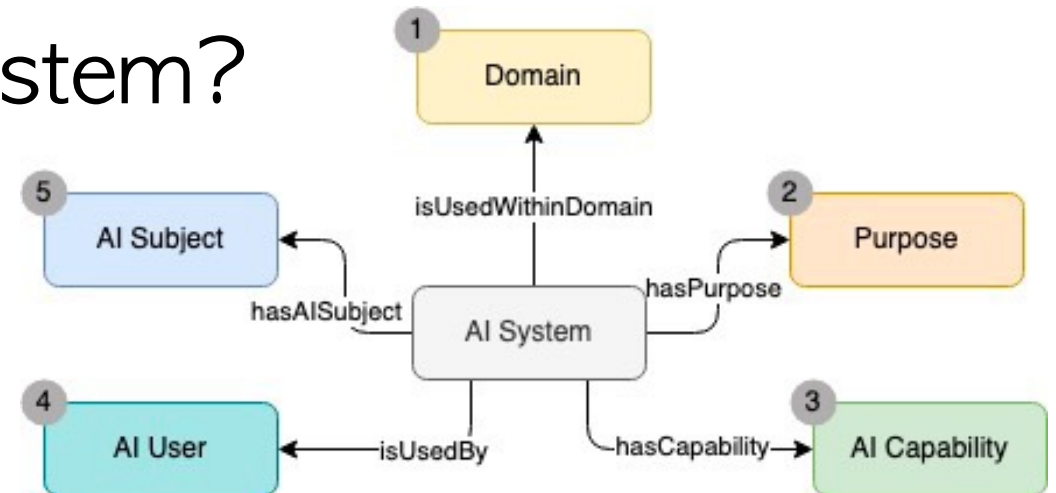


**Who** is responsible for making the decision?

# Information Required for Determining High-Risk AI

By analysis of Annex III, 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**?



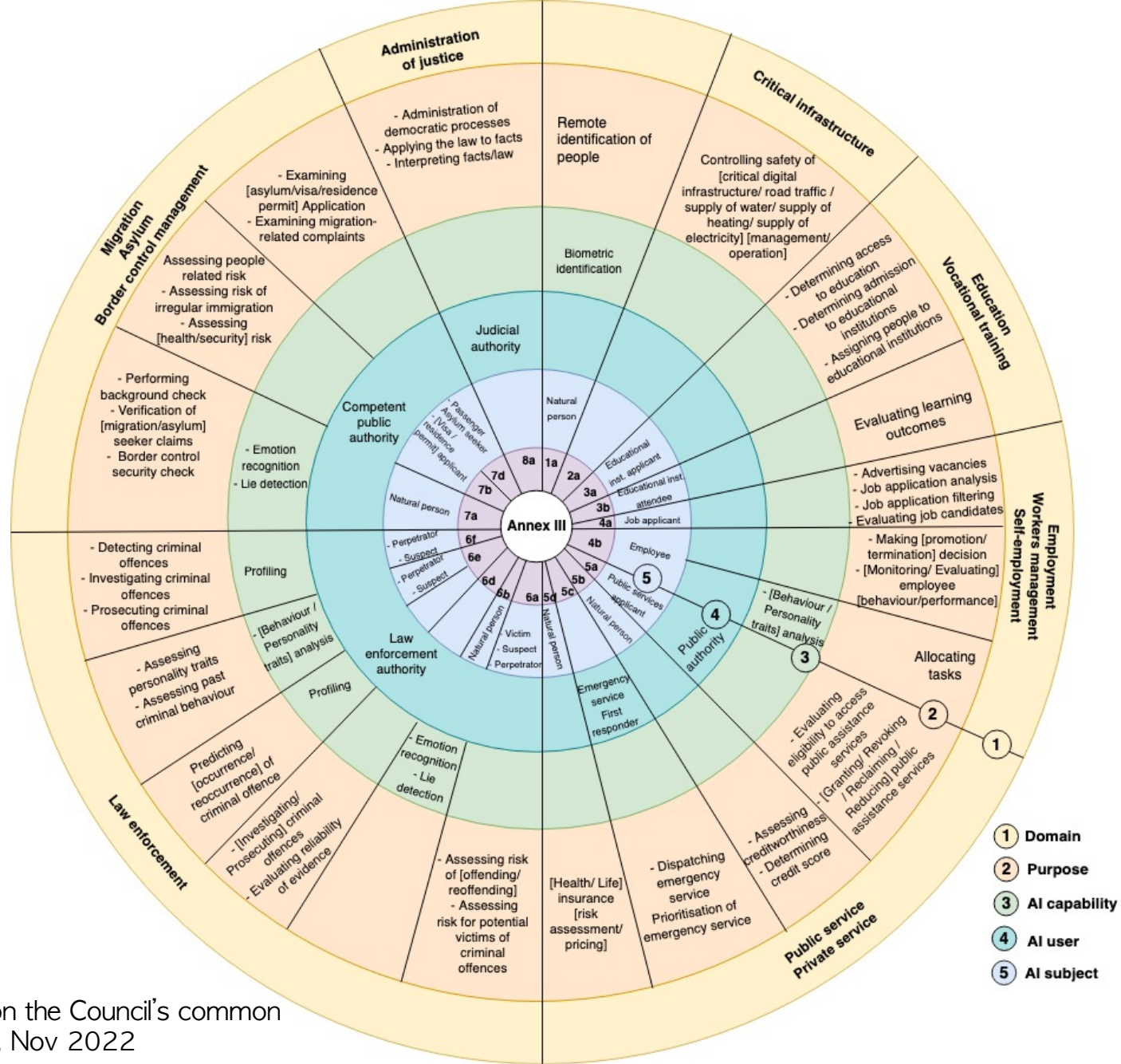
# Example of Identification of High-Risk AI Using the 5 Concepts

- (1) In which **Domain** is the AI system used? Law enforcement
- (2) What is the **Purpose** of the AI system? Assessing past criminal behaviour
- (3) What is the **Capability** of the AI system? Behaviour analysis
- (4) Who is the **User** of the AI system? Law enforcement authority
- (5) Who is the **AI Subject**? Individuals who are suspected of a crime



The AI system is highly likely to be **High-Risk** according to **Annex III, 6e**

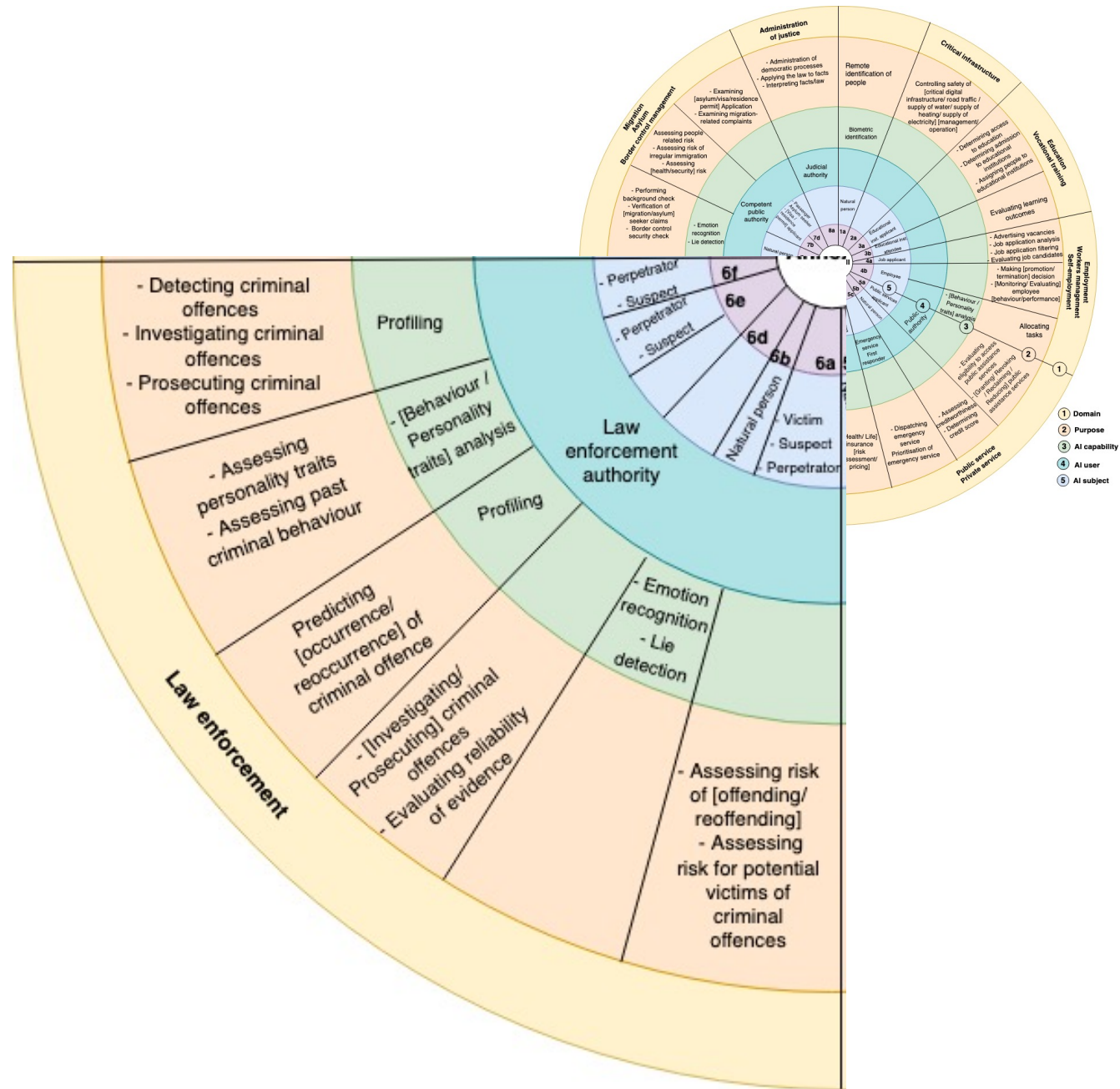
# Specifying Annex III High-Risk Conditions Using the 5 Core Concepts



Based on the Council's common position, Nov 2022



# Annex III High-Risk Conditions Related to Law Enforcement



# VAIR: Vocabulary of AI Risks



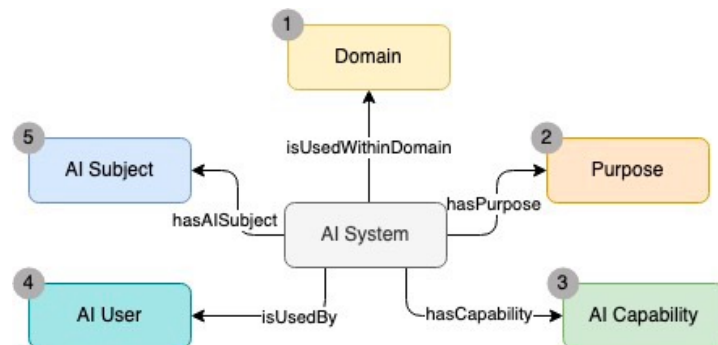
<https://w3id.org/vair>

- Specialisation of AIRO (AI Risk Ontology)



<https://w3id.org/airo>

- VAIR includes instances of concepts represented in AIRO



- 4. Purposes**
  - 4.1 Remote Identification Of People
  - 4.2 Content Generation
  - 4.3 Generating Audio Content
  - 4.4 Generating Image Content
  - 4.5 Generating Video Content
  - 4.6 Knowledge Reasoning
  - 4.7 Applying The Law To Facts
  - 4.8 Interpreting Law
  - 4.9 Interpreting Facts
  - 4.10 Decision Making
  - 4.11 Examining Application
  - 4.12 Examining Asylum Application
  - 4.13 Examining Migration Related Complaints
  - 4.14 Examining Residence Permits Application
  - 4.15 Examining Visa Application
  - 4.16 Assessment
  - 4.17 Assessing Past Criminal Behaviour
  - 4.18 Assessing Admission Test
  - 4.19 Assigning People To Educational Institutions
  - 4.20 Determining Access To Education
  - 4.21 Determining Admission To Educational Institutions
  - 4.22 Assessing Student
  - 4.23 Evaluating Learning Outcomes
  - 4.24 Recruiting

- 7. AI Capabilities**
  - 7.1 Biometric Identification
  - 7.2 RemoteBiometricIdentification
  - 7.3 Personality Traits Analysis
  - 7.4 Emotion Recognition
  - 7.5 Profiling
  - 7.6 Face Recognition
  - 7.7 Computer Vision
  - 7.8 Image Recognition
  - 7.9 Automatic Summarisation
  - 7.10 Dialogue Management
  - 7.11 Information Retrieval
  - 7.12 Machine Translation
  - 7.13 Named Entity Recognition
  - 7.14 Natural Language Generation
  - 7.15 Part Of Speech Tagging
  - 7.16 Question Answering
  - 7.17 Relationship Extraction
  - 7.18 Speech Recognition
  - 7.19 Speech Synthesis
  - 7.20 Pattern Recognition
  - 7.21 Action Recognition
  - 7.22 Gesture Recognition
  - 7.23 Object Recognition
  - 7.24 Music Information Retrieval
  - 7.25 Sound Event Recognition
  - 7.26 Sound Synthesis
  - 7.27 Sound Source Separation
  - 7.28 Speaker Recognition
  - 7.29 Lie Detection
  - 7.30 Sentiment Analysis

- Sources used for population of VAIR:
  - The AI Act
  - ISO/IEC 22989 AI concepts and terminology
  - AI Watch's taxonomy

## Taxonomies in VAIR

### AI

Techniques (19)  
Capabilities (30)  
Types of AI (17)  
Components (34)  
Characteristics (20)  
Outputs (6)

### Risk

Risk Sources (43)  
Consequences (4)  
Impacts (12)  
Impact areas (5)  
Controls (18)

### Uses of AI

Purposes (114)  
Domains (13)

### Stakeholder

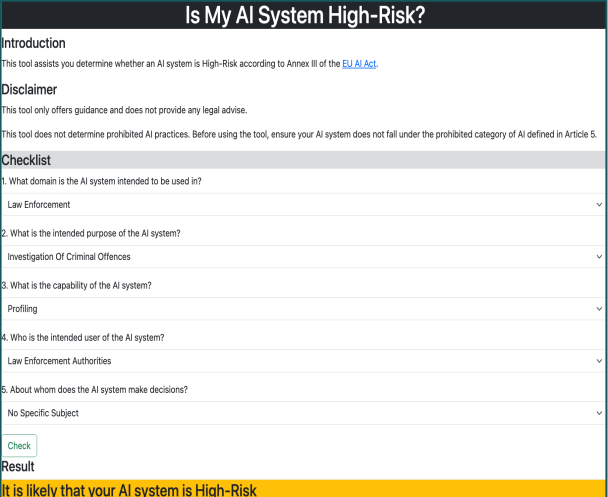
Stakeholder roles (40)

### Document and standard

Documents (12)  
Standard (22)

- A basis for a **checklist for AI risk management**
  - Providing **free and open access to AI risk information**
    - VAIR is available under the CC-BY-4.0 Licence
  - Can be **Extended**
- Describing **rules for determining high-risk** AI as per Annex III
- Maintaining information about AI systems and their risks in a formal and interoperable format

## A Tool for Assisting with Determining High-Risk AI



**Is My AI System High-Risk?**

**Introduction**  
This tool assists you determine whether an AI system is High-Risk according to Annex III of the [EU AI Act](#).

**Disclaimer**  
This tool only offers guidance and does not provide any legal advice.  
This tool does not determine prohibited AI practices. Before using the tool, ensure your AI system does not fall under the prohibited category of AI defined in Article 5.

**Checklist**

1. What domain is the AI system intended to be used in?  
Law Enforcement
2. What is the intended purpose of the AI system?  
Investigation Of Criminal Offences
3. What is the capability of the AI system?  
Profiling
4. Who is the intended user of the AI system?  
Law Enforcement Authorities
5. About whom does the AI system make decisions?  
No Specific Subject

**Result**  
It is likely that your AI system is High-Risk

# When should the decision be re-visited?

## 1) Substantial modification

Changes that affect either the system's conformity with the high-risk AI requirements or the intended purpose (Art. 3(23))

- Changes to the 5 core concepts
  - Domain, Purpose, Capability, AI User, AI Subject
- Other examples
  - Modification of the risk management process (Art. 9)
  - Change of training data (Art. 10)

## 2) Amendments to Annex III made by the European Commission

# Who is responsible for making the decision?

- **AI Provider**
  - Any entity that satisfies the definition of AI provider (Art. 3(2))
  - **Other entities, e.g. AI User**, to whom responsibilities of AI provider are delegated (conditions of Art. 23a(1))
- **Providers of general purpose AI** that might be used as a high-risk AI system or its components

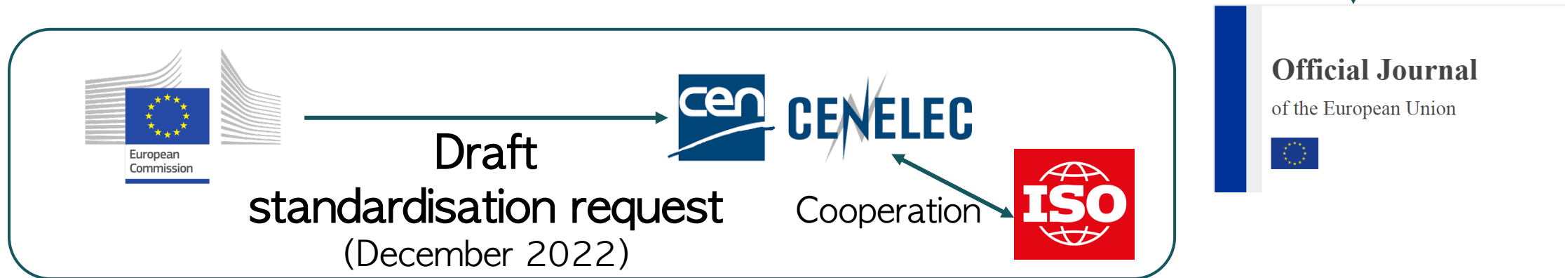
# The AI Act and Harmonised Standards

- The AI Act relies on standards for providing assistance in conformity tasks

## Presumption of Conformity (Art. 40)

compliance with the requirements through conformance to harmonised standards

Indexed in



# Current State of Standardisation at ISO JTC 1/SC 42

- Analysis of activities within ISO JTC 1/SC 42
- Standards that can be used for compliance with high-risk AI requirements
  - Lack of certifiable standards that help with demonstrating compliance with the AI Act
  - Some of the areas that need standardisation:
    - Document generation, e.g. technical document (Art. 11)
    - Record keeping (Art. 12)

Area	AI Act	Standard (ISO development stage as of April 2023)	Type	Coverage
Determine high-risk AI	Art. 6	ISO/IEC TR 24030:2021 AI – Use cases (90.92)	Guidance	AI uses
		ISO/IEC DIS 5339 Guidance for AI applications (40.20)	Guidance	AI uses
Risk management system for AI systems	Art. 9	ISO/IEC 23894 Guidance on risk management	Guidance	AI system
		ISO/IEC TR 24027:2021 Bias in AI systems and AI aided decision making	Technical	AI system
		ISO/IEC TR 24368:2022 Overview of ethical and societal concerns	Guidance	AI system
		ISO/IEC AWI 42005 AI system impact assessment (20.0)	Guidance	AI system
		ISO/IEC CD TR 5469 Functional safety and AI systems (30.60)	Guidance	AI system
Data governance and quality	Art. 10	ISO/IEC 20546:2019 Big data – Overview and vocabulary	Foundational	Big data
		ISO/IEC TR 20547 series Big data reference architecture	Technical	Big data
		ISO/IEC 24668:2022 Process management framework for big data analytics	Organisational	Big data
		ISO/IEC FDIS 8183 Data life cycle framework (50.20)	Guidance	Data
		ISO/IEC [CD/DIS] 5259 series Data quality for analytics and ML (different stages)	Technical	Data
		ISO/IEC CD TS 12791 Treatment of unwanted bias in classification and regression ML tasks (30.20)	Technical	Machine learning
Transparency	Art.13	ISO/IEC AWI 12792 Transparency taxonomy of AI systems (20.00)	Guidance	AI systems
		ISO/IEC AWI TS 6254 Objectives and approaches for explainability of ML models and AI systems (20.00)	Guidance	AI systems ML models
Human oversight	Art. 14	ISO/IEC WD TS 8200 Controllability of automated AI systems (20.60)	Technical	AI system
System quality	Art. 15	ISO/IEC TR 24028:2020 Overview of trustworthiness in AI	Technical	AI system
		ISO/IEC WD TS 25058 SQuaRE – Guidance for quality evaluation of AI systems (20.60)	Technical	AI system
		ISO/IEC PRF TS 25059 SQuaRE – Quality model for AI systems (50.20)	Technical	AI system
		ISO/IEC AWI TS 29119-11 Testing of AI systems (20.00)	Technical	AI system
		ISO/IEC TS 4213:2022 Assessment of machine learning classification performance	Technical	Machine learning
		ISO/IEC TR 24029 Assessment of the robustness of neural networks	Technical	Neural networks
ISO/IEC AWI TS 17847 Verification and validation analysis of AI (20.00)	Technical	AI system		
Quality management system	Art. 17	ISO/IEC DIS 42001 Management system (40.60)	Organisational	Management system



- Enhancing VAIR and providing it as a **checklist for AI risk management**
  - Through engagement with community
  - Sector-specific extensions
- Developing **tools** for assisting AI Providers and AI Users in compliance with the AI Act's requirements
  - Providing guidance on legal requirements, relevant standards, and the information required to be maintained

# To Be High-Risk, or Not To Be

## —Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards

Delaram Golpayegani, Harshvardhan Pandit, Dave Lewis  
[sgolpays@tcd.ie](mailto:sgolpays@tcd.ie), [harshvardhan.pandit@dcu.ie](mailto:harshvardhan.pandit@dcu.ie), [delewis@tcd.ie](mailto:delewis@tcd.ie)

Taxonomy: <https://w3id.org/vair>

Slides



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



**Protect**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 813497 (PROTECT ITN) and is funded by Science Foundation Ireland through the SFI Research Centres Programme through Grant#13/RC/2106\_P2.

