Artificial Intelligence Act

Dr. Tatjana EVAS
European Commission DG CNECT
7 June 2023

Overview

1. AI Act: Fundamentals
2. AI Act: Cybersecurity of AI systems
3. AI Act interplay with a proposal for the Cyber Resilience Act
4. AI Act: the state of play of legislative negotiations and the timeline

1. The AI Act – Fundamentals

The AI Act – Fundamentals

1. **Horizontal Approach**

2. **Product-safety legislation logic + fundamental rights**

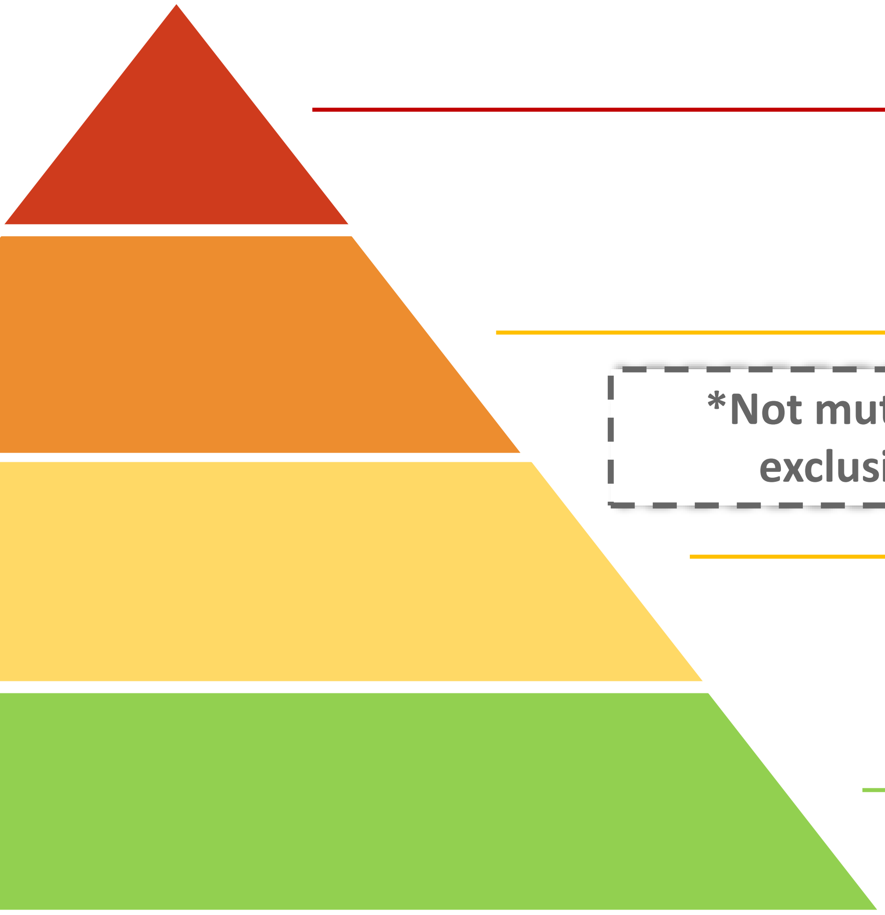
3. **Risk-based Approach**

4. **Set of requirement (including cybersecurity) for high-risk AI systems**

5. **Standards play a key role and must be developed with the broad-base participation of stakeholders and take due account of fundamental rights and Union values**

Risk-based + horizontal approach

Parliament & Council agree



Type equation here.

Unacceptable risk
e.g. social scoring — **Prohibited**

High risk
e.g. recruitment, medical devices — **Permitted** subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

'Transparency' risk
'Impersonation' (bots), deep fake — **Permitted** but subject to information/transparency obligations

Minimal or no risk — **Permitted** with no restrictions

The AI Act – same requirements for all high-risk use cases

High-risk use cases

1. Annex II. A (New legislative framework)
2. Annex II. B (Old approach legislation)
3. Annex III

3 categories but the same set-of requirements apply! for all high-risk AI systems

Difference: the ‘integration technique’ and type of conformity assessment (Annex II/ third-party; Annex III self-assessment)

2. The AI Act – Cybersecurity

Cybersecurity: The AI Act (the Commission proposal)

- **Cybersecurity is an important element of the requirements to ensure that high-risk AI systems are trustworthy!**

- **Key provisions: Article 15 (requirements), Article 42 (2) (presumption of conformity) + (recitals 49 and 51)**

“ (51) ***Cybersecurity plays a crucial role in ensuring that AI systems are resilient*** *against attempts to alter their use, behaviour, performance or compromise their security properties by malicious third parties exploiting the system’s vulnerabilities. Cyberattacks against AI systems can leverage AI specific assets, such as training data sets (e.g. data poisoning) or trained models (e.g. adversarial attacks), or exploit vulnerabilities in the AI system’s digital assets or the underlying ICT infrastructure. To ensure a level of cybersecurity appropriate to the risks, suitable measures should therefore be taken by the providers of high-risk AI systems, also taking into account as appropriate the underlying ICT infrastructure.”*

Cybersecurity: The AI Act (The Commission proposal + EP additions)

Article 15 – paragraph 1

High-risk AI systems shall be designed and developed **following the principle of security by design and by default**. In the light of their intended purpose, **they should achieve** an appropriate level of accuracy, robustness, safety, and cybersecurity, and perform consistently in those respects throughout their lifecycle. **Compliance with these requirements shall include implementation of state-of-the-art measures, according to the specific market segment or scope of application.**

Article 15 – paragraph 4 (3)

The technical solutions to address AI specific vulnerabilities shall include, where appropriate, measures to prevent***, detect, respond to, resolve*** and control for attacks trying to manipulate the training dataset ('data poisoning'), ***or pre-trained components used in training ('model poisoning'),*** inputs designed to cause the model to make a mistake ('adversarial examples' ***or 'model evasion'***), ***confidentiality attacks*** or model flaws***, which could lead to harmful decision-making***.

Cybersecurity: The AI Act (Commission proposal + EP additions)

Article 42 – paragraph 2

High-risk AI systems that have been certified or for which a statement of conformity has been issued under a cybersecurity scheme pursuant to Regulation (EU) 2019/881 of the European Parliament and of the Council (The EU Cybersecurity Act) and the references of which have been published in the Official Journal of the European Union shall be presumed to be in compliance with the cybersecurity requirements set out in Article 15 of this Regulation***, where applicable,*** in so far as the cybersecurity certificate or statement of conformity or parts thereof cover those requirements.

ARTICLE 15: EP ADDS TWO NEW PROVISIONS ON BENCHMARKING AND ENISA

Article 15 (Accuracy, robustness and cybersecurity)

[benchmarking]

1a. To address the technical aspects of how to measure the appropriate levels of accuracy and robustness set out in paragraph 1 of this Article, the AI Office shall bring together national and international metrology and benchmarking authorities and provide non-binding guidance on the matter as set out in Article 56, paragraph 2, point (a).

[ENISA]

1b. To address any emerging issues across the internal market with regard to cybersecurity, the European Union Agency for Cybersecurity (ENISA) shall be involved alongside the European Artificial Intelligence Board as set out Article 56, paragraph 2, point (b)

(recital 49) Performance metrics and their expected level should be defined with the primary objective to mitigate risks and negative impact of the AI system.[…] ***While standardisation organisations exist to establish standards, coordination on benchmarking is needed to establish how these standardised requirements and characteristics of AI systems should be measured.***

The European Commission
Standardization request in support of trustworthy artificial intelligence

COMMISSION IMPLEMENTING DECISION on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence

Reference C(2023)3215
Date 22/05/2023

https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en

Standardization request in support of trustworthy artificial intelligence

The scope and legal requirements

- Basis for future **harmonised standards**
- **10 new areas of European standards and European standardisation deliverables to be drafted** to support the implementation of the proposed **Artificial Intelligence Act**
- To cover **technical areas linked to the requirements** covered in proposal for an AI Act as well as conformity assessment of AI systems and quality management systems. Supporting standards should be also available (e.g. on terminology) when necessary for the implementability of the technical specifications *(Annex I)*
- **Focus is on risks which are common (horizontal) across AI systems**, however, vertical specifications, as appropriate, for some specific AI systems (use cases) or sectors in particularly in the fields of human oversight and accuracy. *(Annex II)*

Standardization request in support of trustworthy artificial intelligence

Key elements

- **Timeline** - deliverables by 30 April 2025 (Article 1)
- **Addressed to CEN/CENELEC**, however work of ETSI to be taken into account and process to be established for leveraging on ETSI experience and work *(Article 1 and Article 2; recitals 9 to 16)*
- **Representation and participation** of the relevant stakeholders, including SMEs, and societal stakeholders *(Article 2 (a))*
- **Fundamental rights and data protection** to be taken into account *(Article 2 (b))*
- **Leveraging on the existing knowledge and ongoing efforts at the EU and international levels**. This however should not bring any prejudice to the full alignment of standards with EU values and specificities *(Article 2 (c); recitals 8 and 16)*

AI
ARTIFICIAL INTELLIGENCE
EXCELLENCE & TRUST

Cybersecurity: Statndardisation Request

SR8: **European standard(s) and/or European standardisation deliverable(s) on cybersecurity specifications for AI systems**

2.8 Cybersecurity specifications for AI systems

- **[AIM]** This (these) European standard(s) or European standardisation deliverable(s) shall provide suitable organisational and technical solutions, to ensure that AI systems are resilient against attempts to alter their use, behaviour, or performance or to compromise their security properties by malicious third parties exploiting the AI systems' vulnerabilities.
- **[COVERAGE]** Organisational and technical solutions shall therefore include, where appropriate, **measures to prevent and control cyberattacks trying to manipulate AI specific assets, such as training data sets (e.g. data poisoning) or trained models (e.g. adversarial examples), or trying to exploit vulnerabilities in an AI system's digital assets or the underlying ICT infrastructure. These technical solutions shall be appropriate to the relevant circumstances and risks.**
- **[CRA]** This (these) European standard(s) or European standardisation deliverable(s) shall take due account of the essential requirements for products with digital elements as listed in Sections 1 and 2 of Annex I to the proposal for a Regulation of the European Parliament and the Council on horizontal cybersecurity requirements for products with digital elements.

3. The AI Act – Interplay with the CRA proposal

Cybersecurity: AIA and CRA interplay

Article 8: The Cyber Resilience Act proposal

(1) Presumption of conformity: high-risk AI systems compliant with the essential requirements of the CRA are deemed to be compliant with the cybersecurity requirement of AIA.

Products with digital elements classified as high-risk AI systems according to Article 6 of the proposed AI Act which fall within the scope of the proposed Cyber Resilience Act Regulation should comply with the essential requirements set out in the proposed Cyber Resilience Act. When those high-risk AI systems fulfil the essential requirements of the Cyber Resilience Act, they should be deemed compliant with the cybersecurity requirements set out in Article 15 of the proposed AI Act in so far as those requirements are covered by the EU declaration of conformity or parts thereof issued under the proposed Cyber Resilience Act.

Cybersecurity: AIA and CRA interplay

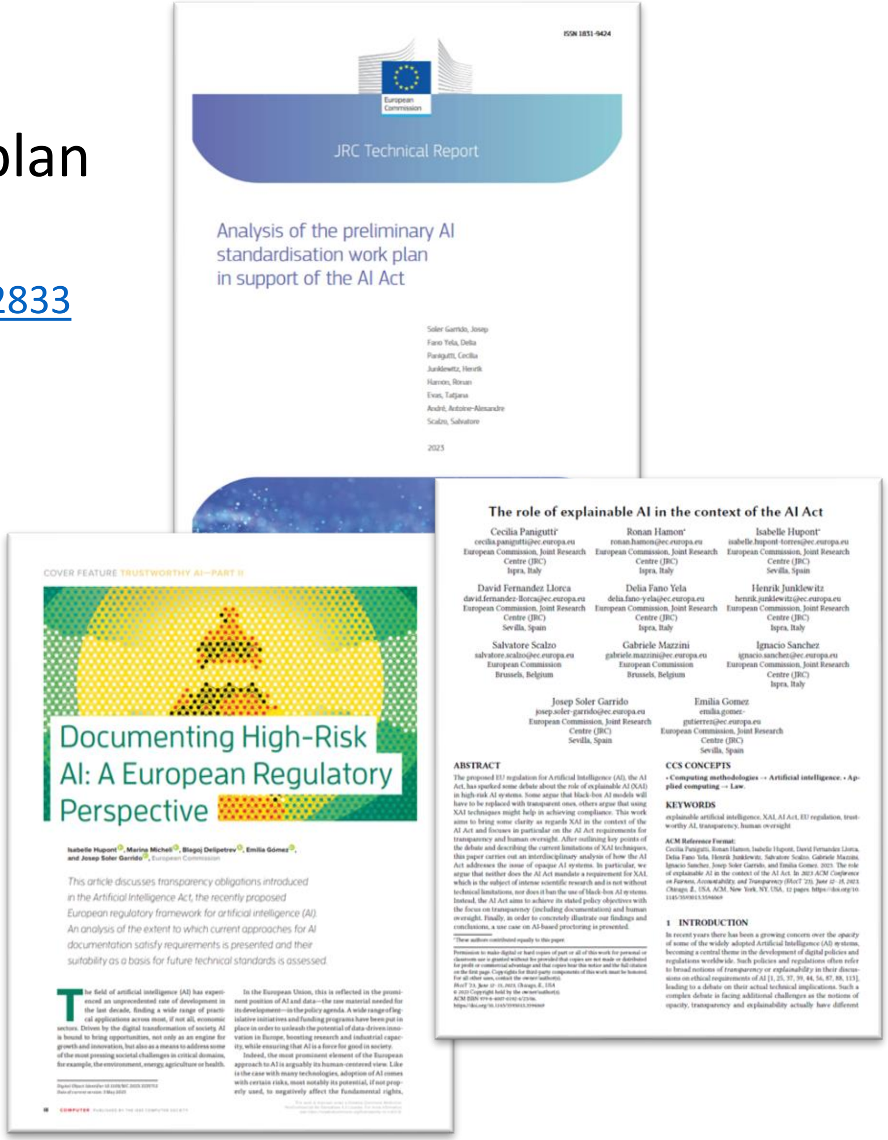
Article 8: The Cyber Resilience Act proposal

(2) Conformity Assessment Procedure. The general rule – the AIA is a reference act, i.e. the procedure under Article 43 AIA is to be followed.

Exception, high risk AI systems qualified as critical products under the CRA + subject to Annex VI (conformity assessment based on internal control), in this case the conformity assessment provisions of the Cyber Resilience Act apply to the essential requirements + for all the other aspects covered by the AI Regulation the respective provisions on conformity assessment based on internal control set out in Annex VI to the AI Regulation apply.

Joint Research Centre – published and upcoming studies

- Analysis of the preliminary AI standardisation work plan in support of the AI Act
  - JRC Report: https://publications.jrc.ec.europa.eu/repository/handle/JRC132833
- AI cybersecurity in the AI Act
  - JRC Report – Forthcoming
- Documenting High-risk AI: A European Regulatory Perspective
  - IEEE Computer: https://ieeexplore.ieee.org/document/10109295
- The role of explainable AI in the context of the AI Act
  - ACM FAccT - 2023 – Forthcoming (June 2023)



ISSN 1831-9424

European Commission

JRC Technical Report

Analysis of the preliminary AI
standardisation work plan
in support of the AI Act

Soler Garrido, Josep
Fano Yela, Delia
Panigutti, Cecilia
Junklewitz, Henrik
Hamon, Ronan
Evas, Tatjana
André, Antoine-Alexandre
Scalzo, Salvatore

2023

COVER FEATURE TRUSTWORTHY AI—PART II

Documenting High-Risk AI: A European Regulatory Perspective

Isabelle Hupont, Marina Micheli, Blagoj Delipetrev, Emilia Gómez, and Josep Soler Garrido, European Commission

This article discusses transparency obligations introduced in the Artificial Intelligence Act, the recently proposed European regulatory framework for artificial intelligence (AI). An analysis of the extent to which current approaches for AI documentation satisfy requirements is presented and their suitability as a basis for future technical standards is assessed.

The field of artificial intelligence (AI) has experienced an unprecedented rate of development in the last decade, finding a wide range of practical applications across most, if not all, economic sectors. Driven by the digital transformation of society, AI is bound to bring opportunities, not only as an engine for [illegible]

In the European Union, this is reflected in the prominent position of AI and data—the raw material needed for its development—in the policy agenda. A wide range of legislative initiatives and funding programs have been put in place in order to unleash the potential of data-driven innovation in Europe, boosting research and industrial capac- [illegible]

The role of explainable AI in the context of the AI Act

Cecilia Panigutti*
cecilia.panigutti@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Ronan Hamon*
ronan.hamon@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Isabelle Hupont*
isabelle.hupont-torres@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

David Fernandez Llorca
david.fernandez-llorca@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

Delia Fano Yela
delia.fano-yela@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Henrik Junklewitz
henrik.junklewitz@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Salvatore Scalzo
salvatore.scalzo@ec.europa.eu
European Commission
Brussels, Belgium

Gabriele Mazzini
gabriele.mazzini@ec.europa.eu
European Commission
Brussels, Belgium

Ignacio Sanchez
ignacio.sanchez@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Josep Soler Garrido
josep.soler-garrido@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

Emilia Gomez
emilia.gomez-gutierrez@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

ABSTRACT

The proposed EU regulation for Artificial Intelligence (AI), the AI Act, has sparked some debate about the role of explainable AI (XAI) in high-risk AI systems. Some argue that black-box AI models will have to be replaced with transparent ones, others argue that using XAI techniques might help in achieving compliance. This work aims to bring some clarity as regards XAI in the context of the AI Act and focuses in particular on the AI Act requirements for [illegible]

CCS CONCEPTS

• Computing methodologies → Artificial intelligence. • Applied computing → Law.

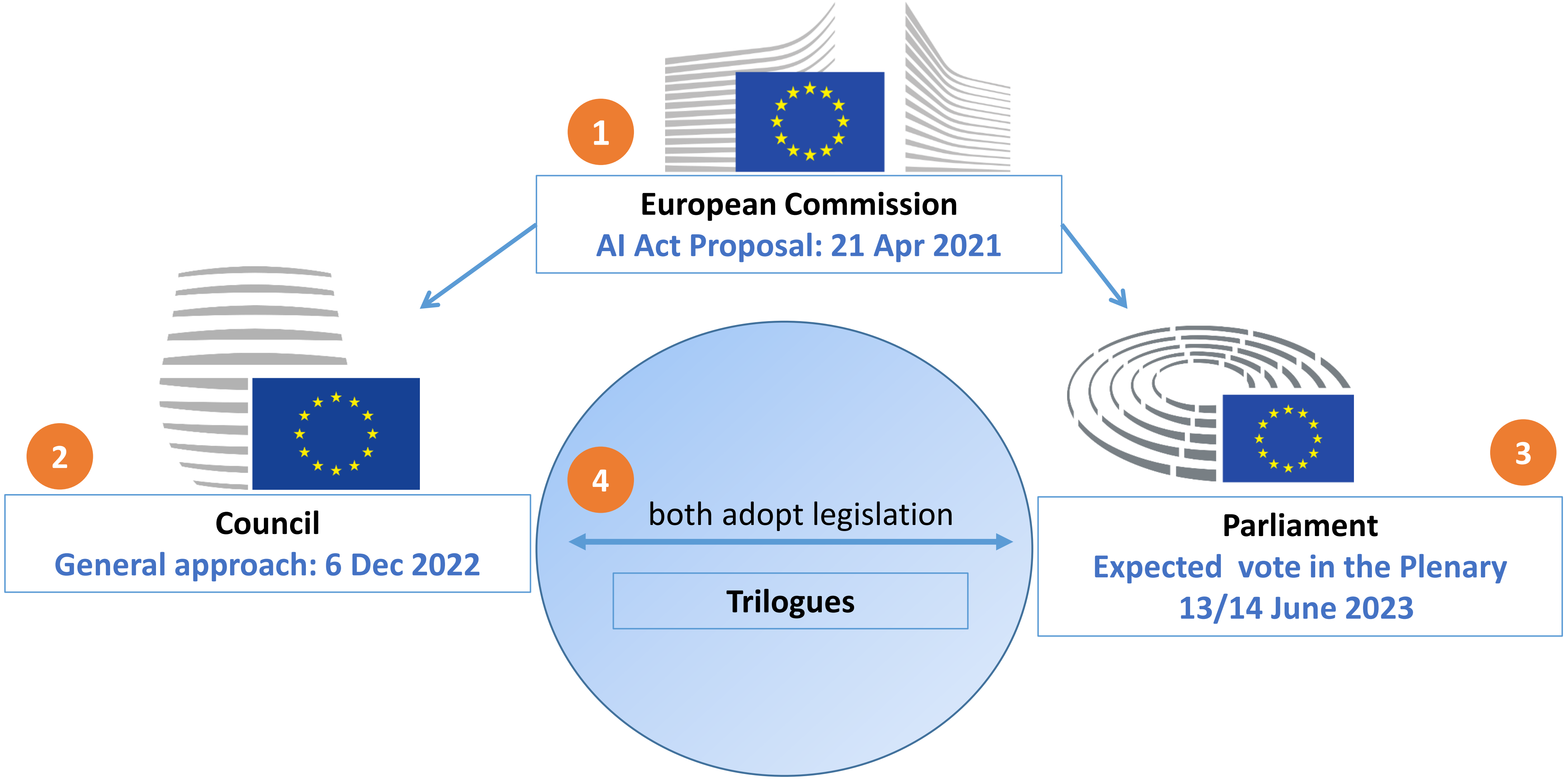
KEYWORDS

explainable artificial intelligence, XAI, AI Act, EU regulation, trustworthy AI, transparency, human oversight

1 INTRODUCTION

4. AI Act – State of Play

AI Act: State of Play (ordinary legislative procedure)



1

European Commission

AI Act Proposal: 21 Apr 2021

2

Council

General approach: 6 Dec 2022

3

Parliament

Expected vote in the Plenary
13/14 June 2023

4

both adopt legislation

Trilogues

Thank you!