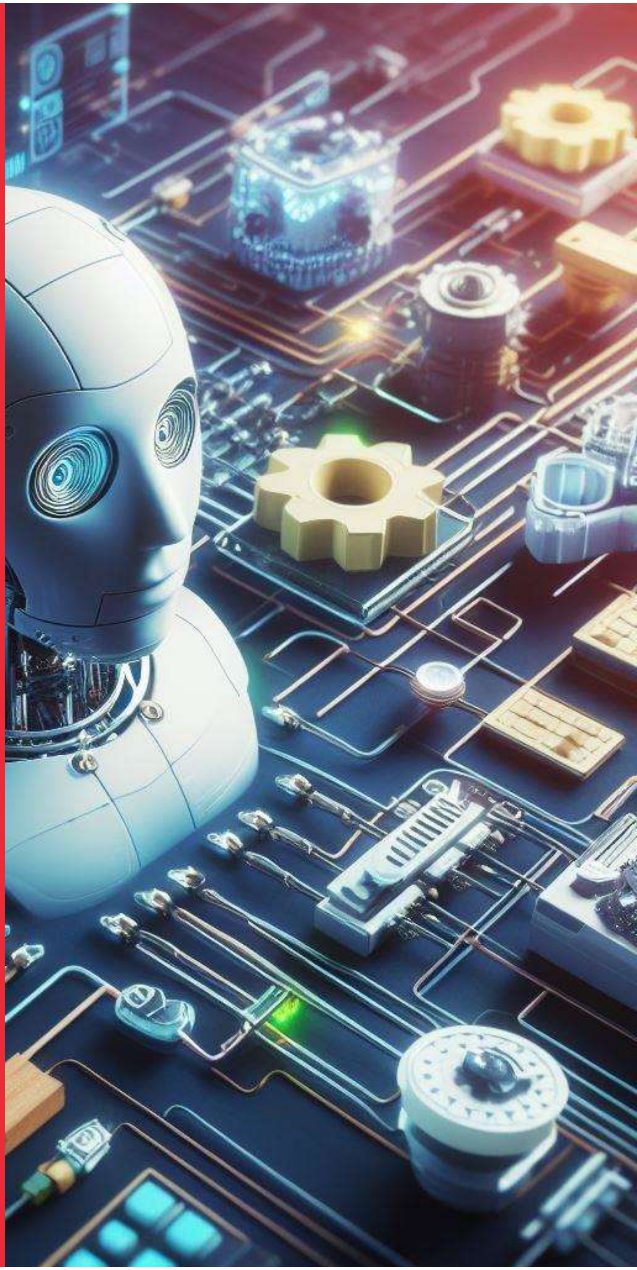
Antwerp Center on Responsible AI

University of Antwerp

The AI Act: What to expect?

prof. dr. Jan Blockx

21 November 2023



EU initiatives re AI

2018: European strategy for AI

+ creation of a High-Level Expert Group

2019: HLEG produces:

Ethics Guidelines for Trustworthy AI

Key requirements

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

2020: White paper on AI

April 2021: Proposal for an AI Act

Chips Act (proposal 2022)

Encouraging AI R&D networks

Digital Markets Act (2022)

Data Act (proposal 2022)

Data Governance Act (2022)

AI Liability Directive (proposal 2022)

R&D financing by EU and MS

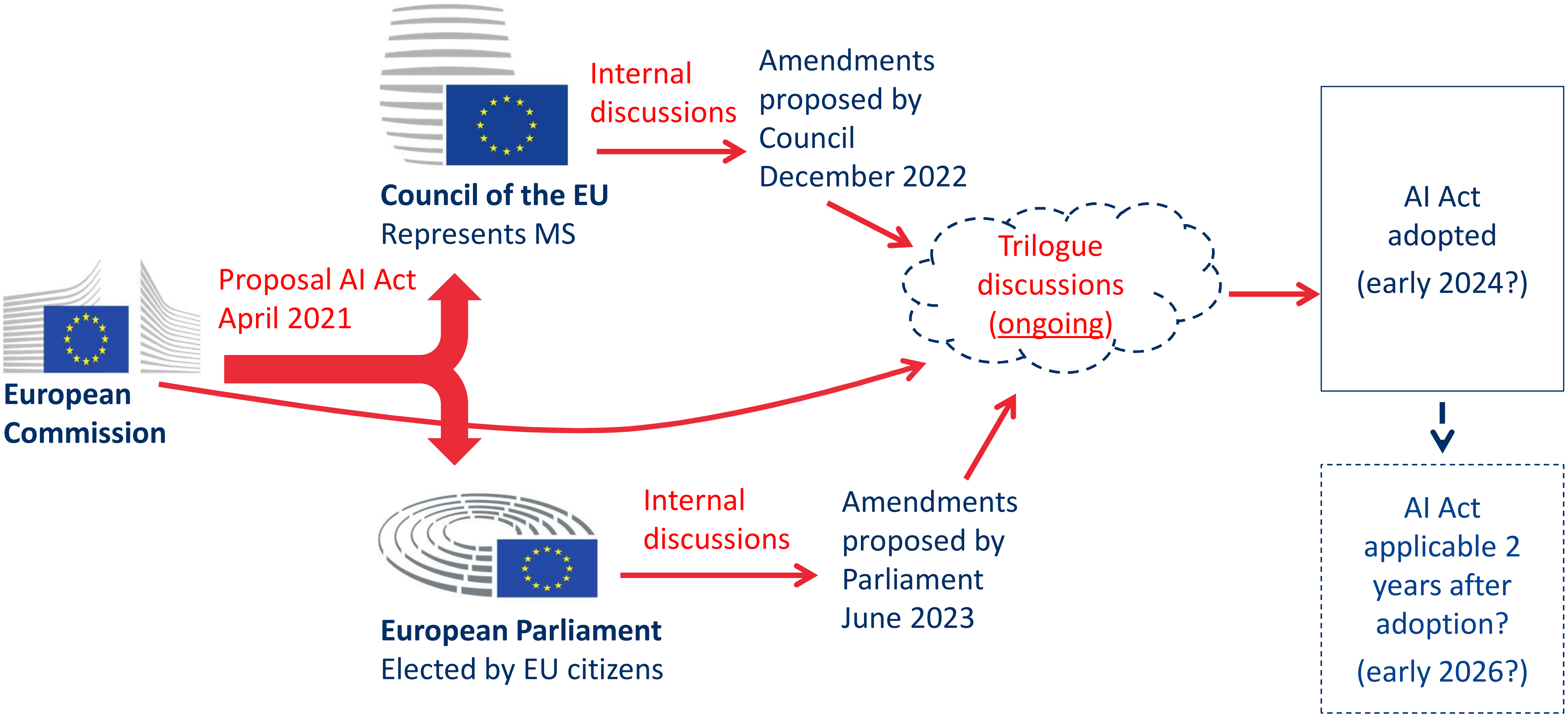
Cybersecurity initiatives

Encouraging ICT skills training

Digital Services Act (2022)

University of Antwerp

AI Act: legislative process



European Commission

Proposal AI Act
April 2021

Council of the EU
Represents MS

Internal
discussions

Amendments
proposed by
Council
December 2022

European Parliament
Elected by EU citizens

Internal
discussions

Amendments
proposed by
Parliament
June 2023

Trilogue
discussions
(ongoing)

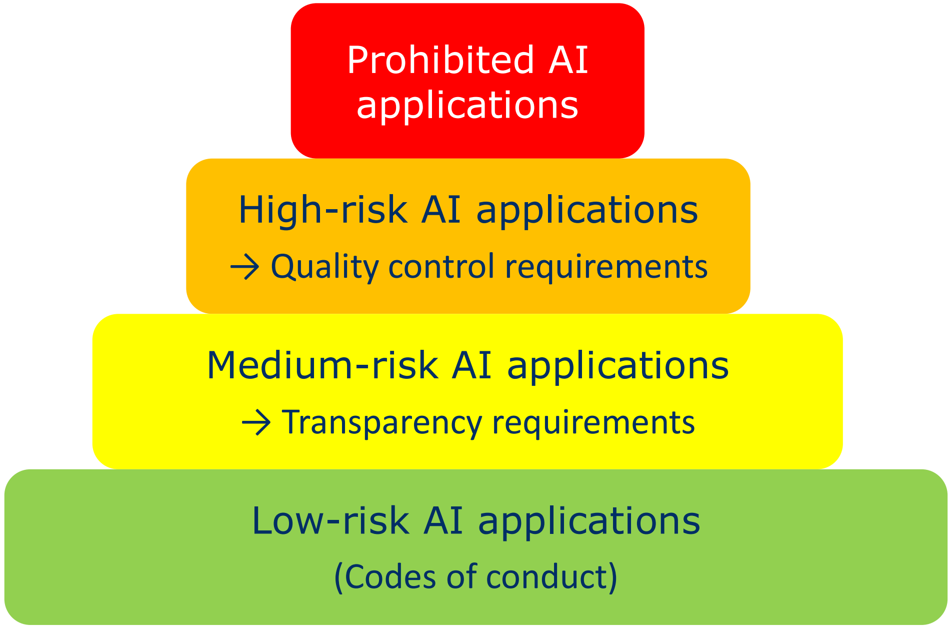
AI Act
adopted
(early 2024?)

AI Act
applicable 2
years after
adoption?
(early 2026?)

EU initiatives re AI: general approach

- Encouraging development & uptake of AI
- Ensuring AI remains under human supervision & is ‘a force for the good’

⇒ A risk-based approach distinguishing:



Prohibited AI applications

High-risk AI applications

→ Quality control requirements

Medium-risk AI applications

→ Transparency requirements

Low-risk AI applications

(Codes of conduct)

Proposed AI Act: definition of Artificial Intelligence

EC proposal

(i) software

(ii) developed for <u>human-defined objectives</u>,

(iii) generating outputs such as content, predictions, recommendations, or decisions to interact with the environment; and

(iv) using <u>one of the following techniques</u>:

- Supervised, unsupervised and reinforcement machine learning
- Logic- and knowledge-based approaches, including inductive (logic) programming, and expert systems;
- Statistical approaches

Proposed AI Act: definition of Artificial Intelligence

Council and Parliament amendments

- Emphasize level of autonomy of system
- Also cover systems with implicit objectives

⇨ Final text will likely be in line with:

OECD Recommendation on AI (2019)

"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"

University of Antwerp

Proposed AI Act: prohibited AI applications

EC proposal

- Manipulation:
  1. Using subliminal techniques
  2. Exploiting vulnerabilities of age, physical or mental disability

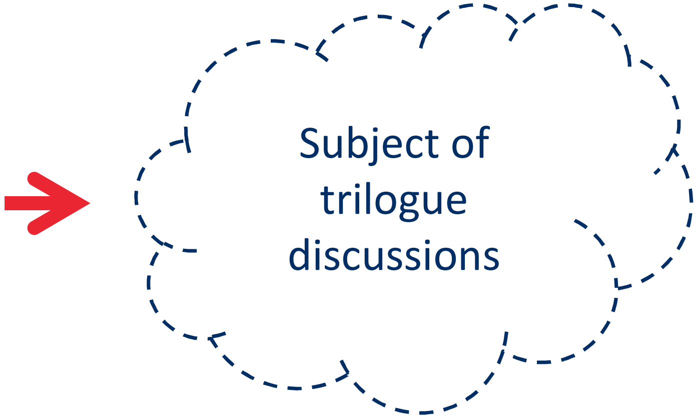
  in order to materially distort a person's behaviour

  in a manner that causes physical or psychological harm
- Surveillance measures:
  3. 'social scoring': evaluating trustworthiness of natural persons by public authorities which lead to either unjustified or disproportionate treatment of individuals or groups, or detrimental treatment in another context
  4. use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement
     - except to identify perpetrators of serious offences (at least three years imprisonment), specific victims (e.g. missing children) or prevent imminent threat to life

Proposed AI Act: prohibited AI applications

Parliament amendments

- No exceptions to ban on real-time biometric identification systems for law enforcement
- Addition of:
  - Criminal risk assessment instruments
  - Emotion recognition for law enforcement, border management, in workplace and education
  - …



Subject of trilogue discussions

Proposed AI Act: high-risk AI applications

EC proposal

1. AI systems that are products or safety components of products already covered by certain Union health and safety harmonisation legislation listed in Annex II (such as toys, machinery, lifts, or medical devices)

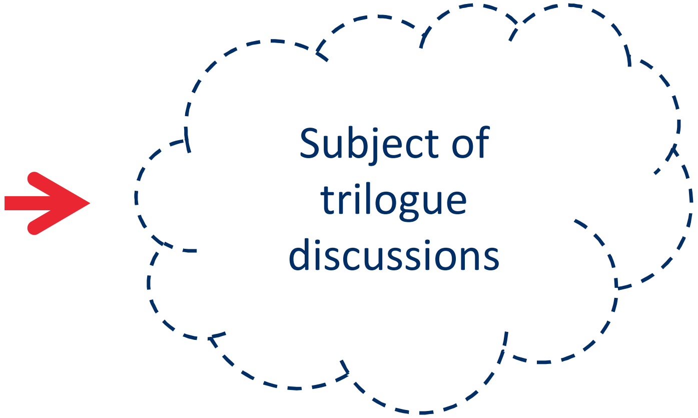
   → but aerial, maritime and road vehicles are excluded

2. AI systems for use in specified fields listed in Annex III:
   - biometric identification;
   - management of critical infrastructure;
   - education;
   - employment;
   - access to essential services and benefits;
   - law enforcement;
   - migration, asylum, border management;
   - administration of justice and democracy.

Proposed AI Act: high-risk AI applications

Parliament amendment

- Additional types of AI systems added to Annex III:
  - Exam fraud detection systems
  - Systems to determine the eligibility of natural persons for health and life insurance
  - Systems aimed at influencing the outcome of an election or voting behaviour
  - Very large social media platforms
  - …



Subject of trilogue discussions

Proposed AI Act: requirements for high-risk AI apps

EC proposal

- Providers of high-risk AI systems:
  - Need to set up a risk management system, assessing
    - Quality of (training) data
    - Accuracy, robustness and cybersecurity of the system
  - Need to log operation of system
  - Need to ensure human oversight
  - Need to ensure documentation and transparency to users
  - Need to register the AI system
- More limited obligations for distributors, deployers, etc.

Parliament amendments include further detail as well as attention for environmental concerns

Proposed AI Act: foundation models

Not a concern yet at the time of the EC proposal

- Parliament proposed to extend most requirements for high-risk AI applications to foundation models
- October **trilogue discussions** moved to a ‘tiered’ approach
  - All foundation models: transparency obligations re modelling and training process
  - Foundation models with ‘advanced’ capabilities: additional requirements, incl. external audits, risk assessment, etc.
- But Council now wants a lighter approach with only codes of conduct re transparency

Parliament disagrees

Proposed AI Act: medium-risk AI applications

1. Those that interact with humans

   → Provider needs to disclose that it is a bot

2. Those that detect emotions or categorize biometric data

   → User needs to disclose this

3. Generation or manipulation of content (‘deep fakes’)

   → User needs to disclose this

Proposed AI Act: enforcement

- EC proposal and Council leaves this largely to national authorities
- Parliament wants a European AI Office

Latest press reports about trilogue

- Most enforcement by national authorities
- AI Office would ensure consistency, in particular re foundation models and general purpose AI

Conclusion

Few AI applications will be prohibited

But quite a few AI applications will be subject to compliance obligations

Including (some) foundation models (?)

All AI that interacts with humans would have to be identified as such

→ Impact on AI use and development in the EU?

University
of Antwerp