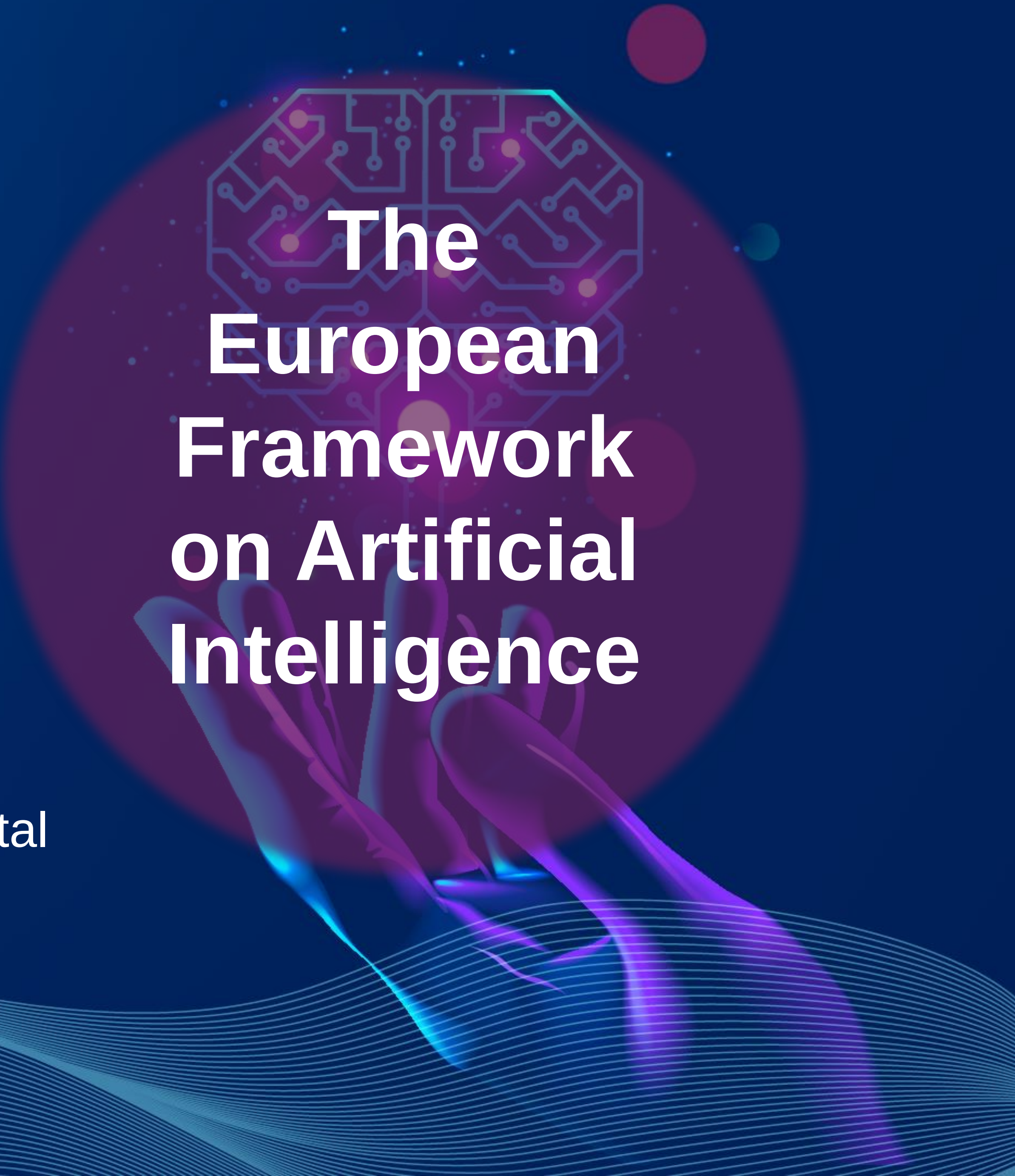
AI
ARTIFICIAL
INTELLIGENCE
EXCELLENCE
& TRUST



The European Framework on Artificial Intelligence

Dr.Tatjana Evas
DG CNECT, European Commission

Trustworthy AI regulations and their industrial/societal implications
4 April 2024

Agenda

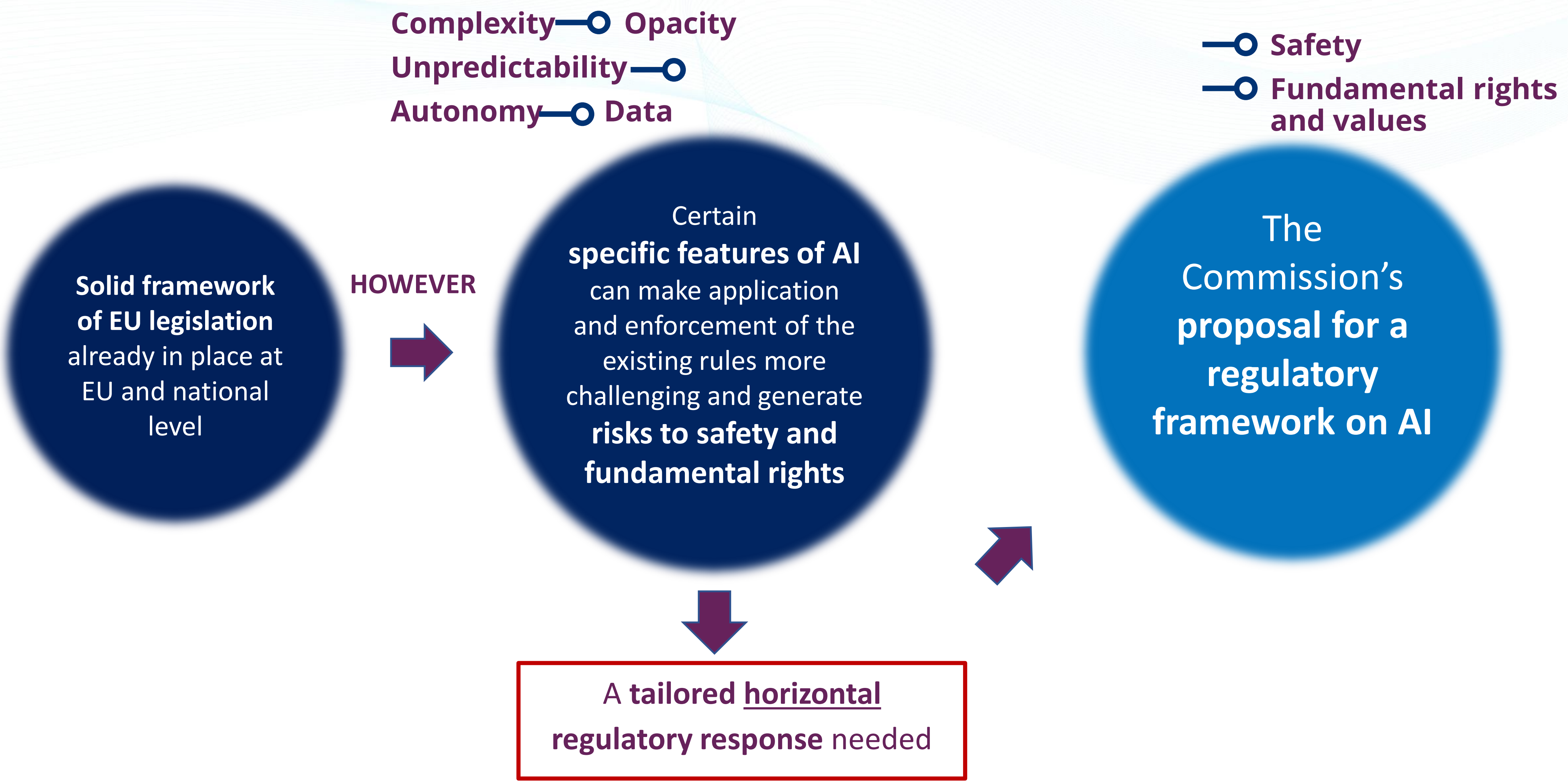
1. Introduction: Goals and Regulatory Framework of the AI Act

2. Key policy choices
   - Horizontal framework
   - Risk-based and technology neutral approach
   - Project safety logic

3. Conclusions: The AI Act is the optimal policy incentive framework for trustworthy AI

The Artificial Intelligence Act

Main Reasons for the regulation



Complexity Opacity

Unpredictability

Autonomy Data

Solid framework of EU legislation already in place at EU and national level

HOWEVER

Certain **specific features of AI** can make application and enforcement of the existing rules more challenging and generate **risks to safety and fundamental rights**

A **tailored horizontal regulatory response** needed

Safety

Fundamental rights and values

The Commission's **proposal for a regulatory framework on AI**

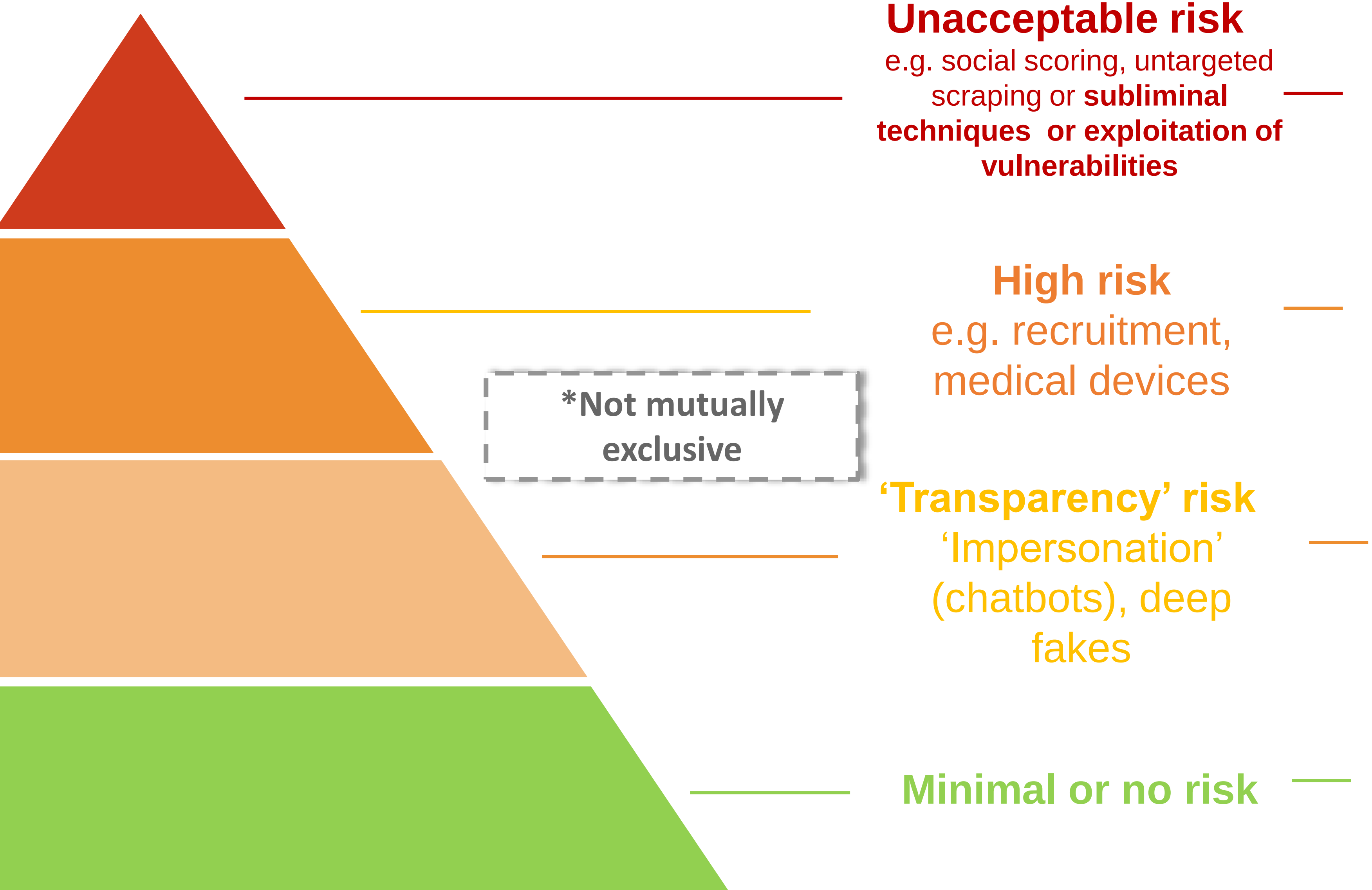
The AI Act: the main policy choices

I. Horizontal framework

II. Risk-based approach

III. Product safety

Risk Based Approach



Unacceptable risk
e.g. social scoring, untargeted scraping or **subliminal techniques or exploitation of vulnerabilities**

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' (chatbots), deep fakes

Permitted but subject to information/transparency obligations

Minimal or no risk

Permitted with no restrictions, voluntary codes of conduct possible

A very limited set of particularly harmful AI uses are banned

Unacceptable risk

Subliminal techniques or exploitation of vulnerabilities	to manipulate people
Social Scoring	for public and private purposes
Biometric categorisation	to deduce or infer for example race, political opinions, religious or philosophical beliefs or sexual orientation, exceptions for labelling in the area of law enforcement
Real-time remote biometric identification	for the purpose of law enforcement, -with narrow exceptions and with prior authorisation by a judicial or independent administrative authority
Individual predictive policing	assessing or predicting the risks of a natural person to commit a criminal offence based solely on this profiling without objective facts
Emotion recognition	in the workplace and education institutions, unless for medical or safety reasons
Untargeted scraping of the internet	or CCTV for facial images to build-up or expand databases

High-risk AI systems will have to comply with certain rules

High-risk use cases defined in Annexes I (embedded AI) and III:

Some examples from Annex III are related to

- **Certain critical infrastructures** such as road traffic, supply of water, gas, heating and electricity
- **Education and vocational training**, e.g. to evaluate learning outcomes
- **Employment, workers management**, e.g. to analyse job applications or evaluate candidates
- **Access to essential private and public services** and benefits, credit scoring
- **Remote biometric identification, categorization, emotion recognition; Law enforcement; border management; administration of justice and democratic processes**

Obligations for providers of high-risk AI systems:

- **Trustworthy AI requirements** such as risk management system, data quality, documentation and traceability, transparency, human oversight, accuracy, cybersecurity and robustness
- **Conformity assessment** before placing the AI system on the market, to demonstrate compliance
- **Quality management systems** to minimise risks for users and affected persons and to ensure compliance
- **Registration in an EU database**

This will be subject to **enforcement** to ensure that the high risk is effectively addressed.

The AI Act: The Main Operational Elements
High-Risk AI systems

New Legislative Framework (NLF) Product Safety Legislation +

↓

Sets

Mandatory Requirements for high-risk AI system before they can be used

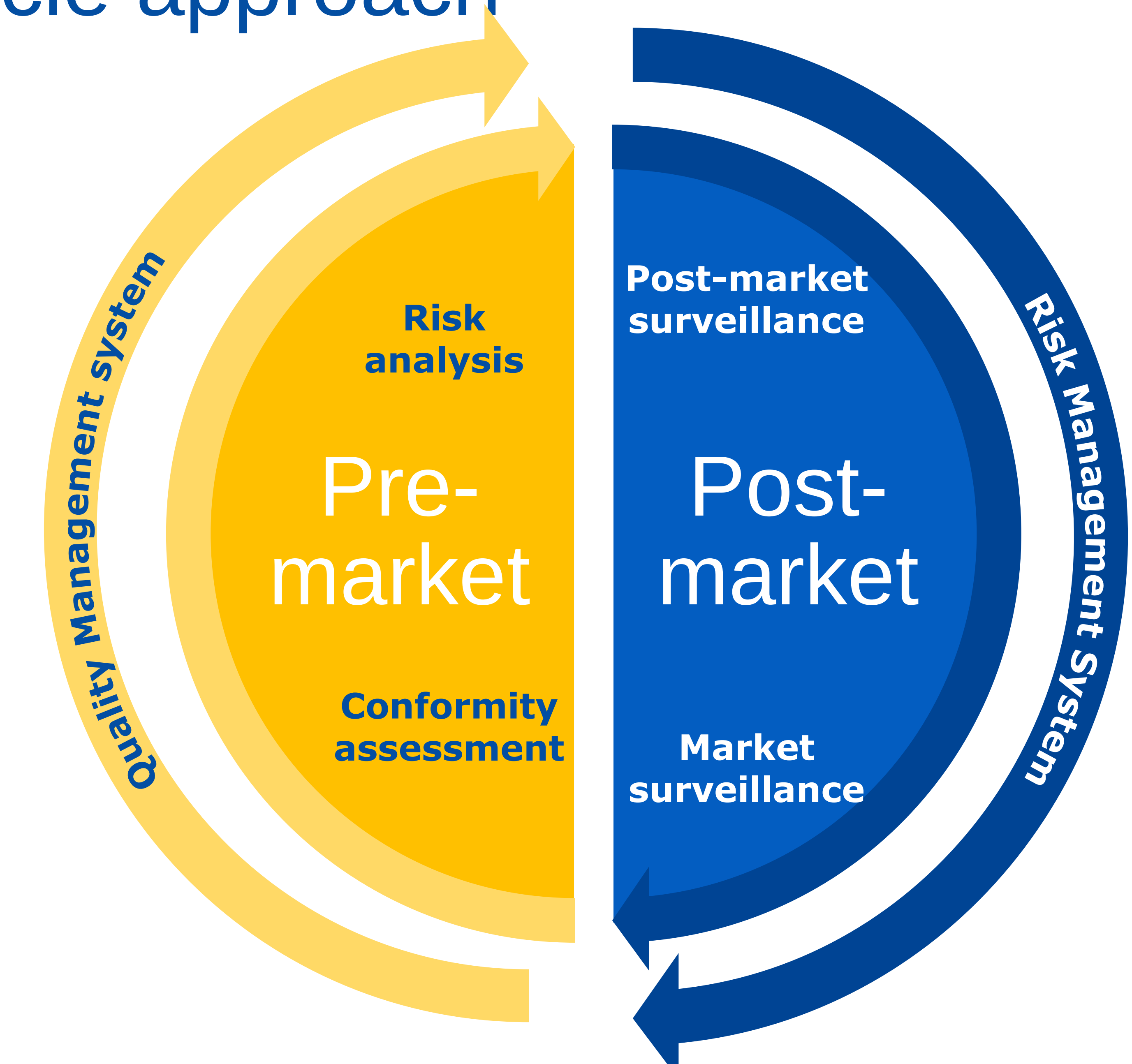
↓

To address AI specific risks triggered by AI characteristics, such as, Complexity, Opacity, Unpredictability, Autonomy and Data

risks to health, safety and fundamental rights

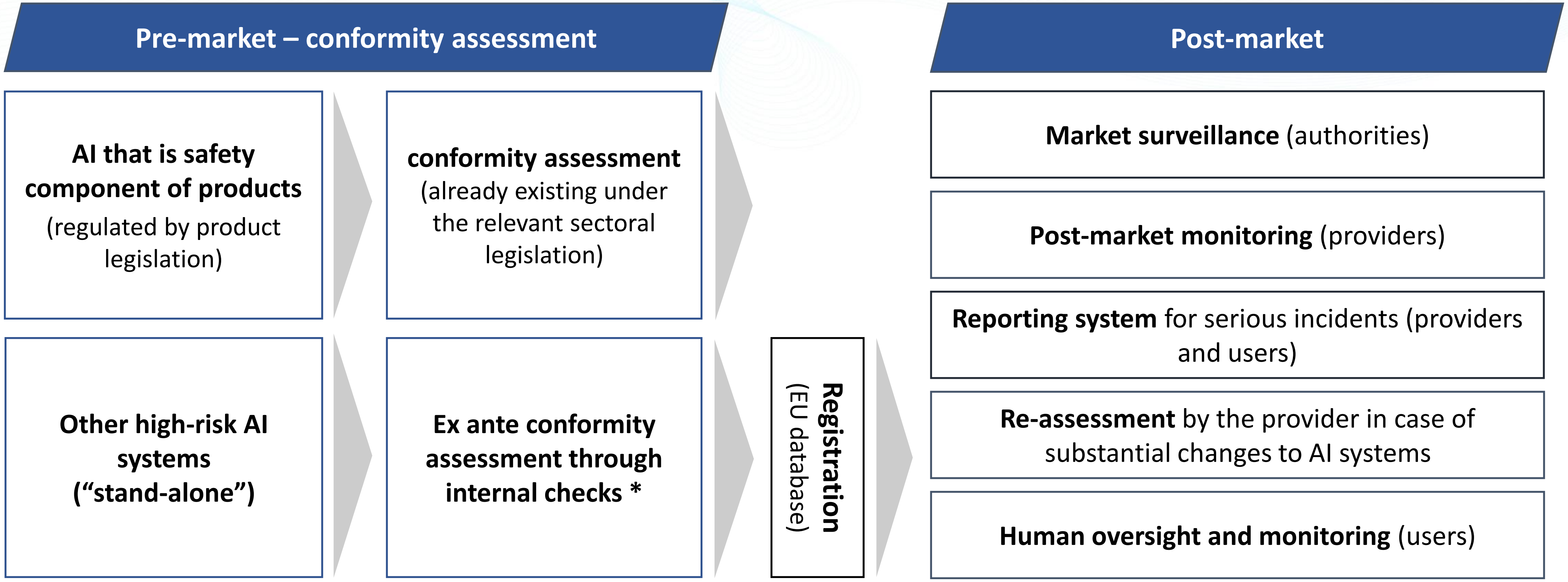
1. **risk management system** for AI systems *[Art. 9 AI Act]*
2. **governance and quality of datasets** used to build AI systems *[Art. 10 Data and data governance]*
3. **record keeping -** built-in logging capabilities in AI systems *[Art. 11 Technical documentation and Art. 12 record-keeping]*
4. **transparency and information** to the users of AI systems *[Art. 13 Transparency and provisions of information to users]*
5. **human oversight** of AI systems *[Art. 14 Human oversight]*
6. **accuracy** specifications for AI systems *[Art. 15 Accuracy, robustness and cybersecurity]*
7. **robustness** specifications for AI systems *[Art. 15 Accuracy, robustness and cybersecurity]*
8. **cybersecurity** specifications for AI systems *[Art. 15 Accuracy, robustness and cybersecurity]*
9. **quality management system** for providers of AI system *[Art. 17]*
10. **conformity assessment** for AI systems *[Art. 19 + Art. 43 Conformity Assessment]*

Life-cycle approach



European Commission

Compliance and enforcement system



Pre-market – conformity assessment

AI that is safety component of products (regulated by product legislation)

conformity assessment (already existing under the relevant sectoral legislation)

Other high-risk AI systems ("stand-alone")

Ex ante conformity assessment through internal checks *

Registration (EU database)

Post-market

Market surveillance (authorities)

Post-market monitoring (providers)

Reporting system for serious incidents (providers and users)

Re-assessment by the provider in case of substantial changes to AI systems

Human oversight and monitoring (users)

New special rules for General Purpose AI models (GPAI)

GPAI models: trained on large data, can competently perform wide range of tasks and be integrated in numerous downstream applications; research, development, and prototyping activities preceding the placement on the market are not covered.

All GPAI
(lower tier)

- Information and documentation requirements, mainly to achieve **transparency for downstream providers**
- Policy to respect copyright and a summary of the content used for training purposes
- **Free and open-source models are exempted** from transparency requirements, when they do not carry systemic risks except from the copyright-related obligations

GPAI with systemic risks
(higher tier)

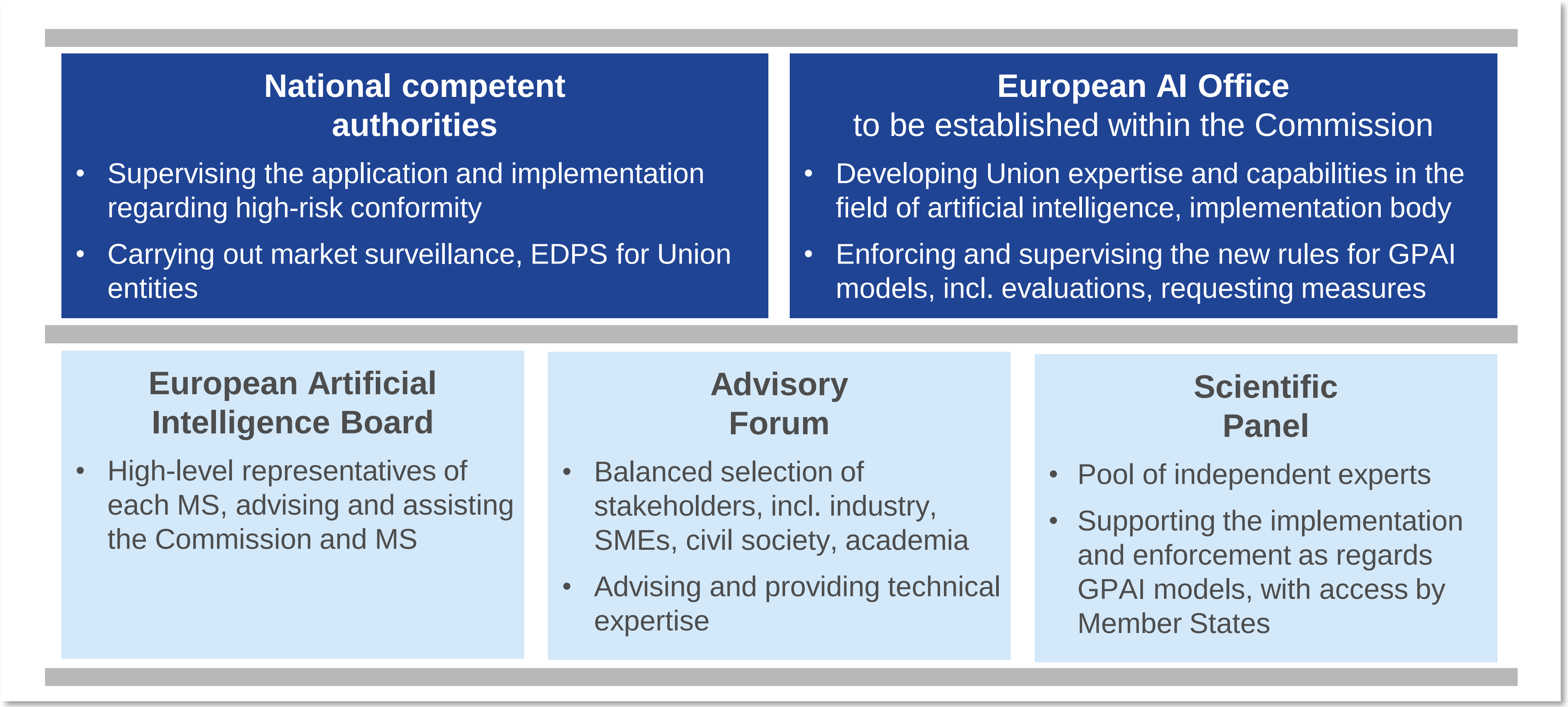
- **at least 10^25 FLOPs** or **designated by the AI Office** (e.g. based on benchmarks for capabilities, user count)
- All obligations from the lower tier **+ state-of-the-art model evaluations** (including red teaming / adversarial testing**), risk assessment and mitigation, incident reporting, cybersecurity and additional documentation**

updateable via delegated acts

- GPAI providers may rely on **Codes of Practice** to demonstrate compliance
- Codes of practice to be developed by industry under coordination of AI Office, the scientific community civil society and other experts also involved; the codes could be approved by COM through implementing act;
- New standardisation deliverable on GPAI to supersede the codes once EU harmonised standards available

A holistic structure ensures effective enforcement

Enforcement by national competent authorities and the AI Office with a supportive structure for close collaboration with Member States and for additional technical expertise



National competent authorities

- Supervising the application and implementation regarding high-risk conformity
- Carrying out market surveillance, EDPS for Union entities

European AI Office
to be established within the Commission

- Developing Union expertise and capabilities in the field of artificial intelligence, implementation body
- Enforcing and supervising the new rules for GPAI models, incl. evaluations, requesting measures

European Artificial Intelligence Board

- High-level representatives of each MS, advising and assisting the Commission and MS

Advisory Forum

- Balanced selection of stakeholders, incl. industry, SMEs, civil society, academia
- Advising and providing technical expertise

Scientific Panel

- Pool of independent experts
- Supporting the implementation and enforcement as regards GPAI models, with access by Member States

The AI Office: Mission and Tasks

Background:

- ❖ Clear need for EU-level governance system for AI (SotEU 2023)
- ❖ Political agreement on AI Act from 8 December introduces role of AI Office

- Responsibility to implement and enforce the AI Act, in particular rules on general-purpose AI models and systems
- Cooperate with all relevant EU bodies and Member States
- Collaboration with stakeholder community
- Cross-sectoral cooperation within the Commission
- Promote uptake of and innovation in AI with societal benefits
- **Coordinate and promote international cooperation on AI**

Definition of Artificial Intelligence System

> *“a machine-based system designed to operate with varying levels of **autonomy** and, that may exhibit **adaptiveness** after deployment and 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 ”*

- The definition is **fully aligned with 2023 revised OECD definition**
- The definition does not cover all software systems enabling automated processes or decisions (ADM) but only **a subset of ADMs.**
- The definition covers systems that are **build with one of the AI techniques** (incl. machine learning, logic and knowledge-based approaches)
- The key concept is a **capability to ‘infer how to generate outputs’** at the building stage of the system.
- The concept of **adaptiveness** which refers to learning refers to the stage of deployment. At the deployment stage adaptiveness could be 0.

European Commission

International activities and international cooperation and convergence is very important

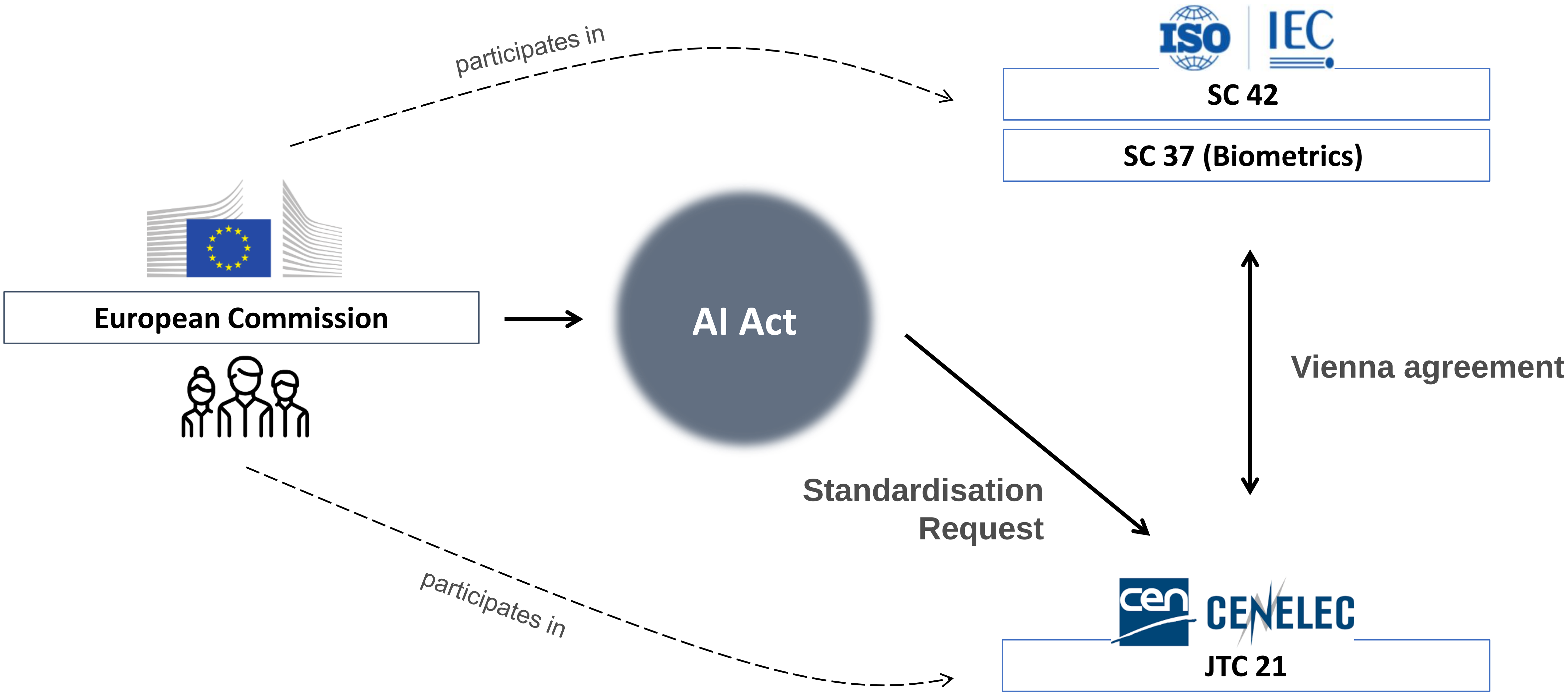
- Council of Europe Convention
- G7 Hiroshima process
- G20
- UN HLAB on AI
- AI Safety Summit
- OECD
- + bilateral cooperation

Definition of Artificial Intelligence System

> *“‘a machine-based system designed to operate with varying levels of **autonomy** and, that may exhibit **adaptiveness** after deployment and 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 ”*

- The definition is **fully aligned with 2023 revised OECD definition**
- The definition does not cover all software systems enabling automated processes or decisions (ADM) but only **a subset of ADMs.**
- The definition covers systems that are **build with one of the AI techniques** (incl. machine learning, logic and knowledge-based approaches)
- The key concept is a **capability to ‘infer how to generate outputs’** at the building stage of the system.
- The concept of **adaptiveness** which refers to learning refers to the stage of deployment. At the deployment stage adaptiveness could be 0.

Standardisation work



AI

ARTIFICIAL INTELLIGENCE

EXCELLENCE & TRUST

Thank you!

Tatjana.Evas@ec.Europa.eu