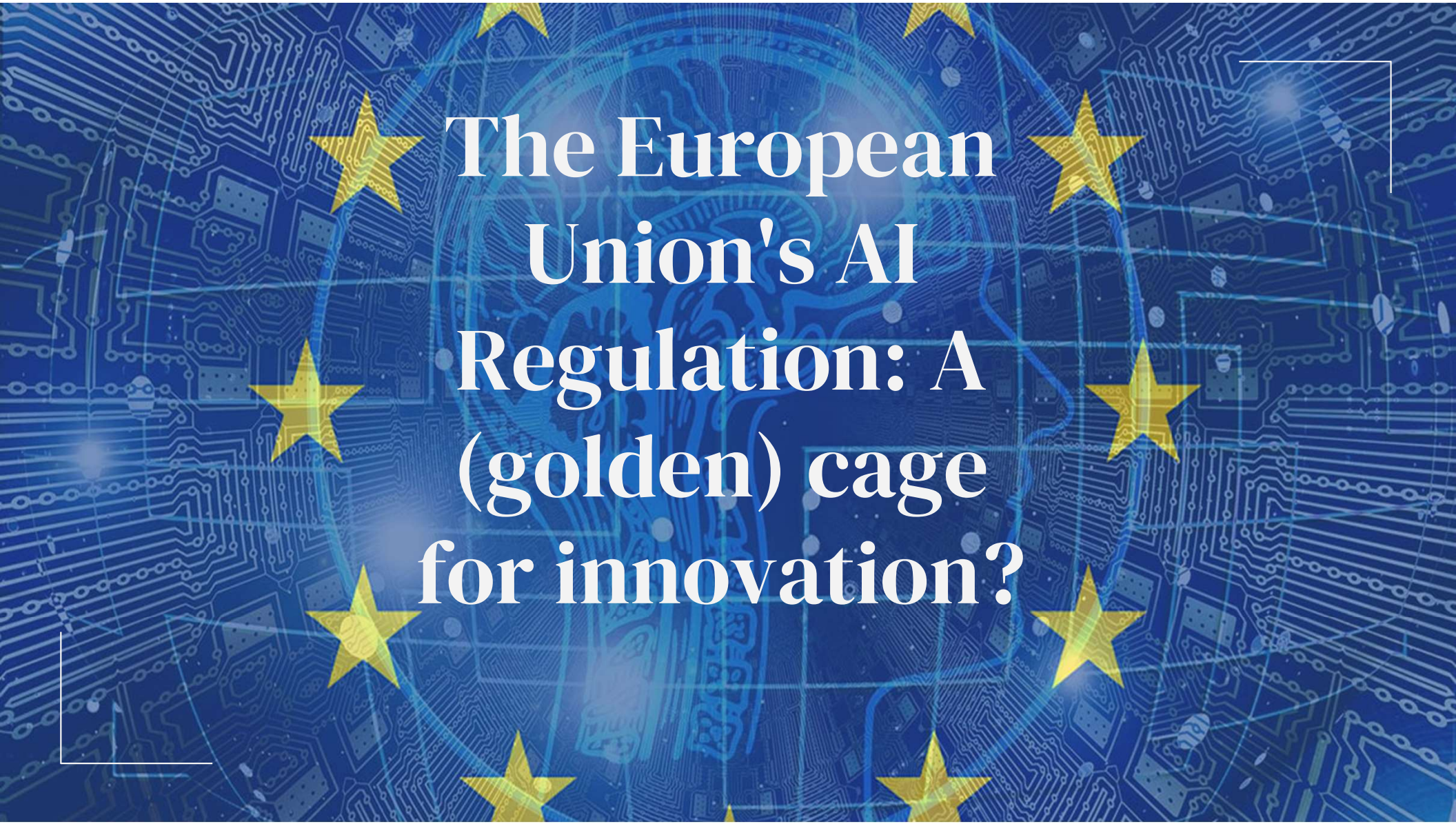
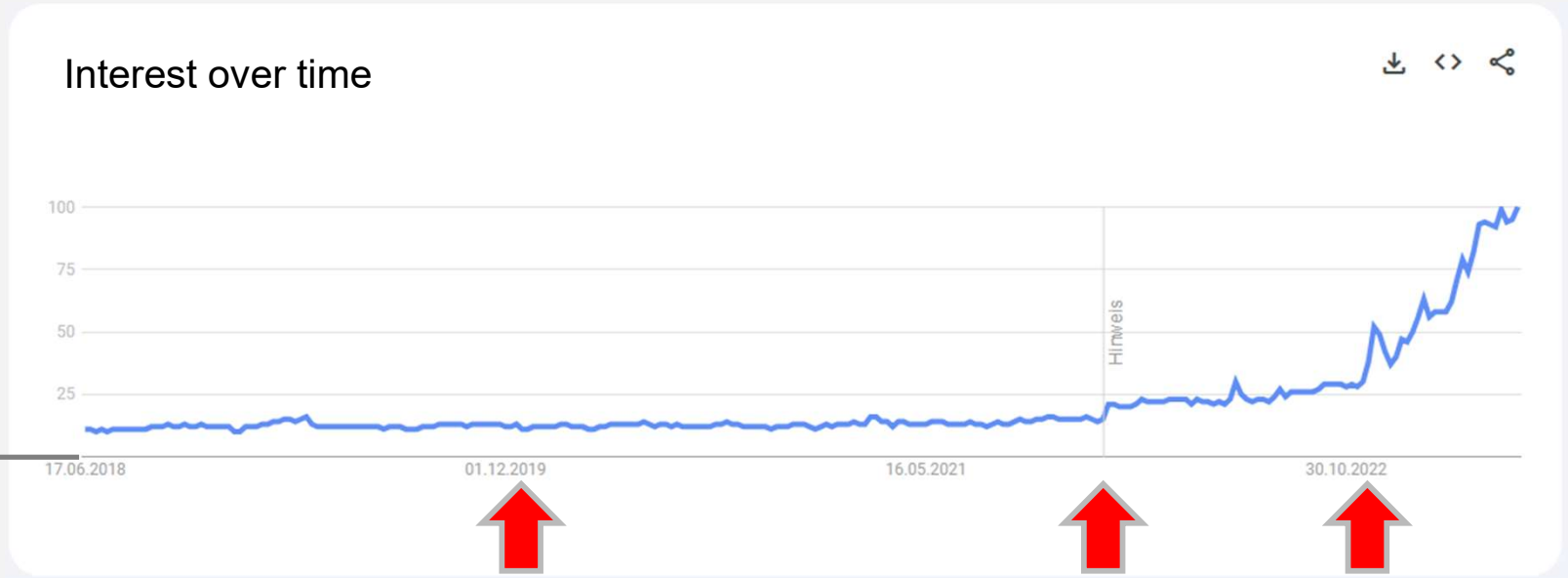


The European Union's AI Regulation: A (golden) cage for innovation?

Interest in AI



27.04.2016

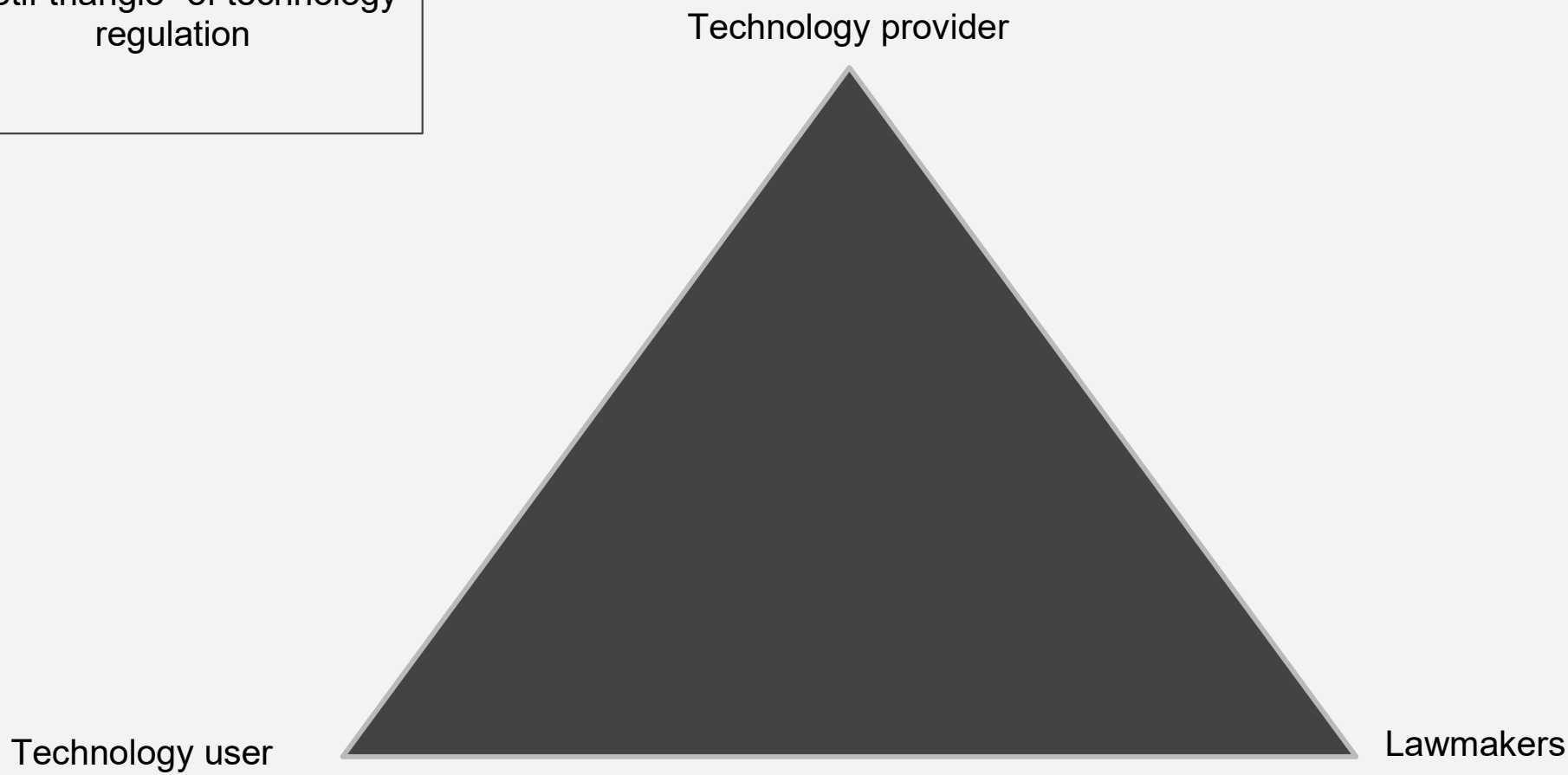
Source: https://trends.google.de/trends/explore?date=today%205-y&q=%2Fm%2F0mkz&hl=de

content

Technology regulation and history of the European AI strategy

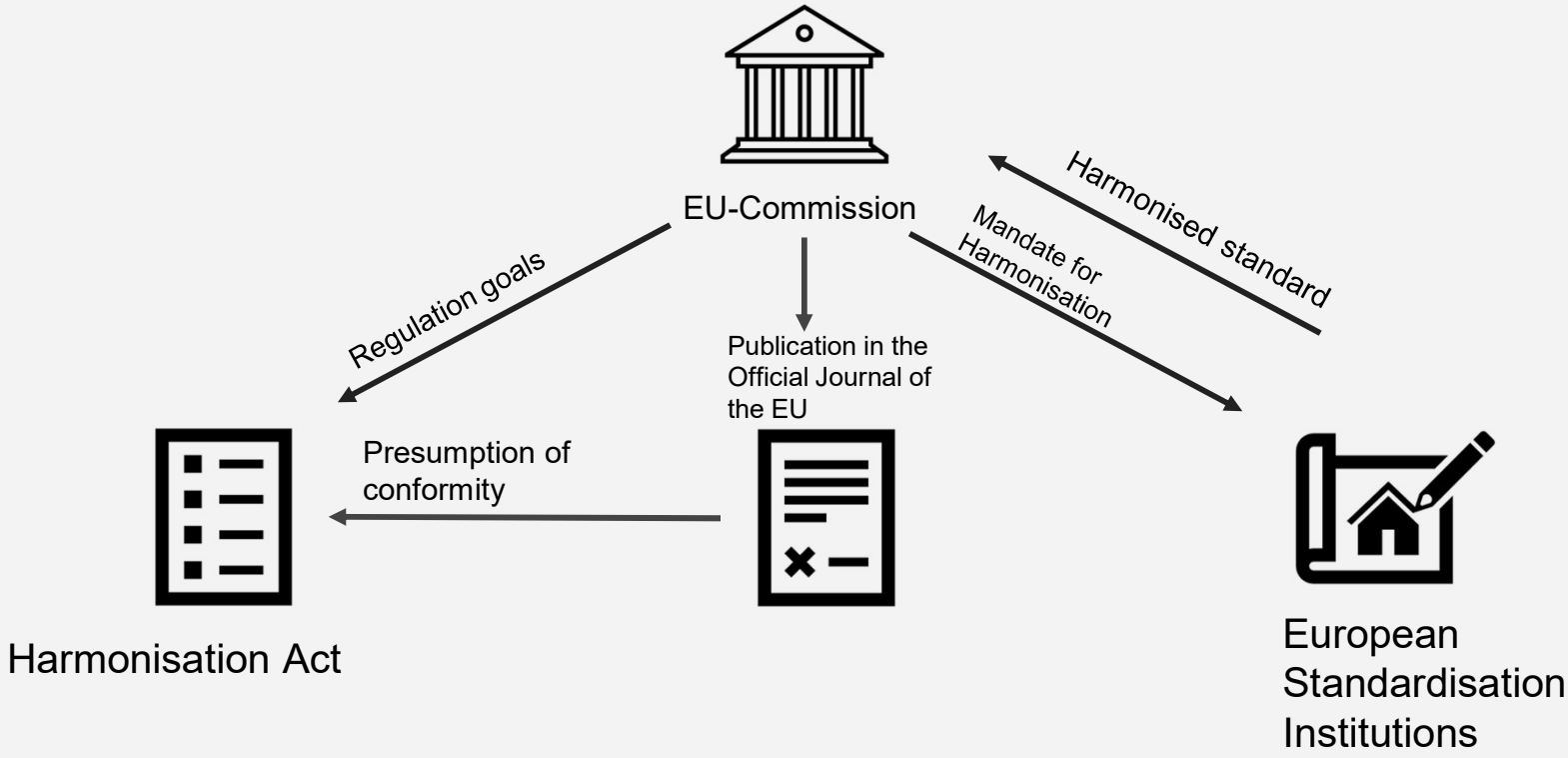
01

"Motif triangle" of technology regulation



European Approach to Regulating AI

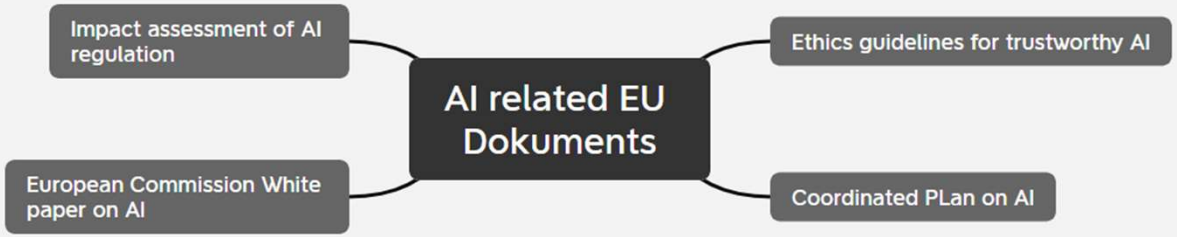
Detail harmonisation, new approach and standardisation

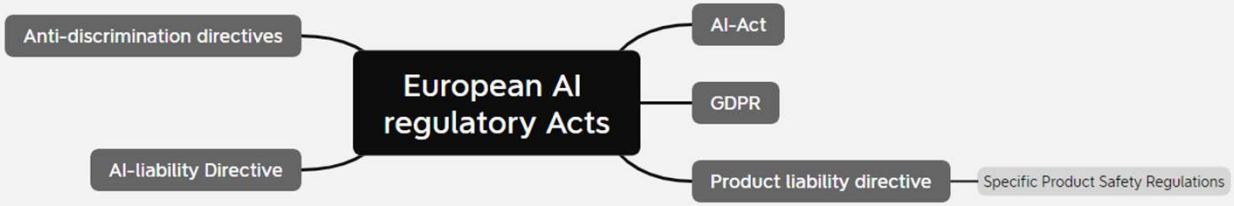


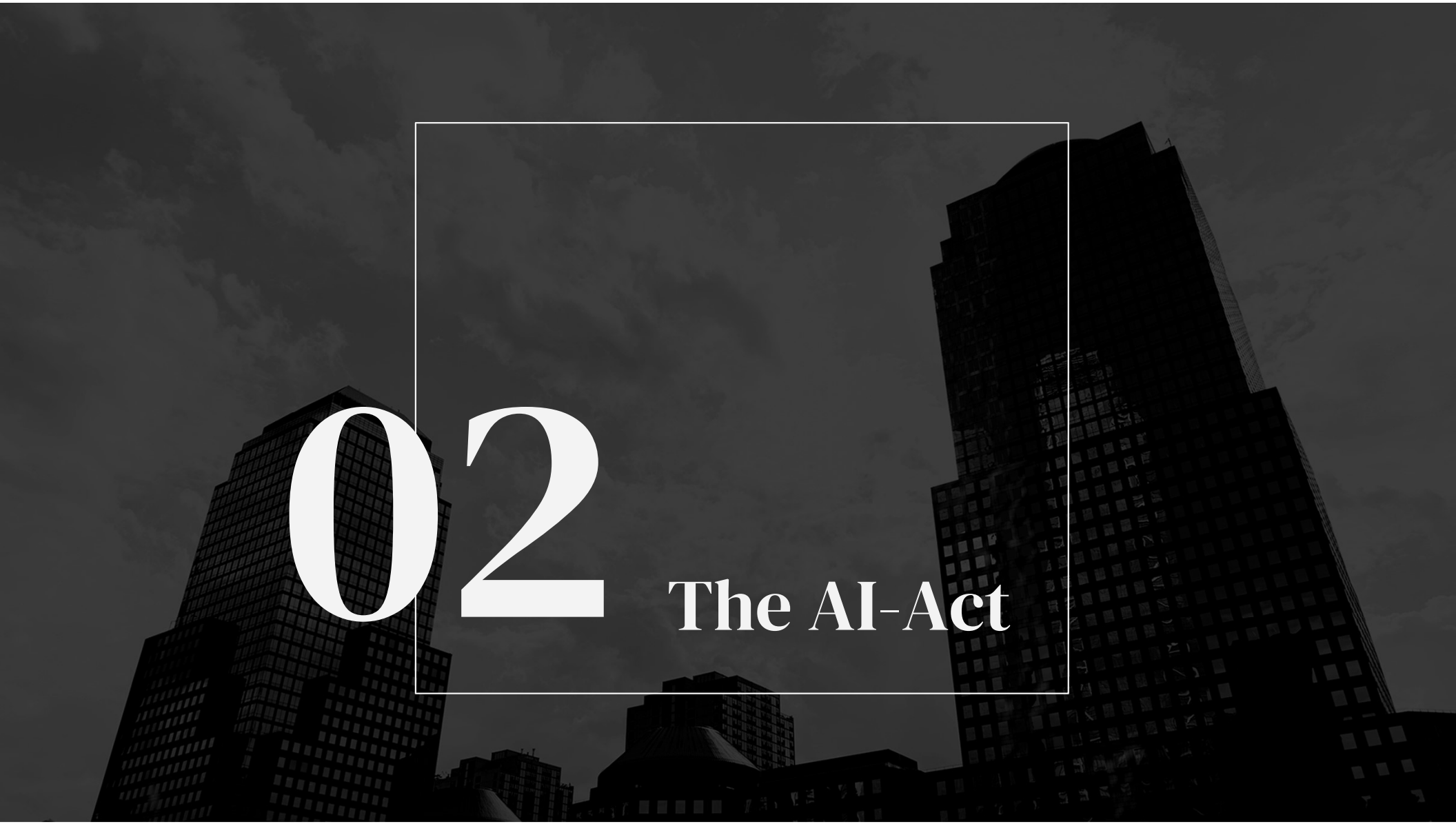
Regulation History

2016	• Adoption of the GDPR
2018	• AI Strategy of the EU Commission • Coordinated plan for AI
2019	• Ethical guidelines for trustworthy AI
2020	• Whitepaper on AI
2021	• Presentation fo the draft AI-Act
2023	• Presentation of Parliament's counter-draft and start of trialogue

European AI-Regulation Documents



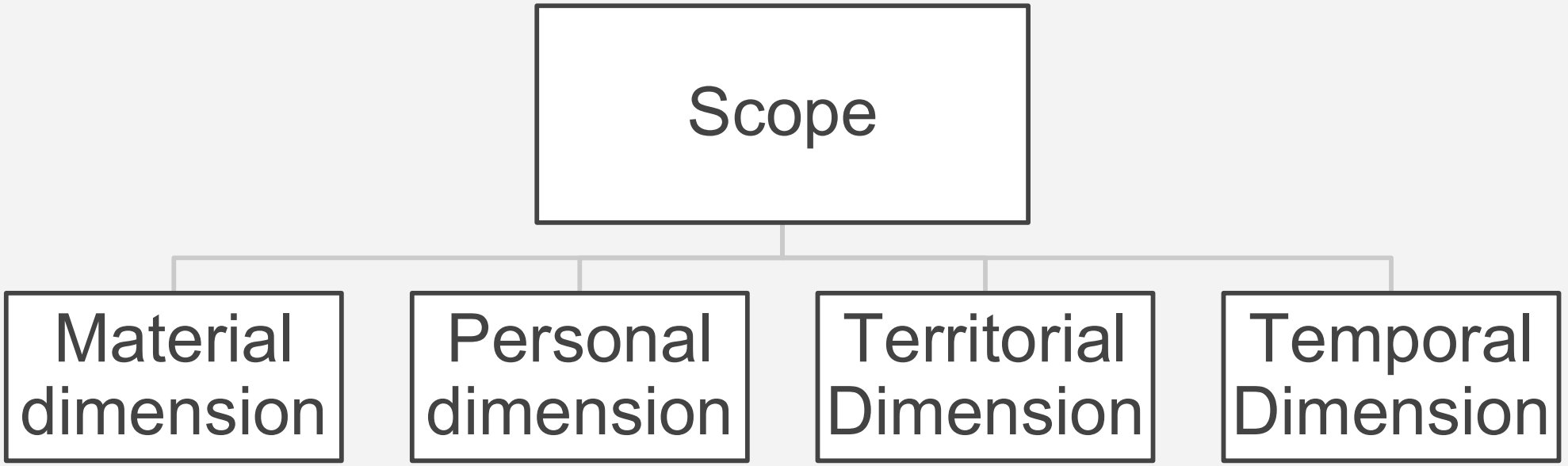




02 The AI-Act

Current State

Scope


Scope
Material dimension
Personal dimension
Territorial Dimension
Temporal Dimension

Material Dimension

Art. 3 Nr. 1:

‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with

Annex I:

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods.

Parlamentary counter-proposal

Article 3
Definitions

For the purpose of this Regulation, the following definitions apply:

(1) 'artificial intelligence system' (AI system) means ***a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments.***

(1a) 'risk' means the combination of the probability of an occurrence of harm and the severity of that harm;

(1b) 'significant risk' means a risk that is significant as a result of the combination of its severity, intensity, probability of occurrence, and duration of its effects, and its the ability to affect an individual, a plurality of persons or to affect a particular group of persons;

(1c) 'foundation model' means an AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks;

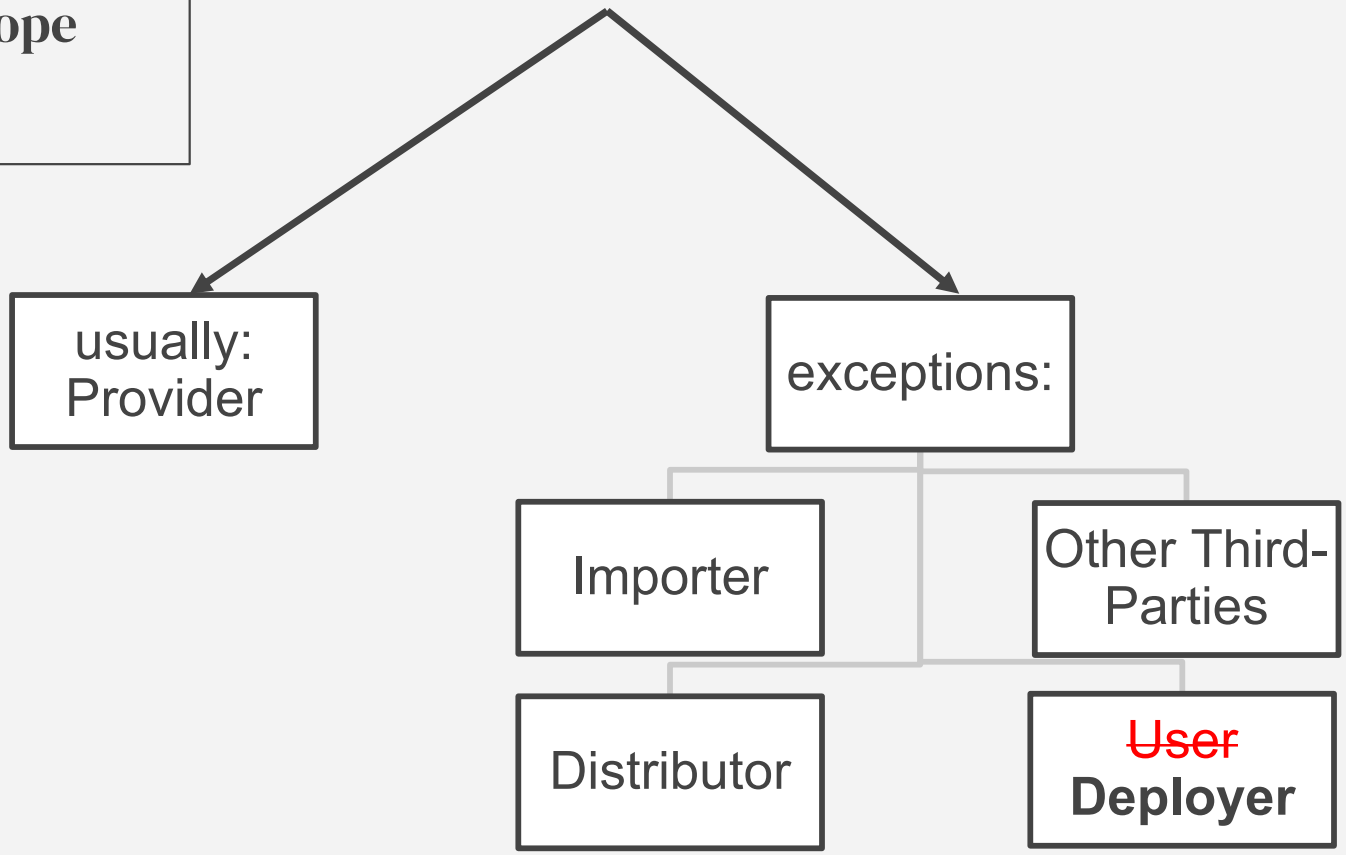
(1d) 'general purpose AI system' means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed;

(1e) 'large training runs' means the production process of a powerful AI models that require computing resources above a very high threshold.

(Probably) Finalised Version

An AI system is a 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 influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

Personal scope



generally: Provider

- In principle, the providers of AI systems are obligated, e.g. Art. 2 para. 1 lit. a, c) 16; 52 para. 1 sen. 1 AIA
- provider' means a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge; Art. 3 Nr.2

Exception 1

Concrete obligation for a concrete actor

Article 26

Obligations of importers

(1) Before placing a high-risk AI system on the market, importers of such system shall ensure that:

(a)the appropriate conformity assessment procedure has been carried out by the provider of that AI system

(b)the provider has drawn up the technical documentation in accordance with Annex IV;

(c)the system bears the required conformity marking and is accompanied by the required documentation and instructions of use.

[…]

Article 27

Obligations of distributors

(1) Before making a high-risk AI system available on the market, distributors shall verify that the high-risk AI system bears the required CE conformity marking, that it is accompanied by the required documentation and instruction of use, and that the provider and the importer of the system, as applicable, have complied with the obligations set out in this Regulation.

[…]

Exception 2

Art. 28 AI-Act (Parliamentary counter proposal)

- Para. 1: Another actor shall be considered a provider in any of the following circumstances:

  a) they put their name or trademark on a high-risk AI system already placed on the market or put into service;

  (b) they make a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service and in a way that it remains a high-risk AI system in accordance with Article 6;

  (ba) they make a substantial modification to an AI system, including a general purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service in such manner that the AI system becomes a high risk AI system in accordance with Article 6.

- In cases a) and ba) they replace original providers within the scope of the regulation, Art. 28 para. 2 AI-Act

Parliamentary counter-proposal

Art. 28 AI-Act

2. Where the circumstances referred to in paragraph 1, point ***(a) to (ba)*** occur, the provider that initially placed the high-risk AI system on the market or put it into service shall no longer be considered a provider ***of that specific AI system*** for the purposes of this Regulation. ***This former provider shall provide the new provider with the technical documentation and all other relevant and reasonably expected information capabilities of the AI system, technical access or other assistance based on the generally acknowledged state of the art that are required for the fulfilment of the obligations set out in this Regulation.***

Paragraph 2 shall also apply to providers of foundation models as defined in Article 3 when the foundation model is directly integrated in an high-risk AI system.

Territorial and Temporal Scope

Article 2
Scope

1. This Regulation applies to:

(a) providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country;

(b) ***deployers*** of AI systems ***that have their place of establishment or who are*** located within the Union;

(c) providers ***and deployers*** of AI ***systems that have their place of establishment or*** are located in a third country, where ***either Member State law applies by virtue of public international law or*** the output produced by the system is ***intended to be*** used in the Union;

(ca) providers placing on the market or putting into service AI systems referred to in Article 5 outside the Union where the provider or distributor of such systems is located within the Union;

(cb) importers and distributors of AI systems as well as authorised representatives of providers of AI systems, where such importers, distributors or authorised representatives have their establishment or are located in the Union;

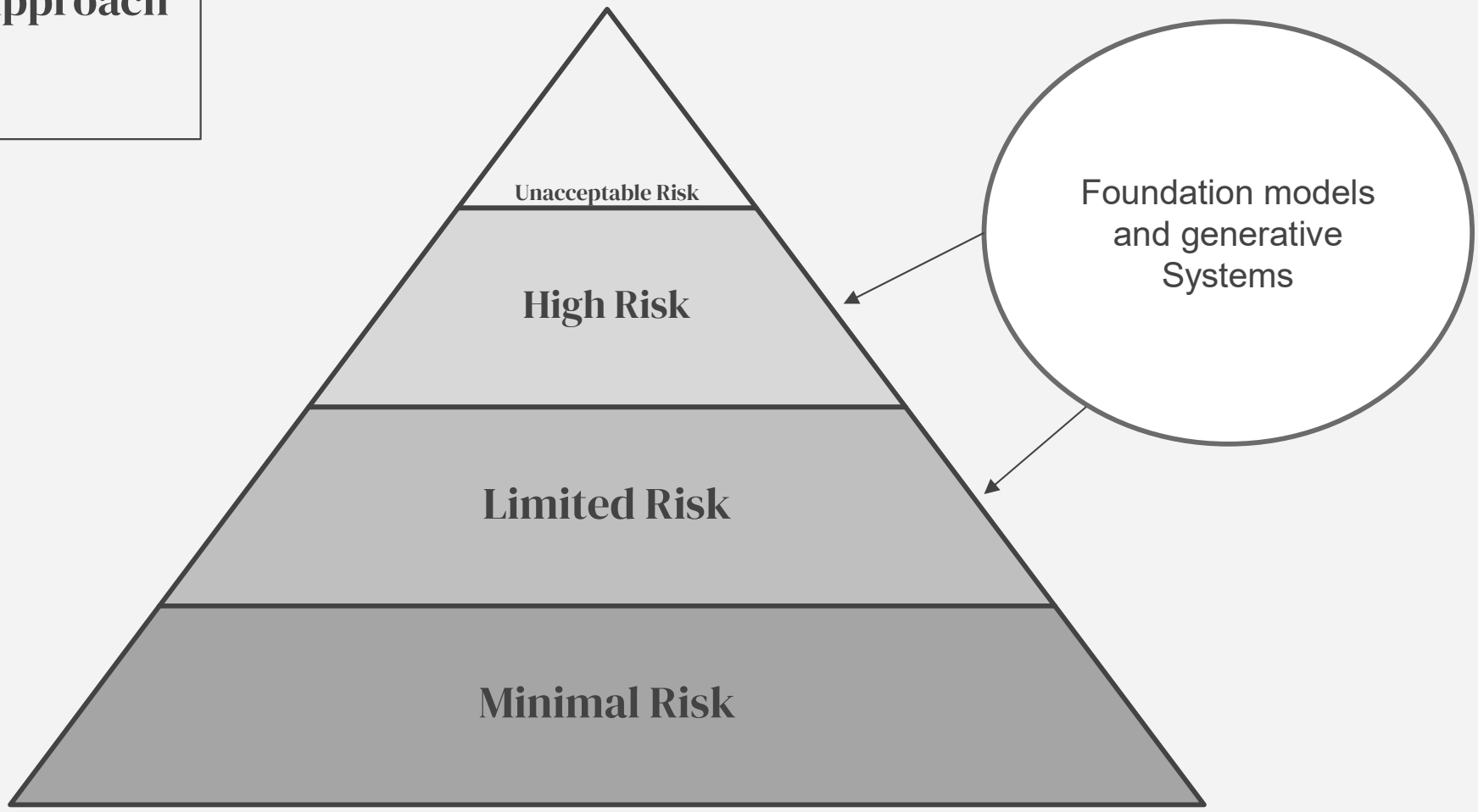
(cc) affected persons as defined in Article 3(8a) that are located in the Union and whose health, safety or fundamental rights were adversely impacted by the use of an AI system that was placed on the market or put into service in the Union;

- **Temporal**: Pursuant to Art. 85 para. 2 AI-Act, this Regulation shall apply 24 months after its entry into force.

General exceptions

- Art. 2 para. 3: AIA shall not apply to systems exclusively developed and used for military purposes
- Trialogue negotiations: an exception for systems developed and published for research purposes

Risk-based Approach



High Risk Systems

- **Art. 6 AI-Act**
  - 1. para. 1: AI systems as safety components of certain products
  - 2. para. 2: systems specifically listed in Annex III

- **Dynamic reference:** according to Art. 7, the Commission is able to modify the use cases mentioned in Annex III
  - Parliament's counter-draft now also allows the Commission to modify and remove high-risk cases.
  - **Pr:** parliament also wants so called „extra layer“: Annex III system shall only be considered high risk **if they pose a significant risk of harm to the health, safety or fundamental rights of natural persons**

Overview of the requirements

- Risk-Management system, Art. 9 AI-Act
- Data and data-governance Art. 10 AI-Act
- Technical documentation Art. 11 AI-Act
- Record-keeping, Art. 12 AI-Act
- Transparency and provision of information to users, Art. 13 AI-Act
- Human oversight, Art. 14 AI-Act
- Accuracy, robustness and cybersecurity, Art. 15 AI-Act

Sanctions

- In the event of a breach of the provisions of the AI Act, sanctions shall be imposed in accordance with Art. 71 AI Act:
  - Para. 3: Fines of up to € 35 million, or in the case of companies, up to 7% of the total worldwide annual turnover in cases::
    - A) non-compliance with the prohibition of the practices listed in Art. 5 (AI systems with unacceptable risk).
    - ~~B) non-compliance of the AI system with the requirements set out in Art. 10.~~
    - **Para. 3a: Fines of up to 20 million or 4% of total worldwide annual turnover of companies for non-compliance with Art. 10 and 13**
  - Para. 5: Fines of up to 7.5 million or 1% of the total worldwide annual turnover of companies for false, incomplete or misleading information in response to a request for information by the competent authorities.
  - Para. 4: Fines of up to 10 million or 2% of total worldwide annual turnover in all other cases.
- Generally, however, the member states determine and enforce the exact sanction orders, para. 1

Regulating Foundation models

03

Foundation model and general purpose systems

Art. 3

(1c) 'foundation model' means an AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks;

(1d) 'general purpose AI system' means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed;

Art. 28b

4. Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system, shall in addition

Art. 28b

Obligations of the provider of a foundation model

- Paragraph 2 contains various general requirements:
  - Specific transparency and testing obligations to demonstrate proportionate risk mitigation of the system (a)
  - use only of data sets suitable for foundation models (b)
  - function-oriented design of the system (c)
  - consideration of ecological dimensions (d)
  - technical documentation (e)
  - Implementation of a quality management system and documentation of compliance (f)
  - Registration of the system in an EU database (g)
- Sentence 2: State of the art here the relevant measure too

Art. 28b

Obligations of the provider of a generative foundation model

- Paragraph 4 then contains a sensible addition for generative systems:
  - A) Must be compliant with the transparency requirement of Article 52(1)
  - B) Training, design and development of the system must provide sufficient safeguards against the generation of infringing content (again, state-of-the-art)
  - C) Sufficient publication of copyright-relevant training data processes

Trialogue compromise	

	• **Two tiered approach**: minimum standards for all FM and additional obligations for FM with systemic-risks
	• **Criteria** for the categorization will be laid down in specific new Annex
	  • Threshold of 10^25 Flops
	  • Number of business users
	  • Number of parameters used

Trialogue compromise	

	• Requirements for **base FM**:
	  • Publish copyright relevant training data usage
	  • Technical documentation of the development
	• Requirements for **systemic risk models**:
	  • Model evaluations, assess and mitigate systemic risks
	  • Adversarial testing
	  • Report to the commission on serious incidents
	  • Ensure cybersecurity
	  • Report on energy sufficiency

Thanks und let's go!

Do you have any questions?

Philipp.Mahlow@student.uibk.ac.at
Philipp Mahlow on LinkedIn
+49 1525 4918645