AI Factory Austria AI:AT

Foundations of Trustworthy AI

Peter Biegelbauer
Co-Lead Legal, Regulatory and Ethics

Michael Löffler
Lead Legal, Regulatory and Ethics

AI Factory Austria AI:AT - PUBLIC Consortium

Disclaimer:
The speakers are solely sharing their personal experiences. Therefore, this free seminar is not a substitute for professional/legal advice.

Beneficiaries

Advanced Computing Austria

AUSTRIAN INSTITUTE OF TECHNOLOGY

Affiliated Entities

TECHNISCHE UNIVERSITÄT WIEN

universität wien

BOKU UNIVERSITY

universität innsbruck

TU Graz

JOHANNES KEPLER UNIVERSITÄT LINZ

Institute of Science and Technology Austria

ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

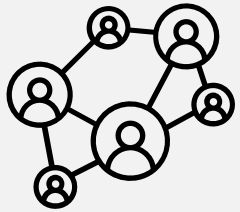
INiTS

eodc

Why do we need AI Factory Austria?

Sovereignty

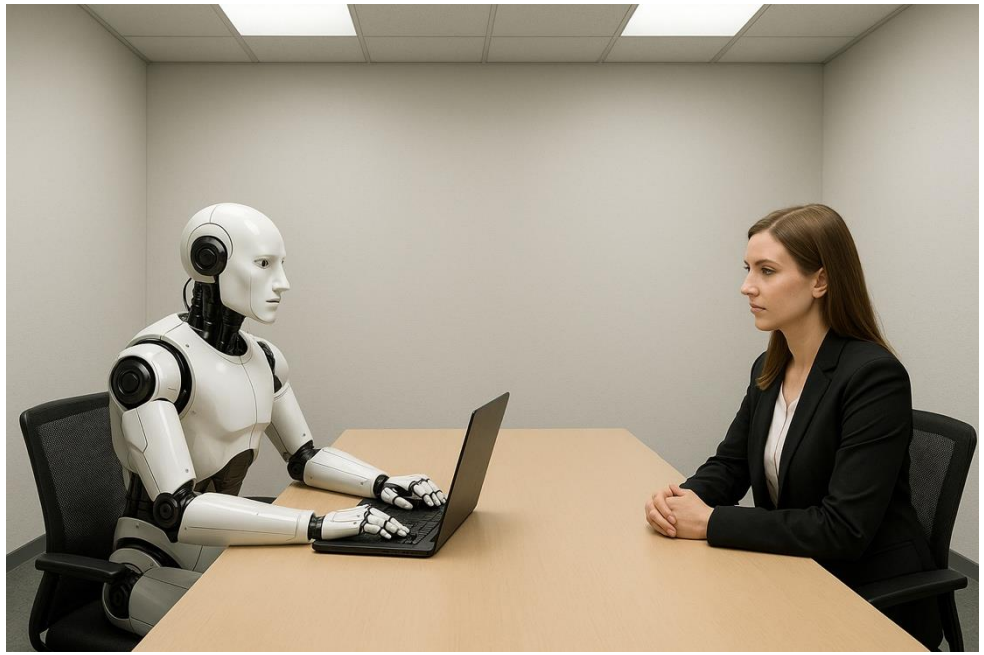
Ethics and Trustworthiness



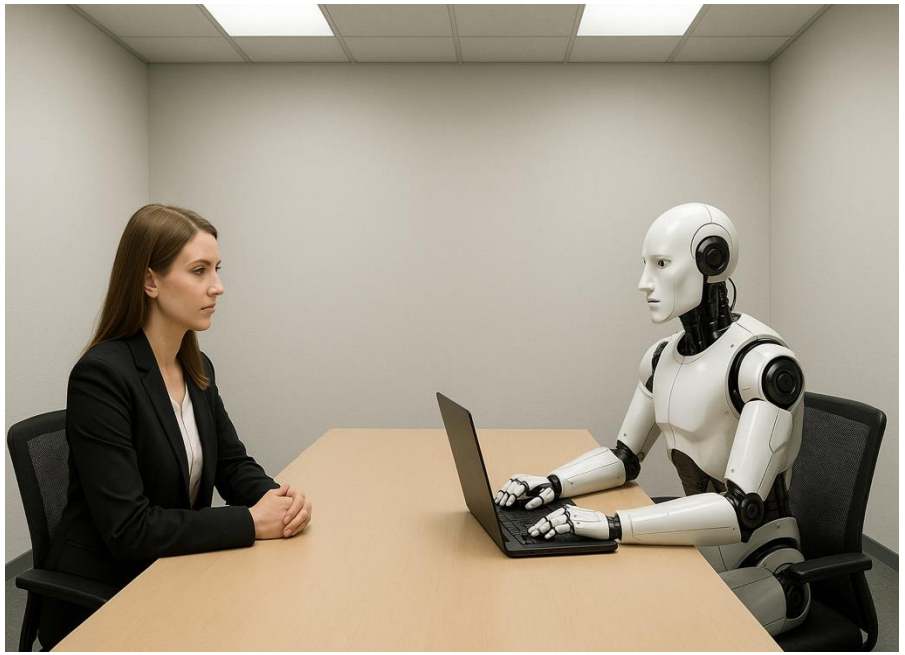
Connecting the Ecosystem

Digitalisation and HR

- Problem: How do I find the best candidate for my organization?
- Solution: Support from an algorithm-based decision-making system
- Promise: Algorithms have no biases



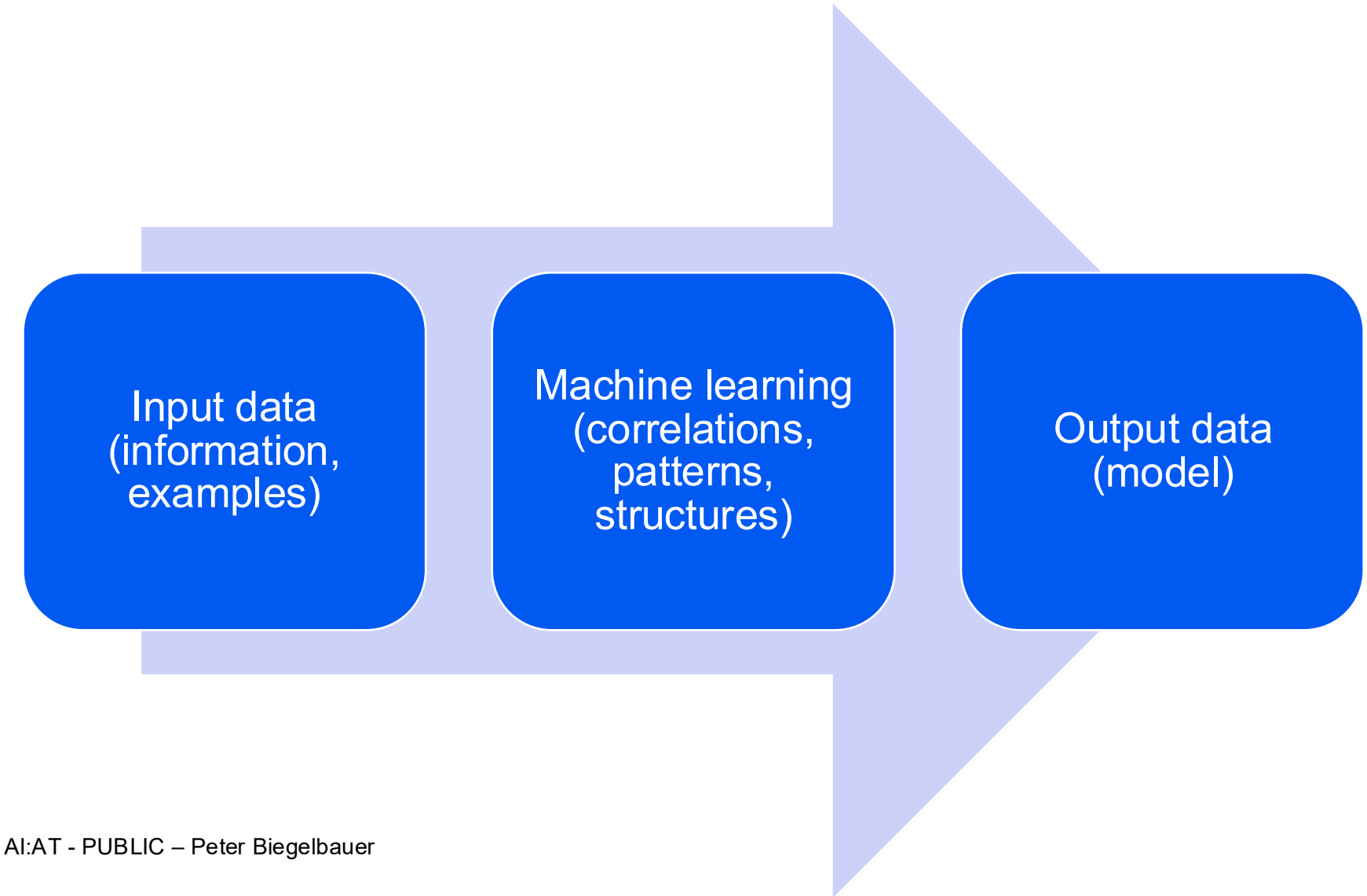
Amazon: Automatic Evaluation of Job Applications



- Amazon's experimental recruitment mechanism followed the patterns existing in the company for the last 10 years (hiring more men than women).
- The algorithm learned to down-rank applications that contained the word “women” until the company discovered the problem.
- https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

How does an algorithmic decision-making system learn?

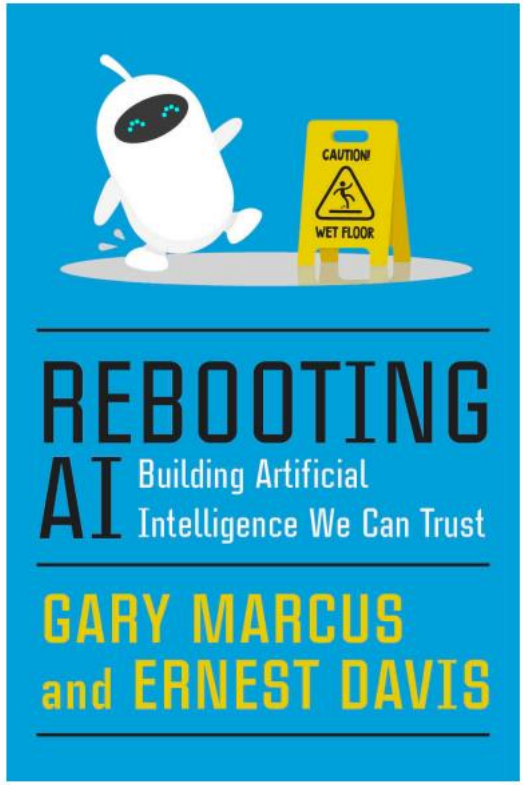
The input is historical data that is projected into the future through the application of machine learning.



Key question:
How can we ensure the benefits of AI to outweigh potential risks?

A healthier life

Bildquelle. https://unsplash.com/de/s/fotos/ai-in-healthcare



Increased productivity and prosperity

Bildquelle. https://pixabay.com/images/search/production%20robots/

Answer:
Through a human-centered approach to AI

"Artificial intelligence should treat all people fairly, empower everyone, perform reliably and safely, be understandable, be secure and respect privacy, and have algorithmic accountability. It should be aligned with existing human values, be explainable, be fair, and respect user data rights. It should be used for socially beneficial purposes and always remain under meaningful human control."

Tom Chatfield (2020). There's No Such Thing As Ethical AI.

AI ethics refers to moral principles and values that should guide the development and use of AI systems.

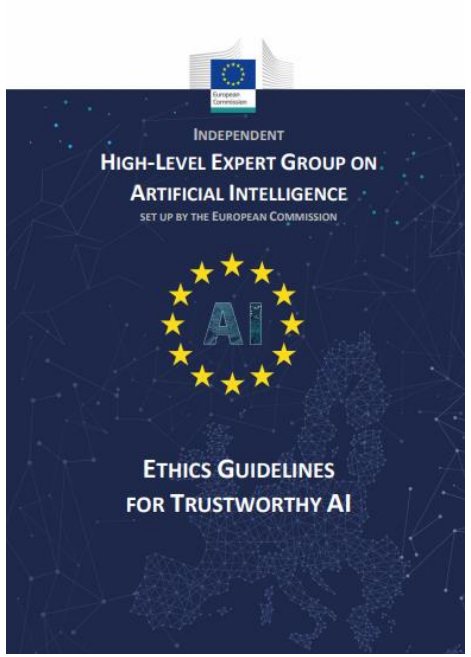
So, why AI ethics?

- **Definition**: Normative framework for the responsible use of AI, beyond mere legality
- **Objective**: Build trust, minimize risks, and maximize social benefits
- **Practical relevance**: Point of reference for regulation, research, and industrial application
- **Distinction**: Ethics as “norms of conduct” beyond mere legal compliance
- **Social dimension**: Consideration of long-term effects on democracy, professinal life, and the environment
- **Dynamics**: Ethics as a continuous discourse in the face of technical innovation and change

Ethical principles in the AI Act

- Integration of the HLEG principles for trustworthy AI into AI Act (orig. draft Art. 4a, now recitals)
- Priority of human action and human oversight
- Technical robustness and security
- Data protection and data quality management
- Transparency
- Diversity, non-discrimination, and fairness
- Social and environmental well-being

The missing principle: Accountability → AI Liability Directive (planned)



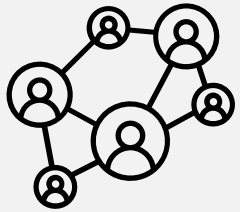
Take-Home Messages

- **Ethics is complex and remains important**
  - There are many unanswered questions: we are in the midst of a journey, not at its end
- **Current AI systems often reflect upon us – the humans (quite a challenge!)**
- **Two levels of AI ethics:**
  - In the short term “AI ethics today”:
    - Bias, deepfakes, facial recognition, copyright, etc.
  - In the long-term questions about “superintelligence”:
    - What will happen if / when AI becomes more intelligent than we are (AGI)?
- **Role of the AI Factory Ethics Team**
  - Provides support for questions regarding AI ethics
  - Develops frameworks and guidelines for trustworthy / responsible AI
  - Supports stakeholders in developing and deploying trustworthy / responsible AI systems

Why do we need AI Factory Austria?

Sovereignty

Ethics and Trustworthiness



Connecting the Ecosystem

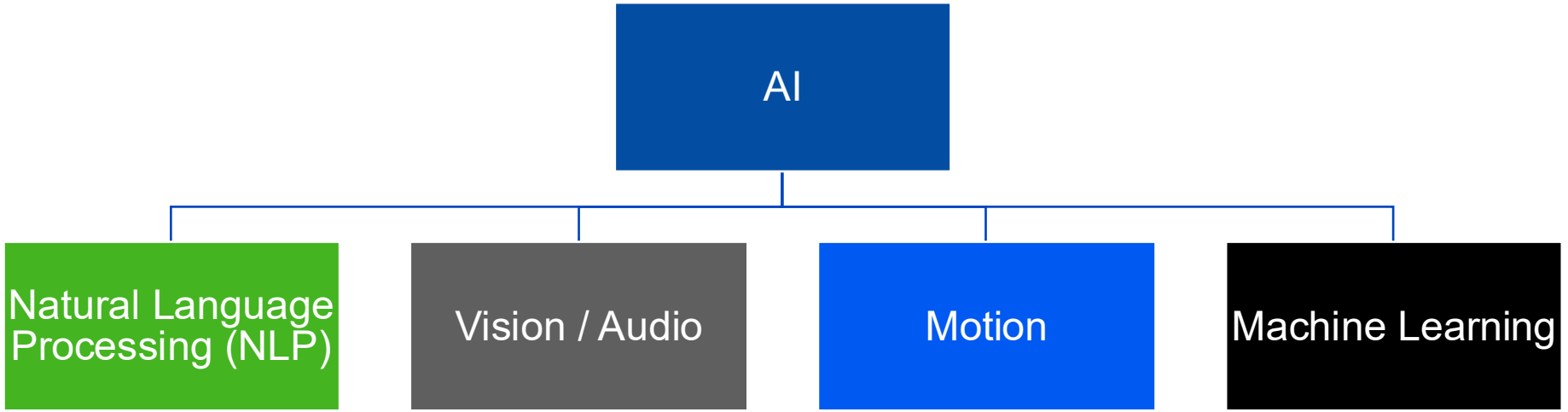
What is "artificial intelligence"?

– System that cannot be distinguished from a human being when chatting ("Turing test")

"Turing test", formulated by Alan Turing in 1950.

Image source: Stable-Diffusion 2.1.

What is "artificial intelligence"?



IBM Technology, AI vs Machine Learning (https://www.youtube.com/watch?v=4RixMPF4xis).

What is "artificial intelligence"?

AI system:[1]

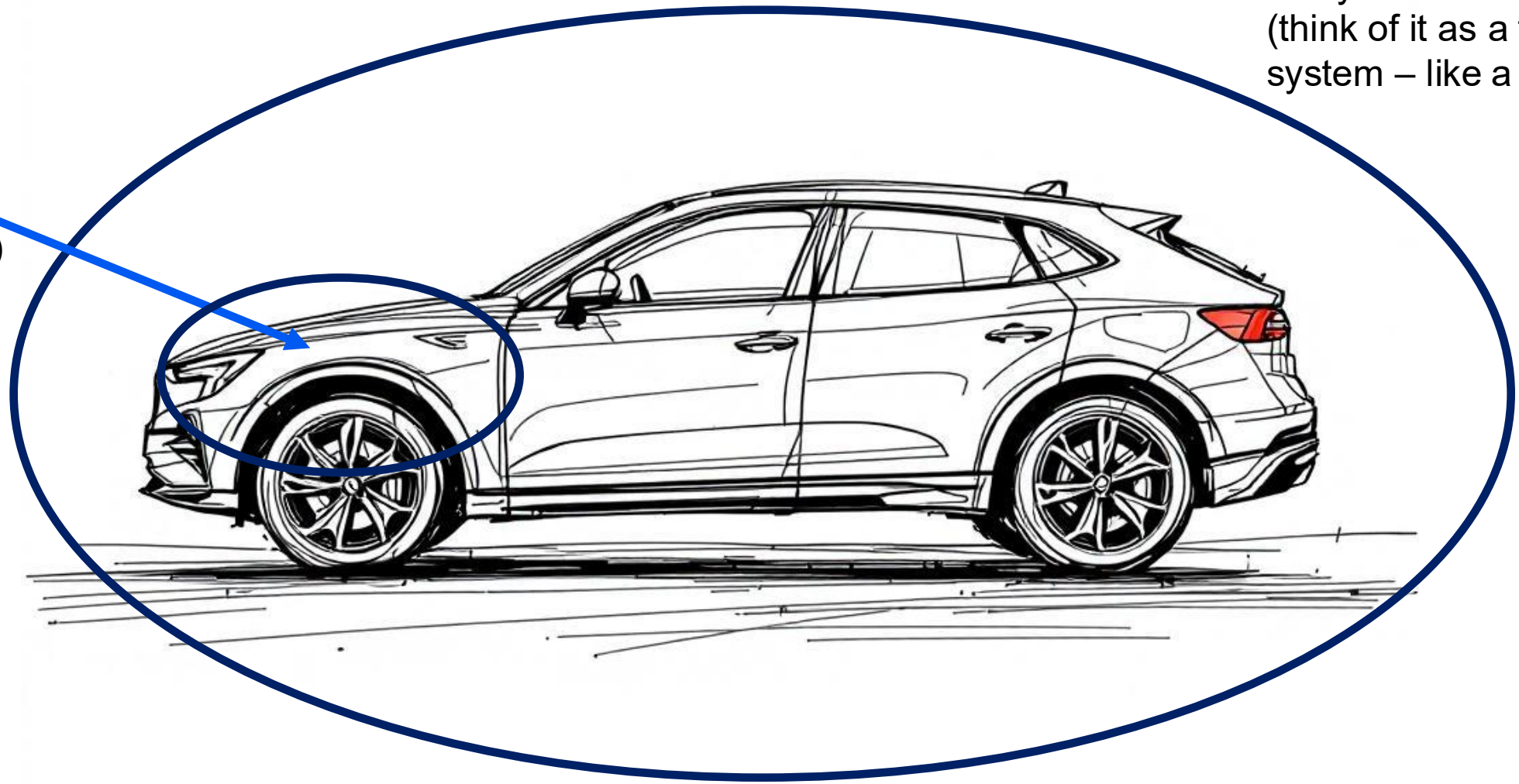
- **machine-based** system,
- designed to operate with **varying levels of autonomy**,
- may exhibit **adaptiveness** after deployment
- for explicit or implicit objectives, infers, from the input it receives, how to **generate outputs** such as predictions, content, recommendations, or decisions,
- can **influence** physical or virtual **environments**.

1: Sec 3 Num 1 AI-Act; European Commission (EC), Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), 06.02.2025.

How do AI Models relate to AI Systems?

AI System
(think of it as a fully usable system – like a whole car)

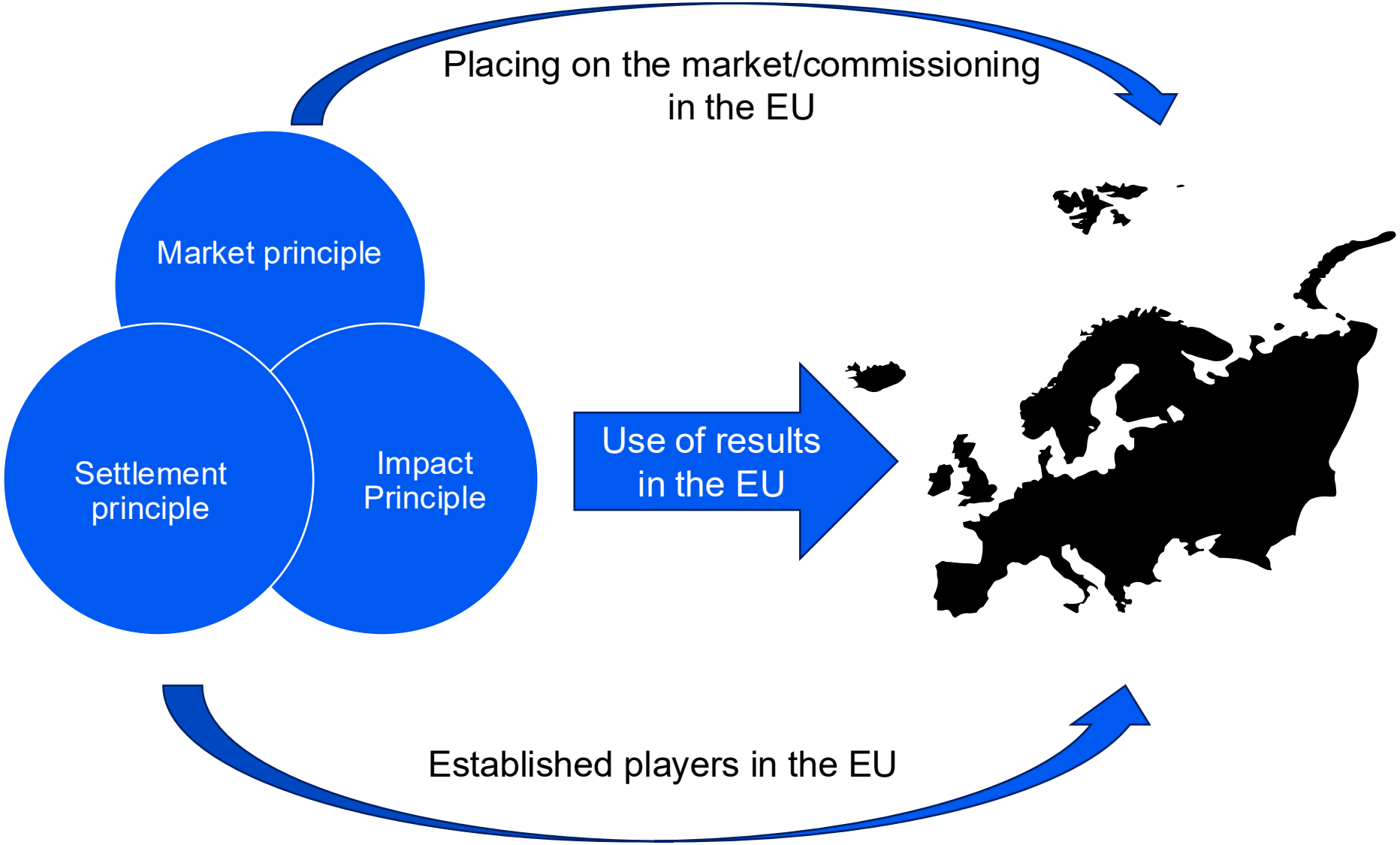
AI Model
(think of it as one essential part – like the engine of a car)



Vgl. *Almada*, Law & Compliance in AI Security & Data Protection, 12.2024.

Bildquelle: Le Chat

When to care about the AI Act?



When is the AI Act not applicable?

– The AI Act does not apply, among other things, to:

...

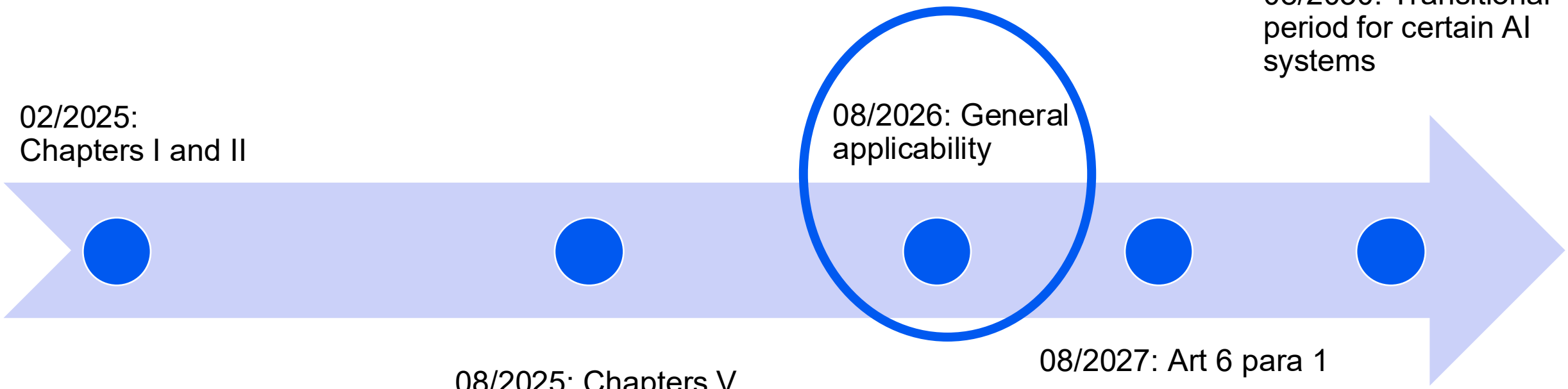
– AI systems or AI models that are developed and put into operation **exclusively** for scientific research and development;

– AI systems provided under free and open source licenses, unless they are Prohibited AI Systems, High-Risk AI Systems, or Art 50 AI Systems;

…

 1: Art 2 Artificial Intelligence Regulation.

When is the AI Act applicable?



When is the AI Act applicable?

Date:	02/2025	08/2025	08/2026	08/2027
Central rules (excerpt):	- AI Competence - Prohibitions	- Governance - general-purpose AI models	- Annex III high-risk AI systems	- Annex I high-risk AI systems

When does the AI Regulation apply?

Commission proposal (Digital Omnibus on AI):

…in discussion!

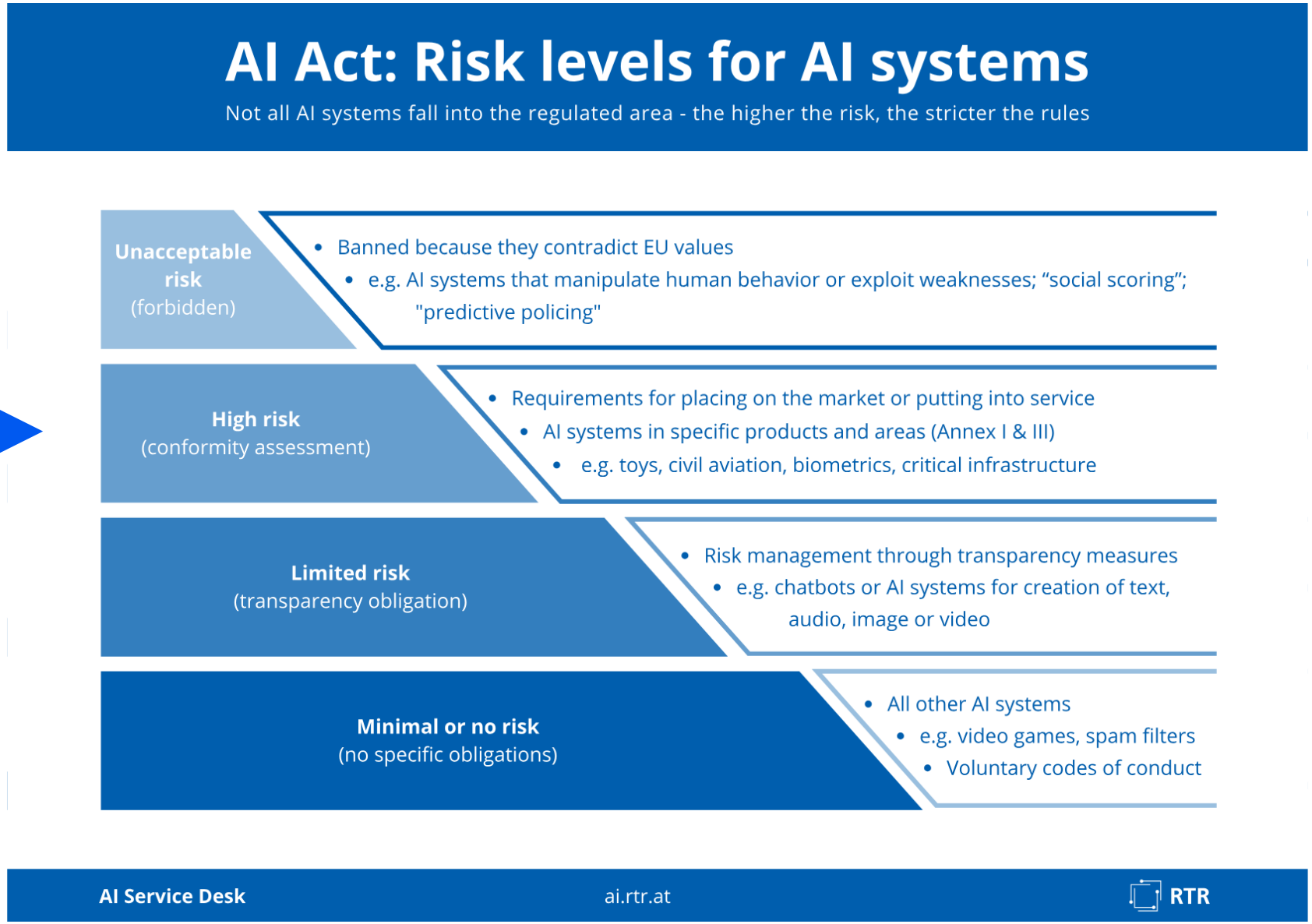
Postponement of the applicability of the regulations to:

- Classification of AI systems as high-risk AI systems
- requirements for high-risk AI systems and
- Obligations of providers and operators of high-risk AI systems and other stakeholders

- These rules will only apply (staggered in time) once the Commission confirms that adequate support measures are in place to comply with the rules.

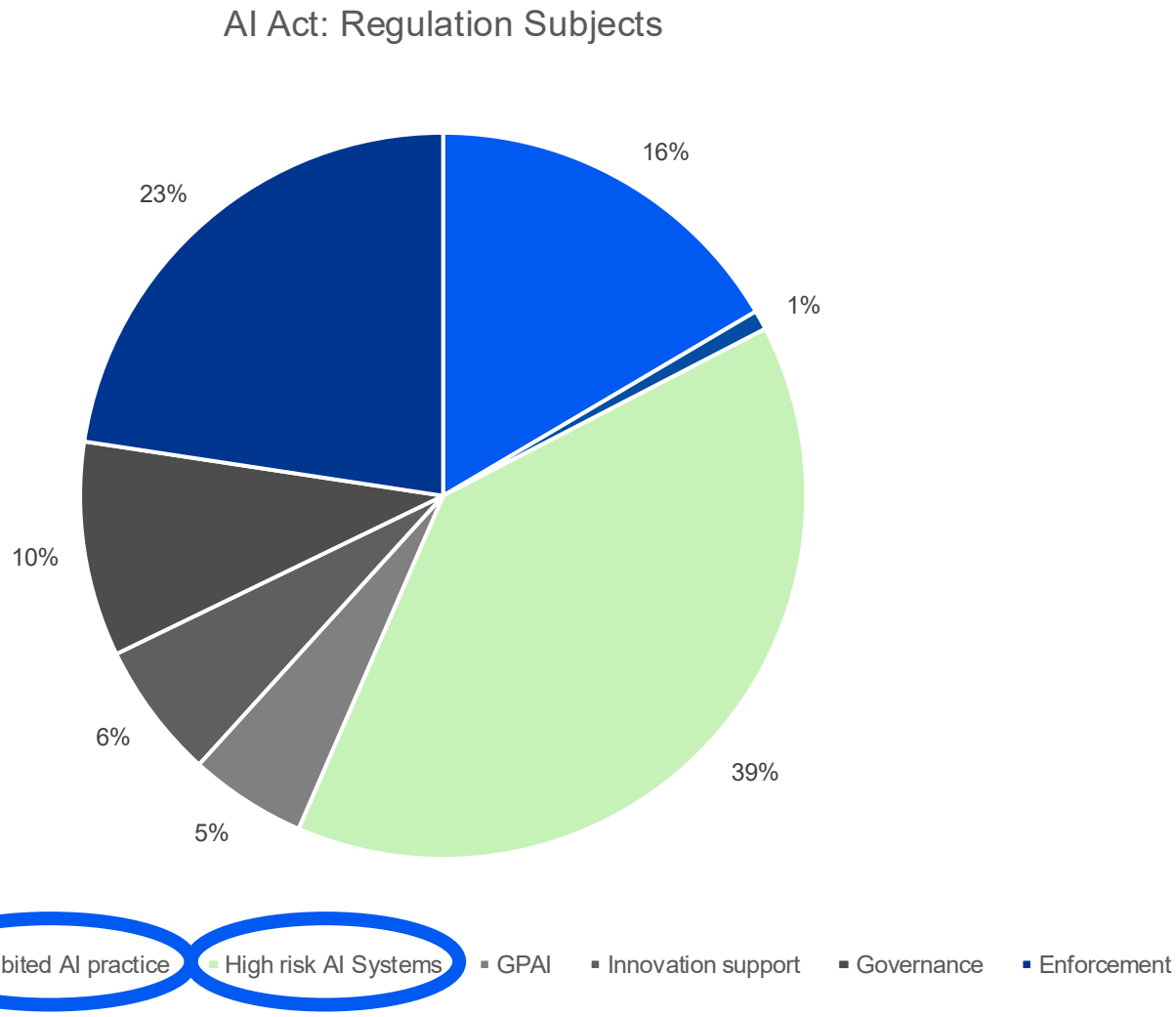
- **At the latest,** the rules should apply:
  - for high-risk AI systems according to Annex III: 12/2027
  - for high-risk AI systems according to Annex I: 08/2028

AI Act Risk Levels



Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH), Risk levels of AI systems, <https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/risikostufen_ki-systeme.en.html>.

What does the AI-Act govern?

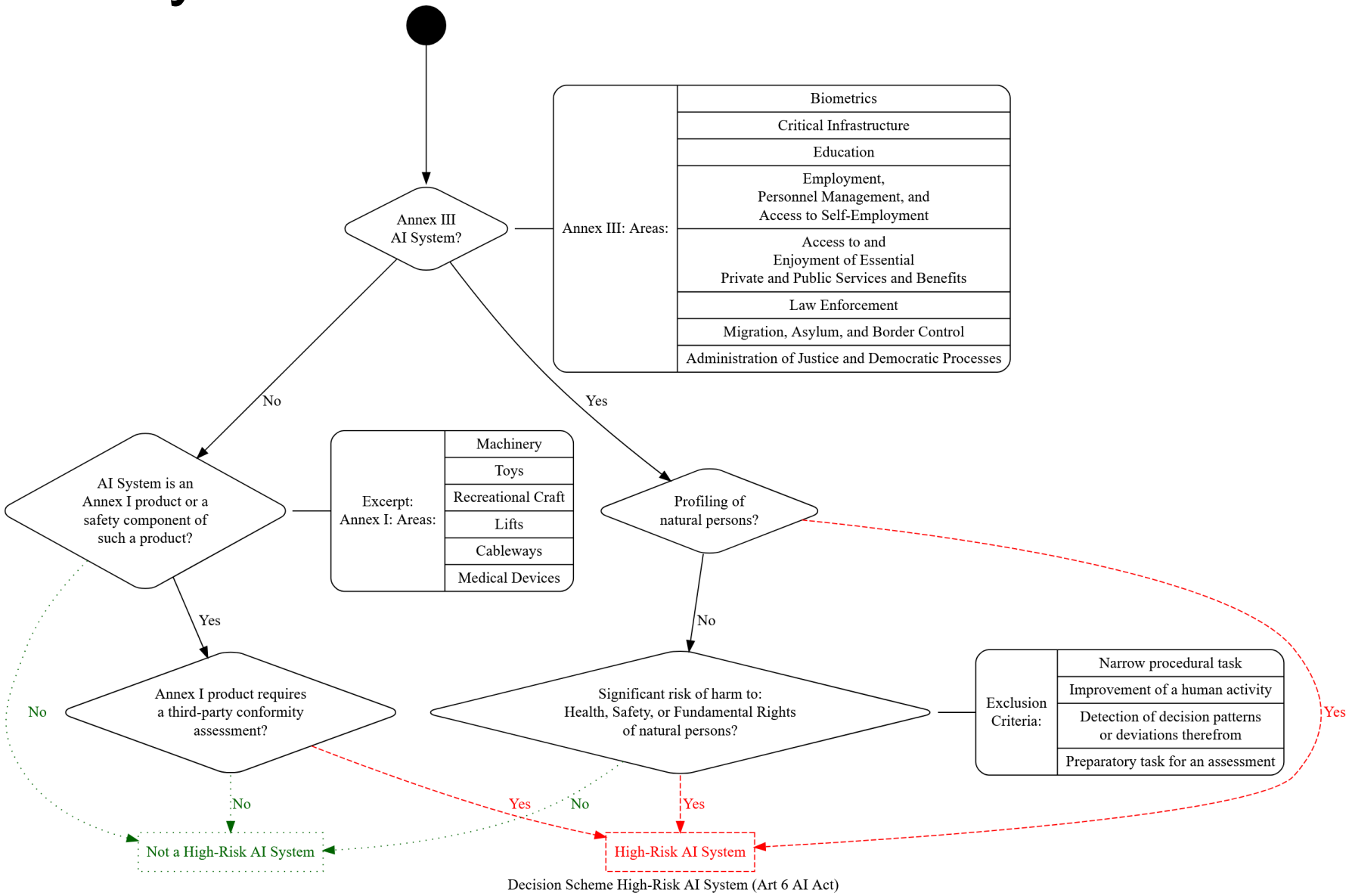


Prohibited AI Practices

- Subliminal **manipulation** with considerable damage
- **Exploiting** vulnerable groups of people
- **Social** scoring with negative impact
- **Profiling** to assess whether a crime will be committed
- Creation of databases for **facial recognition** from images from the Internet
- **Emotion recognition** in the workplace or in educational institutions
- **Biometric categorization** to obtain sensitive data
- **Remote biometric identification** (with strict exceptions for law enforcement purposes)

Art 5 AI Act.

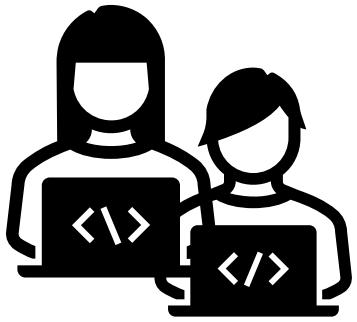
High-risk AI Systems



Decision Scheme High-Risk AI System (Art 6 AI Act)

Art 6 AI Act.

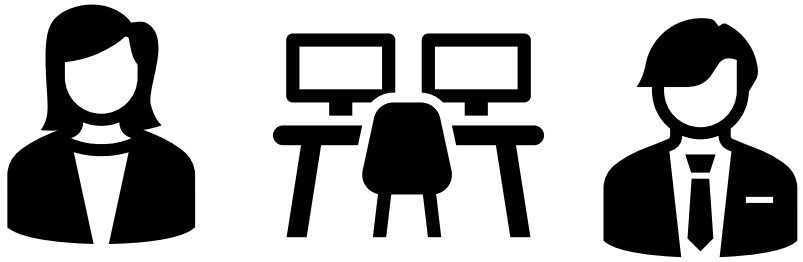
AI Act Operators



Provider
develops an AI system (general-purpose AI model) **and**
places it on the market / puts the AI system into service

Art 3 Sec 3 AI-Act.

AI Act Operators

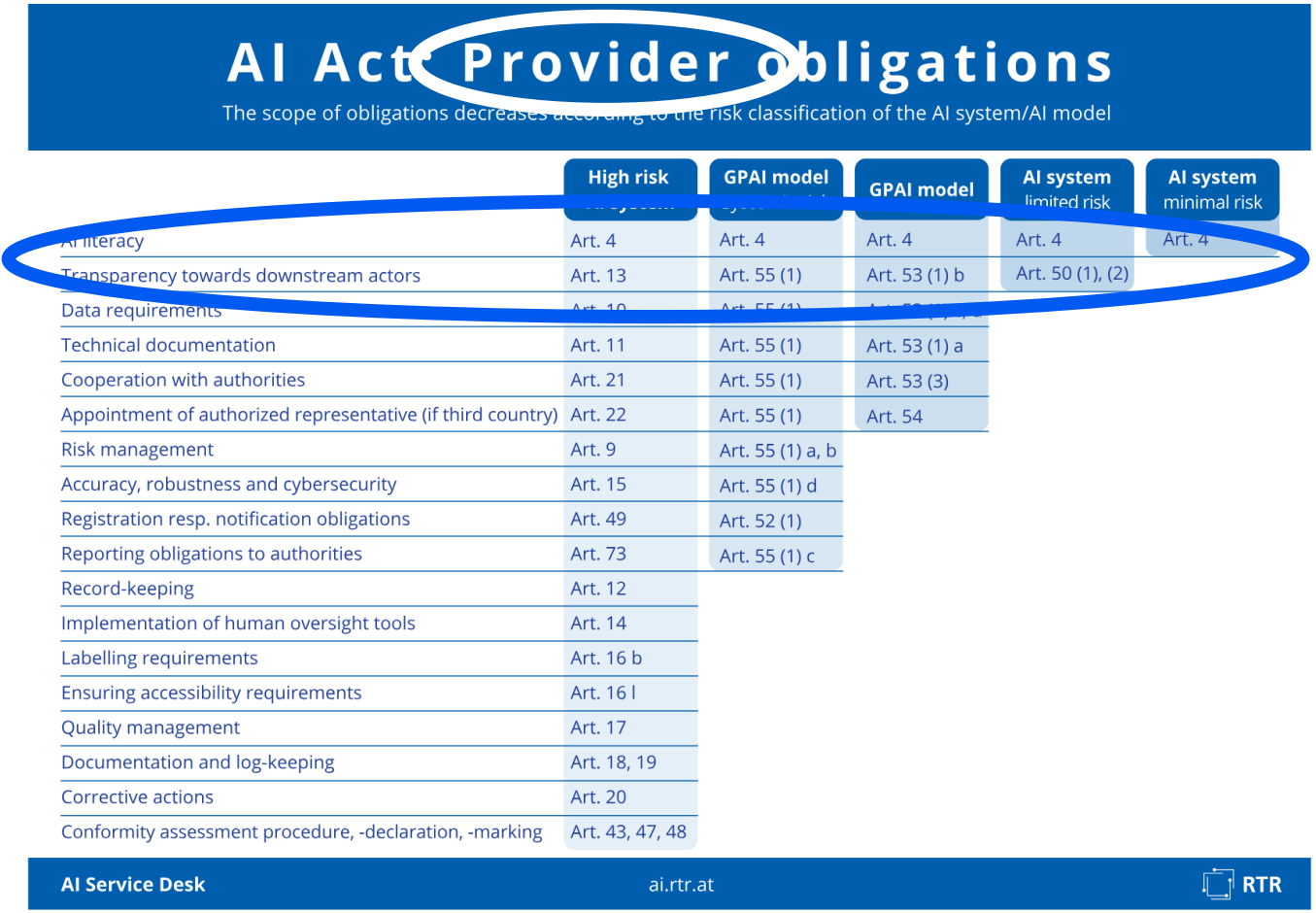


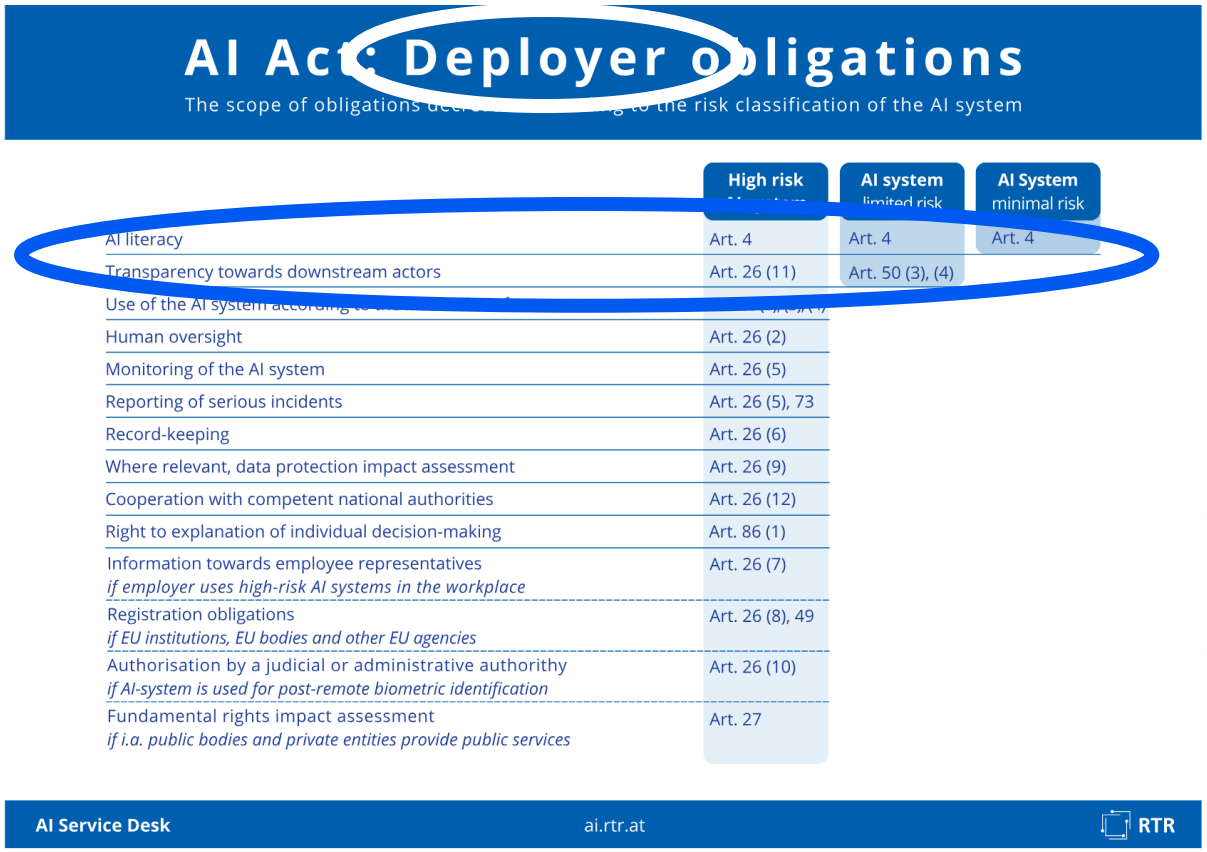
Deployer

using an AI system under its authority

Art 3 Sec 4 AI-Act.

Obligations





Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH), Provider obligations in the AI Act, <https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/Provider_obligations.en.html>.

Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH), AI Act: Deployer Obligations, <https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/Deployer_obligations.en.html>.

Machine-based System

- Is the AMS algorithm an AI system? – See also: https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus

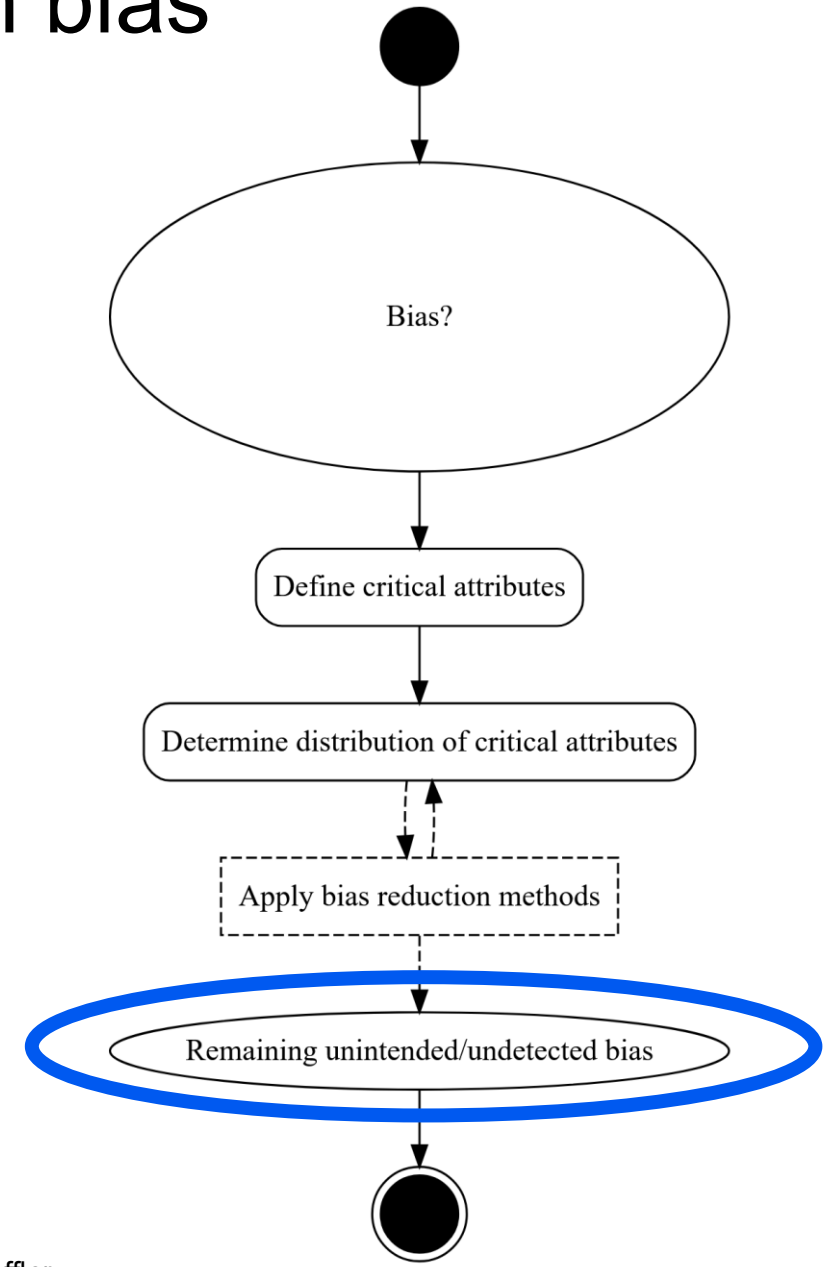
1: ITA, OeAW, The AMS Algorithm (https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus) – accessed on 13.11.2023.

Bias by Design

- the variable **'care responsibilities** yes/no' **only for female jobseekers**[1]
- social **bias** is translated into a technical one – social values that are actually changing are codified[1]

1:ITA, ÖAW, The AMS Algorithm (https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus) – accessed on 13.11.2023.

The problem with bias



Identification and Reduction of Bias

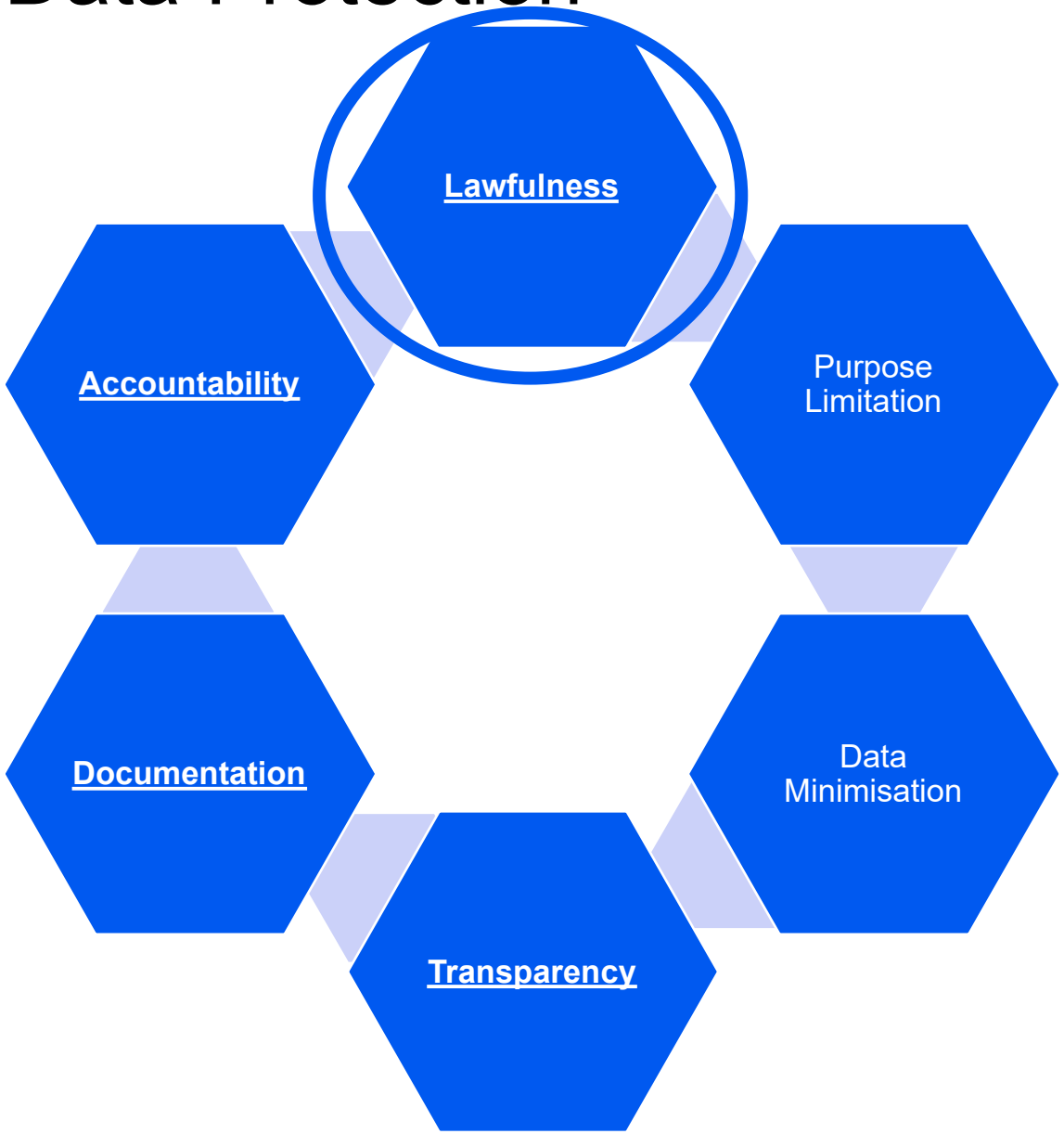
Legal Landscape

Data Protection

"Data can be either useful or perfectly anonymous but never both."[1]

1:*Ohm*, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, UCLA Law Review 2010, 1701.

Data Protection



Data protection law is characterized by the prohibition principle!

“Everything is forbidden unless it is exceptionally permitted.”

Facilitated AI training?

… in discussion!

Commission proposal (Digital Omnibus):

- Permission to use special categories of data (under certain conditions) for the development and operation of AI systems or the training of an AI model.
- The development of AI systems or the training of AI models should represent a legitimate interest (with certain restrictions) and be thus permitted.

What about AI Models that were trained with personal data?

- As of now there is no prevailing opinion on the “lawfullnes” of such AI Models.[1]

1: *Specht-Riemenschneider/Bortnikov/Steffens*, Keine „Fruit of the poisonous KI “-Doktrin im Datenschutzrecht, KIR 2025, 213.
Datenschutzkonferenz, Orientierungshilfe zu datenschutzrechtlichen Besonderheiten generativer KI-Systeme mit RAG-Methode, 10.2025.

Copyright law

Authors have the exclusive right to exploit their works in ways regulated by law -> exploitation right.

1: Vgl. §§ 14, 60 UrhG.

Copyright

- Reproduction right
- Distribution right
- Broadcasting rights
- Right to make available
- etc.

Text- and Data-Mining (TDM)

- Technique to generate information to patterns, trends and correlations;

- Differentiation between research institutions and everyone:
  Everyone: Only if works are lawfully accessible.[1]

Is TDM applicable to AI training?

Vgl. *Schippel*, „Urheberrecht first!“ - Die Pflicht zur KI-Policy nach Art. 53 KI-VO, KuR 2025, 547.

European Union Intellectual Property Office., The development of generative artificial intelligence from a copyright perspective. (2025).

Copyright Challenges

Release dates	August 27, 1964 (Los Angeles)[3] September 24, 1964 (New York City)[3]

Wikipedia: Mary Poppins (film).

The copyright in works of literature, music and the visual arts ends seventy years after death.

Users are (usually) liable for the results used!

Vgl. AG Köln: Anwaltliche Verantwortung für KI-generierte Schriftsätze, Computer und Recht 2025, 621–621.
Lucchi, Generative AI and copyright: training, creation, regulation (2025).

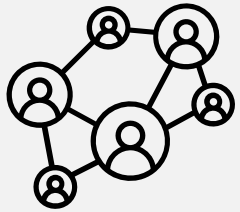
Key takeaways

- The use of AI Systems and models within the Europen Union is governed - by different legal frameworks
- The **AI Act** is one of the most relevant regulations
- Providers and Deployers of AI should (at least) also be aware of **Data Protection** and **Copyright** aspects!

Why do we need AI Factory Austria?

Sovereignty

Ethics and Trustworthiness



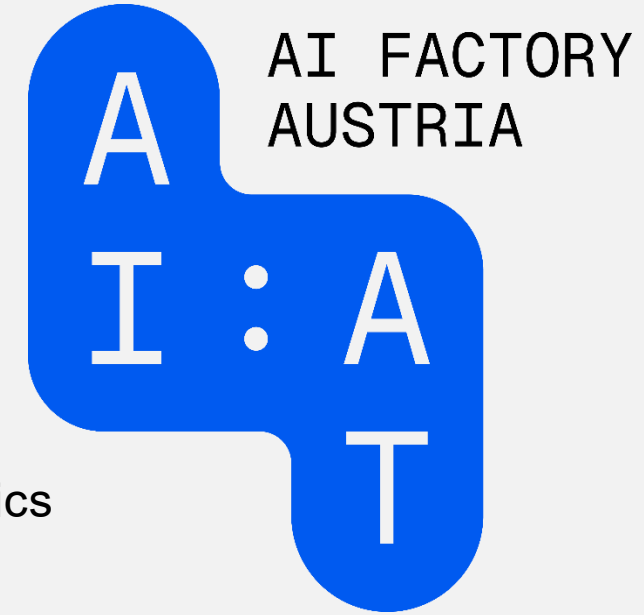
Connecting the Ecosystem

Funded by

under discussion with

AI Factory Austria AI:AT - PUBLIC has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement
No 101253078. The JU receives support from the Horizon Europe Programm of the European Union and Austria (BMIMI / FFG).

Contact



Peter Biegelbauer

Co-Lead Legal, Regulatory and Ethics
AI Factory Austria AI:AT

+43 664 88390033
peter.biegelbauer@ai-at.eu

Michael Löffler

Lead Legal, Regulatory and Ethics
AI Factory Austria AI:AT

+43 664 88390692
michael.loeffler@ai-at.eu

AI Factory Austria AI:AT
Schwarzenbergplatz 2
1010 Wien, Austria

info@ai-at.eu
ai-at.eu

@ai-factory-austria