The AI Act

An introduction

Floortje Jorna, SURF

4 December 2024

Goals

What are the objectives of this presentation?

To tell you the basics of the AI Act

To start the discussion about the possible impact of the AI Act

To make you think about which steps to take next within your institution

We will share the slides afterwards

01
Introduction

Introduction

The AI Act consists of rules for the use of Artificial Intelligence (AI) in the European Union

Why?

Enable innovation and economic development, while protecting public values

Respect freedom of science

Support innovation, respect freedom of science, and should not undermine research and development activity

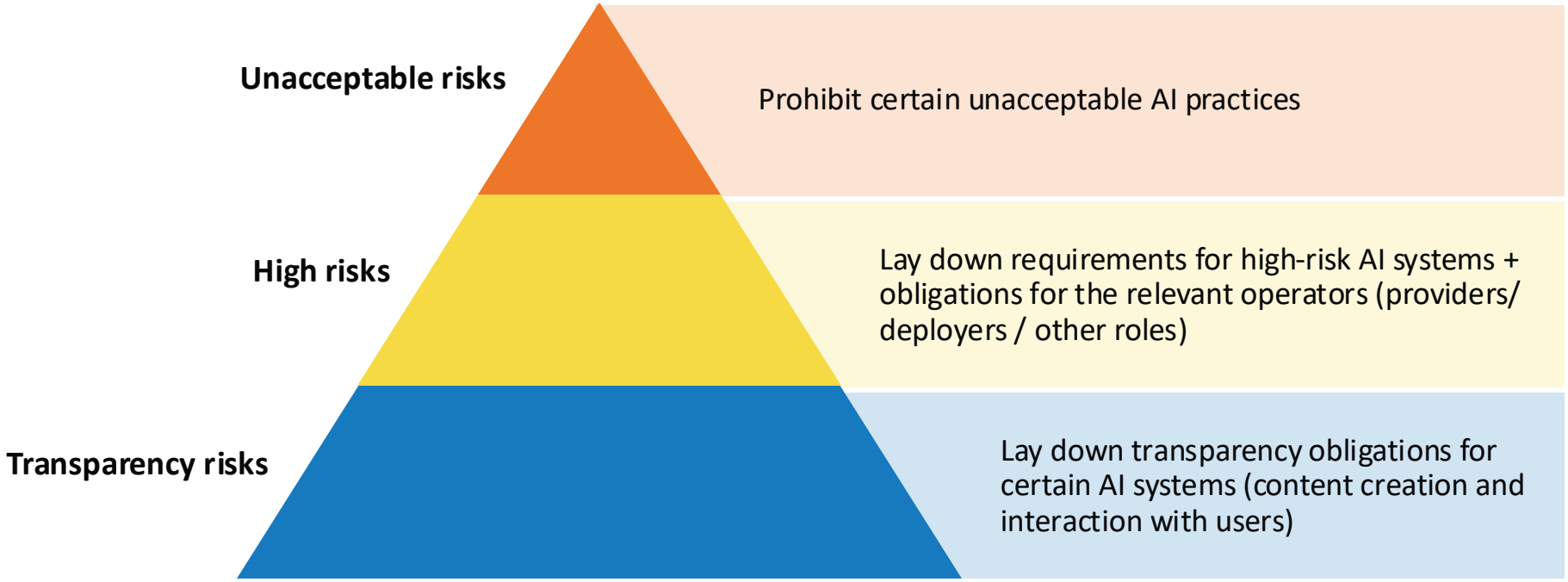
AI systems used in Europe are:

- Safe
- Transparant
- Traceable
- Non-discriminatory
- Environmentally friendly

SURF

Risk-based approach AI systems

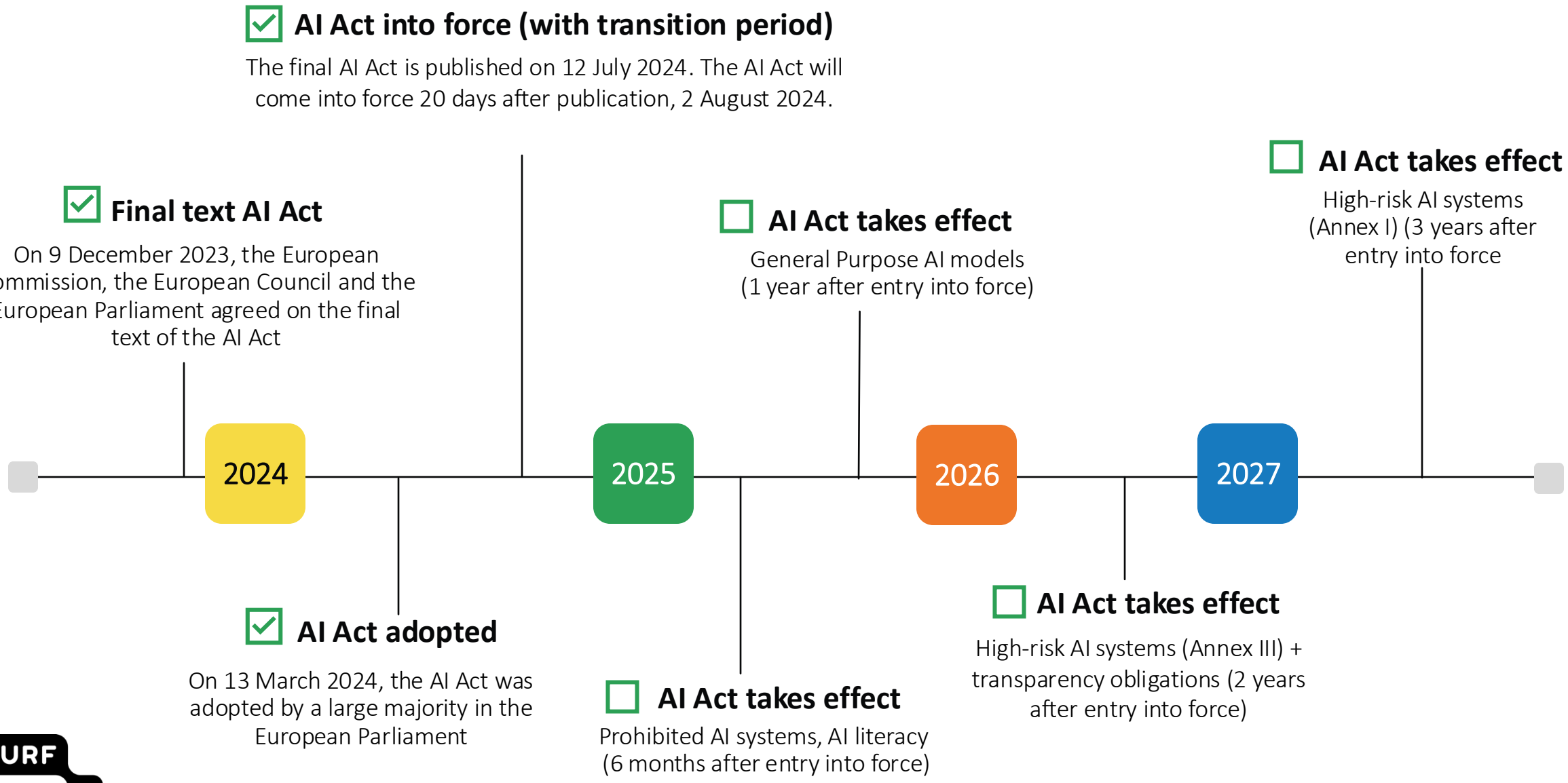
Tailor the type and content of rules to the intensity and scope of the risks that AI systems can generate



Other relevant highlights of the AI Act

- Requirements for every operator in the AI value chain (provider, deployer, etc.)
- AI literacy
- Requirements for General Purpose AI models (GPAI)
- Some research and development is out of scope of the AI Act
- AI regulatory sandboxes

Current status AI Act

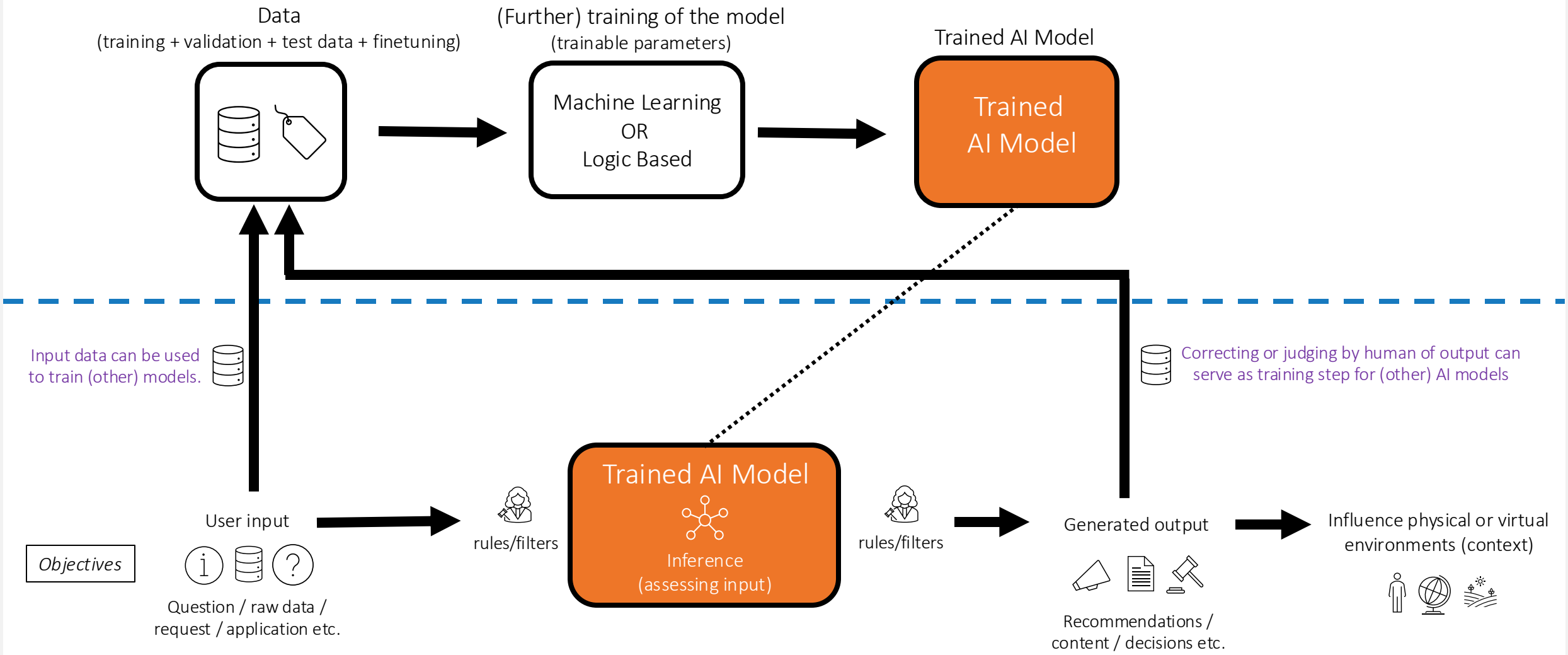


Before we dive into the AI Act…

First of all, let's make sure we all have a basic understanding of the training and deployment of AI models.

We will then build on this foundation to explain the AI Act in more depth.

'Training' an AI model



'Deploying' an AI system

02
Definitions and roles

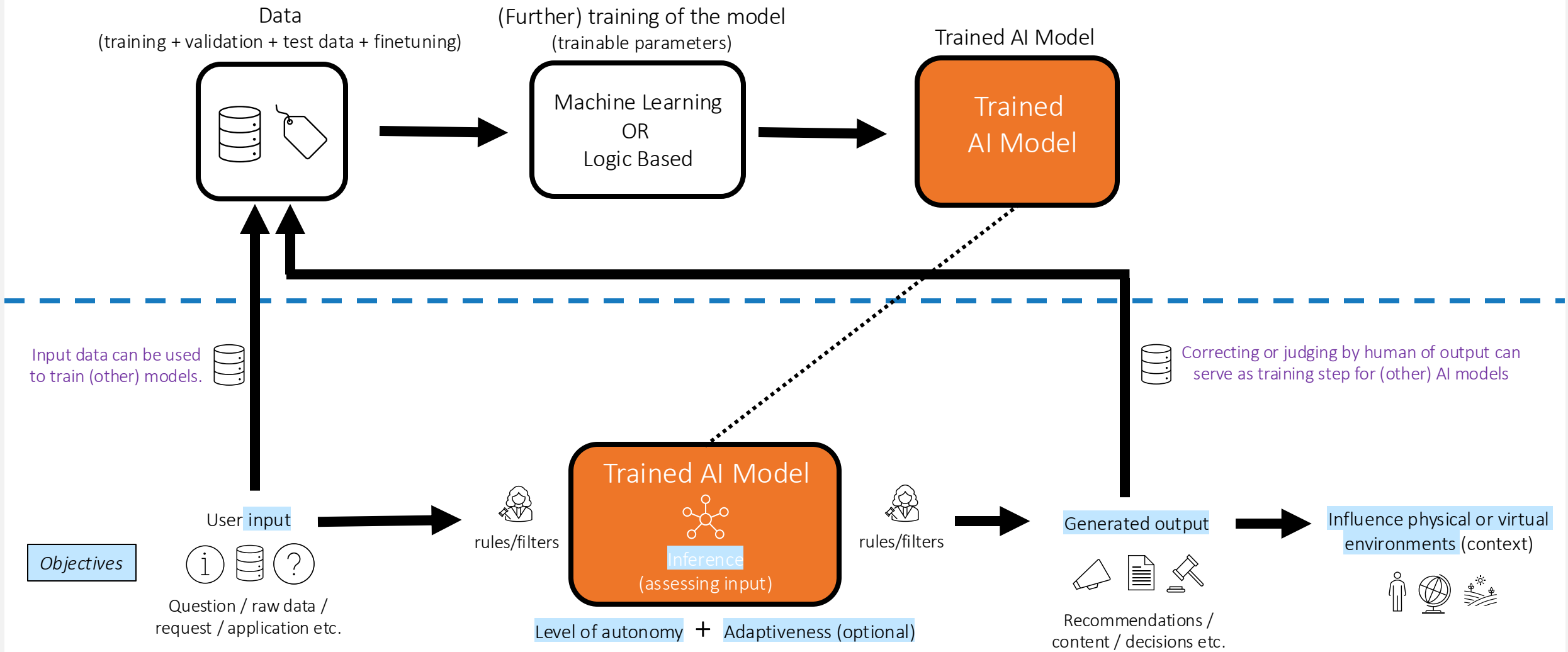
Key definitions

AI system

‘AI system‘ is 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

- *Varying levels of autonomy*
- *Adaptiveness after deployment*
- *Explicit or implicit objectives*
- *Infers from the input*
- *Generate outputs*
- *Influence physical or virtual environments*

'Training' an AI model



'Deploying' an AI system

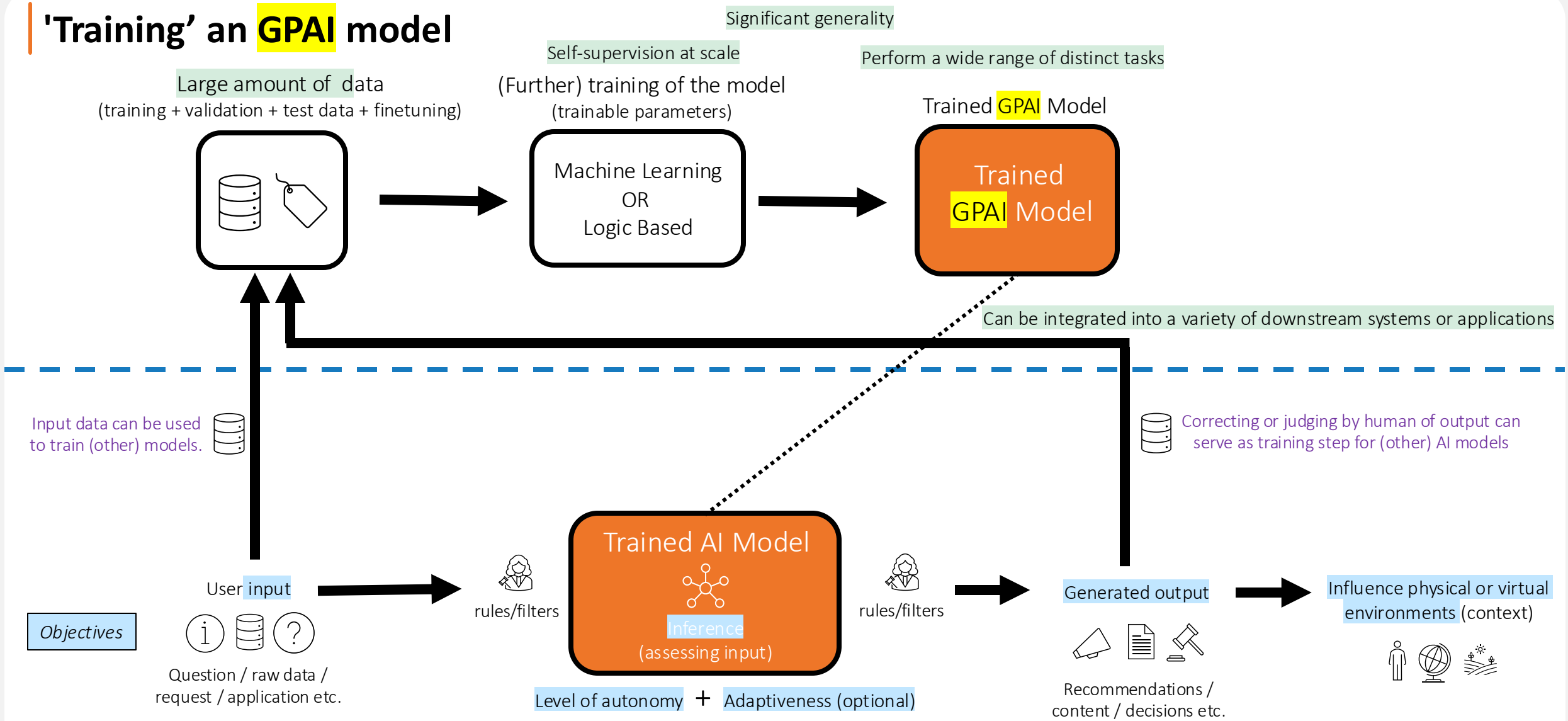
Key definitions

General Purpose AI *model*

‘General purpose AI model’ means an AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications. This does not cover AI models that are used before release on the market for research, development and prototyping activities

- *Trained with a large amount of data*
- *Self-supervision at scale*
- *Significant generality*
- *Perform a wide range of distinct tasks*
- *Can be integrated into a variety of downstream systems or applications*

'Training' an GPAI model



'Deploying' an AI system

Key definitions

General Purpose AI *system*

‘General purpose AI system’ means an AI system which is based on a general purpose AI model , that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

- *An AI system which is based on a general purpose AI model*
- *Has the capability to serve a variety of purposes*

'Training' an GPAI model

Significant generality

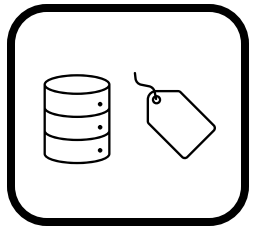
Self-supervision at scale

Perform a wide range of distinct tasks

Large amount of data
(training + validation + test data + finetuning)

(Further) training of the model
(trainable parameters)

Trained AI Model



Machine Learning
OR
Logic Based

Trained
AI Model

Can be integrated into a variety of downstream systems or applications

Input data can be used to train (other) models.

Correcting or judging by human of output can serve as training step for (other) AI models

Based on GPAI model

Variety of purposes

~~Objectives~~

User input

Question / raw data / request / application etc.

rules/filters

Trained GPAI Model

Inference
(assessing input)

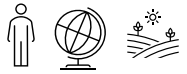
rules/filters

Level of autonomy + Adaptiveness (optional)

Generated output

Recommendations / content / decisions etc.

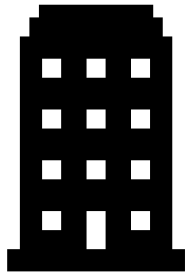
Influence physical or virtual environments (context)



'Deploying' an GPAI system

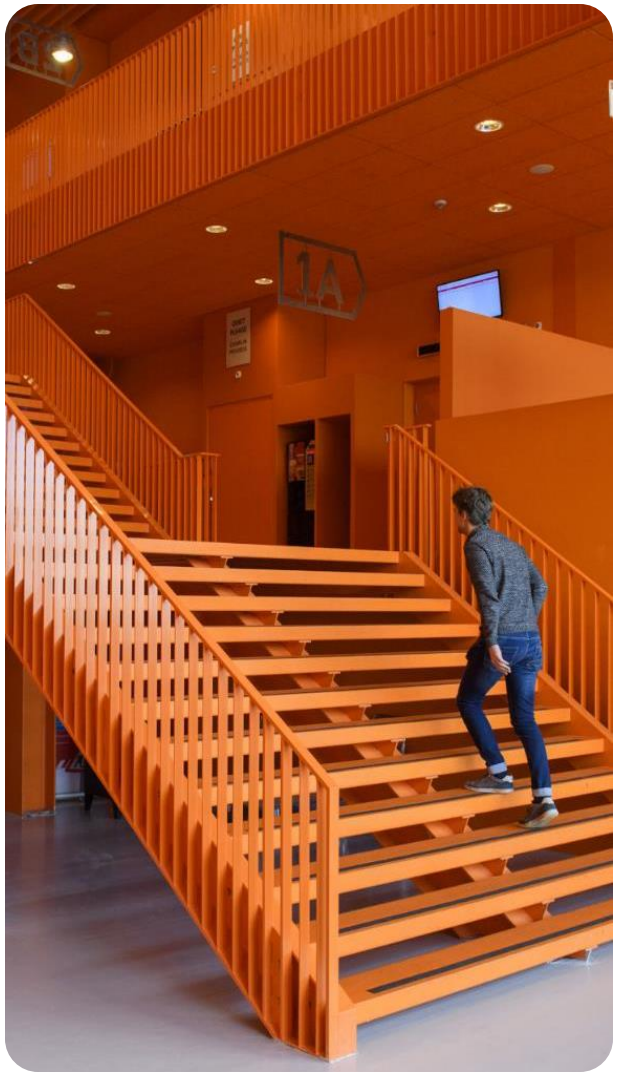
Role: Provider

A natural or legal person, public authority, agency or other body that develops an AI system or a general purpose AI model or that has an AI system or a general purpose AI model developed and places them on the market or puts the system into service under its own name or trademark, whether for payment or free of charge.



Provider

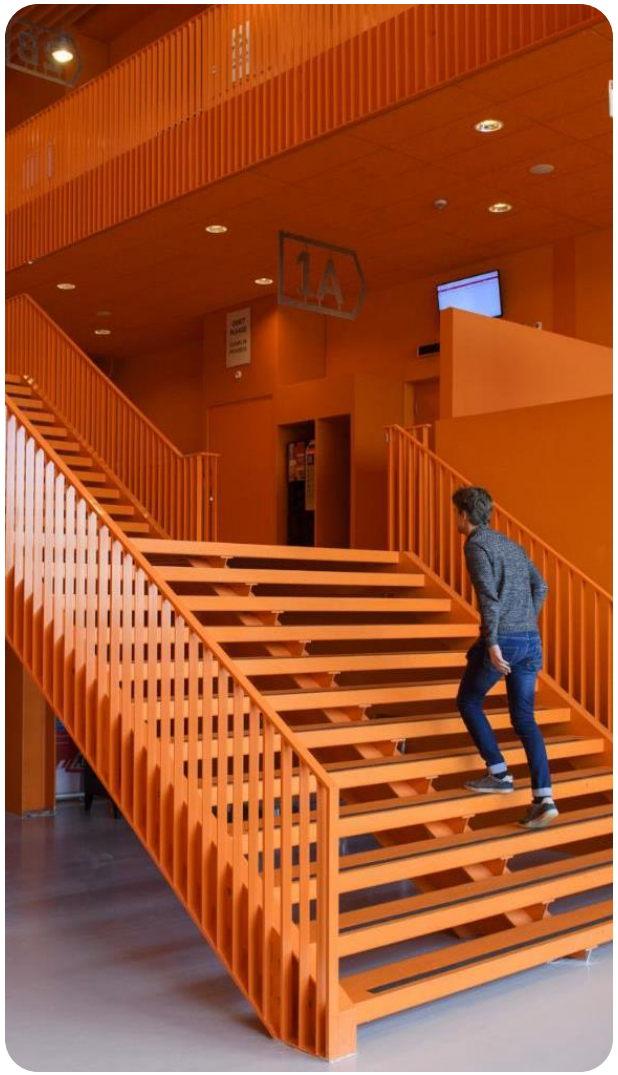
- Develops AI system or GPAI; OR
- Developed and places AI system or GPAI on the market; OR
- Puts the system into service under its own name or trademark



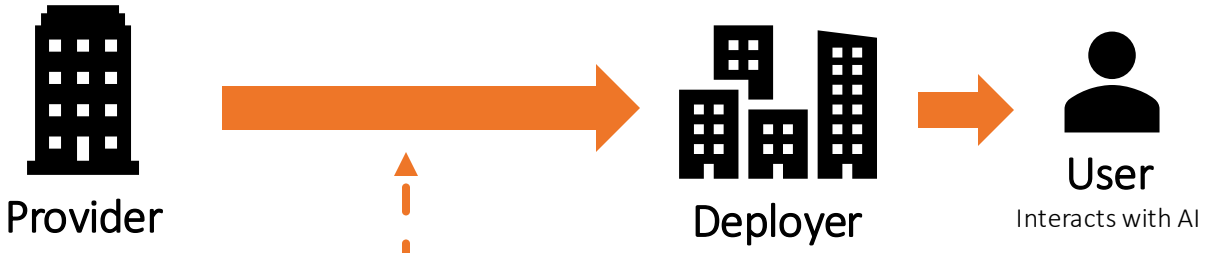
Role: Deployer

Any natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.

- Uses an AI system under its authority
- (Except personal non-professional activity)



Other roles



Authorised representative

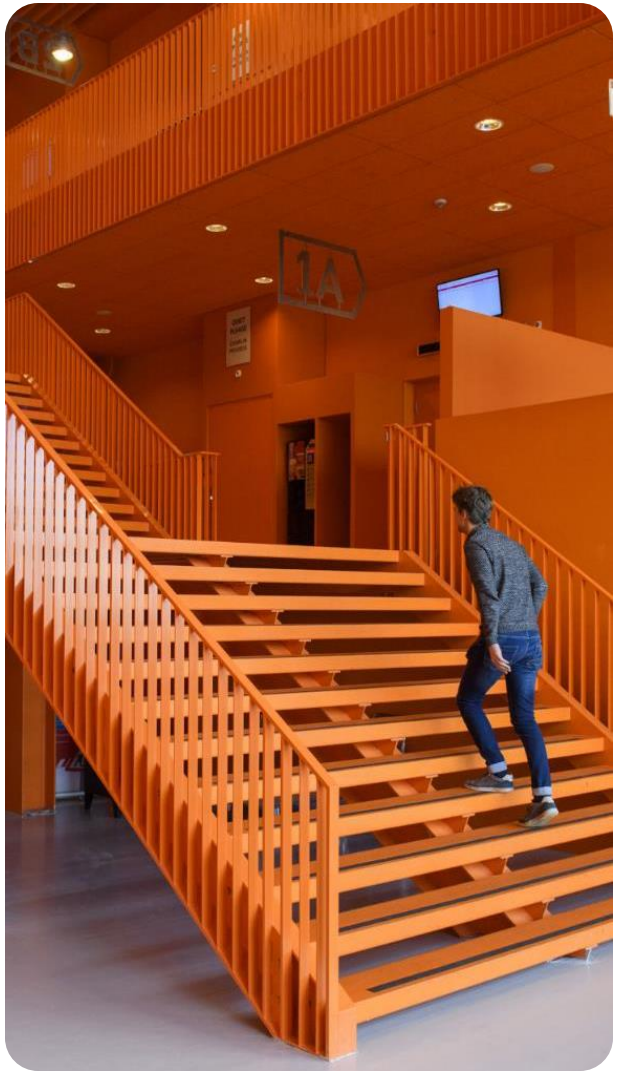
Means any natural or legal person located or established in the Union who has received and accepted a written mandate from a provider of an AI system or a general purpose AI model to, respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation

Distributor

Means any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market

Importer

Means any natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established outside the Union



Important: ways to become a provider!

Provider

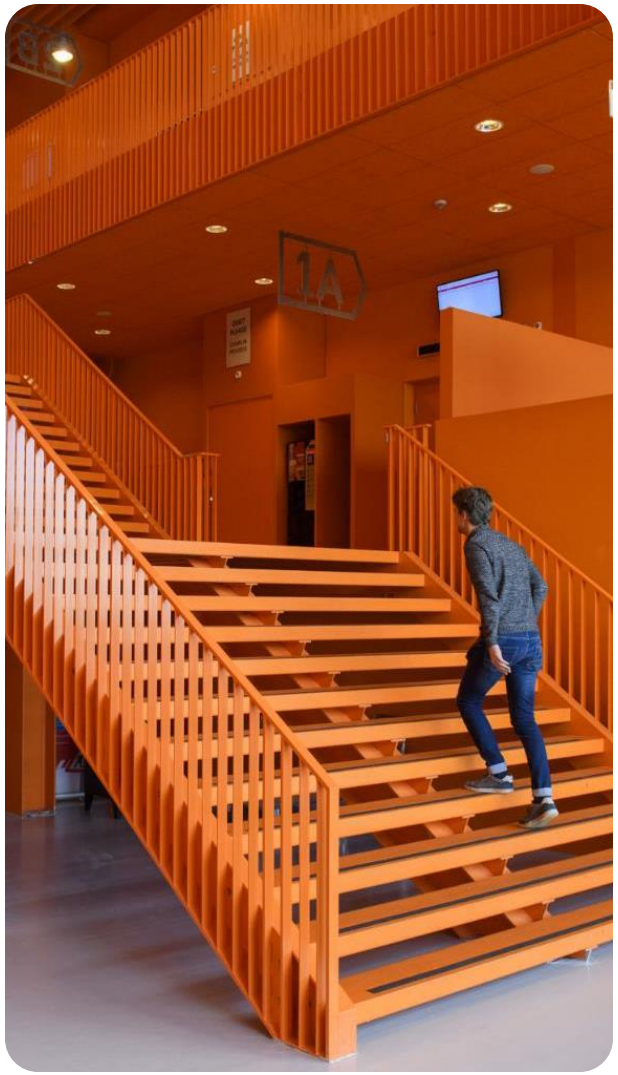
~~Deployer~~
Provider

Put your own name or trademark on a high-risk AI system already put into service or placed on the market

Make substantial modification to a high-risk AI system already put into service or placed on the market

Modify intented purpose of an AI system, or GPAI system, which was not high-risk, but becomes high-risk

Think about the use of ChatGPT for your own intented purposes



03
Prohibited AI systems

What are prohibited AI systems?

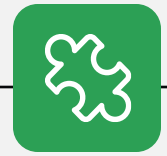
Practices that contradict Union values and fundamental rights

Human dignity

Freedom

Values

Equality



Democracy

Non-discrimination

Dataprotection

Fundamental rights

Privacy

Rights of the child

SURF

Phase out within 6 months

Closed list of prohibited AI systems

1. Using subliminal techniques or purposefully manipulative or deceptive techniques to materially distort behaviour, leading to significant harm
2. Exploiting vulnerabilities of a person or group due to specific characteristics, leading to significant harm
3. Biometric categorisation systems that individually categorise a person based on sensitive information, except for labelling or filtering lawfully acquired biometric datasets in the area of law enforcement
4. Social scoring systems
5. Real-time remote biometric identification systems in the public for law enforcement purposes
6. Predictive policing based solely on profiling or personality traits, except when supporting human assessments based on objective, verifiable facts linked to criminality
7. Facial recognition databases based on untargeted scraping
8. Inferring emotions in workplaces or educational institutions, except for medical or safety reasons

Article 5 AI Act

=

Protect within 6 months

Why prohibited? What is protected?

1. Autonomy, decision-making and free choices
2. Vulnerable individuals and groups
3. Privacy of individuals
   (protects against certain profiling)
4. Equality and justice
   (protects against discriminatory outcomes and exclusion)
5. The rights and freedom of persons, private life
6. The presumption of innocence
7. Right of privacy
   (protects against mass surveillance)
8. Right to be treated equal
   (protects agains biases an unfair treatment)

Recitals 29-44 AI Act

04

High risk AI systems

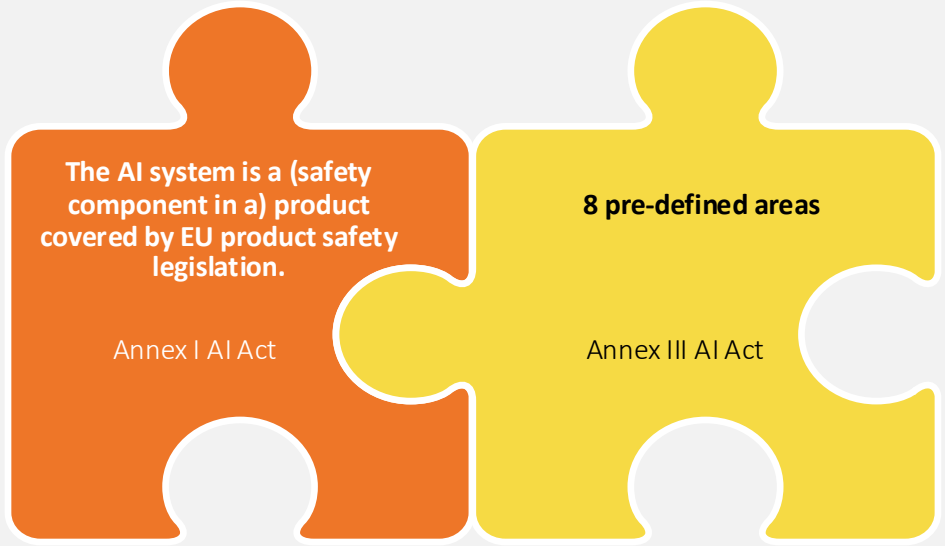
High-risk AI systems

AI systems identified as high–risk should be limited to those that have a significant harmful impact on the health, safety and fundamental rights of persons in the Union and such limitation minimises any potential restriction to international trade, if any.

Two types of high-risk AI systems

For example:

Machinery
Safety of toys
Safety components for lifts
Radio equipment
Pressure equipment
Medical devices



1. Biometrics
2. Critical infrastructure
3. Education and vocational training
4. Work
5. Essential private/public services and benefits
6. Law enforcement
7. Migration, asylum and border control management
8. Administration of justice and democratic processes

In every area, some high-risk AI systems are defined in Annex III

High-risk AI systems in education and vocational training

Used to determine access or admission

Used to evaluate outcomes

Used to assess the appropriate level of education

Used to monitor and detect prohibited behaviour of students during tests

These AI systems are qualified as high-risk, **unless…**

The AI system is intented to:

1. Perform a narrow procedural task
2. Improve the result of a previously completed human activity
3. Intented to detect decision-making patterns
4. Perform a prepatory task to an assessment relevant for the purpose of the use case

These exceptions apply to every AI system listed in Annex III

You can't use these exceptions when the AI system performs profiling of natural persons

Some other high-risk AI systems listed in Annex III

Employment, workers management and access to self-employment

Used for recruitment or selection of natural persons

Used to make decisions affecting terms of work

Biometrics

Used for remote biometric identification systems*

Used for biometric categorisation**

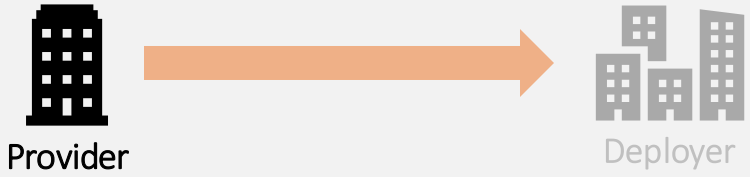
Used for emotion recognition

* Except AI systems whose sole purpose is to confirm that a specific natural person is the person he or she claims to be

**according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics

Provider of a high-risk AI system

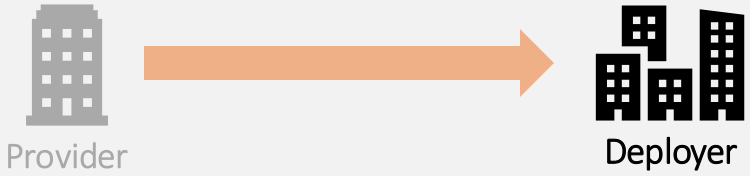
Requirements for Providers



- Risk management system
- Data governance
- Technical documentation
- Record-keeping
- Documentation keeping
- Transparency and provide instructions for use
- Allow deployers to implement human oversight
- Achieve appropriate levels of accuracy, robustness and cybersecurity
- Establish a quality management system
- Conformity assessement (CE)

Deployer of a high-risk AI system

Requirements for Deployers



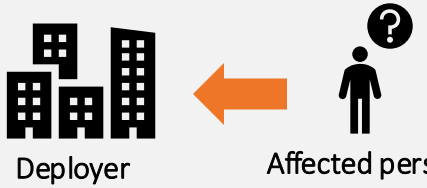
- Use AI systems in accordance with instructions
- Assign natural persons to ensure human oversight
- Inform provider/distributor and authority when AI system presents national risk
- Fundamental rights impact assessement (bodies governed by public law, or private entities providing public services)
- Input data is relevant and sufficiently representative
- Monitor the operation of high-risk AI systems
- Keep automatically generated logs for at least 6 months

A right to explanation of individual decision-making

A right for the affected person

Art. 86 AI Act:

1. Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, with the exception of systems listed under point 2 thereof, and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.



Decision which is taken by deployer

+

On the basis of the output from a Annex III high-risk AI system

+

Legal affects or similar significantly affects

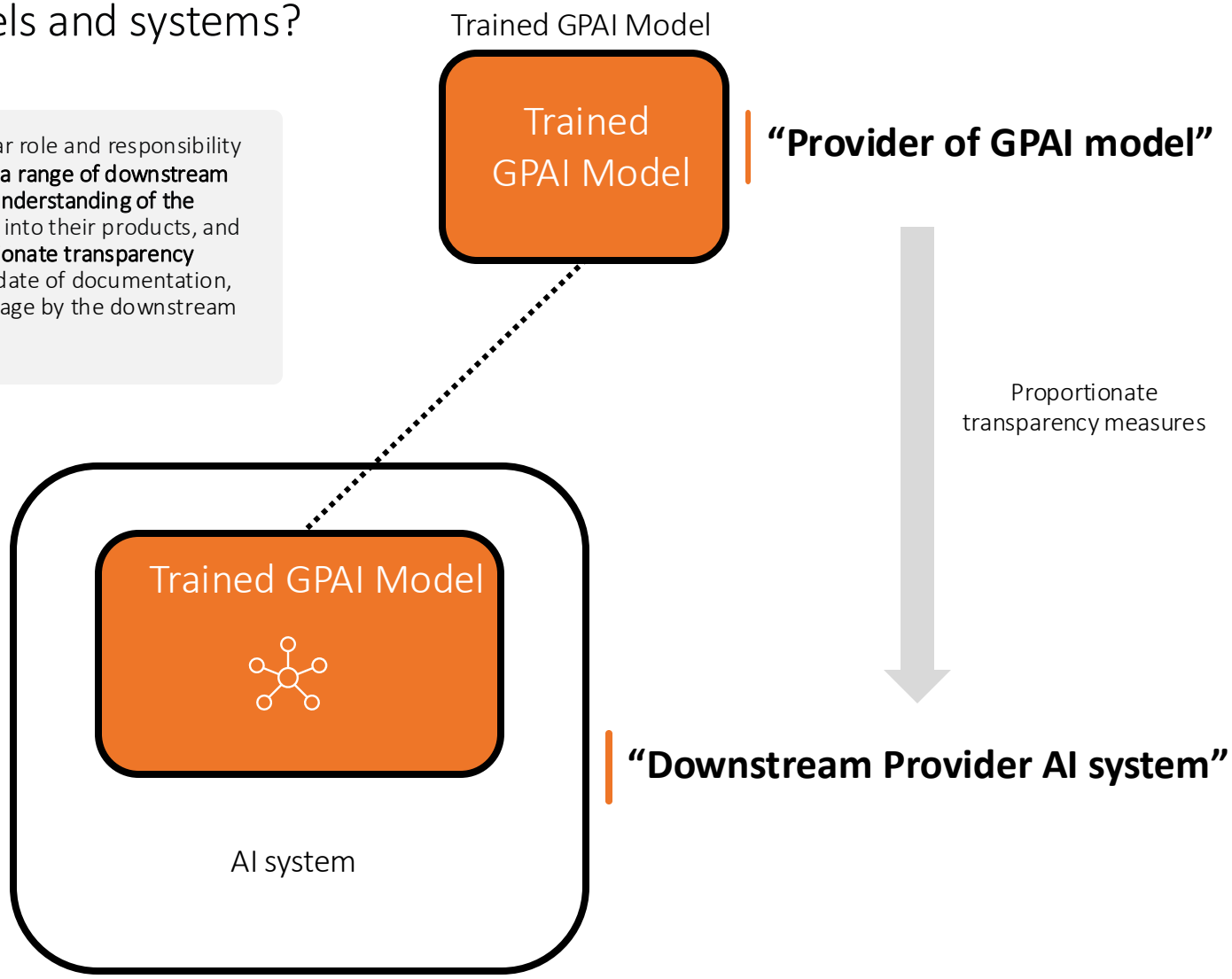
05

GPAI models and systems

General Purpose AI models

Why are there special rules for GPAI models and systems?

Recital 101 AI Act: “Providers of general-purpose AI models have a particular role and responsibility along the AI value chain, as the models they provide may **form the basis for a range of downstream systems**, often provided by downstream providers that **necessitate a good understanding of the models and their capabilities**, both to enable the integration of such models into their products, and to fulfil their obligations under this or other regulations. **Therefore, proportionate transparency measures should be laid down**, including the drawing up and keeping up to date of documentation, and the provision of information on the general-purpose AI model for its usage by the downstream providers.”



Provider of a GPAI model

Requirements for GPAI model providers

Trained GPAI Model

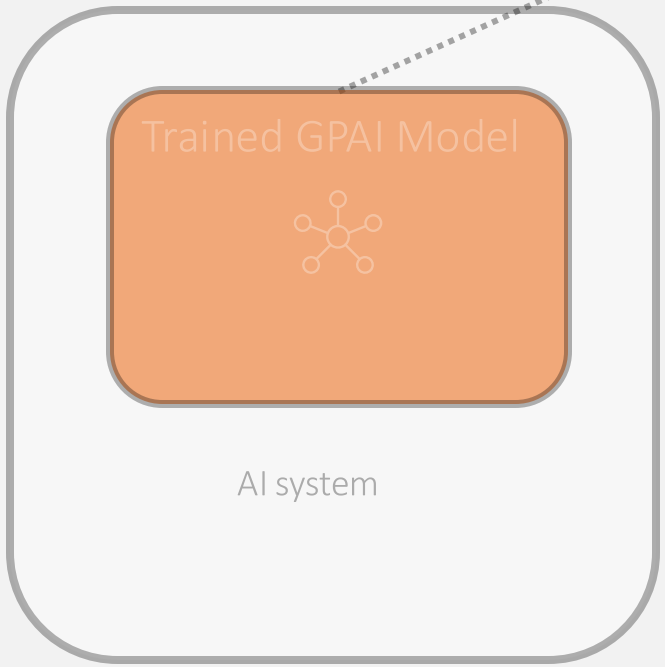
Trained GPAI Model

“Provider of GPAI model”

Requirements:

- Technical documentation (incl. training and testing process)
- Documentation for providers of AI systems who intend to integrate the GPAI model
- Put in place a policy to respect Union copyright law
- Publish a summary about the content used for training

These obligations don’t apply for free and open license GPAI model providers (unless there are systemic risks)



“Downstream Provider AI system”

Systemic risks

Classification of GPAI models with systemic risks

Article 51 AI Act:

1. A general-purpose AI model shall be classified as general-purpose AI model with systemic risk if it meets any of the following criteria:

(a) it has **high impact capabilities** evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks;
(b) based on a **decision of the Commission**, ex officio or following a qualified alert by the scientific panel that a general purpose AI model has capabilities or impact equivalent to those of point a).

2. A general-purpose AI model shall be presumed to have high impact capabilities pursuant to point a) of paragraph 1 when the cumulative amount of compute used for its training measured in **floating point operations (FLOPs) is greater than 10^25.**

GPAI models with systemic risks

Extra requirements for GPAI model providers

Requirements for every GPAI model Provider:

- Technical documentation (incl. training and testing process)
- Documentation for providers of AI systems who intend to integrate the GPAI model
- Put in place a policy to respect Union copyright law
- Publish a summary about the content used for training

+

Extra requirements in case of systemic risks:

- Perform model evaluation, including adversarial testing, to identify and mitigate risks
- Assess and mitigate possible systemic risks
- Track, document and report serious incidents
- Ensure an adequate level of cybersecurity protection

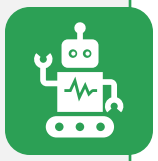
06
Other focus points

Transparency risks

Requirements against impersonation or deception

Recital 132 AI Act:

"Certain AI systems intended to interact with natural persons or to generate content may pose **specific risks of impersonation or deception** irrespective of whether they qualify as high-risk or not. In certain circumstances, the use of these systems should therefore be subject to specific transparency obligations (…)"



Inform when interacting with AI

Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious.

Mark AI generated output

Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.

Notify when using biometric data in some cases

Natural persons should be notified (by the **deployer**) when they are exposed to AI systems that, by processing their biometric data, can identify or infer the emotions or intentions of those persons or assign them to specific categories.

Label deep fakes

Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated.

Obligation: AI literacy

For providers and deployers in the AI value chain

Article 4 AI Act:

"Providers and deployers of AI systems shall take measures to ensure, to their best extent, **a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems** on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used."

Recital 20 AI Act:

"AI literacy should equip providers, deployers and affected persons with the necessary notions **to make informed decisions regarding AI systems**.

Those notions may vary with regard to the relevant context (…) "

Understanding the correct application of technical elements during the AI system's development phase

The measures to be applied during its use

The suitable ways in which to interpret the AI system's output

The knowledge necessary to understand how decisions taken with the assistance of AI will have an impact on affected persons.

Respect freedom of science

Support innovation, respect freedom of science, and should not undermine research and development activity

Recital 25 AI Act:

"This Regulation should support innovation, should respect freedom of science, and should not undermine research and development activity. It is therefore necessary to exclude from its scope AI systems and models specifically developed and put into service for the sole purpose of scientific research and development. Moreover, it is necessary to ensure that this Regulation does not otherwise affect scientific research and development activity on AI systems or models prior to being placed on the market or put into service. As regards product-oriented research, testing and development activity regarding AI systems or models, the provisions of this Regulation should also not apply prior to those systems and models being put into service or placed on the market. (…)"

Exclusion from scope of the AI Act:

Sole purpose of scientific research and development

The AI Act does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development.

Research, testing development activity, prior to placing on the market

The AI Act does not apply to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service. Such activities shall be conducted in accordance with applicable Union law. Testing in real world conditions shall not be covered by that exclusion.

Measures in support of innovation

Regulatory oversight and a safe and controlled space for experimentation, while ensuring responsible innovation and integration of appropriate safeguards and risk mitigation measures.

AI regulatory sandboxes

- Every member state establishes at least one AI regulatory sandbox at national level (incl. guidance, supervision and support)
- Under circumstances, it is allowed to further process personal data for the purpose of developing, training and testing AI systems in the sandboxes

Testing high-risk AI systems outside sandboxes

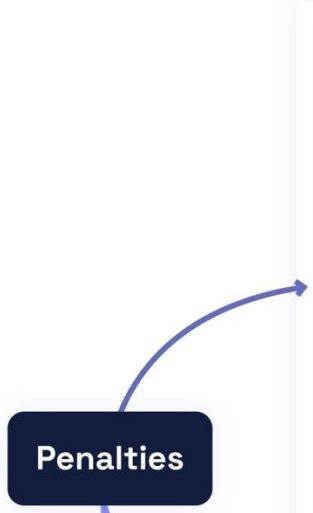
- Allowed if strict conditions are met
- Informed consent to participate in testing outside sandboxes

Special attention for micro, small and medium-sized enterprises

- Inlcuding start-ups
- Priority access to the AI regulatory sandboxes
- Awareness raising and training activities
- New and existing dedicated channels

SURF

Penalties: what if you don't comply?



General fines for operators of AI systems	**Up to 35 000 000 EUR** or, if the offender is a company, up to 7 % of its total worldwide annual turnover for the preceding financial year, whichever is higher	Non-compliance with the prohibition of the artificial intelligence practices under Article 5
	Up to 15 000 000 EUR or, if the offender is a company, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher	Non-compliance with the following: • obligations of providers pursuant to Article 16 • obligations of authorised representatives pursuant to Article 25 • obligations of importers pursuant to Article 26 • obligations of distributors pursuant to Article 27 • obligations of deployers pursuant to Article 29, paragraphs 1 to 6a • requirements and obligations of notified bodies pursuant to Article 33, 34(1), 34(3), 34(4), 34a • transparency obligations for providers and users pursuant to Article 52
	Up to 7 500 000 EUR or, if the offender is a company, up to 1 % of its total worldwide annual turnover for the preceding financial year, whichever is higher	The supply of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request
General fines for operators of AI systems	**Up to 15 000 000 EUR** or, if the offender is a company, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher	In one of the following: • Infringement of the GPAI-relevant provisions • Failure to comply with a request for document or information pursuant to Article 68i or supply of incorrect, incomplete or misleading information • Failure to comply with a measure requested under Article 68k • Failure to make available to the Commission access to the general purpose AI model or general purpose AI model with systemic risk with a view to conduct an evaluation pursuant to Article 68j
For Union institutions, agencies and bodies	**Up to EUR 1 500 000**	Non-compliance with the prohibition of the artificial intelligence practices under Article 5
	Up to EUR 750 000	Non-compliance of the AI system with any requirements or obligations

www.holisticai.com

Holistic AI

https://www.holisticai.com/blog/penalties-of-the-eu-ai-act

SURF

07
Possible impact

Possible impact on institutions

Important context for education and research institutions

Education

- Provider of deployer of prohibited AI systems?
- Provider of deployer of high-risk AI systems?
- Content creation and transparency risks?
- Becoming a provider by using GPAI models?

Research & Innovation

Different obligations and requirements apply

- Development of GPAI models and systems
- AI for research purposes

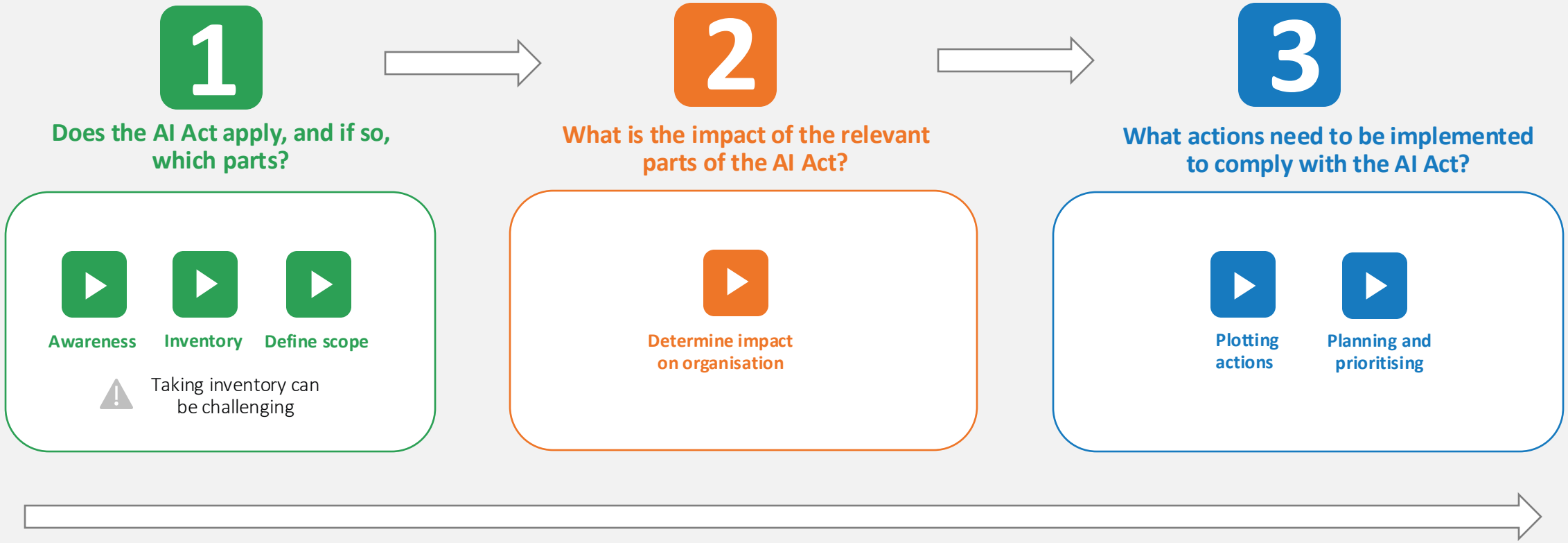
Use of (GP)AI at work

- Possible high-risk AI systems when it comes to employees and AI?
- Content creation and transparency risks?
- Becoming a provider by using GPAI models?

AI literacy

Implementation

How to implement the AI Act within an organisation?

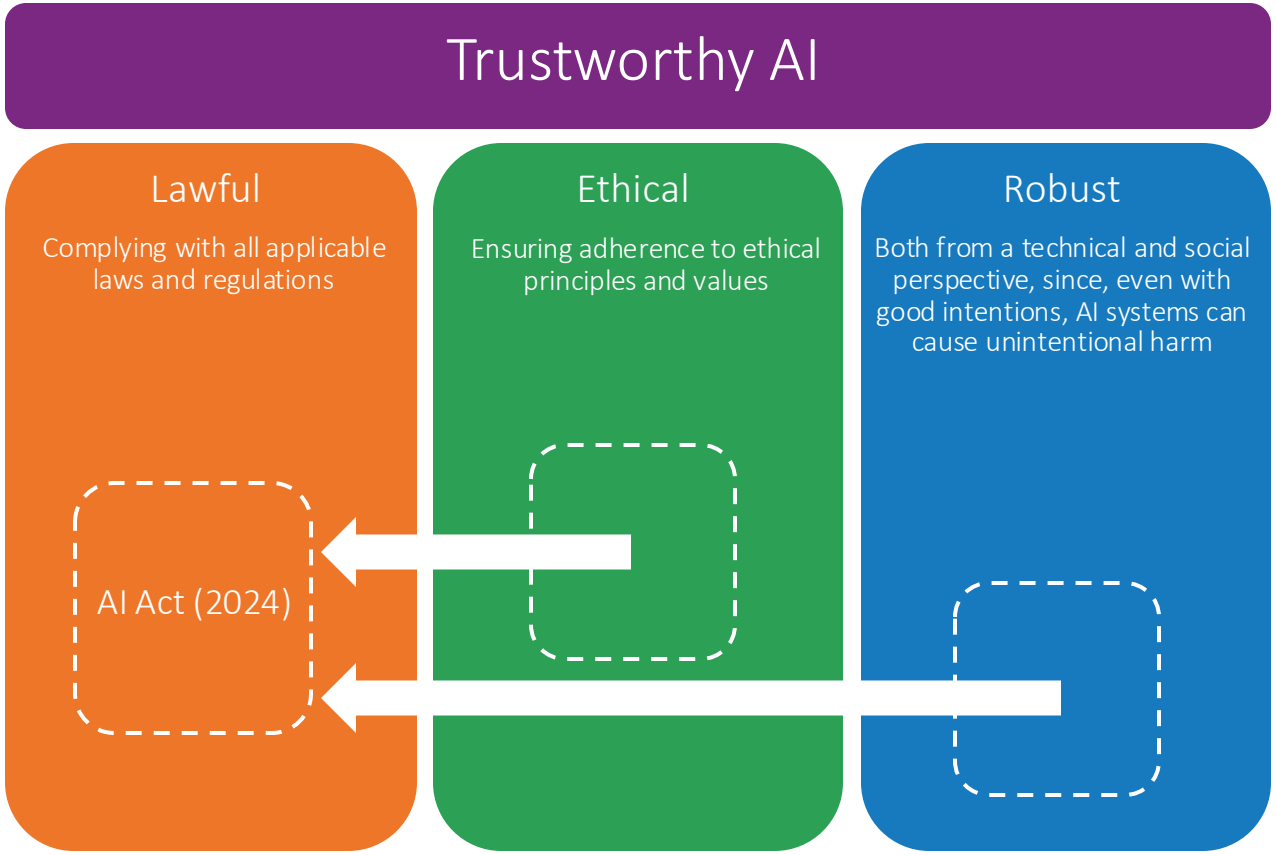


Parallel steps beyond scope AI Act itself: examine how AI takes shape within the organisation

Implementing the AI Act is not a stand-alone activity. AI is broader than the AI Act. The AI Act interfaces with various other legislation (e.g. GDPR, intellectual property, CE legislation and liability legislation). The AI Act interfaces with (AI) governance, data governance, data management, data science, statistics, quality systems, risk management and ethics, among others.

Trustworthy AI

According to High Level Expert Group AI, European Commission (2019)




Questions?


SURF