Deloitte.

The AI Act @ April 2024

Europe's Comprehensive AI Regulation

02.05.2024 | David Thogmartin | Risk Advisory Germany

MAKING AN IMPACT THAT MATTERS since 1845

Executive Summary

The AI Act will enter into force 20 days after translation into all official EU languages, expected to be mid-June 2024

Purpose

To promote human-centric and trustworthy AI – protecting health, safety, fundamental rights, democracy and the rule of law, and the environment from potential harmful effects – while supporting innovation, particularly among European SMEs.

Scope

AI deployed in the European Union. Extra-territorial reach for foundation models & very large online platforms (social media).

Approach

A risk-based approach, categorizing AI systems by use case into categories "unacceptable risk", "high risk", "minimal risk", which drive compliance obligations (prohibited, declaration of conformity, transparency requirements, or voluntary standards).

Foundation Models & General Purpose AI (GPAI)

Given their wide-ranging application are risk-classified using alternate criteria and generally subject to enhanced transparency obligations.

Compliance

Providers (developers, deployers, …) must establish Quality Management Systems and validate high-risk AI systems against trustworthy AI principles prior to issuing a Declaration of Conformity and registering in a public EU database. Post-launch, providers must log issues into the EU database and update conformity assessments throughout the lifecycle.

Timing

The European Parliament approved the final text on 13.3.2024. The AI Act enters into force 20 days after translation into all official EU languages, targeting mid-June 2024.

Enforcement

EU-wide authorities will coordinate across member states and follow larger topics, such as foundation models and GPAI. National supervisors will enforce compliance, appointing "notified bodies" (permitted 3rd party auditors) to assess conformity in specific cases, engaged either by providers prior to issuing Declarations of Conformity or by the supervisor for audits.

Consequences

Fines range from 35 m€ / 7% global turnover (prohibited cases), 15 m€ / 3% (other infringements) to 7,5 m€ / 1,5% (reporting errors), as well as potential non-monetary penalties, such as forced removal of the AI system from the market.

Definition of AI

The final, agreed definition of AI aligns closely to the OECD definition

Various positions prior to the Trilogue

European Commission

- Machine learning
- Rule-based or knowledge-based decision aids
- Traditional statistical models

European Parliament

- Varying levels of autonomy
- Explicit or implicit objectives
- Predictions, recommendations, decisions

Council of EU Member States

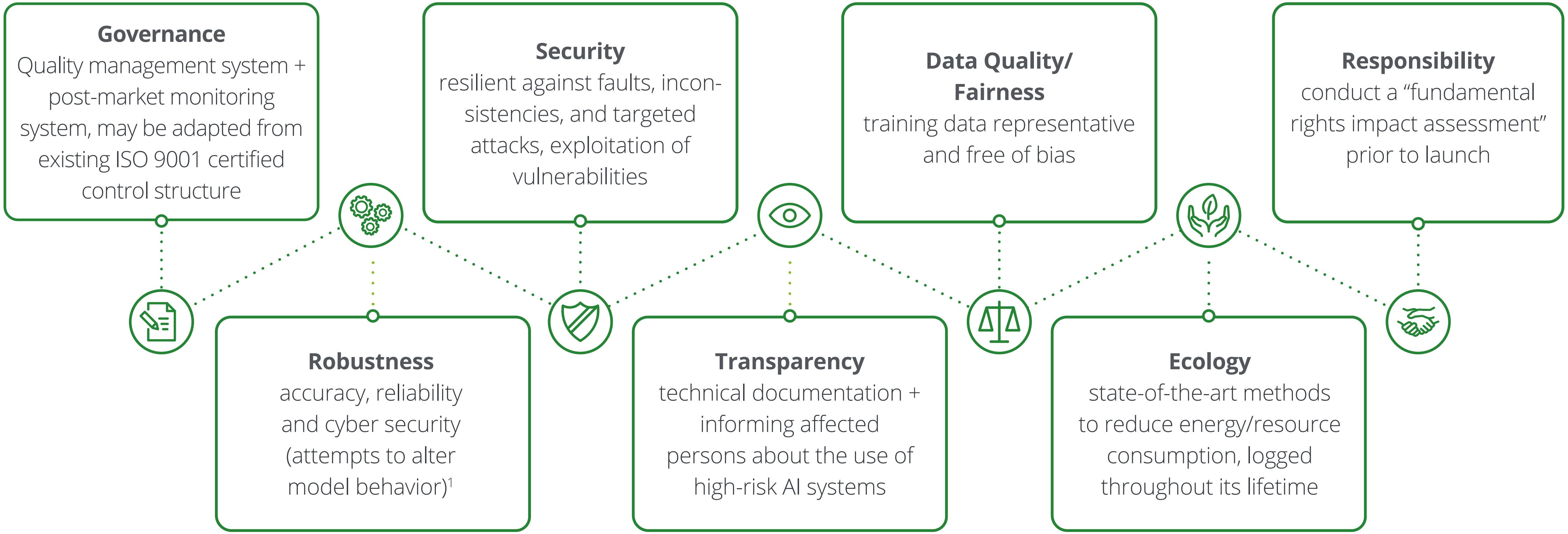
- Narrower definition of AI by listing specific techniques
- Outputs include content (i.e., generative AI)

The negotiated definition will align with the OECD

> **"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 can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."**

Trustworthy AI at the Core

The principles of Trustworthy AI are the paradigm for AI quality and form the bedrock for emerging technical standards



Governance
Quality management system + post-market monitoring system, may be adapted from existing ISO 9001 certified control structure

Security
resilient against faults, inconsistencies, and targeted attacks, exploitation of vulnerabilities

Data Quality/ Fairness
training data representative and free of bias

Responsibility
conduct a "fundamental rights impact assessment" prior to launch

Robustness
accuracy, reliability and cyber security (attempts to alter model behavior)[1]

Transparency
technical documentation + informing affected persons about the use of high-risk AI systems

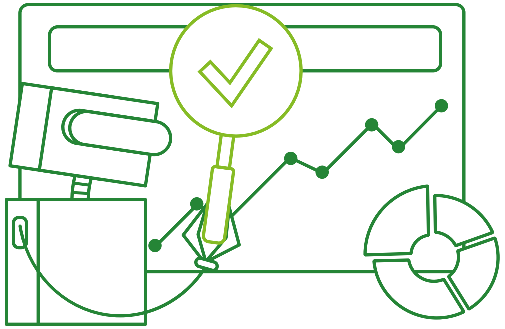
Ecology
state-of-the-art methods to reduce energy/resource consumption, logged throughout its lifetime

Beyond principles, European standards setters (e.g., CEN/CENELEC) will more concretely define how each of these principles translate into technical standards, against which AI systems must demonstrate compliance by law.

[1] ... Data poisoning, model poisoning, adversarial examples, model evasion, confidentiality attacks

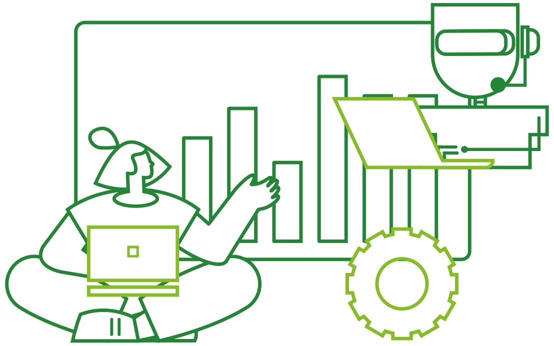
Application Scope

Any AI systems affecting European citizens… not necessarily hosted or operated in Europe



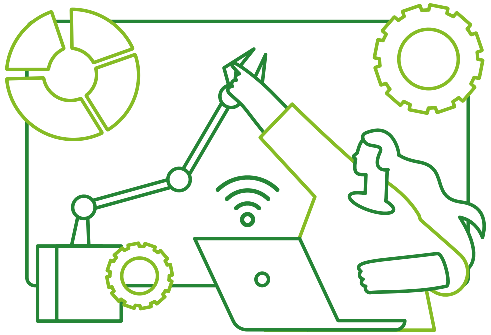
AI Systems…

- AI deployed in the EU
- Hosted outside the EU but deployed in the EU/accessed by European citizens (foundation models, very large online platforms = social media AI)



Providers…

- Developers
- Importers
- Resellers
- Deployers



Exceptions…

Commercial R&D prior to market placement

- Public authorities outside the EU
- R&D prior to market placement
- Academic research
- Personal use
- Open-source licensed [1]

[1] Unless they are (a) used in a high-risk AI system, (b) foundation models

Enforcement

Each Member State shall establish a national supervisor, while the EU AI Office coordinates across borders

EU-wide

European AI Office

- A new body within the European Commission
- To coordinate the implementation of the Act throughout the EU Member States
- To monitor development of foundation models & general purpose AI

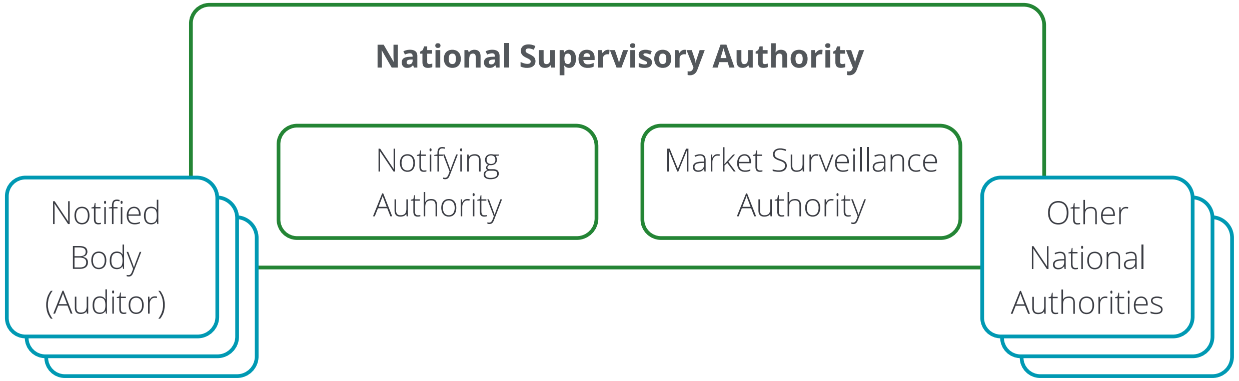
AI Advisory Board

- Composed of stakeholders from the business sector and civil society
- To provide a wide spectrum of viewpoints for consideration in the implementation process

Scientific Advisory Panel

- Consisting of independent experts
- To identify systemic risks, offer guidance on model classifications, ensure enforcement based on latest scientific understanding

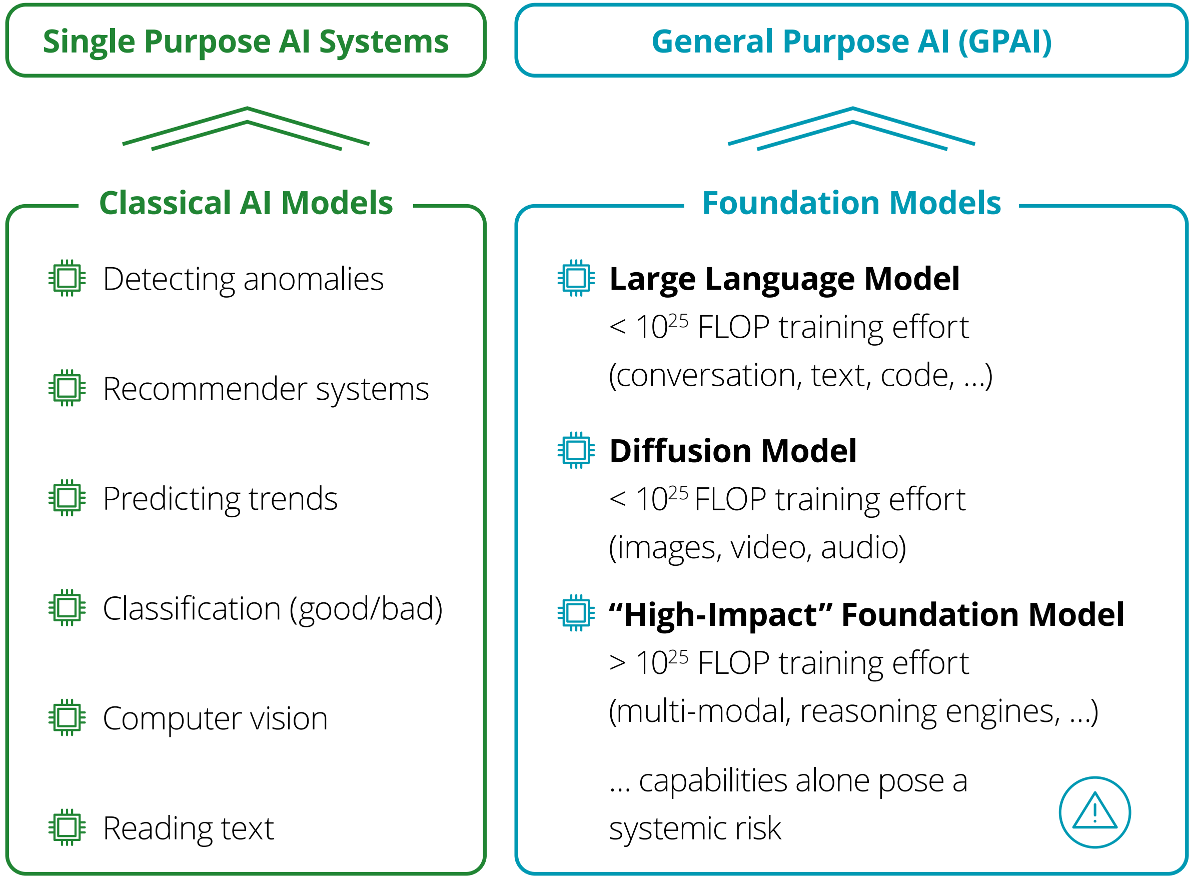
The National Supervisor


National Supervisory Authority
Notified Body (Auditor)
Notifying Authority
Market Surveillance Authority
Other National Authorities

- National supervisor – enforcing compliance
- Other competent authorities – i.e., to supervise other, sector-specific regulatory requirements
- Notifying authority – ensuring conformity assessments conducted properly & timely
- Notified bodies – accredited to conduct conformity assessments

General Purpose AI (GPAI) – Foundation Models

Contrary to SPAI, the risk of GPAI is measured by the power of the foundation model



Single Purpose AI Systems

Classical AI Models

- Detecting anomalies
- Recommender systems
- Predicting trends
- Classification (good/bad)
- Computer vision
- Reading text

General Purpose AI (GPAI)

Foundation Models

- **Large Language Model**
  < 10^{25} FLOP training effort
  (conversation, text, code, ...)
- **Diffusion Model**
  < 10^{25} FLOP training effort
  (images, video, audio)
- **"High-Impact" Foundation Model**
  > 10^{25} FLOP training effort
  (multi-modal, reasoning engines, ...)

... capabilities alone pose a systemic risk ⚠

Providers responsible

Developers of foundation models, deployers if core capabilities substantially altered

Risk categorization

Differentiation in between GPAI and "high-impact" GPAI posing systemic risk

GPAI

Transparency obligations, including technical documentation, training data respecting copyrights, watermarking of AI generated content

Systemic Risk GPAI

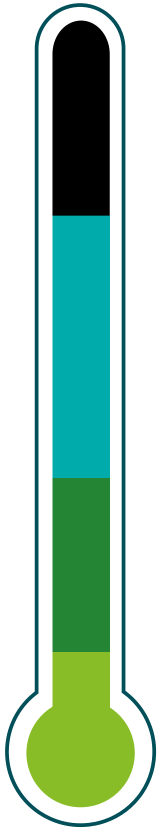
Subject to more stringent obligations similar to high-risk AI systems:

- model evaluations
- assess and mitigate systemic risks
- conduct adversarial testing
- report to the Commission on serious incidents
- ensure cyber security
- report on their energy efficiency
- adherence to codes of practice until harmonized EU standards published

[1] Developers or those commissioning development by IT experts on their behalf

Single Purpose AI (SPAI) – Risk Classification

A differentiated approach depending on the perceived risk to EU citizens



Risk Category		Consequence		Timeline
Unacceptable Risk • Social scoring • Realtime biometric identification • Emotion sensing at the workplace[a]	➡	Prohibited	➡	6 months
High-Risk AI Systems • EU Product Safety list • Critical services, infrastructure • Risk fundamental rights • Specific cases listed in Annex III	➡	Conformity, Governance, Monitoring, Logging	➡	24 to 36 months[b]
Limited-Risk • Chatbots • Deep fakes	➡	Transparency Obligations	➡	12 months (foundation models)
Minimal Risk • Internal models • Procedural tasks	➡	Voluntary Standards	➡	n.a

€ Penalties up to…

35 m€ or 7% global annual turnover

- Unacceptable use cases (Article 5)
- Failure to meet data & data governance requirements for High-Risk AI Systems (Article 10)

15 m€ or 3% global annual turnover

- All other cases than Articles 5 and 10

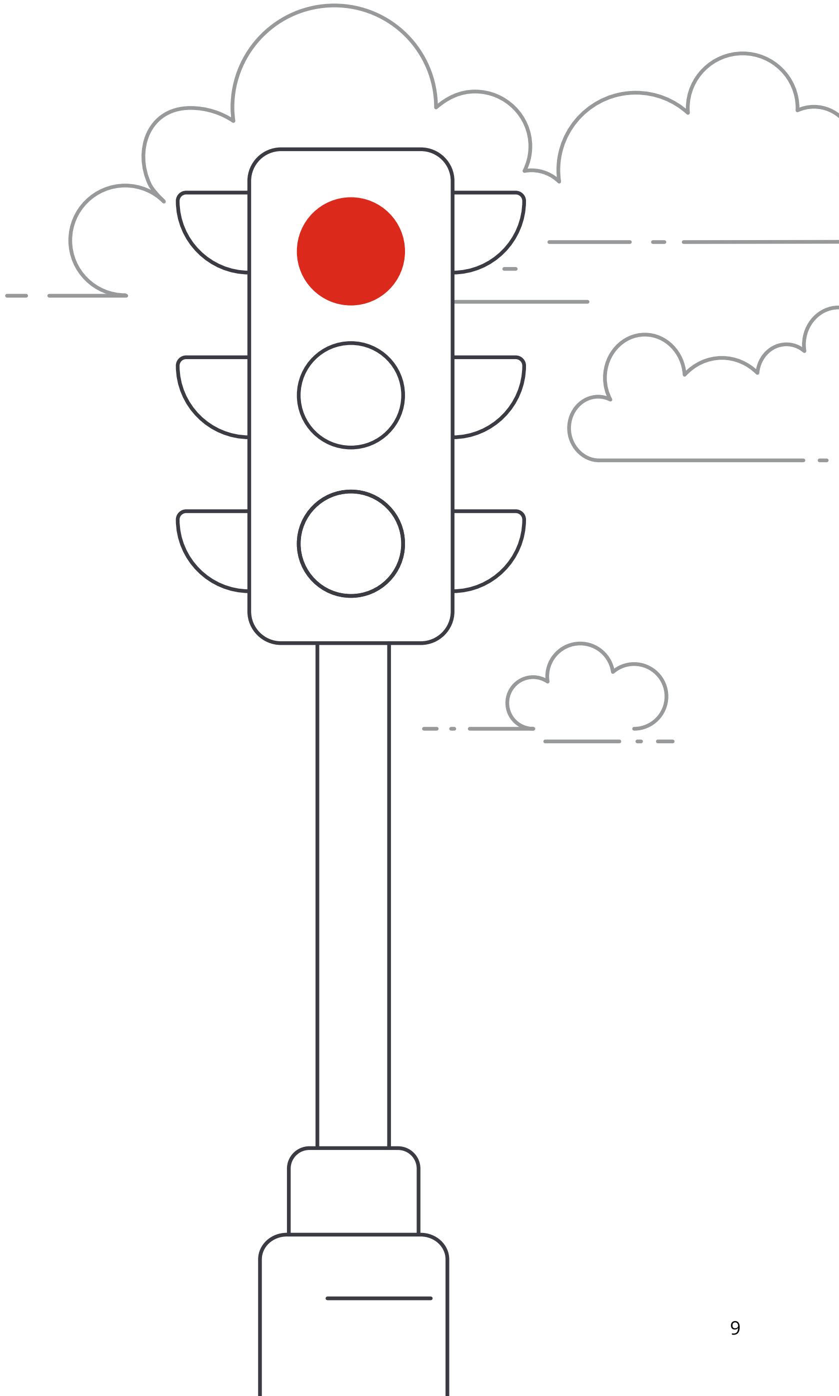
7,5 m€ or 1% global annual turnover

- Incorrect, incomplete, misleading response to a NCA/NB[c] request

[a] Except for fatigue or pain
[b] up to 36 months if the use case appears on the EU Product Safety List
[c] NCA = National Competent Authority, NB = Notified Body

Unacceptable Risk = Prohibited

Specific cases are considered to violate fundamental human rights and are thus forbidden applications of AI



Mass surveillance

Untargeted scraping of facial images from internet or CCTV for databases (privacy); ex-post remote biometric identification[1]

Biometric categorization

Profiling using sensitive characteristics (demographics)[2]

Emotion recognition

At the workplace or in schools

Social scoring

Based on behavior or personal characteristics

Behavioral manipulation

To circumvent free will of individuals – particularly from vulnerable groups[3]

Implementation timeframe

6 months upon entry into force

[1] Only exception: law enforcement upon prior judicial authorization for the targeted search of persons convicted or suspected criminal activity
[2] E.g., socio-economic status, gender, ethnicity, citizenship status, philosophical beliefs, religion, political orientation, sexual orientation
[3] Vulnerable individuals = particularly children, elderly, under-educated

High Risk = Conformity

Specific cases are considered to pose threats to safety or fundamental rights, depending on their implementation

1. Products listed under EU safety legislation[1]

2. Annex III

Corresponding to eight specific areas:

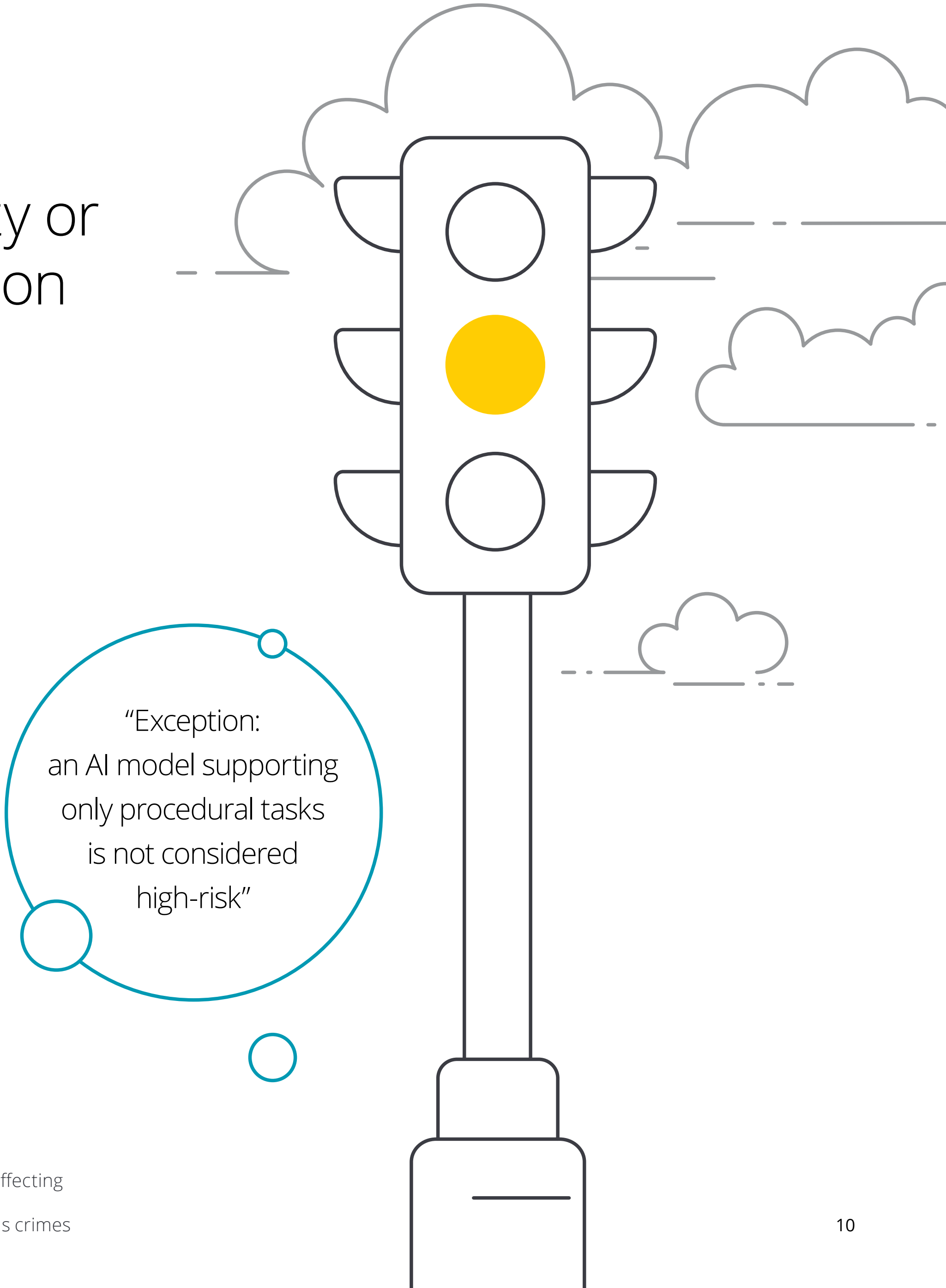
1. Management and operation of critical infrastructure
2. Medical devices
3. Employment, worker management, recruitment
4. Access to essential private & public services/benefits[2]
5. Law enforcement[3]
6. Migration, asylum and border control management
7. Administration of justice[4]
8. Influencing the outcome of elections, the democratic process

3. Real-time remote biometric identification (RBI) under strict conditions and for a limited time and location[5]

Exception: AI models supporting only procedural tasks of otherwise high-risk use cases.

Implementation timeframe

24–36 months upon entry into force, depending on whether on the EU Product Safety list

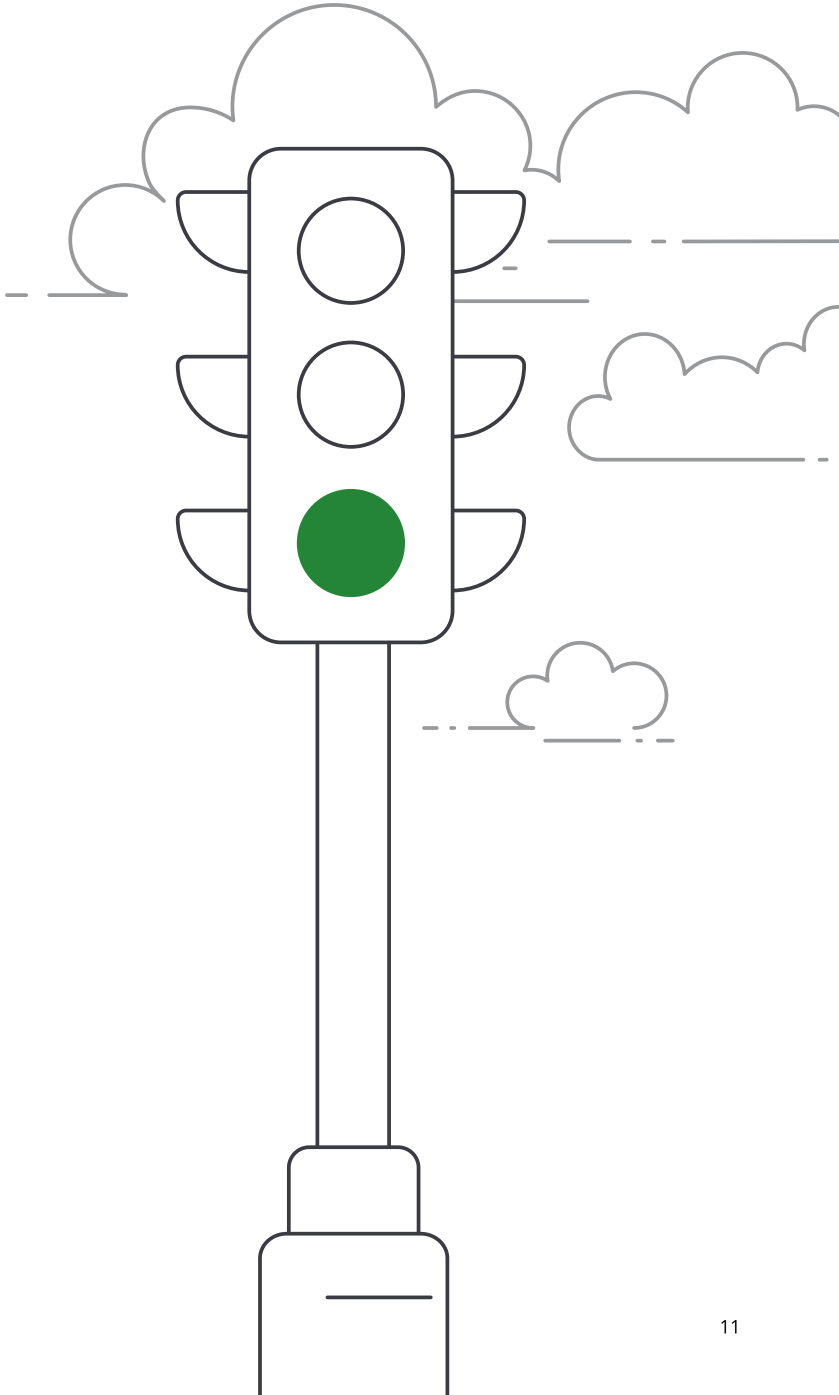


"Exception:
an AI model supporting
only procedural tasks
is not considered
high-risk"

[1] E.g., machinery, toys, aviation, cars, medical devices and lifts.
[2] Such as insurance, credit, housing, utilities, health care, internet access, ... Notable exception: detecting fraud for application to any such service shall not be considered high-risk
[3] Except for administrative proceedings to detect, prevent, prosecute criminal activity
[4] AI may support, but not replace human decision-making for interpretation of law. Exception to AI used for administrative support processes without directly affecting the outcome of justice.
[5] Targeted search of victims, prevention of specific & present terrorism threat, localization or identification of a person convicted or suspected of specific, serious crimes

Limited Risk, Minimal Risk

AI systems which do not negatively impact natural persons, differentiated directly interacting with them or not



Limited Risk – transparency obligations if affecting EU citizens

Subject to transparency obligations, namely informing the user of interaction with an AI. Examples include...

- chatbots
- deep fakes (manipulation of image, audio, video)

Minimal Risk – only voluntary standards iif internal models or limited to procedural tasks

Only subject to voluntary quality standards. Examples include...

- internal rating models
- recommender systems helping internal staff

Obligations

Providers[1] of high-risk AI must demonstrate conformity prior to market placement, maintain throughout the lifecycle

1. Quality Management System

AI Governance

- Data quality and governance
- Transparency
- Robustness, accuracy, cyber security
- Risk management system
- Technical documentation
- Record keeping
- Human oversight

2. Demonstrated Model Conformity

AI Model Validation

- Issuing the Declaration of Conformity to principles and adopted technical standards of Trustworthy AI
- Registering in the EU database
- Affixing the CE marking
- Performing the Fundamental Rights Impact Assessment

[1] Developers or those commissioning development by IT experts on their behalf

The Quality Management System

The cornerstone of quality and risk management of the AI model throughout its lifecycle

Design techniques, control and verification

Development techniques, quality assurance

Examination, test and validation procedures

Technical specifications and applicable standards

Systems and procedures for data management (Article 10)[1]

Risk detection, prevention, mitigation – Risk Management System (Article 9)

Post-market monitoring – logging of serious incidents and malfunction

Communication with authorities (incl. sectoral)

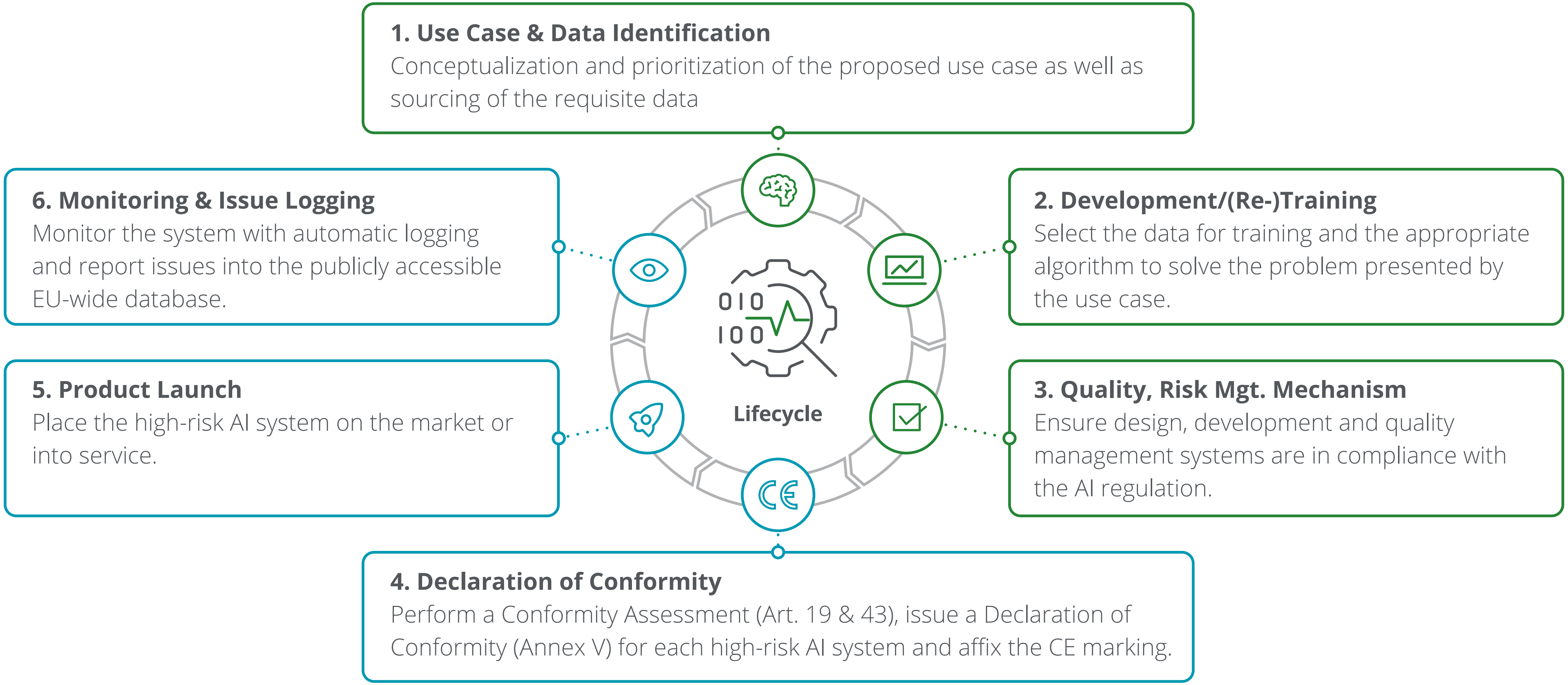
Record keeping, documentation

Resource management and accountability framework

[1] Incl. data acquisition, collection, analysis, labeling, storage, filtration, mining, aggregation, retention

Lifecycle

Conformity to the quality, governance, and documentation standards of the AI Act is a continuous process to be maintained throughout the product lifecycle



1. Use Case & Data Identification
Conceptualization and prioritization of the proposed use case as well as sourcing of the requisite data

2. Development/(Re-)Training
Select the data for training and the appropriate algorithm to solve the problem presented by the use case.

3. Quality, Risk Mgt. Mechanism
Ensure design, development and quality management systems are in compliance with the AI regulation.

4. Declaration of Conformity
Perform a Conformity Assessment (Art. 19 & 43), issue a Declaration of Conformity (Annex V) for each high-risk AI system and affix the CE marking.

5. Product Launch
Place the high-risk AI system on the market or into service.

6. Monitoring & Issue Logging
Monitor the system with automatic logging and report issues into the publicly accessible EU-wide database.

Contact

David Thogmartin
Director
Risk Advisory
Tel: +49 211 8772 2336
dthogmartin@deloitte.de

Torsten Berge
Senior Manager
Business Assurance
Tel: +49 151 58072499
tberge@deloitte.de

Dr. Till Contzen
Partner
Digital Law
Tel: +49 69 71918 8439
tcontzen@deloitte.de

Atrak Yadegari
Director
Strategy, Brand & Reputation
Tel: +49 221 97324 521
ayadegari@deloitte.com

Deloitte.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (DTTL), its global network of member firms, and their related entities (collectively, the "Deloitte organization"). DTTL (also referred to as "Deloitte Global") and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/de/UeberUns to learn more.

Deloitte provides industry-leading audit and assurance, tax and legal, consulting, financial advisory, and risk advisory services to nearly 90% of the Fortune Global 500® and thousands of private companies. Legal advisory services in Germany are provided by Deloitte Legal. Our people deliver measurable and lasting results that help reinforce public trust in capital markets, enable clients to transform and thrive, and lead the way toward a stronger economy, a more equitable society and a sustainable world. Building on its 175-plus year history, Deloitte spans more than 150 countries and territories. Learn how Deloitte's approximately 457,000 people worldwide make an impact that matters at www.deloitte.com/de.

This communication contains general information only, and none of Deloitte GmbH Wirtschaftsprüfungsgesellschaft or Deloitte Touche Tohmatsu Limited (DTTL), its global network of member firms or their related entities (collectively, the "Deloitte organization") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.

Issue 05/2024