The Academic Guide to AI Act Compliance

Marion Ho-Dac, Cécile Pellegrini, Bernard Long

THE ACADEMIC GUIDE TO AI ACT COMPLIANCE

Editors :
Pr. Marion Ho-Dac (University of Artois)
Ass-Pr. Cécile Pellegrini (Lyon Catholic University)
& Dr. Bernard Long (University of Artois)

This project was supported by Lyon Catholic University (UCLy), University of Artois and the French National Research Agency (ANR) under France 2030, MAIA Project ANR-22-EXES-0009

UNIVERSITÉ D'ARTOIS

UCLy
LYON CATHOLIC UNIVERSITY

Acknowledgment:

This project was supported by Lyon Catholic University (Ucly), University of Artois, and the French National Research Agency (ANR) under the France 2030 program, MAIA Project ANR-22-EXES-0009.

Foreword

On 1 August 2024, Regulation [EU] 2024/1689 – known as the AI Act – entered into force in the European Union. The Act represents a first-of-its-kind regulation on artificial intelligence, introducing a swathe of measures aimed at navigating the delicate balance between promoting innovation in this burgeoning technological area and protecting against the profound risks it poses to health, safety and fundamental rights. While building on classical EU safety regulation based on the New Legislative Framework, the AI Act is, in many regards, novel. Amongst its many innovative provisions, the Act outlines a risk-based taxonomic pyramid of AI systems and models, establishes requirements for both providers and downstream deployers, and sets up a networked market surveillance, enforcement and governance scheme. It also employs traditional methods of co- and self-regulation within the novel normative AI ecosystem.

The Act also enjoys wide applicability, regulating the use of AI in areas as diverse as healthcare, education, employment, energy and law enforcement. Geographically, it applies to both European and global AI operators from the moment AI outputs are used in the EU.

Given its singularity and profound influence, the Act has provided a fruitful source of discourse for stakeholders from industry, civil society and academia, while presenting a particularly pertinent challenge for operators [i.e. providers and deployers in particular] in complying with the new regulatory regime it introduces.

Against this backdrop, the Achieving the AI Act Compliance [AAIAC] workshop was co-hosted by Artois University and Lyon Catholic University on 13 & 14 February 2025. The workshop brought together a diverse range of experts, whose presentations provided a comprehensive, compliance-oriented analysis of various important aspects of the AI Act, including its risk-based taxonomy of AI systems and models, the transparency requirements it outlines, and its use of harmonised technical standards as a form of AI co-regulation.

Based on these presentations, those involved are pleased to publish this handbook entitled "The Academic Guide to AI Act Compliance", which is aimed at demystifying the Act's complex network of concepts and requirements and supporting organisations in their compliance journey. In particular, the expert contributions which follow adopt a compliance-focused perspective, which is likely to be of significant interest and utility to concerned industry stakeholders, as well as operators in public administration, academia and civil society.

This handbook serves a multi-faceted purpose:

- First, it is intended to provide a vital reference for those working in compliance and regulatory roles related to the AI Act, providing key guidance on the Act's provisions through a compliance-focused prism.
- Next, the handbook provides a clear and comprehensive analysis of this dense and complex piece of legislation and, as such, ought to be useful to those concerned with policy, including both European and national policymakers and civil society organisations, such as NGOs.
- Lastly, the handbook intends to contribute to academic discourse on both the present state of AI regulation in the EU and its future by dissecting the current regulatory regime outlined by the AI Act.

Table of contents

I - GETTING STARTED. HOW TO ACHIEVE COMPLIANCE - ISSUES & METHODS

Marion Ho-Dac (Univ. Artois) & Cécile Pellegrini (UCLy)

1. The EU AI Act in a Nutshell

The European Union (EU) has developed a new and unique legislative framework for Artificial Intelligence (AI) systems put in place on the internal market[1]. Regulation (EU) 2024/1689, called the "Artificial Intelligence Act", entered into force on 1 August 2024 and will be progressively applicable between 2 February 2025 and 2 August 2027 depending on the provisions concerned[2]. The text is underpinned by a dual rationale. On the one hand, it aims to ensure the free movement of AI-based goods and services while supporting innovation and economic growth in the EU. On the other hand, it seeks to promote trustworthy AI systems, guaranteeing the protection of health, safety and fundamental rights against harmful effects these systems may have on people and on society[3].

The AI Act provides for three main categories of legal provisions: first, a list of prohibited AI practices; second, harmonised rules applicable to marketed AI systems, following a risk-based approach (including provisions on innovation and on general purpose AI models); and third, a comprehensive public enforcement scheme. It consists of more than 100 articles, 180 recitals and 13 annexes. It is thus a massive and complex regulatory framework that both public and private organisations – dealing with AI systems and active on the EU market – will have to master and implement in the upcoming months and years. Therefore, it will be crucial for the AI industry and AI practitioners, including public authorities, to set up an action plan to comply with the AI Act.

2. A Complex and Conflicting Global Landscape on AI Regulation

AI regulation is evolving in a highly complex and sometimes conflicting global landscape. Beyond the EU, we can observe how the AI Act is already shaping global regulatory discussions and industry benchmarks.

On the one hand, there is a growing transnational cooperation to ensure AI governance. A key example is the establishment of AI Safety Institutes in multiple countries, aiming to harmonise risk assessment methodologies and safety standards across borders.

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), *OJ L, 2024/1689, 12 July 2024*

[2] Art. 113, AI Act.

[3] See Art. 1, AI Act.

On the other hand, political developments in major AI-producing nations continue to shape the regulatory agenda. Notably, the withdrawal of the US executive orders on AI by Donald Trump noticeably altered the US stance on AI governance.

Meanwhile, the global AI race is accelerating, with major new players emerging, such as the launch of DeepSeek, the Chinese large-scale AI model.

Against this background, the implementation of the EU AI Act becomes all the more crucial in securing a level playing field for organisations as well as a safety net for citizens in the European market. This raises vital questions regarding organisations' compliance with the AI Act.

3. Compliance in the context of the EU AI Act

In this handbook, compliance refers to the legal obligations imposed on organisations under the AI Act, as a binding regulatory framework that applies within the EU jurisdiction. It also includes all tools, methods, procedures and evaluation schemes to ensure the implementation of those legal obligations by the organisations concerned. In other words, we are talking about legal compliance, including its "organisational" dimension.

Organisations have to comply with a list of requirements laid down in the Act, which raises the following questions:

- Which organisations? And for which AI-based use cases?
- Which requirements?
- What type of compliance process, method or scheme should be used?
- When to comply?

On that last dimension, let us recall that the AI Act's provisions will not apply all at once; rather, there is a gradual implementation timeline. Certain rules, such as prohibitions on specific AI applications and provisions on general-purpose AI models (GPAIM) are already applicable, while other obligations will become enforceable later, such as Annex III on high-risk AI systems, which will enter into force in summer 2026.

Ultimately, these are the key questions we aim to address in this *Guide*. The objective is to provide clarity on what AI Act compliance entails, how organisations should prepare, and what challenges lie ahead.

4. Action Plan for EU AI Act Compliance

To set up an action plan to support AI stakeholders in complying with the AI Act and, more broadly to understand the text, the handbook builds on Article 17 of the AI Act on quality management system (hereafter "QMS"). This provision may be seen as the backbone of the AI Act, even if it only applies to high-risk AI systems; it may also be transposed in a simplified format to other AI models and systems covered by the regulation.

Based on Article 17, AI providers have to "*put a quality management system in place that ensures compliance with this Regulation*". That system shall be documented and include, *inter alia*:

- a strategy for regulatory compliance;
- techniques, procedures and systematic actions for the design and development of AI systems, including their control and verification;
- the implementation of a post-marketing monitoring system; and
- an accountability framework among the organisation and its staff.

These elements perfectly echo more general frameworks on "Management Systems" developed by organisations in the standardisation arena at the international level. Therefore, the handbook builds on the ISO standard on "Compliance Management Systems" (CMS) (ISO 37301:2021). This standard is widely adopted by organisations worldwide, including in the EU, and therefore it can be used as a frame of reference. In addition, the handbook also takes inspiration from the new management systems' standard applied to the AI ecosystem, namely ISO 42001:2023 on "AI Management System", which offers a "*structured way [for organisations] to manage risks and opportunities associated with AI, balancing innovation with governance*".

5. Management System Approach in the AI Act

The management systems' approach is found in several provisions of the AI Act, with two main orientations. On one side, a focus on AI as a product (i.e. system or model) is to be found, *inter alia*, in Article 9 on risk management system, Article 27 on FRIA, Article 43 on conformity assessment and in Article 55. On the other side, other provisions focus on the AI organisation; this is the case of the aforementioned Article 17 on quality management system (QMS).

Therefore, this management systems' approach appears to be a very useful methodology for organisations in analysing the AI Act with a view to achieving legal compliance.

The handbook draws on the structure established by these ISO standards to construct an action plan for AI Act compliance, consisting of four main parts:

- Scope of Compliance Scheme (Chap. 1 to Chap. 4)
- Major High-Risk AI Systems Requirements (Chap. 5 to Chap. 8)
- Compliance Tools and Processes (Chap. 9 to Chap. 12)
- Compliance Evaluation in Practice (Chap. 13 to Chap. 15)

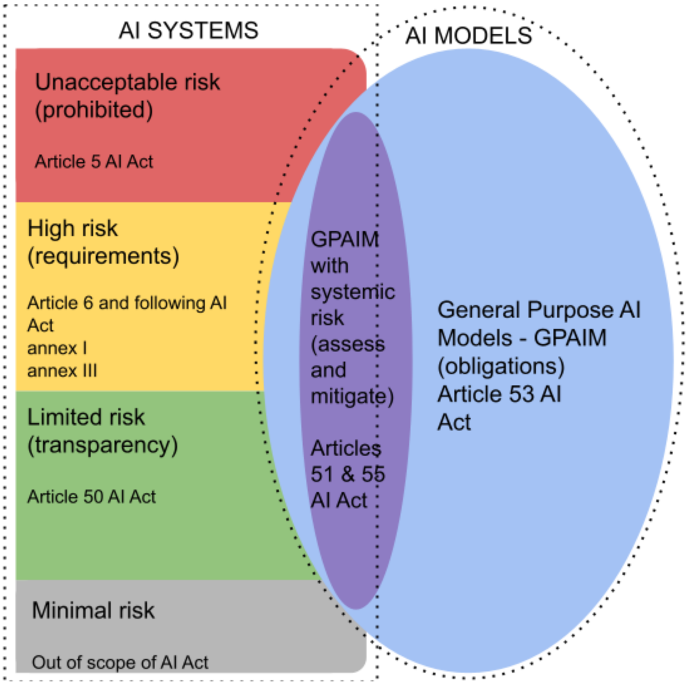
II- SCOPE OF THE COMPLIANCE SCHEME

Chapter 1 - AI Systems & Models Taxonomy Under the AI Act

Juliette Sénéchal (Univ. Lille / Inria)

Introduction

Taxonomy is the process of naming and classifying things, such as animals and plants, into groups within a larger system according to their similarities and differences. Here, I will successively outline the taxonomies relating to AI systems [1] and AI models [2] under the AI Act, specifying in each case the purpose of such taxonomies and the framework of rules associated with each.


AI SYSTEMS
Unacceptable risk (prohibited)
Article 5 AI Act
High risk (requirements)
Article 6 and following AI Act
annex I
annex III
Limited risk (transparency)
Article 50 AI Act
Minimal risk
Out of scope of AI Act
AI MODELS
GPAIM with systemic risk (assess and mitigate)
Articles 51 & 55 AI Act
General Purpose AI Models - GPAIM (obligations)
Article 53 AI Act

1. AI Systems Taxonomy

Pursuant to Article 1(2) of the AI Act, “[the] Regulation lays down: (a) harmonised rules for the placing on the market, the putting into service, and the use of AI systems in the Union”. “AI system” is the main concept regarding the AI Act’s material scope of application. As such, an initial conceptual distinction must be outlined between AI systems within the scope of the Regulation [1.1.] and computer systems outside of its scope [1.2.]. I will finally present the different types of AI system depending on their risk level [1.3].

1.1 *AI Systems under Article 3(1) of the AI Act*

AI system is defined in Article 3 (1) of the AI Act as: “a machine-based system that is 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”.

This definition raises many technical uncertainties, due to the complexity of AI technology, which are partially resolved in Recital 12 of the AI Act (1.1.1) and in the guidelines on AI system’s definition issued by the AI Office (1.1.2)

1.1.1 Recital 12 of the AI Act

1.1.1.1 Negative definition of AI System

This definition, although in line with the definition of the OECD[1], raises many technical questions, particularly about the criteria for distinguishing AI from **simpler, traditional software systems or programming approaches**.

On this first topic, Recital 12 of the AI Act clarifies that “the notion of ‘AI system’ in this Regulation should be clearly defined and should be closely aligned with the work of international organisations working on AI to ensure legal certainty, **facilitate international convergence** and wide acceptance, while providing the flexibility to accommodate the rapid technological developments in this field. Moreover, the definition should be based on key characteristics of AI systems that distinguish it from **simpler traditional software systems or programming approaches and should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations**”.

1.1.2.1 Positive Definition of AI System

In addition to this first negative dimension of the definition, Recital 12 of the AI Act provides the following six additional clarifications:

[1] “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, contents, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment”. See OECD, "Explanatory Memorandum on the Updated OECD Definition of an AI System", OECD Artificial Intelligence Papers, No. 8 (March 2024) Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf

(1) Inference capability - AI systems are defined by their ability to produce outputs (such as predictions, decisions, or content) through inference from inputs or data, by deriving models or algorithms.

Example: A credit scoring AI predicts the likelihood of a customer defaulting on a loan using their financial history.

(2) Underlying technique enabling inference - These systems rely either on machine learning approaches that enable systems to learn from data to achieve certain objectives, or on logic- and knowledge-based methods which draw conclusions from encoded knowledge or symbolic representations.

Example: The credit scoring AI uses gradient boosting (a machine learning technique) to perform the prediction.

(3) Advanced cognitive functions - AI systems transcend basic data processing by incorporating functions such as learning, reasoning, or modelling.

Example: A language model that summarises legal documents mimics understanding and abstraction, cognitive functions typical of humans.

(4) Explicit or implicit objectives - AI systems may operate based on objectives that are either explicitly defined (e.g. goal set by developers), or implicit (emerging from the learning process or data context), which may differ from the intended purpose of the system in a particular use case.

Example: A chatbot is explicitly programmed to provide customer support but also learns (implicitly) to reduce response time through usage data.

(5) Operational autonomy and adaptiveness - AI systems are designed to function with varying degrees of autonomy, meaning they can perform actions without (or with limited) human intervention. Some AI systems are also adaptive, capable of self-learning and evolving after deployment, which introduces new regulatory risks and compliance duties.

Example: A personalised news recommender updates its suggestions based on the user's reading habits without manual reprogramming.

(6) Deployment Contexts (Standalone vs Embedded) - An AI system can operate independently (standalone) or be integrated into a product, whether as an embedded component or as a non-embedded external system but still supporting a product's functionality.

Example: A medical diagnosis AI can be used as standalone desktop software or embedded in a hospital’s imaging device for real-time analysis.

1.1.2The Guidelines of the AI Office

At this stage, the technical uncertainties – as explained above – were such that the AI Office launched a public consultation with a view to clarifying the definition of AI system in the context of the AI Act. This consultation gave birth, on 6 February 2025, to the *Guidelines on the definition of an AI System established by AI Act* (hereinafter "the Guidelines")[2].

It is important to stress that Paragraph 7 of these Guidelines states that "the guidelines are not binding. Any authoritative interpretation of the AI Act may ultimately only be given by the Court of Justice of the European Union (CJEU)". Indeed, the Guidelines are merely an interpretative communication provided by the services of the European Commission. However, even if they are not legally binding in the context of AI system's definition, the Guidelines aim to ensure a consistent interpretation of that key concept across the Union, to guide economic operators, market surveillance authorities, and national courts in the course of implementing the AI Act, and to facilitate convergence with international principles, standards and recommendations, notably those of the OECD, ISO/IEC, and UNESCO.

The Guidelines have reiterated and explained the 7 positive criteria within the definition of AI systems falling within the scope of the AI Act:

1. Machine-based system

Built on hardware and/or software components[3].

Example: the currently most advanced emerging quantum computing systems, which represent a significant departure from traditional computing systems, constitute machine-based systems.

2. Designed with autonomy

Able to operate at varying levels of independence from human control[4]. Paragraph 17 indicates that, "The reference to 'some degree of independence of action' in Recital 12 AI Act excludes systems that are designed to operate solely with full manual human involvement and intervention. Human involvement and human intervention can be either direct, e.g. through manual controls, or indirect, e.g. though automated systems-based controls which allow humans to delegate or supervise system operations".

[2] European Commission, "Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)" (Communication) C(2025) 924 final.

[3] Paragraph (11) of the Guidelines.

[4] Paragraph 14 of the Guidelines refers to the Recital 12 of the AI Act.

3. May exhibit adaptiveness after deployment

Possess self-learning capabilities, enabling behavioural change based on new data or interactions. Here, paragraph 22 of the Guidelines specifies that the concepts of autonomy and adaptiveness represent different dimensions of an AI system's functionality. Recital 12 AI Act clarifies that 'adaptiveness' refers to self-learning capabilities, allowing the behaviour of the system to change while in use. The new behaviour of the adapted system may produce different results from the previous system for the same inputs.

4. Explicit or implicit objectives

Operating according to goals that may be directly programmed or emergent from data and context[5]. The objectives of an AI system are internal to the system, referring to the goals of the tasks to be performed and their results. In contrast, the intended purpose is externally oriented and includes the context in which the system is designed to be deployed and how it must be operated.

Example: "in the case of a corporate virtual assistant, the intended purpose might be to assist a certain department of a company to carry out certain tasks. This might require that the documents that the virtual assistant uses comply with certain requirements (e.g. length, formatting) and that the user questions are limited to the domain in which the system is intended to operate. This intended purpose is fulfilled not only through the system's internal operation to achieve its objectives, but also through other factors, such as the integration of the system into a broader customer services workflow, the data that is used by the system, or instructions for use[6]."

5. Inference capacity

Capable of inferring how to generate outputs (such as predictions or content) from inputs using machine learning or reasoning techniques.

Example: a conversational agent is capable of inferring, based on an input known as a "prompt," an output in the form of a content, which is a probabilistic "answer" to that prompt.

6. Generation of outputs

Produces actionable outputs—predictions, content, recommendations or decisions—that go beyond simple data processing.

[5] Paragraph 14 of the Guidelines.

[6] European Commission, Commission Guidelines on the definition of an AI system pursuant to Regulation (EU) 2024/1689, C(2024) 3349 final, 22 May 2024, §25.

7. Influence on physical or virtual environments

Its outputs are able to affect or interact with real-world or digital contexts.

In particular, the Guidelines have provided an in-depth analysis of the concept of “capacity of inference”.

Inferring how to generate outputs using AI techniques

Concerning the “capacity of inference”, paragraph 31 of the Guidelines states that “the phrase ‘infer how to’, used in Article 3(1) and clarified in Recital 12 of the AI Act, is broader than, and not limited only to, a narrow understanding of the concept of inference as an ability of a system to derive outputs from given inputs, and thus infer the result. Accordingly, the formulation used in Article 3(1) AI Act, i.e. ‘infers, how to generate outputs’, should be understood as referring to the building phase, whereby a system derives outputs through AI techniques enabling inferencing”.

1.1.2.1 First category of AI techniques that enable inference: machine learning approaches

Paragraph 32 of the Guidelines specifies that the first category of AI techniques mentioned in Recital 12 of the AI Act is “machine learning approaches that learn from data how to achieve certain objectives.”

- Large variety of machine learning approaches

The Guidelines state that this category “includes a large variety of approaches enabling a system to ‘learn’, **such as supervised learning, unsupervised learning, self-supervised learning and reinforcement learning”.**

In the case of supervised learning, “the AI system learns from annotations (labelled data), whereby the input data is paired with the correct output. The system uses those annotations to learn a mapping from inputs to outputs and then generalises this to new, unseen data.”

In the case of unsupervised learning, “the AI system learns from data that has not been labelled. The model is trained on data without any predefined labels or outputs. Using different techniques, such as clustering, dimensionality reduction, association rule learning, anomality detection, or generative models, the system is trained to find patterns, structures or relationships in the data without explicit guidance on what the outcome should be.”

According to paragraph 36 of the Guidelines, “**Self-supervised learning is a subcategory of unsupervised learning**, whereby the AI system learns from unlabelled data in a supervised fashion, using the data itself to create its own labels or objectives. AI systems based on self-supervised learning use various techniques, such as auto-encoders, generative adversarial networks, or contrastive learning.”

According to paragraph 37 of the Guidelines, "**AI systems based on reinforcement learning** learn from data collected from their own experience through a 'reward' function. Unlike AI systems that learn from labelled data (supervised learning) or that learn from patterns (unsupervised learning), AI systems based on reinforcement learning learn from experience. The system is not given explicit labels but instead learns by trial and error, refining its strategy based on the feedback it gets from the environment."

- Specificity of deep learning

Paragraph 38 of the Guidelines states, "**Deep learning is a subset of machine learning** that utilises layered architectures (neural networks) for representation learning. AI systems based on deep learning can automatically learn features from raw data, eliminating the need for manual feature engineering. Due to the number of layers and parameters, AI systems based on deep learning typically require large amounts of data to train, but can learn to recognize patterns and make predictions with high accuracy when given sufficient data. AI systems based on deep learning are widely used, and it is a technology behind many recent breakthroughs in AI".

1.1.2.2 Second category of AI techniques that enable inference: logic and knowledge-based approaches

The Guidelines state that in "[i]n addition to various machine learning approaches discussed above, **the second category of techniques mentioned in recital 12 AI Act are 'logic- and knowledge-based approaches** that infer from encoded knowledge or symbolic representation of the task to be solved'. Instead of learning from data, these AI systems learn from knowledge including rules, facts and relationships encoded by human experts. Based on the human experts encoded knowledge, these systems can 'reason' via deductive or inductive engines or using operations such as sorting, searching, matching, chaining. By using logical inference to draw conclusions, such systems apply formal logic, predefined rules or ontologies to new situations."

In particular, "Logic- and knowledge-based approaches include for instance, knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning, expert systems and search and optimisation methods".

1.2 *AI Systems outside of the scope of the AI Act*

Recital 12 of the AI Act explains that the AI system definition should distinguish AI systems from "simpler traditional software systems or programming approaches and should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations." The Guidelines enumerate four categories of systems outside the scope of the AI system definition.

1.2.1 Systems for Improving Mathematical Optimisation

Paragraph 42 of the Guidelines states that "[s]ystems used to improve mathematical optimisation or to accelerate and approximate traditional, well established optimisation methods, such as linear or logistic regression methods, fall outside the scope of the AI system definition. This is because, while those models have the capacity to infer, they do not transcend 'basic data processing'. An indication that a system does not transcend basic data processing could be that it has been used in consolidated manner for many years. This includes, for example, machine learning-based models that approximate functions or parameters in optimization problems while maintaining performance. The systems aim to improve the efficiency of optimisation algorithms used in computational problems. For example, they help to speed up optimisation tasks by providing learned approximations, heuristics, or search strategies."

1.2.2 Basic Data Processing

Paragraph 46 of the Guidelines states that "Basic data processing system refers to a system that follows predefined, explicit instructions or operations. These systems are developed and deployed to execute tasks based on manual inputs or rules, without any 'learning, reasoning or modelling' at any stage of the system lifecycle. They operate based on fixed human-programmed rules, without using AI techniques, such as machine learning or logic-based inference, to generate outputs".

Example of basic data processing: "database management systems used to sort or filter data based on specific criteria (e.g. 'find all customers who purchased a specific product in the last month'), standard spreadsheet software applications which do not incorporate AI enabled functionalities, and software that calculates a population average from a survey that is later exploited in a general context."

1.2.3 Systems Based on Classical Heuristics

Paragraph 48 of the Guidelines states that "Classical heuristics are problem-solving techniques that rely on experience-based methods to find approximate solutions efficiently. Heuristics techniques are commonly used in programming situations where finding an exact solution is impractical due to time or resource constraints. Classical heuristics typically involve rule-based approaches, pattern recognition, or trial-and-error strategies rather than data-driven learning. Unlike modern machine learning systems, which adjust their models based on input-output relationships, classical heuristic systems apply predefined rules or algorithms to derive solutions".

Example of system based on classical heuristics: "a chess program using a minimax algorithm with heuristic evaluation functions can assess board positions without requiring prior learning from data. While effective in many applications, heuristic methods may lack adaptability and generalization compared to AI systems that learn from experience."

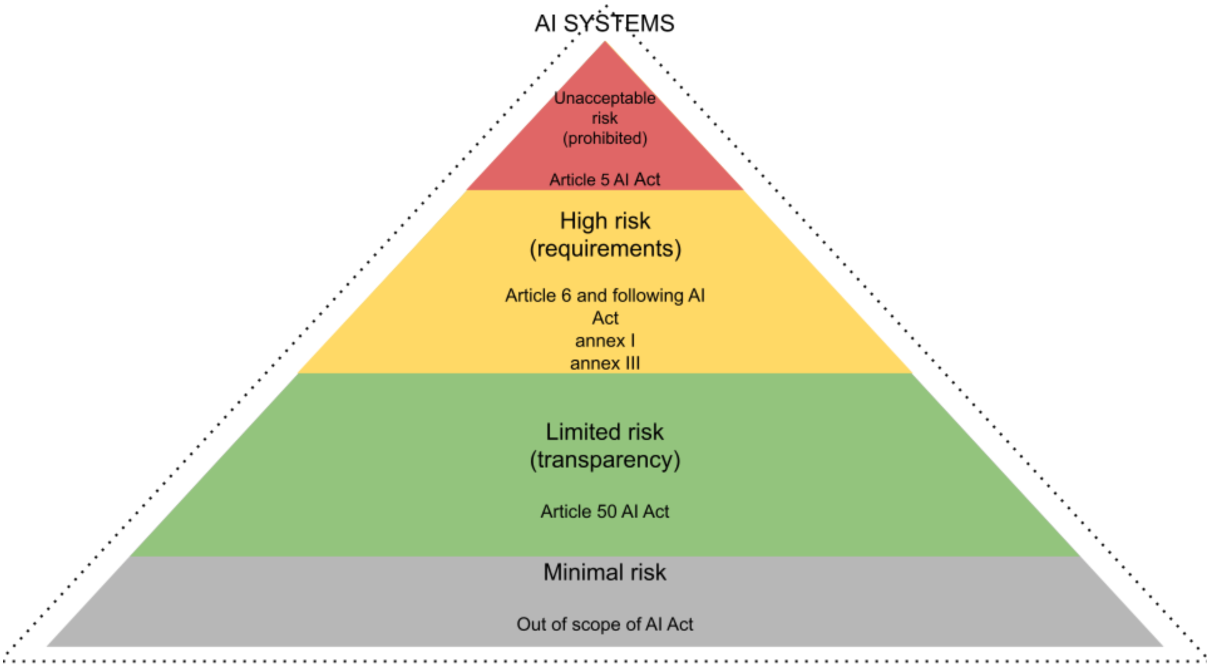
1.2.4 Simple Prediction Systems

Paragraph 49 of the Guidelines states that "All machine-based systems whose performance can be achieved via a basic statistical learning rule, while technically may be classified as relying on machine learning approaches fall outside the scope of the AI system definition, due to its performance".

Example of simple prediction system: "Static estimation systems, such as customer support response time system that are based on static estimation to predict the mean resolution time from the past data and trivial predictors such as demand forecasting for a store to predict how many items of a product the store will sell each day are other examples, that help to establish a baseline or a benchmark, e.g. by predicting average or mean."

1.3 *AI System Taxonomy Based on the Level of Risks: A "Pyramid" of Risks*

The AI Act is often presented as a "risk-based approach regulation", with four levels of risks concerning AI systems: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai



-The unacceptable risk level: article 5 of the AI Act.

All AI systems considered a clear threat to the safety, livelihoods and rights of people are banned. Article 5 prohibits eight practices.

-The high-risk level: article 6 and following of the AI Act

AI use cases that can pose serious risks to health, safety or fundamental rights are classified as high-risk. Annex I and annex III of the AI Act contain the high-risk use-cases.

High-risk AI systems are subject to strict obligations before they can be put on the market.

-The limited risk level: Art 50 of the AI Act

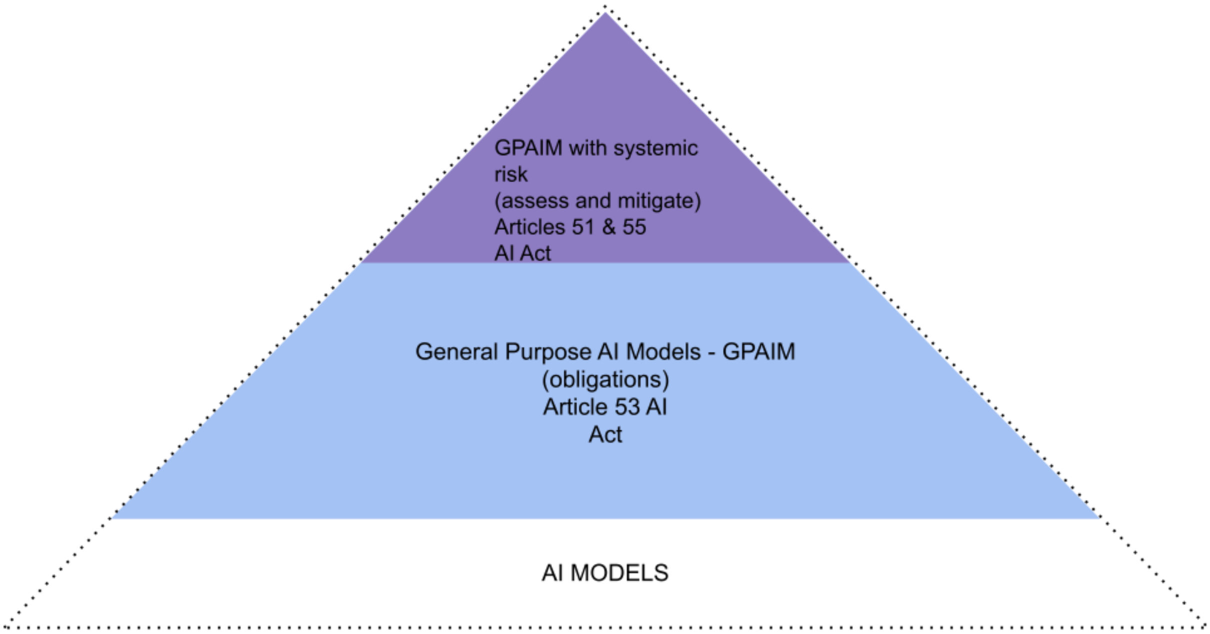
This refers to the risks associated with a need for transparency around the use of AI. Article 50 introduces specific disclosure obligations to ensure the protection of humans.

-The minimal or no risk level: outside the scope of the AI Act.

The AI Act does not introduce rules for AI that is deemed minimal or no risk.

2. AI Models Taxonomy in the AI Act

The AI Act lays down harmonised rules for the placing on the market of general-purpose AI models (hereinafter "GPAI models"). To that end, it includes a taxonomy of AI models, in the form of a conceptual pyramid of models: the most general notion is that of an AI model (2.1). Above this first notion exists a concept of greater precision and specificity: General Purpose Artificial Intelligence model or GPAI model (2.2). At the top of the pyramid, there is a concept of still even greater specificity: GPAI model with systemic risks (2.3).



2.1. AI Model: A Problematic Lack of Definition

The notion of AI model, which is the second core notion of the AI Act, is problematically not defined in the instrument. Recital 97 of the AI Act only states, "*Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems*"[7]. To gain a better understanding of the concept of a model, it is advisable to take into account the definition of model in the "International AI Safety report" of January 2025, "Model: A computer program, often based on machine learning, designed to process inputs and generate outputs. AI models can perform tasks such as prediction, classification, decision-making, or generation, forming the core of AI applications", and its articulation with the definition of system in the same report, "System: An integrated setup that combines one or more AI models with other components, such as user interfaces or content filters, to produce an application that users can interact with".[8]

[7] Y. Bengio (ch.), International AI Safety Report, AI Action Summit, January 2025, https://arxiv.org/abs/2501.17805.

[8] It is also advisable to review the work of the European Data Protection Board. See, for example, European Data Protection Board, "Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models" (17 December 2024). See also Isabel Barbera, "AI Privacy Risks & Mitigations: Large Language Models (LLMs)" (EDPB, April 2025).

2.2. GPAI Model

2.2.1 The definition in the AI Act

Article 3(63) of the AI Act gives the following definition: "'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing 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, except AI models that are used for research, development or prototyping activities before they are placed on the market".

Article 3(64) specifies that "'high-impact capabilities' means capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models".

In order to clarify the meaning of these articles, Recital 97 of the AI Act specifies that:

- "The notion of general-purpose AI models should be clearly defined and set apart from the notion of AI systems to enable legal certainty"
- "The definition should be based on the key functional characteristics of a general-purpose AI model, in particular the generality and the capability to competently perform a wide range of distinct tasks"
- "These models are typically trained on large amounts of data, through various methods, such as self-supervised, unsupervised or reinforcement learning. General-purpose AI models may be placed on the market in various ways, including through libraries, application programming interfaces (APIs), as direct download, or as physical copy".

Article 3(66) of the AI act explains the link between general-purpose AI models and systems: "general-purpose AI system' means an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems".

For instance, the AI conversational agent ChatGPT is a general-purpose AI system, while GPT-4 is a general-purpose AI model embedded in this system. More precisely, GPT-4 is capable of competently performing a wide range of distinct tasks within the meaning of article 3(63) of the AI Act; GPT-4 can generate text, translate languages, write code, summarise content, and answer questions across diverse domains.

Article 53 of the AI Act imposes a general set of obligations on all GPAI providers (including documentation and compliance with Union copyright law).

2.2.2. The definition in the Guidelines of the AI Office: the AI model understood by the AI Office through technical thresholds

On 18 July 2025, the AI Office published guidelines[9] in order to clarify the meaning and scope of the Articles of chapter V of AI Act, applicable from 2 August 2025 to providers of GPAI models, but also of the code of practice provided for in Article 56 of the AI Act and published by the AI Office on 10 July 2025, to which those providers may adhere in order to comply with chapter V of the AI Act dedicated to GPAI models[10].

To help in clarifying the content of the articles of this chapter and of the code of practice, the European Commission, following on from its previous guidelines, intends to address the specific issue of the definition of a GPAI model and focuses primarily on technical thresholds in order to answer to this question.

While acknowledging the technical vagueness of the answer, paragraph 17 of the Guidelines of the AI Office specifies that the requirement of "significant generality" and the ability to "competently perform a wide range of distinct tasks" provided for in the legal definition of a GPAI model in Article 3(63) may be fulfilled when the cumulative amount of computation used for its training, measured in floating point operations (FLOP), exceeds 10^{23}. "An indicative criterion for a model to be considered a general-purpose AI model is that its training compute is greater than 10^{23} FLOP and it can generate language (whether in the form of text2 or audio3), text-to-image or text-to-video. This threshold corresponds to the approximate amount of compute typically used to train a model with one billion parameters on a large amount of data", within the meaning of Recital 98 of the AI Act.

This threshold in only an indicative one. Paragraph 20 states: "If a general-purpose AI model meets the criterion from paragraph 17 but, exceptionally, does not display significant generality or is not capable of competently performing a wide range of distinct tasks, it is not a general-purpose AI model. Similarly, if a general-purpose AI model does not meet that criterion but, exceptionally, displays significant generality and is capable of competently performing a wide range of distinct tasks, it is a general-purpose AI model."

[9] European Commission (AI Office), Guidelines on the scope of obligations for providers of general-purpose AI models under the AI Act. (18 July 2025), Available at:
https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act

[10] European Commission (AI Office), The General-Purpose AI Code of Practice (10 July 2025), Available at: : https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

Example of a model within scope - “A model is trained this on a broad range of natural language data (i.e., text) curated and scraped from the internet and other sources (as is currently typical for language models) using 10^{24} FLOP. The criterion from paragraph 17 indicates that the model should be a general-purpose AI model because it can generate text and its training compute is greater than 10^{23} FLOP. Training on a broad range of natural language further indicates that the model should display significant generality and should be capable of competently performing a wide range of distinct tasks. Therefore, the model likely is a general-purpose AI model.”

This threshold of 10^{23} FLOP is in addition to the threshold of 10^{25} FLOP, legally enshrined in Article 51 of the AI Act for GPAI models posing systemic risks.

The FLOP is a technical criterion, only used in the AI Act for the legal characterisation of GPAIM and GPAIM with systemic risks (but not for AI systems), and is an expression of tech-oriented law, the definition of which can be found in Article 3(67) of the AI Act: “‘*floating-point operation’ means any mathematical operation or assignment involving floating-point numbers, which are a subset of the real numbers typically represented on computers by an integer of fixed precision scaled by an integer exponent of a fixed base*.”

Paragraph 16 of the Guidelines also gives some clarifications regarding the notion of “training compute”: “Training compute has the advantage of combining number of parameters and number of training examples into a single number that is reasonably straightforward for providers to estimate. This number is typically proportional to the number obtained by multiplying these two numbers, allowing a single threshold to be set rather than separate thresholds for model size and training data size. While training compute is an imperfect proxy for generality and capabilities, the Commission considers setting an indicative criterion which includes a training compute threshold to be the most suitable approach at present. Nevertheless, the Commission’s approach may change in the future as technology and the market evolve.”

Finally, in response to the question “what is the estimated computing resources required for training a model?”, paragraph 124 of the Guidelines states that the amount of computation used to train or modify a model can be estimated in two ways: by tracking the use of the graphics processing unit (GPU) (hardware-based approach) or by counting the expected number of floating point operations per second (FLOP) based on the model architecture (architecture-based approach) and accompanying this assertion with several formulas intended for AI model providers.

2.3. GPAI Model with Systemic Risks

Article 51, paragraph 1 of the AI Act states: “A general-purpose AI model shall be classified as a general-purpose AI model with systemic risk if it meets any of the following conditions:

(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 from the scientific panel, it has capabilities or an impact equivalent to those set out in point (a) having regard to the criteria set out in annex XIII.”

Article 51, paragraph 2 of the AI Act creates an additional presumption: “A general-purpose AI model shall be presumed to have high impact capabilities pursuant to paragraph 1, point (a), when the cumulative amount of computation used for its training measured in floating point operations is greater than 10^{25}.”

Article 3(65) of the AI Act specifies what a systemic risk is: “‘systemic risk’ means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain”. Article 3(67) specifies what a floating-point operation is: “‘floating-point operation’ means any mathematical operation or assignment involving floating-point numbers, which are a subset of the real numbers typically represented on computers by an integer of fixed precision scaled by an integer exponent of a fixed base”.

Annex XIII of the AI Act specifies the criteria for the designation of general-purpose AI models with systemic risk referred to in Article 51: “For the purpose of determining that a general-purpose AI model has capabilities or an impact equivalent to those set out in Article 51(1), point (a), the Commission shall consider the following criteria:

(a) the number of parameters of the model;

(b) the quality or size of the data set, for example measured through tokens;

(c) the amount of computation used for training the model, measured in floating point operations or indicated by a combination of other variables such as estimated cost of training, estimated time required for the training, or estimated energy consumption for the training;

(d) the input and output modalities of the model, such as text to text (large language models), text to image, multi-modality, and the state-of-the-art thresholds for determining high-impact capabilities for each modality, and the specific type of inputs and outputs (e.g. biological sequences);

(e) the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to;

(f) whether it has a high impact on the internal market due to its reach, which shall be presumed when it has been made available to at least 10 000 registered business users established in the Union;

(g) the number of registered end-users."

Article 55 introduces additional duties for providers of GPAI models with systemic risk. These include robust risk identification, monitoring, and mitigation throughout the model's lifecycle. In this regard, Article 56 provides for the elaboration of a voluntary code of practice, intended to assist GPAI providers in meeting these obligations. The final version of this code, initially due by 2 May 2025, was published on 10 July 2025[11]. This code clarifies operational guidance on transparency, traceability, copyright compliance.

For models that may pose a systemic risk, the code of practice aims to describe how providers of such models can ensure compliance with the obligations relating to the assessment and mitigation of systemic risks throughout the model's lifecycle, in accordance with Article 55 of the AI Act. In this regard, the code of practice provides a taxonomy of systemic risks in Appendix 1 of the Safety and Security Chapter of the code of practice.

Conclusion

The material scope of the AI Act - what is an AI system that falls within the scope of the AI Act? What is an AI model that falls within the scope of the AI Act? - is not an easy question to answer, as it depends on numerous criteria - some conceptual and some quantitative or technical - which are subject to numerous interpretations. Operators that may be subject to the AI Act would be well advised, in case of doubt, to consult with AI Act authorities and bodies in charge of its implementation, to obtain guidance on the classification of the systems and models they develop, supply, or deploy, in addition to using the numerous guidelines published by the AI Office on this subject.

[11] European Commission (AI Office), The General-Purpose AI Code of Practice (10 July 2025), Available at: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

Bibliography

- I. Barbera, "AI Privacy Risks & Mitigations: Large Language Models (LLMs)" (EDPB, April 2025)

- Y. Bengio (ch.), International AI Safety Report, AI Action Summit, January 2025, https://arxiv.org/abs/2501.17805

- F. G'sell, Regulating under Uncertainty: Governance Options for Generative AI (October 06, 2024). Available at SSRN : https://ssrn.com/abstract=4918704 or http://dx.doi.org/10.2139/ssrn.4918704

- European Commission (AI Office), Guidelines on the scope of obligations for providers of general-purpose AI models under the AI Act, (18 July 2025) Available at : https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act

- European Commission (AI Office), The General-Purpose AI Code of Practice (10 July 2025), Available at: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

- European Commission (AI Office), Guidelines on AI system definition (6 February 2025), Available at : https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application

- European Data Protection Board, "Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models" (17 December 2024)

- OECD, "Explanatory Memorandum on the Updated OECD Definition of an AI System", OECD Artificial Intelligence Papers, No. 8 (March 2024) Available at : https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf

Chapter 2 - Regulating General-Purpose AI Models: A Dual Disruption

Arnaud Latil (Sorbonne University)

Introduction

The EU AI Act represents a landmark effort to address the opportunities and risks posed by AI models, particularly those with broad capabilities like ChatGPT, Bard, Claude and other transformative AI models developed by companies like OpenAI, Meta or Mistral. However, the regulation of general-purpose AI models (GPAI models) is one of the most critical legal and ethical issues.

Indeed, GPAI models raise significant issues for the economy and the whole of society. They constitute a significant asset in the value chain of AI, particularly in fine-tuning,[1] as they provide a powerful general foundation that can be efficiently adapted to perform specialised tasks through additional training. They are also important in terms of security as most AI systems need these general-purpose models to operate. As an essential building block for numerous AI applications, both for professionals and consumers, these models are under intense scrutiny.

The provisions concerned are provided by Chapter V of the AI Act, under Articles 51 to 55, and Annexes XI, XII and XIII. A major distinction is made by these provisions between, on the one hand, GPAI models that raise systemic risks, and, on the other hand, other GPAI models. GPAI models with systemic risks are considered as more dangerous, justifying more stringent obligations.

The circumstances of the introduction of AI model-related provisions within the AI Act are well known; the release of OpenAI's ChatGPT in November 2022, right in the middle of the negotiations on the Regulation, profoundly changed the approach to this issue. The 2021 AI Regulation proposal designed a single framework for classifying AI systems: prohibited systems, high-risk systems, medium-risk systems, and others. This categorisation has become known as the AI risk pyramid. However, the introduction of GPAI models disrupted this single framework.

[1] Fine-tuning refers to the process of adapting a GPAI model to perform a specific task or operate within a particular domain by training it further on a more targeted dataset.

In a nutshell, the provisions concerning AI models create two upheavals with the original 2021 AI Regulation proposal and, in a broader sense, with the text's overall logic. The first concerns the general taxonomy of different forms of AI covered by the text [1]. The second disruption, even more profound, concerns the way the risks associated with them are considered. Indeed, the introduction of GPAI-related provisions leads to a new hierarchy of risks within the AI Act [2].

Note - The European Commission is publishing guidelines to clarify key concepts underlying the AI Act 's provisions on GPAI models.[2] These guidelines aim to complement the GPAI Code of Practice[3] which sets out commitments to which GPAI models' providers may adhere to ensure compliance with their obligations under the AI Act.

GPAI Guidelines are expected to cover clarifications *inter alia* on the concept of GPAI model, the characterisation of GPAI model provider, including when a downstream modifier is a provider, the open-source exemptions and the calculation of computational resources used to train or modify a model.

1. A Disrupted Taxonomy

The first disruption arises from the introduction of the notion of "model", as distinct from "system" [1.1]. This distinction marks the return of technical considerations among the definitional elements of AI [1.2].

1.1 *Models vs. Systems*

The definition of an AI system[4] according to Article 3(1) of the AI Act is as follows:

"AI system' means a machine-based system that is 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".

[2] See European Commission, Guidelines on the application of the AI Act to providers of general-purpose AI models, published 25 July 2025, available at: https://digital-strategy.ec.europa.eu/en/policies/guidelines-gpai-providers.

[3] As per Article 56 of the AI Act. The Code of Practice is avalilable here: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

[4] On AI system definition, see in this Guide, *supra*, J. Sénéchal p. 14.

Conversely, the definition of a "model" pursuant to Article 3(63) is less precise:

"'General-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing 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, except AI models that are used for research, development, or prototyping activities before they are placed on the market".[5]

This latter definition is largely tautological, as a "*general-purpose model*" is defined by its generality. The definition adds that it includes cases "*where such an AI model is trained with a large amount of data using self-supervision at scale*" and "*is capable of competently performing a wide range of distinct tasks*", which does not improve the definition much.[6]

This legal definition is now complemented by GPAI Guidelines. They aim to clarify two main elements: first, the conditions for sufficient generality and capabilities and, second, criteria for the differentiation between distinct models and model versions.[7]

An exception for "*research, development, or prototyping activities before they are placed on the market*" appears to be important. This exception is nonetheless very strict because GPAI for research, development or prototyping that are "*placed on the market*" are regulated under the AI Act. GPAI Guidelines from the European Commission will provide clarification on the scope of this exemption.

The French Data Protection Authority (CNIL), which partially oversees AI regulation in France, provides the following definition of a "model":

"AI model is a mathematical construction generating a deduction or prediction from input data. The model is estimated from annotated data during the learning (or training) phase of the AI system".[8]

This definition highlights two key aspects. First, models are "*mathematical constructions*". Different types of models exist: linear regression, deep neural networks, decision trees, etc. The second element, even more intriguing, is that a model is "*estimated.*" In consequence, a model remains a "*mathematical estimation*". For example, if I prompt

[5] See also Recitals 97 to 99 of the AI Act.

[6] On GPAI model definition, see in this Guide, *supra*, J.Sénéchal, p. 25.

[7] European Commission, Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act), 18 July 2025, Sections 2.1–2.2, paras. 13–24.

[8] CNIL. (s.d). Glossary of artificial intelligence (AI): Model (IA). "The AI model is the mathematical construction ... training phase of the AI system." Retrieved from CNIL website.

the following sentence: "*The sky is...*," the machine will likely respond "*blue*" because it is the most probable answer. The machine will "*estimate*" that the next word is "*blue*" based on its learning.

Note - Regulating models implies focusing on technological parameters rather than use cases, marking a fundamental distinction between "systems" and "models." AI systems are regulated based on their concrete applications, whereas models are defined by their technical attributes.

1.2 *The Return of Technical Considerations*

This distinction in defining AI system and GPAI models reintroduces technical criteria into AI regulation.

The 2021 AI Regulation proposal initially included technical elements in Annex I, but these were later removed in favor of technological neutrality.

However, the AI Act now differentiates between two types of GPAI models — those with systemic risks and those without — using technical benchmarks. Article 51(2) of the AI Act establishes a classification presumption based on computing power, measured in floating point operations per second (FLOPs). Technological considerations also lie in para. 1 of Article 51, which outlines that GPAI with systemic risks should be evaluated by the Commission "*on the basis of appropriate technical tools and methodologies, including indicators and benchmarks*". Additionally, Annex XIII lists' criteria such as the number of parameters, dataset quality and size, and computational requirements. Finally, other considerations – not technical in nature – are also taken into account, such as the number of registered users.

While this technical approach is certainly necessary given the very nature of GPAI models, i.e. their generality, it carries the risk of obsolescence, as technological advancements may soon render benchmarks such as FLOPs inadequate.

The GPAI Guidelines drafted by the European Commission aim to provide methods for estimating these technical capabilities for the purpose of legal characterisation of AI models under the AI Act, while evolving in line with technological needs.

2. **A New Hierarchy of Risks**

The AI Act initially established a risk-based classification of AI systems (prohibited, high-risk, medium-risk, and unregulated). However, GPAI-related provisions introduce a new risk hierarchy by distinguishing between models that pose systemic risks and those that do not (2.1). This distinction also has geopolitical implications (2.2).

2.1. Defining Systemic Risks in AI

The notion of "*systemic risks*" has its origins in banking and financial law. More specifically, the 2008 economic crisis raised the need to address failures that can affect the whole market. In particular, the demise of the so-called "*too big to fail*" doctrine highlights that some risks are too important to be left to the unregulated market. In the US, the 2010 Dodd-Frank Act now imposes capital requirements to mitigate systemic financial risks. International regulatory framework for banks "Basel III" has extended these requirements. Following this, the concept of systemic risks has been adapted to other sectors, such as AI.

Under the AI Act, Articles 53 to 55 distinguish between two kinds of GPAI models based on whether or not they raise systemic risks. Article 51(1) provides that:

"a general-purpose AI model shall be classified as a general-purpose AI model with systemic risk if it meets any of the following conditions:

(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 from the scientific panel, it has capabilities or an impact equivalent to those set out in point (a) having regard to the criteria set out in Annex XIII".

Article 51(2) states that "a general-purpose AI model shall be presumed to have high impact capabilities pursuant to paragraph 1, point (a), when the cumulative amount of computation used for its training measured in floating point operations [FLOPs] is greater than 10[^25]". Additionally, Article 3(64) of the AI Act defines the concept of 'high-impact capabilities' as "capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models".

In short, based in the AI Act, a GPAI model is classified as having systemic risk if it demonstrates high-impact capabilities, assessed through appropriate technical benchmarks and tools. Any model trained using over 10[^25] FLOPs is presumed to have high-impact capabilities. The European Commission may also designate a model as presenting systemic risk if it has equivalent high-impact capabilities, based on factors such as model size, training data and computation, input-output modalities, performance benchmarks, adaptability and autonomy, as well as the model's market reach and scale of deployment within the Union.

Article 3(65) of the AI Act defines "systemic risk" as:

“risk specific to high-impact capabilities of general-purpose AI models, having a significant impact on the EU market due to their reach, or due to actual or foreseeable negative effects on public health safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.”

Annex XIII indirectly provides elements for determining systemic risks based on the criteria for characterising GPAI models with systemic risk, but it does not specify the risks themselves. These criteria may be interpreted as operationalising the concept of systemic risk by identifying technical and market indicators, such as model size, training intensity, multi-modal functionalities, and user reach, that signal a model’s disruptive potential. Together, they help determine whether a model’s influence is sufficiently extensive and autonomous to justify its classification as posing systemic risk.

Comparative Perspectives - The **Digital Services Act** (DSA)[9] explicitly defines systemic risks, setting a threshold at 45 million users and categorising risks into four types:

- Dissemination of illegal content
- Negative impact on fundamental rights
- Threats to democratic processes and public security and
- Gender-based violence, public health and minors’ protection, and serious harm to physical or mental well-being[10].

A broader perspective is found in the *International AI Safety Report* (2025)[11], which categorises systemic risks as labor market disruptions, global AI R&D imbalances, market concentration, environmental concerns, privacy issues, and copyright infringements.

The **GPAI Code of Practice** released in June 2025 provides for essential complementary information on the definition and characterisation of systemic risks in GPAI context. Its third Chapter on Safety and Security “*outlines concrete state-of-the-art practices for managing systemic risks, i.e. risks from the most advanced models. Providers can rely on this chapter to comply with the AI Act obligations for providers of general-purpose*

[9] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act), OJ L 277, 27.10.2022, pp. 1–102, Art. 34(1)– (2).

[10] Art. 34 (1) a to d) of the DSA.

[11] International AI Safety Report 2025, chaired by Y. Bengio and supported by an Expert Advisory Panel representing 30 countries including the UN, OECD, and EU, published 29 January 2025, Department for Science, Innovation and Technology (DSIT) research paper number 2025/001.

AI models with systemic risk".[12] In Chapter 3, Appendix 1.1., the Code lays down a list of five main types of risk for the purpose of identifying systemic risks:

(1) Risks to public health

(2) Risks to safety

(3) Risks to public security

(4) Risks to fundamental rights

(5) Risks to society as a whole

Although distinct, these risk types may overlap in some cases.
Based on these types of risks, a list of "*specified systemic risks*" is provided in Appendix 1.4:

"(1) **Chemical, biological, radiological and nuclear:** Risks from enabling chemical, biological, radiological, and nuclear (CBRN) attacks or accidents. This includes significantly lowering the barriers to entry for malicious actors, or significantly increasing the potential impact achieved, in the design, development, acquisition, release, distribution, and use of related weapons or materials.

(2) **Loss of control:** Risks from humans losing the ability to reliably direct, modify, or shut down a model. Such risks may emerge from misalignment with human intent or values, self-reasoning, self-replication, self-improvement, deception, resistance to goal modification, power-seeking behaviour, or autonomously creating or improving AI models or AI systems.

(3) **Cyber offence:** Risks from enabling large-scale sophisticated cyber-attacks, including on critical systems (e.g. critical infrastructure). This includes significantly lowering the barriers to entry for malicious actors, or significantly increasing the potential impact achieved in offensive cyber operations, e.g. through automated vulnerability discovery, exploit generation, operational use, and attack scaling.

(4) **Harmful manipulation:** Risks from enabling the strategic distortion of human behaviour or beliefs by targeting large populations or high-stakes decision-makers through persuasion, deception, or personalised targeting. This includes significantly enhancing capabilities for persuasion, deception, and personalised targeting, particularly through multi-turn interactions and where individuals are unaware of or cannot reasonably detect such influence. Such capabilities could undermine democratic

[12] As explained on the GPAI Code of Practice dedicated webpage of the European Commission: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

processes and fundamental rights, including exploitation based on protected characteristics."

Additionally, Appendix 1.2 provides for considerations concerning the nature of systemic risks (i.e. essential characteristics of the nature of systemic risks and contributing characteristics) in order to inform systemic risk identification, and Appendix 1.3. details potential (and non-exhaustive) systemic risk sources based on four categories: model capabilities, model propensities, model affordances, and contextual factors.

2.2. The Geopolitical Challenges of Systemic Risks

The regulation of GPAI models underscores geopolitical considerations. Indeed, the European Commission exercises broad authority over GPAI,[13] drawing directly from its experience with GDPR enforcement. This centralisation prevents Member States from adopting overly lenient approaches toward major tech companies. Provisions on GPAI must be seen as a will to undertake systemic risks on an EU level. This position is justified by the very nature of systemic risks that cannot be addressed on a state level.

Furthermore, regulating GPAI sends a strong signal globally, i.e. that the deployment of these models in Europe must align with the region's social and economic frameworks. As the most important category of risks, it is essential to build a strong and consistent legal framework in this regard, to avoid fragmented approaches among states challenging European unity. In that respect, the drafting of the GPAI Code of Practice pursuant to Article 56 of the AI Act and its future implementation by the AI industry are strong signals of the EU's will to to establish a regulatory framework that both protects its core values based on fundemantal rights and respects innovation and underlying economic interests.

Conclusion

As the AI Act does not provide for a definitive list of systemic risks, compliance with systemic risk provisions requires a broad approach. Democracy, rule of law, environment, safety, fundamental rights, and systemic risks are extremely concrete, but, at the same time, are far abstracted from individuals' actions. Moreover, systemic risks may lie in a large range of activities and, thus, mitigating them will be one of the key challenges for compliance with the AI Act. In this context, the complementary normative frameworks provided for by the European Commission (such as the GPAI Guidelines)

[13] See in particular Article 88 of the AI Act on "Enforcement of the Obligations of Providers of General-Purpose AI Models", stating that the European Commission has the sole authority to oversee and enforce rules related to GPAI models.

and the multistakeholder approach (such as the GPAI Code of Practice) following the AI Act's dedicated provisions should play a crucial role.

Bibliographic references

- Kutterer Cornelia, “Regulating Foundation Models in the AI Act : From “High” to “Systemic” Risk, AI-Regulation Papers, Chair Legal And Regulatory Implications of Artificial Intelligence, 12.01.2024.
- Ebers Martin, « Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act », European Journal of Risk Regulation. Published online 2024 : 1-20. Doi
- European Commission, “Commission Guidelines on prohibited artificial practices”, 6 February 2025.
- European Commission, Third Draft of the General-Purpose AI Code of Practice, Commitments by providers of general-purpose ai models with systemic (https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts), march 2025.
- Keller, A., Martins Pereira, C., & Lucas Pires, M. (2023). The European Union's Approach to Artificial Intelligence and the Challenge of Systemic Risk. In H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. Martins Pereira, E. Vaz Sequeira, & L. Barreto Xavier (Eds.), Multidisciplinary Perspectives on Artificial Intelligence and the Law (415-439). Springer Verlag. https://doi.org/10.1007/978-3-031-41264-6_22
- Palmieri, Alice and Kollnig, Konrad and Tamò-Larrieux, Aurelia, Systemic Risks of Dominant Online Platforms : A Scoping Review (October 23, 2024). Available at SSRN : https://ssrn.com/abstract=5002743 or http://dx.doi.org/10.2139/ssrn.5002743
- Rangone N, Megale L., Risks Without Rights ? The EU AI Act's Approach to AI in Law and Rule-Making. European Journal of Risk Regulation. Published online 2025 :1-16. Doi :10.1017/err.2025.13.
- Risto Uuk and Carlos Ignacio Gutierrez and Daniel Guppy and Lode Lauwaert and Atoosa Kasirzadeh and Lucia Velasco and Peter Slattery and Carina Prunkl, « A Taxonomy of Systemic Risks from General-Purpose AI », 22 nov. 2024.
- Viral V Acharya and Matthew Richardson, The Dodd-Frank Act, systemic risk and capital requirements, *in* Thorsten Beck (dir.), *The Future of Banking*, Center for Economic Policy Research, 2011.

Chapter 3 - AI Operators under the AI Act

Marco Pasqua (Catholic Univ. Sacred Heart of Milan)

Introduction

The AI Act[1] defines AI operators in Article 3(8) as encompassing "provider, product manufacturer, deployer, authorised representative, importer or distributor". Each of these is also defined in detail in Article 3 (3) to (7) and, before that, refered to in Article 2(1). This latter provision does not delineate the personal scope of the Act in isolation; rather, it interconnects it with the territorial scope of application in various ways. Consequently, both aspects must be considered together. To whom does the AI Act apply and what is its geographical scope?

From a personal scope perspective, the AI Act aims to regulate the entire AI value chain, encompassing providers, deployers, manufacturers, importers, distributors, representatives and other relevant actors dealing with an AI system or model. Each of these operators plays a distinct role within the regulatory framework, with specific responsibilities that warrant closer examination.

Regarding the territorial scope, the AI Act does not limit its application to operators established or located in EU Member States. Instead, it adopts a broad extraterritorial approach, extending its reach globally. This has significant implications for compliance, requiring careful attention to ensure adherence to its provisions beyond the EU's borders.

Each AI operator is examined individually in the following sections, starting with the two main actors (i.e. providers and deployers) and, then, the secondary players[2].

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) PE/24/2024/REV/1 OJ L, 2024/1689, 12.7.2024 (hereinafter 'AI Act').

[2] *Cf.* Section 3 — The actors in the product supply chain — under Commission notice the 'Blue Guide' on the implementation of EU product rules 2022 (Text with EEA relevance) 2022/C 247/01 C/2022/3637 OJ C 247, 29.6.2022, p. 1–152.

1. Main AI Operators Of The AI Value Chain: AI Providers And Deployers

1.1. Providers Established or Located Within the EU or in a Third Country

Article 2[1][a] of the AI Act establishes that the Regulation applies to providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the EU, irrespective of whether those providers are established or located within the EU or in a third country.[3]

This extraterritorial reach is designed to ensure a level playing field among AI industry actors at the global scale, and to guarantee effective protection of individuals' rights and freedoms across the EU. The principle of non-discrimination between domestic and foreign providers is explicitly stated in Recital 21, which underscores that the rules of the AI Act should apply equally to all providers, irrespective of their place of establishment.[4] By subjecting both EU-based and foreign providers to its provisions, the AI Act aims to prevent regulatory arbitrage, ensuring that non-EU providers do not gain an undue competitive advantage by operating under less stringent regulations.

The AI Act defines a provider in Article 3[3] as "[any] 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 it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge"[5].

The distinction between "placing on the market" and "putting into service" is particularly relevant in understanding the obligations of providers:

- placing on the market refers to the making available of an AI system or a general-purpose AI model on the EU market,[6] typically for distribution or use in the course of a commercial activity.[7]

- putting into service pertains to the supply of an AI system for first use directly to the deployer or for own use in the EU for its intended purpose.[8]

It is important to highlight that whether an AI system is offered for payment or free of charge does not affect its classification under the AI Act.[9] However, research,

[3] Article 2[1][a] of the AI Act.
[4] Recital [21] of the AI Act.
[5] Article 3[3] of the AI Act.
[6] Article 3[9] of the AI Act.
[7] Article 3[10] of the AI Act.
[8] Article 3[11] of the AI Act.
[9] Article 3[3] of the AI Act.

development and testing conducted before an AI system or a general-purpose AI model is placed on the market or put into service in the EU generally falls outside the scope of the AI Act.[10]

One of the unclear aspects of the AI Act is the definition of when a provider has actually 'developed' an AI system. The Act does not provide a precise threshold for when development is considered complete. Therefore, providers should follow clear development steps, such as functional availability of the system, documented testing, and version labelling, as objective indicators of completion. Additionally, the Act only minimally addresses scenarios involving multiple providers, such as cases where an AI system incorporates elements developed by another provider. A solution could be for providers to draft a "responsibility attribution statement" to transparently allocate compliance obligations among contributors, distinguishing between lead and component providers.

For high-risk AI systems, Article 25(1) of the AI Act expands the scope of responsibility along the AI value chain. Operators such as distributors, importers, deployers or other third parties may be classified as providers in any of the following circumstances: *(a) they brand existing high-risk AI systems with their name or trademark, regardless of contractual arrangements; (b) they substantially modify high-risk AI systems, maintaining their high-risk classification under Article 6; (c) they change the intended purpose of AI systems (including general-purpose AI) in a way that reclassifies them as high-risk under Article 6.*[11]

Furthermore, according to Article 25(3) of the AI Act, for high-risk AI systems that serve as safety components of products covered by EU harmonisation legislation (Annex I, Section A), the product manufacturer is considered to be the provider and must comply with Article 16 if: *(a) the AI system is placed on the market alongside the product under the manufacturer's name or trademark; or (b) the AI system is put into service under the manufacturer's name or trademark after the product has already been placed on the market.*[12]

1.2. Deployers Established or Located Within the EU

Article 2(1)(b) provides that the AI Act applies to deployers who have their place of establishment or are located within the EU.[13]

[10] Article 2(6) and Recital (25) of the AI Act.

[11] Article 25(1) of the AI Act.

[12] Article 25(3) of the AI Act.

[13] Article 2(1)(b) of the AI Act.

The AI Act defines a deployer in Article 3(4) as 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”.[14]

However, the AI Act does not explicitly define what constitutes usage ‘*under its authority*’. To distinguish this concept from the mere use of AI systems, a certain degree of control over the system appears necessary. Not all professional uses of AI automatically classify an entity as a deployer. Rather, the AI system must be intended to be controllable for specific purposes by someone other than the provider.

For example, using a customer service chatbot on a website for professional activities does not necessarily mean the user is a deployer, unless that entity has the ability to control or configure the chatbot for their own purposes within certain limits. This distinction is critical because deployers are subject to legal obligations, whereas end-users — who merely interact with AI systems without control over them — should primarily be protected from AI-related risks rather than be held accountable for compliance.

The AI Act applies to all deployers established or located in the EU. This includes entities with their legal or administrative headquarters in the EU (establishment) and entities operating within the EU, even if their headquarters are outside the EU (located).

1.3. Providers and Deployers Established or Located in a Third Country Where the Output Produced by the AI System is Used in the EU

Article 2(1)(c) of the AI Act extends its scope beyond providers and deployers established or located within the EU, applying to providers and deployers of AI systems that have their place of establishment or are located in a third country where the output produced by the AI system is used in the EU.[15]

This provision broadens the territorial reach of the AI Act to address potential risks associated with AI systems developed or deployed outside the EU but whose impact is felt within its territory. Recital 22 further clarifies that certain AI systems should be covered by the AI Act even if they are not placed on the market, put into service or directly used within the EU, in order to prevent circumvention by providers and deployers based in third countries.[16]

[14] Article 3(4) and Recital (13) of the AI Act.

[15] Article 2(1)(c) of the AI Act.

[16] Recital (22) of the AI Act.

However, the AI Act does not explicitly define what constitutes the '*use of AI system output'* in the EU. Recital 12, which discusses the notion of AI systems, offers guidance by listing examples of AI output such as predictions, content, recommendations or decisions.[17] What is clear is that for an AI system's output to fall within the scope of the AI Act, it must be '*intended to be used in the Union'* according to Recital 22 of the Act, meaning that its use must be directed towards the EU.

For example, an AI provider based in the United States develops a medical diagnosis system and licenses it to a healthcare company [AI deployer] in Canada. The Canadian deployer offers remote diagnostic services to patients located in several EU Member States. Since the output of the AI system [diagnoses and recommendations] is intended to be used for EU-based clients [i.e. end-users], the AI Act applies — even though both the provider and deployer are outside the EU. The result is different if the deployer's services are reserved for North American customers; the Regulation is not applicable. It should be noted that European citizenship, i.e. possession of the nationality of an EU Member State, does not trigger the application of the Regulation. Therefore, the fact that a European citizen living in Canada uses these medical diagnoses does not require said provider and deployer to comply with the AI Act.

2. Secondary Operators In The AI Value Chain

2.1. Importers and Distributors

Article 2[1][d] of the AI Act establishes that the AI Act applies to importers and distributors of AI systems.[18]

An importer is defined in Article 3[6] as a "*natural or legal person located or established in the EU that places on the market an AI system that bears the name or trademark of a natural or legal person established in a third country*"[19], while the distributor is, according to Article 3[7], "a natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the EU market"[20].

The Article 2 provision seeks to establish shared responsibility among all actors in the AI value chain, importers and distributors included, ensuring that AI systems entering the

[17] Recital [12] of the AI Act.

[18] Article 2[1][d] of the AI Act.

[19] Article 3[6] of the AI Act.

[20] Article 3[7] of the AI Act.

EU market comply with safety and regulatory requirements[21]. Although Articles 3(6) and 3(7) of the AI Act suggest that every AI system requires an importer and a distributor, the AI Act ultimately limits this requirement to high-risk AI systems, exempting importers and distributors of GPAI models, even those associated with systemic risks[22].

Importers and distributors assume obligations similar to those of the original provider when they engage in specific activities outlined in Article 25, such as placing their name or trademark on a high-risk AI system or making substantial modifications to such a system. This approach reinforces accountability within the supply chain, ensuring that even if the original provider is located outside the EU, an identifiable entity within the EU remains responsible in cases where an AI system is misused or altered in a manner that introduces significant risks.

2.2. Product Manufacturers

Article 2(1)(e) of the AI Act provides that the AI Act applies to product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark.[23] This regulatory approach aims to establish clear lines of responsibility, ensuring that manufacturers remain accountable for the AI systems embedded in their products.

However, the AI Act does not provide a definition of "*product manufacturers*" nor does it specify what constitutes an '*AI system together with a product*' (i.e. an integrated AI system)[24]. While the wording of Article 2(1)(e) suggests that any product manufacturer incorporating AI into its products falls within the AI Act's scope, specific obligations for these manufacturers are limited to high-risk AI systems. More precisely, manufacturers are subject to AI Act requirements only when the high-risk AI system is also covered – as a safety component of a regulated product – by the EU harmonisation legislation listed in Section A of Annex I.[25] In such cases, as outlined in Articles 25(3) and 43(3) of the AI Act, the product manufacturer is considered the provider of the high-risk AI system and must ensure its compliance with the AI Act's requirements.

This suggests that the scope of Article 2(1)(e) is primarily restricted to high-risk AI systems integrated into products. However, this does not imply that manufacturers developing

[21] Recital (83) of the AI Act.

[22] The Regulation does not lay down specific obligations for importers and distributors of GPAI models as such, unlike the obligations applicable to high-risk AI systems. However, they may still be indirectly concerned if they are involved in placing on the market or integrating a GPAI model into an AI system.

[23] Article 2(1)(e) of the AI Act.

[24] The term is not defined in the Regulation itself but refers to sector-specific EU harmonisation legislation (such as for e.g. the Machinery Regulation, the Medical Devices Regulation, etc.).

[25] Article 25(3) and Article 43(3) of the AI Act.

products with integrated AI systems that do not qualify as high-risk are entirely excluded from the AI Act's scope. In such cases, these manufacturers are likely to be considered AI system providers and, as clarified in Recital 87, must comply with the corresponding obligations.[26] AI systems may function either as a stand-alone entity or as a component of a product, regardless of whether they are physically embedded or merely serve the product's functionality without being integrated. While product manufacturers may not be directly subject to the AI Act's obligations when integrating non-high-risk AI systems into their products, they are still required to comply with provider obligations under the AI Act.

Additionally, even if an AI system embedded in a product is not classified as high-risk under the AI Act, it must still meet safety requirements when placed on the market or put into service. Recital 166 clarifies that, in such instances, the EU General Product Safety Regulation[27] serves as a 'safety net' to ensure consumer protection.[28]

2.3. Authorised Representatives

Article 2(1)(f) of the AI Act establishes that the AI Act applies to authorised representatives of providers which are not established in the EU.[29] The AI Act defines an authorised representative in Article 3(5) as "*a natural or legal person located or established in the EU 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 the AI Act*".[30]

While the definition of authorised representatives indicates that every non-EU provider of an AI system must appoint one, Article 22(1) specifically mandates this requirement only for providers of high-risk AI systems.[31] Providers of non-high-risk AI systems are not obligated to appoint an authorised representative under the AI Act. This is fully coherent

[26] Recital (87) of the AI Act.

[27] Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC (Text with EEA relevance) PE/79/2022/REV/1 OJ L 135, 23.5.2023, p. 1–51 (hereinafter, 'EU General Product Safety Regulation').

[28] Recital (166) of the AI Act.

[29] Article 2(1)(f) of the AI Act.

[30] Article 3(5) of the AI Act.

[31] Article 22(1) of the AI Act.

as the Act's legal requirements for low-risk AI systems are limited to transparency obligations pursuant to Article 50 of the Regulation.[32]

2.4. Affected Persons Located in the EU

Article 2(1)(g) of the AI Act provides that the AI Act applies to affected persons that are located in the EU.[33] This provision underscores the "human-centric" focus of the AI Act, emphasising the protection of individuals within the EU Member States.[34]

While earlier drafts from the European Parliament provided a clear definition of 'affected persons' as "*any natural person or group of persons who are subject to or otherwise affected by an AI system*",[35] the final text of the AI Act omits this definition, leaving open the question of when a person can be considered 'affected' and whether legal persons are also included under this term. On that latter point, the response could be seen as negative, as the provision explicitly refers to "*natural person*"; on the other side, the decision to exclude this concept from the Regulation seem to leave the issue open.

The inclusion of "affected persons" in the scope of the AI Act should be viewed in the broader context of the EU's commitment to safeguarding individuals from the potential risks associated with AI systems and GPAI models, as stated in Article 1 and related Recitals of the AI Act. Therefore, Article 2(1)(f) should not be interpreted as encompassing any AI system or GPAI model that could affect someone within the EU. Such a broad interpretation would undermine the purpose of the other provisions in Article 2(1), as these would be unnecessary if the mere potential impact on persons in the EU were sufficient to trigger the AI Act's scope. The impact of this criterion for applying the Act is therefore limited to the rights that those affected could derive from the Regulation, either directly (e.g. right to information under Article 26 (11) or right to explanation under Article 86) or through the potential horizontal direct effect of the text.[36]

Conclusion

The AI Act introduces a comprehensive regulatory framework for various AI operators, ensuring accountability across the entire AI value chain. Compliance with the AI Act requires careful attention to the personal scope, verifying which operators are included

[32] On transparency / Article 50, see in this Guide, *infra*, F. Guillaumé, p.80.

[33] Article 2(1)(g) of the AI Act.

[34] Article 1(1) of the AI Act.

[35] See Article 3(1)(8a) of the Report on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, 22 May 2023 (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

[36] On that dimension, see M. Ho-Dac, "The EU AI Act and the Challenge of Protecting Fundamental Rights", *Common Market Law Review*, vol. 62 (2025) issue 5.

and which are not, as well as to the territorial scope, since the AI Act applies not only to operators within the EU but also to providers and deployers outside the EU whose AI systems impact the EU market and its consumers and citizens.

For providers, compliance necessitates adherence to strict obligations, particularly for high-risk AI systems. Providers must ensure that their AI systems comply with the AI Act's risk classification criteria, particularly when modifying or integrating AI into products already regulated under EU harmonised legislation. The ambiguity regarding when an AI system is considered 'developed' requires further clarification, and organisations should take a cautious approach by aligning their practices with the AI Act's fundamental principles.

Deployers must be aware of their obligations, particularly concerning their control over AI systems. The AI Act does not impose obligations on end-users, but deployers who actively manage AI systems under their authority should implement robust compliance measures. Ensuring transparency in decision-making processes and maintaining oversight over AI functionalities will be critical to mitigating legal risks.

The AI Act's broad reach, covering both EU-based and specific non-EU providers and deployers whose AI systems impact the EU, necessitates a thorough evaluation of operational activities and their potential regulatory implications. Organisations should establish clear frameworks to determine their status under the AI Act, considering both the nature of their role in the AI value chain and the geographical reach of their (overall) AI systems.

For importers and distributors, the AI Act reinforces shared responsibility, requiring them to verify compliance when bringing AI systems into the EU market. These actors should establish due diligence mechanisms to ensure that high-risk AI systems meet regulatory requirements before distribution or commercialisation. The same applies to product manufacturers, who must comply with provider obligations if they integrate high-risk AI systems into their products.

Authorised representatives play an increasingly important role in regulatory compliance, particularly for high-risk AI systems. Providers operating outside the EU must ensure that their representatives within the EU can effectively fulfil their regulatory obligations. Organisations should anticipate increased scrutiny of the liability of EU representatives, given the growing emphasis on their role in EU digital laws.

Finally, while the AI Act adopts a human-centric approach based on EU values, the lack of a clear definition of 'affected persons' raises questions about the extent of individual rights under the AI Act. Operators should prioritise transparency and accountability in AI decision-making, especially in cases where their AI systems impact individuals in the EU, even if deployed from a third country.

Bibliography

- Christakis T, Pinto S and Raj P, *Navigating the EU AI Act : A comprehensive meta-guide to leading tools* (AI-regulation.com, 26 February 2025).
- Ho-Dac M, 'Premier décryptage du règlement européen sur l'intelligence artificielle (AI Act) : Vers un standard mondial de l'IA de confiance ?' 2024. hal-04665170.
- Mandarà E, 'Il Regolamento UE sull'intelligenza artificiale' in Iaselli M (ed), *AI ACT – Principi, regole ed applicazioni pratiche del Reg. UE 1689/2024* (Maggioli Editore 2024) 45.
- Marafioti L, 'Caratteri essenziali e ambito di applicazione del Regolamento' in Cassano G and Tripodi EM (eds), *Il Regolamento europeo sull'intelligenza artificiale – Commento al Reg. UE n. 1689/2024* (Maggioli Editore 2024) 369.
- Peaden VF, 'Who's Who under the EU AI Act : Spotlight on Key Actors' (Bakerdonelson, 14 March 2024).
- Van Eecke P and Regenhardt B, 'Article 2 – Scope' in Pehlivan CN, Forgó N and Valcke P (eds), *The EU Artificial Intelligence (AI) Act : A Commentary* (Kluwer Law International 2024) 33.

Chapter 4 - AI Legislative Frameworks Coordinated with the AI Act: What Does It Mean for Implementation?

Béatrice Schütte (Univ. of Helsinki)

Introduction

While the AI Act is seen as the first attempt to comprehensively regulate AI, it certainly does not provide exhaustive regulation of all matters related to this family of technologies. The Act addresses the definition of the term 'AI system', establishes parameters for the classification of risks and mandates obligations for stakeholders in the value chain, among other provisions.

Due to AI's manifold use cases and the consequences related to these uses, the AI Act must often be read in conjunction with other pieces of legislation. For instance, if an AI system processes personal data, the rules enshrined in the General Data Protection Regulation (GDPR)[1] have to be observed, as per Article 2(7) of the AI Act. Some of the legal frameworks with which the AI Act must be coordinated have been enacted earlier, such as the GDPR, whereas others, like the revised (EU) Directive 2024/2853 on liability for defective products, were adopted at a later stage. Particular vigilance is therefore required with regard to the temporal scope of the texts to be coordinated.

The AI Act is classified as a specialised product safety legislation[2], which means that general product safety legislation, such as the new General Product Safety Regulation[3] and, where applicable, sector-specific product safety legislation, must be considered if an AI system qualifies as a product.

In view of never-ending technological development, the EU legislator faces the hard task of ensuring that the legal framework is consistent. Otherwise, there is a risk of legal

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119, 4.5.2016, p. 1–88.

[2] M. Kop, EU Artificial Intelligence Act: The European Approach to AI. Stanford – Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021, 2

[3] Regulation (EU) 2023/988, General Product Safety Regulation, 10 May 2023, OJ L 135, 23 May 2023, pp. 1–51 (applicable from 13 December 2024).

uncertainty, which might then jeopardise the envisioned goals of a level playing field in AI regulation across the EU, as well as fostering trustworthy AI and its acceptance along the value chain.

This contribution presents selected frameworks applicable to AI and examines them in light of their legal interplay with the AI Act. The objective is to analyse their normative alignment to ensure legal consistency with the AI Act. These frameworks relate to product safety, intellectual property, cybersecurity, data and liability. All of these can be linked to critical properties of AI that have been identified by the European Commission in its Safety and Liability Report published in 2020, namely autonomy, opacity, data dependency and connectivity.[4] Further, this contribution outlines critical issues in relation to the compatibility of the respective frameworks.

1. The AI Act and Product Safety

As regards product safety, both general and sector-specific legislation are relevant. In general terms, the AI Act is part of the New Legislative Framework that structures legislation on product safety in the EU internal market. The fact that AI systems can be products can be inferred, for instance, from Article 6 of the AI Act, referring to AI systems being products themselves.[5] In addition, Annex I of the AI Act provides a further hint, citing sector-specific product safety legislation. This is also clear from the similarity between the term 'product' in the Market Surveillance Regulation and the AI Act.[6]

To be in line with technological progress, the EU legislator has revised legislation on general product safety. The previous General Product Safety Directive from 2001 was replaced by the General Product Safety Regulation (GPSR)[7], as the European Commission acknowledged that the earlier Directive's applicability to new technologies was not straightforward.[8]

[4] European Commission, 'Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics', COM (2020) 64 final.

[5] See also Recital 51 of the AI Act.

[6] Art. 74 (1) b) of the AI Act : « any reference to a product under Regulation (EU) 2019/1020 shall be understood as including all AI systems falling within the scope of this Regulation".

[7] Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC, OJ L 135, 23.5.2023, p. 1–51.

[8] European Commission, "COMMISSION STAFF WORKING DOCUMENT: IMPACT ASSESSMENT accompanying the document 'Proposal for a Regulation of the European Parliament and of the Council

AI systems that are at the same time products as per the GPSR must comply with the rules set out in both regulations, and, if applicable, also with applicable sector-specific safety rules, such as the Machinery Directive.[9]

The GPSR contains an updated definition of products in its Article 3(1), now referring to interconnected products. Unfortunately, as opposed to the revised Product Liability Directive (PLD)[10], the new regulation itself failed to clarify whether standalone software can also be a product. Only later did the European Commission explain that products under the GPSR can be intangible, and that this notion includes software.[11]

The GPSR includes critical factors relating to new technologies in its safety criteria, such as evolving, learning and predictive functionalities in Article 6(1)(h). The GPSR adds cybersecurity to the factors that must be considered in assessing the safety of products, as per Article 6(1)(g). This reflects the elevated importance of products with digital elements and the fact that they will likely undergo changes during their life cycles.

As opposed to the AI Act, the GPSR does not distinguish between different categories of products' risks. As such, any product must meet the relevant criteria – for instance in terms of cybersecurity – to be considered safe, whereas the cybersecurity requirement in Article 15 of the AI Act only applies to high-risk AI systems. As a consequence, AI systems that are at the same time products under the GPSR must meet certain cybersecurity requirements, regardless of their risk classification.

2. The AI Act and Intellectual Property

In the text of the first proposal for the AI Act, copyright was not mentioned. Only the Explanatory Memorandum had a reference to a resolution on AI and copyright issued by the European Parliament[12]. This topic only became relevant with the advent of generative AI (GenAI) – particularly large language models (LLMs) – and sent the

on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council", SWD(2021) 168 final, p. 12.

[9] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast) (Text with EEA relevance) OJ L 157, 9.6.2006, p. 24–86

[10] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC (Text with EEA relevance), *OJ L, 2024/2853, 18.11.2024,* ELI: http://data.europa.eu/eli/dir/2024/2853/oj.

[11] European Commission, Product Safety Legislation, https://ec.europa.eu/safety-gate/#/screen/pages/productSafetyLegislation

[12] European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI).

legislator back to the drawing board in 2022.[13] GenAI can produce "original" creations such as texts, images, and videos following user instructions or queries given to the model (so-called 'prompts') to obtain a certain result.[14] This leads to the question of who the copyright holder of such AI-generated works is. To date, the predominant view is that only humans can hold intellectual property rights.[15] However, in the future it will likely be necessary to regulate the question of who holds the copyright of AI-generated works.

The final version of the AI Act does not directly regulate intellectual property-related questions either. However, in relation to general purpose AI (GPAI) systems, it sets out in Article 53(1)(c) that providers of GPAI models must establish a policy to comply with EU copyright law and related rights. The AI Act further states in Recital 105 that the use of copyright-protected content in the training of GPAI models must be authorised by the copyright holder. In this context, the AI Act also refers to Directive 2019/790 on Copyright in the Digital Single Market (CDSM Directive).[16] The Directive lays down rules aiming to further harmonise Union law applicable to copyright and related rights in the framework of the internal market, taking into account, in particular, digital and cross-border uses of protected content.[17]

At this point, the implementation by GPAI providers of a copyright policy to comply with Union law, in particular by identifying and respecting the reservations expressed by rightholders, is supported by the drafting of the code of practice[18] provided for in Article 56 of the AI Act.

3. The AI Act and Cybersecurity

Due to their data dependency and connectivity, AI systems can be vulnerable to cyber-attacks. For high-risk AI systems, Article 15 of the AI Act sets out that they must be

[13] Andres Guadamuz, 'The EU's Artificial Intelligence Act and copyright' (2025), The Journal of World Intellectual Property 28 (1), 214.

[14] European Commission, 'Artificial intelligence and copyright: use of generative AI tools to develop new content' (2024), News Blog, available at https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/artificial-intelligence-and-copyright-use-generative-ai-tools-develop-new-content-2024-07-16-0_en.

[15] See e.g. Belinda Bennett & Angela Daly (2020) 'Recognising rights for robots: Can we? Will we? Should we?', Law, Innovation and Technology, 12:1, 72, 73.

[16] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125.

[17] CDSM Directive, A. 1.

[18] https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

"*designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity…*". Cybersecurity measures taken must be appropriate to the relevant circumstances and the risk. These requirements can be criticised as vague; however, one must keep in mind that the technology evolves rapidly and the Regulation is meant to be applied to a plethora of AI systems.

Beyond the AI Act, one highly-relevant piece of legislation at the EU level is the *Cybersecurity Act [CSA]*[19]. The CSA was enacted before the AI Act. It does not mention AI as a technology, but refers instead to information and communication technologies, which is a broader notion and encompasses AI. Cybersecurity is defined in Article 2[1] of the CSA as "*the activities necessary to protect network and information systems, the users of such systems, and other persons affected by cyber threats*".

The CSA establishes ENISA, the European Agency for Cybersecurity. In addition, it establishes a framework for the establishment of European cybersecurity certification schemes to ensure an adequate level of cybersecurity for ICT products, ICT services and ICT processes in the EU. The Regulation also aims to avoid the fragmentation of the internal market with regard to cybersecurity certification schemes.[20] As per Article 51 of the CSA, a European cybersecurity certification scheme shall be designed to achieve a number of objectives, among others, the protection of stored, transmitted, or otherwise processed data against accidental or unauthorised storage, processing, access, disclosure, destruction, loss or alteration. While both the CSA and the AI Act specify cybersecurity as a requirement or as a goal to achieve, they are silent as to how to achieve cybersecurity, likely in order to keep the rules future-proof.

The operationalisation of cybersecurity requirements requires the establishment of technical standards prepared by stakeholders.[21] This will be the case for the requirements of Article 15 of the AI Act, which must be supplemented by one or more normative deliverables, including a harmonised standard on cybersecurity, robustness and accuracy. The European AI standardisation committee CEN-CENELEC JTC 21, responsible for AI standards in the context of the AI Act, is working with the cybersecurity committee to ensure the harmonisation of technical requirements and procedures in AI/cybersecurity standardisation.

[19] Regulation [EU] 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA [the European Union Agency for Cybersecurity] and on information and communications technology cybersecurity certification and repealing Regulation [EU] No 526/2013 [Cybersecurity Act] [Text with EEA relevance], OJ L 151, 7.6.2019, p. 15–69.

[20] CSA, Article 1[1].

[21] On technical standards in AI context, see this Guide, infra, O. Kanevsakia, p. 110.

Relevant in relation to cybersecurity is also the Cyber Resilience Act (CRA),[22] which lays down rules to ensure the cybersecurity of products with digital elements, as per its Article 1. It has entered into force in December 2024 and will become applicable in 2027.

The CRA applies to all products connected directly or indirectly to other devices or networks except for specified exclusions such as certain open-source software or products that are already covered by existing rules, for instance medical devices, aviation and cars.[23] The CRA tackles the inadequate level of cybersecurity in many products and addresses the challenges faced by both businesses and consumers in determining which products are cybersecure and in setting them up securely.[24] For the definition of the term cybersecurity, the CRA refers to the CSA.

The AI Act does not contain any definition of cybersecurity, nor does it refer to the CSA for it. However, in Recital 77, the AI Act refers to the CRA with regard to AI systems that fall under the scope of the notion '*products with digital elements*', stating that when these products comply with the cybersecurity requirements established therein, they should also be considered compliant with the requirements set out in Article 15 of the AI Act.

4. The AI Act and Data

4.1. General background on EU data protection law in the AI Act context

As AI systems are highly data dependent, meaning that they have to be trained with the help of data and in most cases also process data during their deployment, the AI Act must also be coordinated with relevant legislation on data.[25] The GDPR is particularly relevant if personal data are concerned. In addition, the Data Governance Act (DGA)[26] and the Data Act (DA)[27] may also be applicable. All three regulations are part of the

[22] Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act) (Text with EEA relevance), *OJ L, 2024/2847, 20.11.2024*, ELI: http://data.europa.eu/eli/reg/2024/2847/oj.

[23] European Commission, Cyber Resilience Act, available at https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act.

[24] *Ibid.*

[25] On the requirements regarding data in the AI Act, see this Guide *infra*, J.-M. Van Gyseghem, p.75.

[26] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJ L 152, 3.6.2022, p. 1–44.

[27] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA relevance), *OJ L, 2023/2854, 22.12.2023*, ELI: http://data.europa.eu/eli/reg/2023/2854/oj.

European Strategy for Data, which strives to create a single market for data in order to ensure Europe's global competitiveness and data sovereignty.[28] This section will focus on the GDPR.

The GDPR applies to the processing of personal data by wholly or partially automated means and to the non-automated processing of personal data that forms or is intended to form part of a filing system.[29] As the GDPR is technology-neutral[30], it does not refer explicitly to AI but many of its provisions are nonetheless relevant thereto.[31] The Regulation focuses on the effects of data processing and on the potential impact on risks for fundamental rights, rather than on the technologies used.[32] As regards the AI life cycle, the GDPR applies to the development of AI systems and to their use for analysis and decision-making about individuals.[33]

Like the AI Act, the GDPR reflects the adoption of a risk-based approach, with the principle of accountability at its core based on Article 24, which sets out that controllers '*shall implement appropriate technical and organizational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation*'.[34] Despite its risk-based approach, the GDPR does not define risk, but leaves risk evaluation and mitigating measures to the discretion of data controllers and processors, taking a bottom-up approach.[35]

As regards the compatibility of the GDPR and the AI Act, concerns have been voiced that the GDPR's strong focus on privacy and control over personal data could clash with the AI Act's need to access data for certain AI systems. At the same time, the AI Act might

[28] European Commission, 'A European strategy for data', available at: https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

[29] See GDPR, Article 2(1).

[30] GDPR, Recital 15.

[31] Sartor G, Lagioia F (2020) The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. European Parliamentary Research Service Study, II.

[32] Mitrou L (2018) 'Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'?', 26.

[33] Ibid., p. 27.

[34] Gellert R (2021) The role of the risk-based approach in the General Data Protection Regulation and in the European Commission's proposed Artificial Intelligence Act: business as usual? *J Ethics Leg Technol* 3(2), 20.

[35] Dunn P, De Gregorio G (2022) The ambiguous risk-based approach of the artificial intelligence act: links and discrepancies with other union strategies. IAIL 2022 Imagining the AI Landscape after the AI Act, 3; Gellert R (2021) The role of the risk-based approach in the General Data Protection Regulation and in the European Commission's proposed Artificial Intelligence Act: business as usual? J Ethics Leg Technol 3(2), 20.

be able to reinforce the GDPR's objectives.[36] The GDPR has been criticised for failing to adequately address Big Data and the data environment it creates and for its inability to close the gaps left by the AI Act in relation to, for instance, social media.[37]

4.2. The interplay between the GDPR and the AI Act

AI and personal data are interrelated. Many AI systems process personal data[38]; AI is fed *inter alia* by personal data and produces inferred data.[39] Therefore, AI systems and models must be designed and programmed so as to conform with the GDPR.

Articles 5 and 6 of the GDPR establish the core principles of data processing. Most importantly, data processing must be lawful, that is, it must have an actual legal basis. Furthermore, data must be processed fairly, and the interests that are involved must be balanced. The other core principles are transparency, purpose limitation, data minimisation, accuracy, and storage limitation.[40] These principles must be integrated into the programming and training of AI systems. As far as the transparency requirement is concerned, relevant information must be provided to the user of the AI system or to any other person whose data is processed by the system. The information in question can be provided through the interface of the system or by any other means of communication.

The relationship between the AI Act and the GDPR is of importance also in relation to automated decision-making (ADM). According to Article 22(1) of the GDPR, data subjects have the right to not be subjected to decisions taken solely by automated processing. Exceptions to this rule are found in Article 22(2) GDPR, i.e. relating to contractual necessity, authorised by EU or Member State law, subject to sufficient safeguards being in place, or the data subject's explicit consent. That being said, for ADM to be permitted under the scope of the GDPR, human intervention is required, which can also be concluded from Article 22(3) GDPR. Similarly, Article 14 of the AI Act requires human oversight in high-risk AI systems. If ADM takes place in the domains and use cases mentioned in Annex III of the AI Act, and the system used is an AI system, it will be

[36] Butt, J. S, 'The General Data Protection Regulation of 2016 (GDPR) Meets Its Sibling the Artificial Intelligence Act of 2024: A Power Couple, or a Clash of Titans?' Acta Universitatis Danubius Juridica, vol. 20, no. 2, 2024

[37] Zarsky TZ. (2017) Incompatible: the GDPR in the age of big data. Seton Hall Law Rev 47(4 (2)), 996; Schütte B, "AI Regulation in the EU: The Future Interplay Between Frameworks, in Gill-Pedro E, Moberg A (eds.), YSEC Yearbook of Socio-Economic Constitutions 2023,

[38] Sartor G, Lagioia F (2020) The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, op. cit.

[39] Mitrou L (2018) 'Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'?', 19.

[40] Feiler L et al (2018) 'The EU General Data Protection Regulation (GDPR): a commentary', 73, 74.

classified as high-risk, meaning that both Article 22 GDPR and Article 14 AI Act apply at the same time.

With regard to Article 22 GDPR, it has been stated that human intervention must be "*meaningful*".[41] This is a vague term, however it has been further clarified that "*meaningful*" has to go beyond mere rubber-stamping, and the human involvement requires a person having the authority and competence to change the decision.[42] Human oversight under the AI Act requires measures appropriate to the risks, level of autonomy and context of use of the respective AI system.[43] In addition, Article 14 (4) AI Act states that persons to whom human oversight is assigned must be able to understand the capabilities and limitations of the respective AI system, to duly monitor its operation, to be aware of automation bias, to be capable of interpreting the system's output and to be prepared to either override the system's decision or to decide to discontinue its use. As such, both Regulations require a necessary level of competence to oversee the AI system and to intervene meaningfully in relation to an ADM system.

5. The AI Act and Civil Liability

The AI Act does not contain any rules on civil liability. To date, there is no dedicated harmonised legislation at the EU level to deal with liability for damage specifically caused by AI systems.

In the AI context, the revised Product Liability Directive 2024/2853 (PLD) is applicable to a certain extent. The Directive establishes no-fault liability for certain economic operators in case a defective product has caused damage to a person, and it applies to digital products including AI. The concept of an "*economic operator*" is wider than that of the "*producer*" under the 1985 PLD, reflecting the increasing complexity of value chains: nowadays, it is common for the different components of a product to be provided by different economic operators.[44]

For a successful claim, a product must also be defective. Pursuant to Article 7 (1) a product is defective "*where it does not provide the safety that a person is entitled to expect or that is required under Union or national law*". To this end, according to Article

[41] Article 29 Data Protection Working Party, "Guidelines on Auto- mated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679" (2018).

[42] Article 29 Data Protection Working Party, "Guidelines on Auto-mated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679" (2018); Lazcoz G, de Hert P, "Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities" (2023) *Computer Law and Security Review*, 11.

[43] Article 14 (3) AI Act.

[44] See also Schütte B, "AI Regulation in the EU: The Future Interplay Between Frameworks", in Gill-Pedro E, Moberg A (eds.), *YSEC Yearbook of Socio-Economic Constitutions 2023*, 35.

7[2] of the PLD, all circumstances must be taken into account, including a product's presentation and characteristics and its reasonably foreseeable use. With regard to the AI Act and other related legislation, particularly relevant are a product's ability to learn, the effects of interconnectedness with other products and cybersecurity requirements. The latter must be read in conjunction with the AI Act, the CRA and the GPSR, which can be concluded from Article 7 [1] PLD, referring to safety required under Union law.

The concept of defectiveness has been expanded in view of more complex value chains and products changing during their life cycles. Defects can manifest even after the product is placed on the market, for instance due to faulty updates, as well as due to the interconnectedness of products, for instance in smart home environments in which a defect in one product can spread to the other products that interact with it.[45]

An example that combines the application of the PLD and the AI Act would be as follows: An AI system used by a financial institution to assess creditworthiness [Annex III, point 5[b] AI Act] malfunctions due to a design flaw, systematically underestimating applicants' income. As a result, a consumer is wrongly denied access to a loan needed for urgent home repairs, and the delay leads to structural damage to her property. In this case, the AI system could be considered defective under Article 7[1] PLD, as it failed to provide the level of safety required under Union law — particularly where the AI provider did not comply with the AI Act's requirements on data quality, risk management, and transparency

In 2022, the EU legislator had published a proposal for an AI Liability Directive [AILD][46], which was withdrawn in February 2025.[47] This was a liability framework in the name only, as it did not establish any rules on the substance of liability but focused instead on procedural aspects; it included provisions on the disclosure of evidence at the defendant's disposal and the presumption of causation. These rules aimed to address difficulties in obtaining compensation for injured parties when they have been harmed by an AI system. Critical properties like autonomy and opacity can make it difficult to pinpoint the origin of the damage. In addition, there is often an information asymmetry between the provider or deployer of an AI system and the injured party.

However, with the withdrawal of the proposal, it remains to be seen whether there will ever be any harmonised rules on liability for AI-related harm. Any claim must be settled

[45] *Ibid.*

[46] Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence [AI Liability Directive], COM/2022/496 final.

[47] https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-ai-liability-directive?p3373

according to the applicable national tort law. The burden of proof of the prerequisites of tort liability rests with the claimant. In any case, it is very likely that obligations stipulated in the AI Act will be relevant in determining whether the defendant was at fault. For instance, a high-risk AI system provider's failure to comply with obligations set out in chapter III, section 2 of the AI Act should mean that they did not observe the appropriate standard of care.

Conclusion

This contribution illustrated some aspects of coordination between the AI Act and related legislation at the EU level. Regulating new technologies is challenging already due to the fact that technological progress is constantly accelerating, while the legislator can only react.

A concise regulatory framework is crucial to foster the uptake of AI and the acceptance of these technologies across society. If stakeholders along the value chain are unable to oversee their rights and obligations, potential risks might become uninsurable, and they may end up being deterred from using any of these technologies. The legislator further faces the challenge of coordinating new legislation with already existing frameworks, while also keeping it compatible with any future pieces of legislation. This is also important as AI systems can easily fall under the scope of more than one EU Regulation or Directive.

This contribution shows that even the applicability of more than one framework does not provide for seamless regulation. For instance, the strict rules of the GDPR might in some cases also be a stumbling block in the training of AI systems.

Moreover, the coordination between the AI Act and cybersecurity legislation seems patchy at this point. Actual cybersecurity standards are only established for products with digital elements, which covers only a fraction of all AI systems. Predominantly, the question of whether cybersecurity requirements are met has to be assessed on a case-by-case basis.

At this point, persons who suffer AI-related harm can only rely on harmonised rules when their harm falls under the scope of the revised PLD. Harmonised rules on AI liability in general seem very far away after the withdrawal of the AILD proposal.

To facilitate compliance, the EU legislator must constantly monitor ongoing technological developments and be ready to re-evaluate key concepts. In view of the ongoing twin transitions, i.e. the green and digital transitions, it will become increasingly necessary to coordinate not only different frameworks on technology regulation, but also to include regulation on sustainability. At this point, references to sustainability in the technology legislation are scarce, as are references to technology in the frameworks on sustainability.

Bibliography

- Bennett B, Daly A (2020) 'Recognising rights for robots : Can we ? Will we ? Should we ?', Law, Innovation and Technology, 12:1, 60-80, DOI: 10.1080/17579961.2020.1727063
- Butt, J S, 'The General Data Protection Regulation of 2016 (GDPR) Meets Its Sibling the Artificial Intelligence Act of 2024 : A Power Couple, or a Clash of Titans?.' Acta Universitatis Danubius Juridica, vol. 20, no. 2, 2024
- Dunn P, De Gregorio G (2022) The ambiguous risk-based approach of the artificial intelligence act: links and discrepancies with other union strategies. IAIL 2022 Imagining the AI Landscape after the AI Act. https://ceur-ws.org/Vol-3221/IAIL_paper7.pdf
- Feiler L et al (2018) The EU General Data Protection Regulation (GDPR): a commentary. Globe Law & Business
- Gellert R (2021) The role of the risk-based approach in the General Data Protection Regulation and in the European Commission's proposed Artificial Intelligence Act : business as usual ? J Ethics
- Leg Technol 3(2) : 15–33. https://jelt.padovauniversitypress.it/2021/2/2
- Guadamuz A, 'The EU's Artificial Intelligence Act and copyright' (2025), The Journal of World Intellectual Property 28 (1), 213-219
- Kop M (2021) EU Artificial Intelligence Act : The European Approach to AI. Stanford – Vienna
- Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford
- University, Issue No. 2/2021. Available at https://law.stanford.edu/publications/eu-artificial-intelligence-actthe-european-approach-to-ai/
- Lazcoz G, de Hert P, 'Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities' (2023) Computer Law and Security Review.
- Mitrou L (2018) 'Data Protection, Artificial Intelligence and Cognitive Services : Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'?' https://doi.org/10.2139/ssrn.3386914
- Sartor G, Lagioia F (2020) The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. European Parliamentary Research Service Study. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530
- Schütte B, 'AI Regulation in the EU: The Future Interplay Between Frameworks', in Gill-Pedro E, Moberg A (eds.), YSEC Yearbook of Socio-Economic Constitutions 2023, 17-49
- Zarsky TZ (2017) Incompatible: the GDPR in the age of big data. Seton Hall Law Rev 47(4(2)):995–1020. https://scholarship.shu.edu/cgi/viewcontent.cgi?article=1606&context=shlr

III- MAJOR HIGH-RISK AI SYSTEMS REQUIREMENTS

Chapter 5 - Risk Management System Under The AI Act

Amélie Favreau (Univ. Grenoble)

Introduction

Pursuant to Article 9, the AI Act mandates the implementation of a robust Risk Management System (RMS) by AI providers to ensure the safety and reliability of high-risk AI systems.

1. Contextualising Risk Management in AI

As highlighted in the AI Safety Report published during the 2025 AI Summit[1], we are witnessing a technological revolution that is fundamentally reshaping our societies and economies:

"We are in the midst of a technological revolution that will fundamentally alter how we live, work, and interact with one another. Artificial Intelligence (AI) promises to transform many aspects of our society and economy… Along with this rapid progress, experts are becoming increasingly aware of the current harms and potential future risks associated with the most capable types of AI."[2]

AI holds unprecedented potential. Yet, as its capabilities advance rapidly, experts are growing increasingly concerned about the risks and potential harms posed by highly capable AI systems.

The AI Act adopts a proactive approach to addressing these concerns by requiring high-risk AI systems to implement a structured RMS. Under the Act, *risk* is defined in Article 3 as "*the combination of the probability of the occurrence of harm and the severity of that harm*". However, the Act does not prescribe a specific risk assessment methodology, allowing for both qualitative and quantitative approaches. The precise and operational framework, including analysis and metrics benchmarks, will be established through harmonised standards currently being developed by the CEN-CENELEC JCT 21.[3]

[1] Y. Bengio et alii., "International AI Safety Report" (DSIT 2025/001, 2025); https://www.gov.uk/government/publications/international-ai-safety-report-2025.

[2] Y. Bengio et alii., "International AI Safety Report", *op. cit.* p. 24-25.

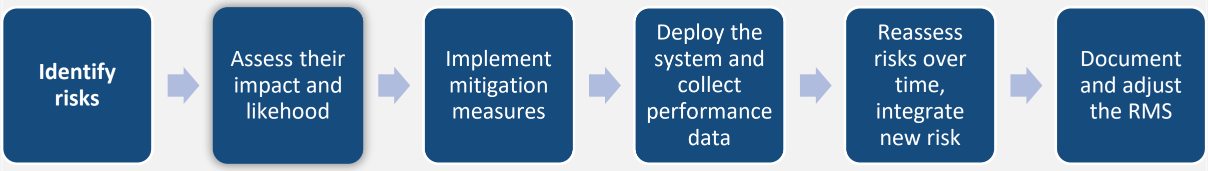
[3] On AI technical standards, see in this Guide, *infra,* O. Kanevskaia, p. 104 and also on RMS standardisation, see G. Bernard, esp. p. 162.

Although not explicitly defined in the Act, risk management is understood as a systematic process involving the identification, assessment, mitigation, and monitoring of risks. This understanding aligns with the principles set out in international standards such as ISO 31000:2018, which frame risk management as a continuous and iterative process.[4] Furthermore, risk management in the context of AI can be seen as an extension of the principle of accountability, ensuring that AI providers remain responsible for the safe operation of their systems.[5]

In practice, it is recommended that each stage of the process be thoroughly documented —including risk identification, assessment, mitigation measures, and accepted residual risks — and that the risk management plan be reviewed and updated following any technical modification, change in intended use, or significant contextual shift.

2. Understanding the Risk Management Requirements under the AI Act

According to the AI Act, Article 8(1) establishes that risk management must be taken into account when ensuring compliance with regulatory requirements. It reinforces that Article 9 provides the foundational framework for understanding risk management obligations. The interpretation of risk management must be context-specific, guided by the "*intended purpose*" and the "*state of the art*," while also considering "*reasonably foreseeable misuse*". This inherently subjective element places significant responsibility on AI providers to determine appropriate risk mitigation strategies.


Identify risks
Assess their impact and likehood
Implement mitigation measures
Deploy the system and collect performance data
Reassess risks over time, integrate new risk
Document and adjust the RMS

[4] ISO (2018) "ISO 31000:2018 - Risk Management Guidelines."

[5] Veale, M., & Borgesius, F. Z., "Demystifying the Draft EU Artificial Intelligence Act." (2021) 4 Computer Law Review International 97.

3. Key Provisions of Article 9: A Structured Approach

Article 9 serves as the cornerstone of AI risk management, regulating the acceptability of residual risks through a structured process.[6] The provision can be broken down into three key components.

3.1 A dynamic and continuous process

The first part establishes risk management as a dynamic, planned, and continuous process throughout the lifecycle of an AI system. This aligns with the broader principle of adaptive governance in AI.[7] The iterative nature of risk management ensures that AI systems remain aligned with evolving societal expectations, legal requirements, and technological developments.

- Adaptive governance

Adaptive governance in AI refers to the ability of governance structures to evolve and adjust in response to new information, changing circumstances, and emerging risks. This approach is essential in the AI context, where technological advances and societal impacts are rapid and unpredictable. Adaptive governance involves continuous learning, stakeholder engagement, and flexible regulatory frameworks that can be adjusted as needed.

For example, consider a company using AI for predictive maintenance that identifies a new type of failure not anticipated during initial testing. Following the collection of field data after market release, a model update is then initiated and documented in accordance with Section 9.2(b).

- Lifecycle risk management

Risk management in AI must be embedded throughout the entire lifecycle of an AI system, from design and development to deployment, monitoring, and decommissioning. This lifecycle approach ensures that risks are identified and mitigated at every stage, thereby reducing the likelihood of unforeseen problems.

3.2 Risk identification and assessment

The risk management process begins with risk identification, requiring AI providers to recognise known and reasonably foreseeable risks related to health, safety, and

[6] J. Schuett, “Risk Management in the Artificial Intelligence Act” (2024) 15 European Journal of Risk Regulation 367–385.

[7] Cf. Mittelstadt, B., "Principles Alone Cannot Guarantee Ethical AI" (2019) 1 Nature Machine Intelligence 501.

fundamental rights. This is followed by risk evaluation and estimation, which can be interpreted in two ways. One interpretation sees risk assessment as a direct continuation of risk identification, while another suggests that it introduces a distinct category of emerging risks. This distinction is critical for understanding whether additional risk management measures are required beyond those addressing initially identified risks.

- Known risks

Known risks are those that have been previously identified and documented. These risks are typically addressed through established mitigation strategies and are continuously monitored to ensure their effectiveness. Examples of known risks in AI systems include data privacy violations, algorithmic bias, and system failures.

- Emerging risks

Emerging risks are those that are not yet fully understood or that have not previously been encountered. These may arise from novel technologies, shifting societal expectations, or evolving threat landscapes. Emerging risks require proactive identification and innovative mitigation strategies to prevent potential harm.

For example, consider a natural language processing AI system deployed as a medical chatbot that was initially based on a validated static model. After a software update integrating a self-learning model, the system began providing increasingly personalised — but unverified — medical advice. This newly introduced behavior, not anticipated in the original design, can be considered an emerging risk. It necessitates the introduction of post-deployment verification tests and technical safeguards such as real-time validation filters or human-in-the-loop controls.

3.3 Post-market monitoring and mitigation

The next stage involves post-market monitoring, whereby AI providers must systematically collect, document, and analyse performance data to identify and mitigate new risks that emerge over time. Risk management measures must then be implemented to address these identified risks in accordance with Articles 9(4) and 9(5). The interpretation of Article 9(2)(b) plays a crucial role in determining whether mitigation measures are needed for emerging risks resulting from evolving operational conditions.

Performance data analysis. Post-market monitoring includes the collection and analysis of performance data to identify trends, anomalies, and potential risks. This data may include system logs, user feedback, and environmental factors that could impact the functioning of the AI system. Performance data analysis enables AI providers to proactively detect and resolve issues, thereby improving system safety and reliability.

Mitigation measures. Mitigation measures are actions taken to reduce the likelihood or impact of identified risks. These may include design modifications, software updates, user training, and communication strategies. The choice of appropriate mitigation measures depends on the nature of the risk, the intended use of the AI system, and applicable regulatory requirements.

Practical Analysis Grid: Article 9 - Key Obligations

Steps	Concrete Actions	Stage of Application	Documentation
Identification	Mapping known risks	Design phase	Risk sheet
Assessment	Severity/probability analysis	Development phase	Risk matrix
Mitigation	Implementation of technical or organisational measures	Before deployment	Mitigation plan
Monitoring	Post-market data collection	In production	Monitoring log
Review	RMS update after incidents or changes	Post-incident or update	Risk review report

4. Risk Prioritisation and Mitigation Hierarchy

The second part of Article 9(2) acknowledges that no AI system is ever completely risk-free. Risk prioritisation is therefore necessary, whereby the most severe and probable risks must be addressed first. This principle is central to risk-based regulation[8] and aligns with the broader philosophy of prioritising risks that pose the greatest impact on individuals and society.

The AI Act establishes a hierarchy of risk mitigation measures, emphasising that risk elimination through design and development should take precedence over *ex post* measures such as user instructions. The iterative nature of risk reduction means the process must be repeated until all risks are minimised to an acceptable level. If residual risks remain, the decision to accept them must be carefully documented.

[8] Cf. Black, Julia (2010) *Risk-based regulation: choices, practices and lessons learnt. In: Risk and Regulatory Policy: Improving the Governance of Risk. OECD, Paris, France, pp. 185-224.*

4.1. Risk-based regulation

Risk-based regulation focuses on identifying and addressing the most significant risks posed by AI systems. This approach allows organisations to allocate resources efficiently and prioritise actions that have the greatest impact on safety and societal well-being. It involves a systematic evaluation of risks, consideration of their probability and severity, and the implementation of proportionate mitigation measures.

4.2. Residual risks

Residual risks are those that remain after all feasible mitigation measures have been applied. These risks are considered acceptable based on a cost-benefit analysis and the societal value of the AI system. Residual risks must be thoroughly documented, and stakeholders must be informed of their existence and potential impacts. The acceptability of a residual risk should not rest solely on the unilateral judgment of the provider.

For example, consider a company developing an AI system for image-assisted medical diagnosis that conducts robustness testing during development using a variety of synthetic datasets to detect potential biases. Before deployment, real-world testing is carried out in a partner hospital to assess the system's accuracy on actual cases, with systematic human oversight. These tests are based on predefined metrics (eg. false positive rate, diagnostic coverage rate, response time).

The tests results are included in the technical documentation required by Article 9 of the AI Act, serving as the basis for the compliance documentation under Article 60.

It must be justified by:

(1) the state of the art in technology,

(2) societal expectations (values, fundamental rights and principles), and

(3) comprehensive documentation including assessment, trade-offs, and validation.

This reinforces the requirement for transparency and justification.[9]

5. **Testing Procedures and Compliance**

The final part of Article 9 of the AI Act addresses testing procedures, which play a crucial role in ensuring regulatory compliance. Testing must occur at multiple stages of AI system development and deployment, including real-world testing, when necessary, in accordance with Article 60. These tests must be carried out against predefined

[9] Schuett, *op cit.*

performance metrics to validate the system's safety and reliability. The ability to perform rigorous testing ensures that AI systems continuously meet regulatory and ethical standards.

5.1. Predefined Performance Metrics

Performance metrics are quantifiable measures used to assess the effectiveness, efficiency, and safety of AI systems.[10] These may include accuracy, precision, recall, response time, and system availability. Predefined metrics allow AI providers to objectively evaluate system performance and identify areas for improvement.

5.2. Real-World Testing

Real-world testing involves evaluating AI systems in their intended operational environments. This type of testing is critical for identifying risks and issues that may not appear in controlled lab conditions. Real-world testing allows providers to validate system performance under actual conditions and make necessary adjustments before full-scale deployment.

6. **Regulatory Policy Considerations in Implementing AI Risk Management**

Implementing a robust RMS under the AI Act presents several challenges and offers significant opportunities.

6.1. AI risk management challenges for organisations

Compliance costs. Compliance costs refer to the financial resources required to meet regulatory obligations. These may include the development and implementation of risk management systems, continuous monitoring and evaluation, documentation efforts, and potential penalties for non-compliance. High compliance costs can be particularly burdensome for SMEs, which may lack the necessary resources for ongoing risk evaluation and documentation.

In that respect, it worth noting that the EU AI Office is about to launch an AI Act Service Desk. It "will be an information hub with simple, straightforward information on the application of the AI Act and the possibility to receive targeted answers to questions. It will include the European Commission's Single Information Platform, as foreseen in the AI Act, which will provide online interactive tools to help stakeholders determine whether they are subject to legal obligations and understand the steps they need to

[10] On performance metrics, see in this *Guide*, infra, G. Bernard, p.161.

take to comply".[11] This initiative is part of the AI Continent Action Plan launched in April 2025.[12]

Innovation vs. Regulation. Balancing innovation with regulatory oversight is a critical challenge in AI governance. Overly strict regulations can hinder innovation by imposing heavy burdens and limiting the exploration of new technologies. Conversely, insufficient regulation may lead to unaddressed risks and potential harm. Striking the right balance requires collaboration among regulators, industry stakeholders, and other actors to develop flexible and adaptive regulatory frameworks. This rationale of multi-stakeholder regulatory governance is at the heart of the AI Act, as it involves organisations both in regulatory compliance, such as AI risk management, and in rule-making, as demonstrated by the preparation of technical standards.

6.2. AI risk management opportunities for organisations

Accountability and ethical development. Accountability in AI refers to the responsibility of AI providers to ensure that their systems operate safely, ethically, and in compliance with legal requirements. Ethical development practices involve anticipating the potential impacts of AI systems on individuals and society, and taking proactive steps to mitigate harm. Accountability and ethical development are essential to building trust in AI and promoting responsible adoption.

Public trust and global AI governance. Transparent risk management practices contribute to building confidence that AI applications align with societal values. Public trust in AI systems is critical to their widespread acceptance and adoption. Moreover, the AI Act can serve as a potential model for global regulatory harmonisation, complementing initiatives like the OECD AI Principles.[13] Regulatory alignment can foster international collaboration and support a more unified approach to AI governance at the global scale.

Conclusion

Risk management is central to the responsible development of AI, ensuring that high-risk AI systems operate safely and ethically. As highlighted in the AI Safety Report[14], policymakers must carefully balance AI's opportunities and risks, exercising caution in regulatory responses. The AI Act's risk management framework offers a structured yet

[11] https://digital-strategy.ec.europa.eu/en/funding/commission-launches-call-tender-part-efforts-establish-ai-act-service-desk

[12] COM(2025)165.

[13] OECD, "AI Principles" (2019).

[14] Op. cit.

flexible approach, emphasising iterative risk mitigation, prioritisation, and continuous monitoring.

Additionally, the Risk Management System (RMS) required under Article 9 must not be viewed in isolation: it is structurally linked to Article 17 on Quality Management Systems.[15]

[15] On QMS, see in this *Guide*, supra, M. Ho-Dac & C. Pellegrini, esp. p. 12.

Bibliography

- AI Action Summit, International AI Safety Report: The International Scientific Report on the Safety of Advanced AI (January 2025)
- Black, J. (2010) Risk-based regulation: choices, practices and lessons learnt. In: Risk and Regulatory Policy: Improving the Governance of Risk. OECD, Paris, France, pp. 185-224.
- ISO (2018). "ISO 31000:2018 - Risk Management Guidelines."
- Mittelstadt, B. (2019). "Principles Alone Cannot Guarantee Ethical AI." Nat Mach Intell 1, 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
- OECD (2019). "OECD AI Principles."
- Ceyhun Necati Pehlivan, Nikolaus Forgo & Peggy Valcke, eds., *The EU Artificial Intelligence (AI) Act: A commentary* (Wolters Kluwer, 2025)
- J. Schuett, Risk Management in the Artificial Intelligence Act, European Journal of Risk Regulation (2024), 15, 367–385.
- Veale, M., & Borgesius, F. Z. (2021). "Demystifying the Draft EU Artificial Intelligence Act.", (2021) 4 Computer Law Review International 97.

Chapter 6 - Data Governance and Management Practices under the AI Act

Jean-Marc Van Gyseghem (Univ. Namur)

Introduction

The AI Act[1] is often represented as a pyramid, with minimal-risk applications at the base and prohibited, unacceptable-risk applications at the top. Between these two extremes lie limited-risk and high-risk AI systems — the primary focus of the AI Act's regulatory framework.

This Chapter focuses on analysing Article 10 of the AI Act, dedicated to "data and data governance".

1. Interplay between the AI Act and the GDPR

Before tackling the issue of data management, it's important to consider how the AI Act fits into the European legal framework and, primarily, its relation to the General Data Protection Regulation (GDPR).[2]

It is undisputed that the AI Act must be regarded as cross-sectoral legislation, given its impact on multiple sectors and subject matters. As highlighted in Recital 3 of the Regulation, "AI systems can be easily deployed in a wide range of sectors of the economy and in many parts of society, including across borders, and can easily circulate throughout the Union." The same is true of the GDPR, which is likewise intended to be cross-sectoral, due to the very purpose of the Regulation, namely to regulate the (free) circulation of personal data irrespective of the sector involved. This cross-sectoral nature does not preclude the emergence of sector-specific laws, or *lex specialis*, that may establish additional or more detailed rules in certain fields.

[1] Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) OJ L 119, 4.5.2016, p. 1–88.

The question that arises concerns the interplay between two transversal legislations, noting that the AI Act refers to the GDPR in various provisions but also in its Recitals. Its Recital 10 states *inter alia*, "Harmonised rules for the placing on the market, the putting into service and the use of AI systems established under this Regulation should facilitate the effective implementation and **enable the exercise of the data subjects' rights and other remedies guaranteed under Union law on the protection of personal data** and of other fundamental rights". There are therefore necessarily areas of interaction between the two regulations with preeminence afforded to the GDPR. The AI Act can therefore be considered as *lex specialis* at the level of these areas of interaction.

The AI Act also borrows terminology from the GDPR, such as "impact assessment." However, despite sharing the term, the concept serves a distinct purpose in each regulation. Under the GDPR, a Data Protection Impact Assessment (DPIA) assesses risks to the rights and freedoms of individuals arising from the processing of personal data. By contrast, the AI Act introduces the Fundamental Rights Impact Assessment (FRIA), which has a broader scope. The FRIA aims to identify and mitigate risks that AI systems — particularly high-risk systems — may pose to fundamental rights, health, safety, and other societal interests, beyond data protection alone. However, a link is established between the two by Article 26, 9: "Where applicable, deployers of high-risk AI systems shall use the information provided under Article 13 of this Regulation to comply with their obligation to carry out a data protection impact assessment under Article 35 of Regulation (EU) 2016/679". This is confirmed by Article 27.4 stating that: "If any of the obligations laid down in this Article is already met through the data protection impact assessment conducted pursuant to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, the fundamental rights impact assessment referred to in paragraph 1 of this Article shall complement that data protection impact assessment."

Under the AI Act, deployers of high-risk systems must carry out a Fundamental Rights Impact Assessment (FRIA).

In practice, controllers and deployers should consider conducting the FRIA and the GDPR DPIA within a single, integrated assessment framework. Such a joint analysis helps to avoid unnecessary duplication of efforts, mitigates the risks of inconsistent findings, and promotes a coherent approach to both data protection obligations and the broader fundamental rights safeguards envisaged by the AI Act.

In addition to the FRIA, the AI Act establishes specific provisions on data governance and dataset quality, notably in Article 10, which details key obligations for high-risk AI systems.

2. The scope of Article 10 on data and data governance

Article 10 of the AI Act focuses on the high-risk AI systems[3]. This focus on high-risk systems is reflected in two key aspects. First, Article 10's placement within the Regulation — in Section 2 of Chapter III — underscores its exclusive relevance to high-risk AI systems. It immediately follows the provision on risk management systems, emphasising its role in the overall risk-based approach of the Regulation. It should be noted that the concept of data governance in the AI ecosystem has its roots in the guidelines of the Independent High-Level Expert Group on Artificial Intelligence (HLEG), which identified, among the 7 key requirements that AI systems should meet to be considered trustworthy, "privacy and data governance"[4].

The Expert Group emphasised this requirement by stating that in addition to guaranteeing full respect for privacy and data protection, adequate data governance mechanisms must also be put in place, taking into account data quality and integrity, and guaranteeing legitimate access to data.[5] This means that ensuring compliance with privacy and data protection laws, such as the GDPR, is not sufficient on its own; — organisations must also implement effective data governance mechanisms to guarantee data quality, integrity, and legitimate access within AI systems.

They provide a detailed explanation of how privacy and data governance principles apply to AI systems. They highlight the importance of preventing harm to privacy and emphasise that effective data governance must ensure data quality, integrity, relevance, and lawful access, all of which are critical to maintaining trustworthy AI systems. As stated in the Guidelines: "Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy."[6]

2.1 Privacy and data protection

AI systems must guarantee privacy and data protection throughout a system's entire lifecycle. This covers the personal data initially provided by the user, as well as the information generated about the user over the course of interaction with the system (e.g. outputs that the AI system generates for specific users or how users responded to particular recommendations). Digital records of human behaviour may allow AI systems

[3] Referred to in Art. 6(1) of the AI Act and listed in Annex I.

[4] AI HLEG, *Ethics Guidelines for Trustworthy AI*, 2019, p. 14.

[5] Op. cit. p. 17.

[6] *Ibid.*

to infer not only individuals' preferences, but also sensitive attributes such as their sexual orientation, age, gender, and religious or political views. To foster trust in the data collection process, it is essential to ensure that personal data is not used in ways that result in unlawful or unfair discrimination against individuals.

2.2 Quality and integrity of data

The quality of datasets is paramount to the performance and reliability of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. These issues must be identified and addressed before any dataset is used for training. In addition, the integrity of the data must be ensured. Introducing malicious or manipulated data into an AI system — especially self-learning systems — may alter its behavior in unintended or harmful ways. Processes and data sets used must be tested and documented at each step, such as planning, training, testing and deployment. This should also apply to AI systems that were not developed in-house but acquired elsewhere.

2.3 Data access

Regarding data access, HLEG's Guidelines provide: "In any given organisation that handles individuals' data (whether someone is a user of the system or not), data protocols governing data access should be put in place. These protocols should outline who can access data and under which circumstances. Only duly qualified personnel with the competence and need to access individuals' data should be allowed to do so."[7]

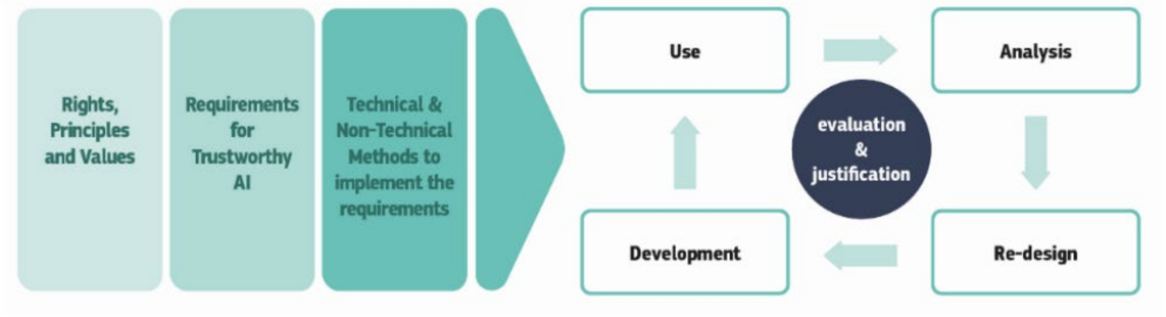
This clearly indicates that particular attention must be paid to privacy and, more specifically, to the protection of personal data as laid down by the GDPR. In this spirit, HLEG sees governance as a way of preventing any breach of privacy. It also governs "the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy."[8]

It can therefore be observed that the notion of data governance supports the protection of privacy. This "pairing" forms the common thread running through Article 10 and highlights the close connection between the GDPR and the AI Act. To implement these key requirements, the working group has developed a continuous process including data governance[9] :

[7] *Ibid.*

[8] *Ibid.*

[9] *Op. cit.*, p. 20.


Rights, Principles and Values
Requirements for Trustworthy AI
Technical & Non-Technical Methods to implement the requirements
Use
Analysis
evaluation & justification
Development
Re-design

3. **Data governance and management**

3.1 Context

Article 10 of the AI Act essentially aims to regulate techniques involving the training of AI models by setting out detailed data management and governance requirements. Recital 67 emphasises the critical importance of high-quality datasets for the performance and safety of high-risk AI systems, particularly when model training techniques are used. It highlights that training, validation, and testing datasets must be relevant, sufficiently representative, as accurate and complete as possible, and free from bias. Recitals, such as Recital 67, also stress the need for appropriate data governance and management practices, especially when personal data is involved, to ensure compliance with the GDPR, including transparency regarding the original purpose of data collection.

As a direct consequence of the AI Act's reference to the GDPR, the obligations applicable to data controllers and processors — particularly those concerning transparency of processing (Articles 13 and 14 of the GDPR) — continue to apply. This information must include, among other elements, the purposes of the processing, the rights of the data subjects, how long the data will be kept and whether the data will be transferred outside the European Economic Area. It is also worth recalling that many of the security requirements embedded in the AI Act find their roots in the GDPR's data protection and security principles.

Article 10 also encompasses a transparency dimension. Recital 67 explicitly requires transparency about the original purpose of the personal data collection.

3.2 Data governance and management practices

Paragraph 2 of Article 10 gives a list of practices in "data governance and management practices appropriate for the intended purpose of the high-risk AI system" that shall concern in particular:

- the relevant design choices

- data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection
- relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation
- the formulation of assumptions, in particular with respect to the information that the data are supposed to measure and represent
- an assessment of the availability, quantity and suitability of the data sets that are needed;
- examination in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations
- appropriate measures to detect, prevent and mitigate possible biases identified according to the examination seen in previous bullet
- the identification of relevant data gaps or shortcomings that prevent compliance with this Regulation, and how those gaps and shortcomings can be addressed.

This is a catalog of measures to be adopted, from design to the detection of biases and gaps in the data. Some of these measures require documentation, documentation that will have to be drawn up.

Checklist for Data Governance, AI Act, Art. 10(2)

Article 10 of the AI Act specifies a set of governance and management practices that must be operationalised for high-risk AI systems. In particular, organisations should:
– verify the provenance and conditions of data collection;
– implement and document measures for bias detection and mitigation;
– ensure transparency of data preparation processes (annotation, labelling, cleaning, enrichment);
– identify and remediate data gaps or shortcomings that may undermine compliance.
These requirements should be approached as a dynamic compliance checklist, to be reviewed and updated continuously across the AI system's lifecycle (design, training, testing, and deployment).

3.3 Data quality

Paragraph 3 of Article 10 of the AI Act states that: "Training, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. Those characteristics of the data sets may be met at the level of individual data sets or at the level of a combination thereof."

It sets out the quality rules for data used for training, validation and testing purposes of the AI system. We detect here an indirect reflection of the data quality principle enshrined in the Article 5.1 of the GDPR, which requires that the data must be “accurate and, where necessary, kept up to date”[10]. Not only are we talking about datasets that are relevant, but they must also be "free of errors and complete with regard to the intended purpose".

In line with the GDPR, and even though this is explicitly mentioned only in Article 10(5) of the AI Act, providers must assess upstream whether the data they intend to use for training, validation, and testing is genuinely necessary, applying the “data protection cascade principle” where anonymous data is preferred, pseudonymised data is used if necessary, and raw personal data is processed only as a last resort.

It should be noted that the use of the words "to the best extent possible" tends to indicate that this is an “obligation of means” rather than an obligation of result since it is not the result that must be promised but, rather, the fact that every effort must be made to achieve the best possible result, namely the absence of errors and relevance of the data. This wording appears to favour market participants, as, in case of liability, the burden of proof generally lies with the party owing the obligation. This is even more generous given that the European Commission has abandoned its draft directive on AI liability and that the AI Act contains no provisions on damage compensation, which is obviously a regrettable omission.

3.4 Contextualisation of data sets

Paragraph 4 of Article 10 of the AI Act states, "Data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, contextual, behavioural or functional setting within which the high-risk AI system is intended to be used."

As Recital 67 of the AI Act is not more explicit regarding this contextualisation requirement, it is difficult for market participants to understand the exact scope of this paragraph. This is particularly unfortunate given that this requirement has an important consequence. Recital 122 specifies that "without prejudice to the use of harmonised standards and common specifications, providers of a high-risk AI system that has been trained and tested on data reflecting the specific geographical, behavioural, contextual or functional setting within which the AI system is intended to be used, should be

[10] Art. 5.1 GDPR.

presumed to comply with the relevant measure provided for under the requirement on data governance set out in [the AI Act]". This presumption is inserted in Article 72(1) of the Regulation, on post-market monitoring by providers of high-risk AI systems, which requires that the monitoring system "shall actively and systematically collect, document and analyse relevant data [...] on the performance of high-risk AI systems throughout their lifetime, and which allow the provider to evaluate the continuous compliance of AI systems with the requirements set out in Chapter III, Section 2 [of the AI Act]." However, the European legislator made clear that AI systems cannot be developed in isolation from the context in which they will be used. This implies that an AI system must use training, validation and test data sets that are relevant to the context in which the system will be used.

3.5 Processing of special categories of data

Paragraph 5 of Article 10 sets out:"To the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems in accordance with paragraph (2), points (f) and (g) of this Article, the providers of such systems may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. In addition to the provisions set out in Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, all the following conditions must be met in order for such processing to occur:

(a) the bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data;

(b) the special categories of personal data are subject to technical limitations on the re-use of the personal data, and state-of-the-art security and privacy-preserving measures, including pseudonymisation;

(c) the special categories of personal data are subject to measures to ensure that the personal data processed are secured, protected, subject to suitable safeguards, including strict controls and documentation of the access, to avoid misuse and ensure that only authorised persons have access to those personal data with appropriate confidentiality obligations;

(d) the special categories of personal data are not to be transmitted, transferred or otherwise accessed by other parties;

(e) the special categories of personal data are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whichever comes first;

(f) the records of processing activities pursuant to Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680 include the reasons why the processing of special categories of personal data was strictly necessary to detect and correct biases, and why that objective could not be achieved by processing other data."

Processing of Special Categories of Data, AI Act, Art. 10[5]

Article 10[5] of the AI Act provides for the *exceptional* processing of special categories of personal data, but only where such processing is demonstrably and strictly necessary for the purposes of bias detection and correction in high-risk AI systems. Recourse to these data is admissible solely when alternative solutions—such as anonymised or synthetic data—are inadequate. In such cases, providers remain bound to implement state-of-the-art safeguards, including pseudonymisation, restricted access controls, and secure deletion measures.

Providers should record and justify why reliance on sensitive data was unavoidable, apply rigorous data-minimisation techniques, and establish clear deletion protocols once the bias has been corrected or the retention period has elapsed.

The AI Act addresses the processing of special categories of personal data within the meaning of the GDPR. It is important to recall that Article 9[1] of the GDPR[11] provides an exhaustive list of personal data categories subject to a general prohibition on processing. These include *"data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health, or data concerning a natural person's sex life or sexual orientation."*[12] However, Article 9[2] of the GDPR allows for this prohibition to be lifted in specific circumstances, as set out in the list of exceptions provided by the Regulation.

Paragraph 5 also affirms the *lex generalis* nature of the GDPR through the use of the words "in addition to the provisions [of the GDPR]".

The minimisation principle imposed by the GDPR is also reaffirmed. Thus, the AI Act confirms the principle that personal data should only be processed if"bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic

[11] GDPR also considers personal data relating to criminal convictions and offences to be special categories of data.

[12] GDPR, article 9.1

or anonymised data".[13] It also requires that only data relevant and necessary for the intended purpose be processed[14].

Data minimisation is crucial, and only data necessary for the intended purpose should be processed. It should be noted that the principle of minimisation also makes possible the reduction of the risks associated with the processing of personal data, and thus improves data protection

Furthermore, and in view of the sensitivity of these data, they can only be used to detect and correct biases. We also note that these two purposes (bias monitoring and detection and bias correction) constitute additional grounds to the exceptions provided for in Article 9(2) of the GDPR to the prohibition on processing referred to in Article 9(1) of the GDPR. This implies, by virtue of the principle of restrictive interpretation of any exception, that any purpose other than the detection and correction of bias is proscribed. This is confirmed by Recital 63, which states that the Regulation "cannot be regarded as constituting a legal basis for the processing of personal data, including special categories of personal data, where applicable, unless expressly provided otherwise [in the AI Act]".

These two purposes also aim, per Recital 70, to "protect the rights of others against discrimination that could result from biases in AI systems". The Recital adds that the processing of special categories of personal data is carried out "exceptionally, to the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems".

In addition to this minimisation aspect, organisational and technical measures must be taken to protect data from misuse. This includes further processing, transmission, transfer or other kinds of access by other parties"[15].

Among these measures, particular attention should be paid to the impact assessment required under **Article 27 of the AI Act**, which is the responsibility of deployers. This assessment complements — and does not replace — the Data Protection Impact Assessment (DPIA) required under **Article 35 of the GDPR**, where applicable.[16]. Note that Article 35 of the GDPR will be applicable in the majority of cases: "a type of processing in particular using new technologies, and taking into account the nature, scope, context

[13] Synthetic data is artificial data created by algorithms to avoid the processing of real data, while anonymised data is data by which the identification of the individual to whom it relates is impossible or requires unreasonable means see RGPD, article 1, 1)).

[14] GDPR, art. 5.1 (c).

[15] AI Act, Article 10(5)(d).

[16] AI Act, Article 27(4).

and purposes of the processing, [which] is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data..."[17]

In practice, it seems appropriate to carry out a joint impact analysis between that required by Article 35 of the GDPR and that provided for in Article 27 of the AI Act in order to avoid either redundancy or document inflation. The data processed for bias detection and correction must be "deleted once the bias has been corrected or the personal data retention period has expired, whichever comes first".[18]

The AI Act also requires that the record of processing activities referred to in Article 30 of the GDPR include “the reasons why the processing of special categories of personal data was strictly necessary to detect and correct biases, and why that objective could not be achieved by processing other data”[19].

Conclusion

The AI Act represents a significant step forward in the regulation of artificial intelligence in Europe. By imposing strict governance and data management requirements, it aims to ensure that AI systems are safe, transparent and accountable. By complying with Article 10, this guarantees a high level of protection of the fundamental rights and freedoms of the data subject. However, the work of AI stakeholders is not straightforward, as the AI Act imposes obligations on them that are in addition to those imposed by other legislation, primarily the GDPR.

But it is only by imposing the data governance rules set out in Article 10 that we can move towards the ethical and responsible use of AI.

We can, of course, lament that the AI Act does not contain any provisions on liability. This is all the moreso given that the proposal for a directive on AI liability has been withdrawn by the Commission.

[17] GDPR, Article 35[1].

[18] AI Act, Article 10[5][e].

[19] AI Act, Article 10[5][f].

Bibliography

- Veale, Michael, & Zuiderveen Borgesius, Frederik, “Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach”, Computer Law Review International, 2021.
- European Commission, Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence, April 2019.
- Kaminski, Margot E., “The Right to Explanation, Explained”, Berkeley Technology Law Journal, 2022, Vol. 37(1), pp. 1–63.

Chapter 7 - Transparency Under the AI Act

Florence Guillaume (Univ. Neuchâtel)

Introduction

Transparency is a fundamental requirement in the regulation of artificial intelligence (AI). It involves understanding how AI systems operate, ensuring that users and regulators have access to relevant information, and holding those involved in the development and deployment of AI systems accountable for the decisions made by these systems.

Following a risk-based approach, the European legislator crafted the AI Act, which imposes legal obligations on operators of high-risk AI systems, as well as on operators of certain types of AI systems that pose particular risks in their interactions with humans.

In this chapter, the question addressed is as follows: 'What does it mean to have a transparent AI system under the AI Act?'

To explore this issue, this chapter will first examine the concept of transparency (1.). Next, an analysis of the transparency requirements for high-risk AI systems will be conducted (2.). Finally, the chapter will discuss the transparency requirements for transparency-risk AI systems (3.).

1. Concept of Transparency

Before delving into the transparency requirements imposed by the AI Act, it is necessary to first revisit the foundations of this concept by exploring how transparency serves as a cornerstone for ensuring human-centric and trustworthy AI (1.1), fostering accountability in AI (1.2), addressing the limitations in explainability or interpretability of AI systems (1.3), and acting as a tool for implementing the risk-based approach outlined in the AI Act (1.4).

1.1 Transparency for Ensuring Human-Centric and Trustworthy AI

The development of AI is expected to improve people's lives, but it also presents significant risks for individuals and society. AI systems can be used to influence decisions in critical areas, such as healthcare (e.g., surgical robots, medical diagnosis), finance (e.g., algorithmic trading, credit scoring), and predictive justice (e.g., risk assessment in criminal matters). Because of their far-reaching impacts, AI systems must be designed in a way that does not violate fundamental rights.

This necessary limitation on the development of AI has evolved from a well-known fictional concept – '*A robot may not injure a human being or, through inaction, allow a* *human being to come to harm*[20] – into a regulatory priority. The first of the Three Laws of Robotics is on the verge of becoming not only an ethical norm but also a legal concept.

Most ethical guidelines, best practice standards, and legal regulations that set out mandatory obligations or non-binding recommendations to minimise the risks associated with AI systems consistently highlight transparency as a primary principle.

For example, transparency is one of the key characteristics of trustworthy AI identified in the 2019 EU Ethics Guidelines for Trustworthy AI[21], as well as in the 2023 U.S. Artificial Intelligence Risk Management Framework[22]. Both reports agree that transparency is crucial for mitigating the risks associated with AI systems, particularly those arising from the opaque nature of their decision-making processes.

By making the decision-making processes of AI systems understandable, transparency serves as a safeguard that ensures that AI promotes human well-being, remains aligned with human values, and does not pose a danger to humanity and self-determination. It is part of a set of core characteristics inherent in trustworthy AI systems. Transparency also helps to build trust in human-AI interactions by enabling the verification of compliance with ethical, social, and legal standards.

1.2 Transparency for Promoting Accountability

Trustworthy AI depends on the ability to verify the compliance of an AI system and to hold someone accountable for outcomes that are incorrect or that otherwise lead to negative impacts. In other words, transparency serves as a tool to ensure that private entities developing or deploying AI systems are held accountable for their activities.

On one hand, transparency facilitates the oversight of AI systems by ensuring access to relevant information. On the other hand, it requires private entities to disclose details about their AI systems, including data sources, the decision-making logic of algorithms, and any potential biases or risks associated with their use. For AI system operators to be held accountable, this information must be made accessible in a transparent way.

[20] The first of Asimov's Three Laws of Robotics, introduced in "Runaround", a short story first published in 1950.

[21] High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy AI" (8 April 2019) Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[22] National Institute for Standards and Technology, Artificial Intelligence Risk Management Framework (January 2023) Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Although this underlying logic is understandable, the transparency requirement still presents practical challenges for private entities, particularly in safeguarding trade secrets.

Determining the appropriate level of transparency requires balancing the interests of AI system operators and regulatory demands. The exact scope of the transparency requirement for a particular AI system depends on the applicable regulatory framework, whether it involves non-binding recommendations outlined in best practice standards or mandatory obligations prescribed in legal regulations.

Note - It is important to clarify that transparency does not require AI systems to be open source. In the context of AI, transparency refers to understandability; a transparent AI system provides clear documentation about its functionality, decision-making process, and associated risks. However, this does not imply that its source code is publicly accessible. In contrast, an open-source AI system explicitly makes its source code available to the public.

The U.S. National Science and Technology Council emphasised in 2016 that transparency and accountability go hand in hand when pursuing fairness and safety in algorithmic systems.[23] The European Parliament also underlined three years later that *'[t]ransparency and accountability are both tools to promote fair algorithmic decisions by providing the foundations for obtaining recourse to meaningful explanation, correction, or ways to ascertain faults that could bring about compensatory processes.'*[24] This shows that transparency is also promoted as a core principle for safeguarding fundamental rights and the responsible development of AI.

Transparency thus not only enhances accountability, but also empowers individuals to make informed decisions and challenge unfair or discriminatory outcomes made by algorithms.

1.3 Transparency for Addressing the Opaque Nature of AI Systems

Transparency is a concept that defies a one-size-fits-all definition.

In the context of AI, transparency can be understood as the effort to make decision-making processes clear and comprehensible by providing stakeholders with relevant

[23] National Science and Technology Council, "Preparing for the Future of Artificial Intelligence" (October 2016), p. 2.

[24] European Parliament, "A governance framework for algorithmic accountability and transparency" (4 April 2019), p. 76. See also *ibid.*, p. 1: "*if it is not known what an organisation is doing, it cannot be held accountable and cannot be regulated.*"

information. This aligns with the definition found in the ISO standards, which describe transparency for AI systems as "*the property of a system [under which] stakeholders receive relevant information about the system. This can include information on items such as system features, limitations, data, system design and design choices*".[25] Notably, this is one of the few existing definitions of transparency specifically related to AI.

Providing a more precise definition is challenging, as the meaning of this concept and the associated requirements vary, depending on the context in which they are applied. Moreover, the term '*transparency*' is understood differently from a legal perspective compared to a technical one. This difficulty in clearly defining the contours of the notion creates the risk of generating a multiplicity of meanings that complicates the implementation of transparency.

While it can be acknowledged that transparency is a general obligation applicable to all AI systems, it remains true that the transparency requirements for a particular AI system will depend on the level of risk it poses to individuals and society, as well as the amount of information needed to assess that risk. As previously discussed, transparency also plays a crucial role in monitoring AI systems throughout their lifecycles and identifying any breaches of safety obligations by AI system operators. Moreover, certain AI systems may benefit from exemptions from transparency obligations, particularly when they serve a public interest purpose, such as national security and the detection or prevention of criminal offenses.

In sum, the concept of transparency in the context of AI is not a single, well-defined concept but rather an umbrella term encompassing several interrelated principles, each contributing to its overall meaning and purpose. These sub-concepts include explainability, traceability, and accountability among others. Together, they form the building blocks of transparency, providing a more comprehensive framework for understanding AI systems. By defining transparency through these sub-concepts, regulatory frameworks can ensure that AI systems are not only transparent but are also aligned with value frameworks like trustworthy and responsible AI, thereby promoting fairness and accountability.

1.4 Transparency for Implementing the Risk-Based Approach in the AI Act

The AI Act does not define transparency but outlines mandatory obligations for operators based on the risk level of their AI systems. This approach is reflected in Recital 27 of the AI Act, which specifies: '*Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as*

[25] ISO/IEC 22989:2022. See also ISO/IEC FDIS 12792.

well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights.'

Transparency is therefore comprised of three main sub-concepts:

- Traceability
- Explainability
- User communication

These three sub-concepts are incorporated into the specific provisions of the AI Act related to the transparency requirements,[26] and these impose obligations on various operators (mainly providers[27] and deployers[28]). The very same sub-concepts were already the building blocks of transparency found in the 2019 *Ethics Guidelines for Trustworthy AI*.[29]

It is interesting to note that the transparency requirement as prescribed in the AI Act primarily focuses on algorithmic transparency. This requirement mainly involves disclosing how algorithms operate, including the logic behind their decision-making processes, the data used to train them, and the potential biases they may carry.

This level of transparency is necessary to ensure fairness and non-discrimination in algorithmic outcomes, which is particularly crucial given the automated decision-making that characterises AI systems. Data transparency receives less emphasis in the AI Act, limiting its focus to informing individuals about data collection and processing for protection of privacy purposes.[30] This aspect is indeed primarily governed by the General Data Protection Regulation (GDPR),[31] which mandates clear and accessible information about data processing activities.

Transparency is one of the main tools prescribed by the EU legislator for implementing the risk-based approach adopted in the AI Act. This requirement indeed plays a crucial role in risk assessment for AI systems. To fully understand the functionality and implications of a particular AI system, it is essential to have insight into its internal workings (e.g. the training data used). This knowledge is necessary for identifying

[26] See *infra* sections 2 and 3.

[27] The definition of 'providers' can be found in Art. 3(3) of the AI Act.

[28] The definition of 'deployers' can be found in Art. 3(4) of the AI Act.

[29] *Op cit.*, note 2.

[30] On data transparency, see in the Guide, *supra*, J -M Van Gyseghem, esp.p. 78.

[31] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), OJ L 119, 4.5.2016, p. 1-88.

potential risks, such as biases or security vulnerabilities, and to assess the severity and likelihood of these risks for users.

The AI Act establishes a set of mandatory rules for AI systems of each risk category (i.e. medium, high and unacceptable) and prescribes appropriate measures to ensure their safety and compliance. As is clearly stated in Recital 26 of the AI Act, '*A risk-based approach [means that it is] necessary to prohibit certain unacceptable AI practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators, and to lay down transparency obligations for certain AI systems*.' In contrast, there are neither mandatory obligations nor non-binding recommendations for AI systems at the base of the risk pyramid (i.e. AI systems deemed to be of minimal or no risk).

Thus, under the AI Act, the transparency requirement primarily applies to two categories of AI systems:

- **first, high-risk AI systems** (e.g., surgical robots, biometric identification, predictive justice), which are subject to stringent transparency requirements to enable the verification of legal compliance by the relevant operators
- **second, 'certain AI systems'** that are subject to specific transparency requirements due to their particular interaction with individuals (e.g., chatbots, generative AI).

This distinction ensures that the level of transparency is proportionate to the risks involved, balancing safety and accountability.

2. Transparency Requirements for High-Risk AI Systems

AI systems falling into the category of '*high-risk AI system*' (2.1) are subject to strict transparency requirements before being placed on the market and throughout their operation. This means that when a high-risk AI system is placed on the market, put into service, or used in the EU,[32] a high level of transparency must be maintained. This entails the following key obligations for operators: explainability (2.2), user awareness (2.3), traceability (2.4), and human oversight (2.5).

2.1. Concept of High-Risk System in Relation to Transparency

The concept of a "*high-risk AI system*" is defined in Article 6 of the AI Act.[33] Systems identified as high-risk should be limited to those that have a significant harmful impact

[32] See Art. 1(2) of the AI Act.

[33] On the concept of "high-risk AI systems" under the AI Act, see this Guide, supra, J. Senechal, p. 24.

on the health, safety, and fundamental rights of individuals, based on the domains and uses cases listed in Annexes I and III of the AI Act.

For AI systems classified as high-risk, transparency is useful not only for risk assessment – ensuring the correct classification of a particular AI system in this category – but also for risk management. Transparency allows verification that a particular AI system maintains an acceptable level of risk in accordance with legal provisions. It also facilitates the monitoring of the implementation of appropriate safety measures and ensures compliance with the mandatory obligations prescribed for the relevant operators.

2.2. Explainability

First, the transparency requirement imposes an obligation of explainability. According to Article 13(1) of the AI Act, "*High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately.*" This means that deployers must be able 'to understand how the AI system works, evaluate its functionality, and comprehend its strengths and limitations.[34]

According to the *Ethics Guidelines for Trustworthy AI*, explainability refers to the ability to clarify both the technical processes of an AI system and the associated human decisions. It requires that AI decisions be understandable to deployers through timely and appropriately-tailored explanations that match their level of expertise. Furthermore, to ensure business model transparency, explanations should be provided on the system's impact on organisational decision-making, design choices, and deployment rationale.[35]

In short, this means that high-risk AI systems must be designed to provide a clear explanation of how the system functions, why it produces a particular output over another, and what the capabilities and limitations of the system are.

The exact type and level of transparency depend on the circumstances but must, in all cases, be appropriate to enable the provider and the deployer to ensure compliance with their obligations under Section 3 of the AI Act.[36] Among these obligations is the requirement that providers implement a quality management system (QMS) that "*shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions.*"[37] This obligation is designed to ensure that high-risk AI

[34] Recital 72 of the AI Act.

[35] *Op cit.*, note 2.

[36] Art. 13(1) of the AI Act.

[37] Art. 17(1) of the AI Act.

systems comply with the stringent requirements related to risk management, transparency, and accountability.

The obligation of explainability aims to enhance stakeholders' understanding of an AI system by ensuring they comprehend its behavior. This prevents blind reliance on AI and enables challenges to AI decisions both from a technical perspective (through technical audits of decisions) and a legal perspective (through the possibility of human review and the ability to appeal against decisions).

Note: It is worth notindoes not explicitly grant the right to challenge AI decisions, including the right to obtain human intervention. Yet, this right is closely connected to the requirements of transparency and explainability.

This link is particularly evident in the OECD's Recommendation on AI (2019), which states in its Article 1.3 that the information provided should enable individuals adversely affected by an AI system to challenge its outcomes. Since explainability aims, among other things, to enable those affected by an AI system to understand its outputs and to therefore assess the lawfulness of how a particular result was generated, one might question whether the AI Act implicitly establishes a right to challenge AI decisions rendered by high-risk AI systems.

This right could be based on Article 86 of the AI Act, which establishes 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.

The Court of Justice of the European Union (CJEU) has been asked to provide a preliminary ruling on whether this provision can serve as a legal basis for subjecting an automated decision to human judicial review in the context of real judicial proceedings. The court will also determine whether a judge has the right to request information in order to understand how the automated decision was made. This decision will help clarify the scope of the right to an explanation of decisions made by a high-risk AI system.

See: ECJ, C-203/22, Dun & Bradstreet Austria GmbH, 27 February 2025, ECLI:EU:C:2025:117.

2.3. User Awareness

Second, the transparency requirement imposes an obligation of user awareness. According to Article 13(2) of the AI Act, 'High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to deployers.'

This means that high-risk AI systems must be accompanied by appropriate information in the form of instructions for use, which must contain a specific set of information listed in Article 13(3) of the AI Act, including the characteristics, capabilities and limitations of performance of the AI system. To improve the legibility and accessibility of the information provided in the instructions for use, illustrative examples should be included where appropriate, particularly regarding the limitations and the intended and prohibited uses of the AI system.[38] Furthermore, providers must ensure that all documentation, including the instructions for use, provides clear, comprehensive, accessible, and easily understandable information, tailored to the needs and expected knowledge of the target deployers.[39]

Note: It seems to follow from the AI Act that the user awareness requirement, which complements the explainability requirement, establishes a specific obligation within the broader general transparency obligation. The obligation of user awareness is important because it ensures that deployers are fully informed before integrating an AI system into their workflows.

2.4. Traceability

Traceability, which is the third obligation set out under the transparency requirements, is provided for in Article 12(1) of the AI Act. Under this provision, '*High-risk AI systems shall technically allow for the automatic recording of events (logs) over the lifetime of the system*.' This refers to a record-keeping obligation or, in other words, the ability to document, monitor, and track AI system operations by means of logs throughout their lifecycle.

Article 12(2) and (3) of the AI Act specify the types of events that must be automatically recorded. This shows that traceability requires comprehensive documentation of data sets, data gathering, data labeling, and algorithms to enhance transparency and enable the identification of errors.[40]

Having clear information on how high-risk AI systems are developed and how they perform throughout their lifecycle ensures traceability and enables the verification of legal compliance.[41] This improves auditability and explainability, helping prevent future mistakes.

[38] Recital 72 of the AI Act.

[39] Ibid.

[40] For training data transparency, see also Art. 53(1)(d) and Recital 107 of the AI Act.

[41] Recital 71 of the AI Act.

2.5. Human Oversight

The fourth obligation associated with the transparency requirement involves identifying appropriate human oversight measures before a high-risk AI system is placed on the market, put into service or used in the EU. According to Article 14(1) of the AI Act, "*High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that can be effectively overseen by natural persons during the period in which they are in use*."

In particular, where appropriate, these measures should ensure that the system operates under built-in constraints that it cannot override on its own and that it remains responsive to the human operator. Additionally, they should guarantee that the individuals responsible for human oversight possess the required competence, training, and authority to effectively perform this role.[42]

Although the requirement for human oversight is not strictly part of the obligations related to transparency, the two requirements are closely interconnected. Transparency establishes the necessary conditions for effective human oversight by enabling individuals to understand how the AI system operates, evaluate its functionality, recognise its strengths and limitations, and verify its compliance with legal standards. The specific requirement for human oversight thus enhances the reliability and accountability of high-risk AI systems.

3. **Transparency Requirements for Transparency-Risk AI Systems**

AI systems falling into the category of "transparency-risk AI system" (3.1) are subject to specific transparency requirements (3.2) before being placed on the market, put into service or used in the EU.[43]

3.1 Concept of Transparency-Risk AI System

The concept of a "transparency-risk system*"* – not provided for as such in the AI Act – includes AI systems designed to interact with natural persons where the AI system poses particular risks of impersonation, manipulation, or deception.[44] These types of AI systems – which are also referred to as "*limited-risk AI systems"* in the literature on the AI Act – may encompass both high-risk and non-high-risk AI systems.

According to Article 50 of the AI Act, this category notably includes AI systems intended to interact directly with natural persons, such as chatbots or virtual personal assistants.[45]

[42] Recital 73 of the AI Act.

[43] See Art. 1(2) of the AI Act.

[44] Recital 132 of the AI Act.

[45] Art. 50(1) of the AI Act.

It also encompasses AI systems, including general-purpose AI systems, which generate synthetic audio, images, video, or text content, such as deep fake generators or text-to-speech applications.[46] Emotion recognition AI systems or biometric categorisation AI systems are also part of this category.[47] For instance, facial expression analysis tools and voice emotion recognition systems are good examples of such AI systems. Finally, AI systems that generate or manipulate image, audio or video content constituting a deep fake are also covered, as well as AI systems that generate or manipulate text intended for publication to inform the public on matters of public interest, such as news-generating bots or automated news summarisers.[48]

For AI systems classified as limited-risk, transparency plays a crucial role in fostering user trust by informing end-users that they are interacting with an AI system. This ethical dimension of transparency enables individuals to understand how AI systems function, including the automated decisions that affect them, or at the very least, makes them aware of the limitations and potential biases of their algorithmic counterparts. This aims to ensure informed consent, empowering individuals to choose whether to accept or refuse the involvement of an AI system or interaction with it.

Furthermore, as with AI systems classified as high-risk, transparency is also useful for both risk assessment and risk management. It allows stakeholders to evaluate potential risks more accurately and ensures that appropriate safety measures are implemented and maintained throughout the system's lifecycle.

3.2 User Communication

When an AI system falls into the foregoing category – i.e. *transparency-risk system* – the AI Act provides for specific transparency obligations on its operators. **If the AI system is also classified as high-risk, these specific requirements are added to those already applicable due to its high-risk status.**

Article 50 of the AI Act essentially establishes a communication obligation, aimed at ensuring that natural persons communicating or interacting with an AI system are aware that they are engaging with AI. In other words, AI systems that interact with humans must clearly disclose their artificial nature.

Note - The communication obligation does not apply to AI systems used for law enforcement or public safety, such as detecting or preventing cyber-attacks. These

[46] Art. 50[2] of the AI Act.

[47] Art. 50[3] of the AI Act.

[48] Art. 50[4] of the AI Act.

systems are exempt from the transparency requirements of Article 50 of the AI Act, allowing their use without disclosing details about their operation.

Conclusion

The AI Act requires operators of high-risk AI systems to comply with strict transparency requirements to ensure safety, fairness, and trustworthiness in AI. **These requirements include explainability, user awareness, traceability, and human oversight.** In addition, AI systems designed to interact with natural persons and that pose particular risks of impersonation, manipulation, or deception are subject to **specific transparency obligations**. These measures are intended to ensure that individuals are clearly informed when they are communicating or interacting with an AI system.

Several practical challenges remain for AI systems to achieve full compliance with transparency requirements. One major challenge is the high cost of transparency, an aspect which could hinder the growth of private entities unable to bear the expenses of implementing the legal transparency measures. Another significant challenge is the "*black box*" effect, where even the developers themselves struggle to explain how and why an AI system made a particular decision. This issue is particularly acute with LLMs (Large Language Models), whose capabilities and behavior are largely unpredictable. Additionally, concerns about trade secrecy pose a barrier to transparency because private entities that invest heavily in developing advanced AI systems are motivated to protect their proprietary AI models. Finally, the technical implementation of transparency poses a significant challenge for private entities, particularly due to the lack of clear guidelines on how to implement legal requirements within AI systems. Technical standards should play a crucial role in that regard, in particular future harmonised standards currently under drafting within CEN-CENELEC JTC21.

Regulating a rapidly evolving technology, whose technical aspects are still difficult to fully grasp, presents a significant legislative challenge. The rapid and continuous advancement of AI systems heightens this complexity further still, and thus necessitates a dynamic and adaptable regulatory framework capable of keeping pace with technological progress. Legislators must strike a delicate balance between encouraging innovation and fostering transparency and accountability in AI.

Bibliography

- European Commission, High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (8 April 2019), https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (accessed 3 March 2025).
- European Parliament, « A governance framework for algorithmic accountability and transparency » (4 April 2019), https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2019)624262 (accessed 3 March 2025).
- Executive Office of the President, National Science and Technology Council (NSTC), Committee on Technology, 'Preparing for the Future of Artificial Intelligence' (October 2016),
- https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (accessed 3 March 2025).
- Florence G'Sell, « An Overview of the European Union Framework Governing Generative AI Models and Systems » (20 May 2024), https://ssrn.com/abstract=4762804 (accessed 3 March 2025).
- Q. Vera Liao and Jennifer Wortman Vaughan, « AI Transparency in the Age of LLMs : A Human-Centered Research Roadmap » (8 August 2023, version 2), arXiv:2306.01941, https://arxiv.org/abs/2306.01941 (accessed 3 March 2025).
- Emmanuel Netter, « La transparence en droit européen du numérique » (2023) 26 *Revue de droit d'Assas* 103.
- U.S. Department of Commerce, National Institute of Standards and Technology, « Artificial Intelligence Risk Management Framework (AI RMF 1.0) » (January 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (accessed 3 March 2025).

Chapter 8 - AI literacy under Article 4 of the AI Act

Nathalie Nevejans (Univ. Artois)[1]

Introduction

The AI Act was adopted on 13 June 2024 and took effect on 1 August 2024, with certain obligations coming into force on 2 February 2025. This is true of the AI training obligation recently introduced by the Act in Article 4 dealing with AI literacy. This provision represents a new challenge for organisations. AI systems providers and deployers shall ensure that all persons responsible for the operation and use of their AI systems have an adequate level of AI knowledge, for example through training. This means that from 2 February 2025, organisations have to implement internal structures and measures to promote the development of skills related to artificial intelligence (AI). However, the contours of the requirements are rather vague. This chapter will look at the legal value of the principle of AI literacy (1.), the measures related to AI literacy (2.) and finally the scope of the measures laid down in Article 4 (3.).

1. Legal Value of the AI Literacy Principle

Article 4 of the AI Act stems from an amendment proposed by the European Parliament during the negotiations on the Regulation.[2] It states that "*[p]roviders 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*". It is essential to analyse the ethical value of AI literacy (1.1.) and then to identify the duty-holder and beneficiaries of AI training measures (1.2.).

[1] This research was supported by the Responsible AI Chair project from the French National Research Agency (ANR-19-CHIA-0008).

[2] Art. 4(b), Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM(2021)0206 – C9-0146/2021 – 2021/0106(COD).

1.1 *AI literacy, an ethical value principle*

Recommendations on AI training are not new in the EU. In 2019, the European Commission's High-Level Expert Group on AI (HLEG) published '*Ethics guidelines for trustworthy AI*'[3], in which education and training are among the recommendations. Indeed, the guidelines state that:

"[c]ommunication, education and training play an important role, both to ensure that knowledge of the potential impact of AI systems is widespread, and to make people aware that they can participate in shaping the societal development. This includes all stakeholders, e.g. those involved in making the products (the designers and developers), the users (companies or individuals) and other impacted groups (those who may not purchase or use an AI system but for whom decisions are made by an AI system, and society at large). Basic AI literacy should be fostered across society. A prerequisite for educating the public is to ensure the proper skills and training of ethicists in this space".

At an early stage in its deliberations, the European Commission therefore recognised the need to educate AI stakeholders (designers and developers), but also affected persons and civil society as a whole, about the potential consequences of AI systems. However, this encouragement was only of ethical value, i.e. non-binding.

Against this background, the question arises arises as to what extent the AI control measures provided for in Article 4 of the AI Act are binding. *A priori*, the wording of Article 4 is obligatory. This is clear from the use of the word '*shall*'. However, other arguments suggest that this provision is in fact more of an ethical text and therefore not binding.

The first argument concerns the place of AI literacy in the AI Act. Article 4 can be found in Chapter I under the general provisions. These are therefore positioned before the prohibited practices in Chapter II, the high-risk systems in Chapter III, and the transparency obligations for providers and deployers of specific AI systems in Chapter IV. This position in the AI Act raises two sets of remarks.

It is evident that Article 4 establishes a general principle that exceeds the scope of the risk pyramid. This principle is regarded as a general guiding principle, implying its applicability to all risk levels, not solely to high-risk categories. Article 4's position within the AI Act suggests that the intention is to address all risks. Consequently, it can be deduced that Article 4 is designed to encompass all AI solutions, rather than a specific

[3] High-Level Expert Group on Artificial Intelligence, "Ethics Guidlines for Trustworthy AI" (8 April 2019) Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

category of risks. However, it is essential to emphasise that the specific requirements of each risk level must be considered, in addition to the general requirements outlined in Article 4.

Note - In the context of high-risk systems, the legislator also alludes to training measures. Consequently, Article 9(5) requires the provider to arrange training for the deployer within the domain of risk management. It is important to note that training is only mandated in instances where eliminating risks or implementing mitigation measures are not feasible, in accordance with standard risk management practices.[4]

Conversely, the strategic position of AI literacy within the AI Act, which precedes the classification of risks, appears to suggest the applicability of Article 4 to AI systems with minimal or no risk. The confirmation of this analysis (for instance, through case law) would imply the dual virtue of this text, namely:

- The establishment of training requirements for low- or zero-risk systems,
- And the provision of a basis for such systems in the event of a claim for compensation in the event of damage.[5]

The second argument pertains to the absence of an administrative penalty in Article 4. This text, akin to the remainder of the AI Act, relates to the conformity of AI. Nevertheless, Article 99(4) on 'Penalties' does not refer to Article 4. Consequently, it can be deduced that the AI Act does not provide for an administrative fine,[6] which is explained by the general nature of the obligation.

It can be concluded that AI literacy is more of an ethical principle encouraging action than a real obligation. This is most likely due to the European Parliament's wish to assign AI ethics a prominent place and to include low-risk systems in this respect.[7] In order to

[4] For example, the Machinery Directive also provides – in Annex III on the essential health and safety requirements relating to the design and construction of machinery or related products – a series of three obligations intended to avoid risks, the last level corresponding to users' information on residual risks, due to the incomplete effectiveness of the protective measures adopted, and to the required training. See Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast) (Text with EEA relevance) OJ L 157, 9.6.2006, p. 24–86

[5] See *infra* section 4.2.

[6] See *infra* section 4.1.

[7] The European Parliament had even listed a series of ethical principles in Article 4a on the 'General principles applicable to all AI systems'. Paragraph 1 of this text thus required all operators falling under the

promote the wider adoption of ethical and trustworthy AI, the EU has adopted a two-pronged soft law approach.

On the one hand, the promotion of AI proficiency is mentioned in Article 95[2][c] of the AI Act relating to *Codes of conduct for the voluntary application of certain requirements*. According thereto,

"The AI Office and the Member States shall facilitate the drawing up of codes of conduct concerning the voluntary application, including by deployers, of specific requirements to all AI systems, on the basis of clear objectives and key performance indicators to measure the achievement of those objectives, including elements such as, but not limited to… promoting AI literacy, in particular that of persons dealing with the development, operation and use of AI".[8]

Recital 165 is explicit in stating that non-high-risk AI systems providers should be encouraged to establish codes of conduct, accompanied by appropriate governance mechanisms, to promote the voluntary application of all or part of the mandatory requirements applicable to high-risk AI systems. Recital 20 also states that *"[i]n cooperation with the relevant stakeholders, the Commission and the Member States should facilitate the drawing up of voluntary codes of conduct* to advance AI literacy *among persons dealing with the development, operation and use of AI*".[9]

On the other hand, Recital 165 emphasises that "Providers and, as appropriate, deployers of all AI systems, high-risk or not, and AI models should also be encouraged to apply, on a voluntary basis, additional requirements related, for example, to the elements of the Union's Ethics Guidelines for Trustworthy AI, environmental sustainability, AI literacy measures…".[10]

2. Duty-Holders and Beneficiaries of the AI Training Measures

There are two aspects to the question of who is subject to AI training.

Firstly, the list of persons subject to training measures has evolved over time. In the European Parliament's 14 June 2023 first version, Article 4b [1] also referred to the Union and Member States. It stated that:

regulation to make every effort to develop and use AI systems or general-purpose AI systems in accordance with the general principles, which consist of the human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, and finally social and environmental well-being.

[8] Emphasis added.

[9] Emphasis added.

[10] Emphasis added.

“[w]hen implementing this Regulation, the Union and the Member States shall promote measures for the development of a sufficient level of AI literacy, across sectors and taking into account the different needs of groups of providers, deployers and affected persons concerned, including through education and training, skilling and reskilling programmes and while ensuring proper gender and age balance, in view of allowing a democratic control of AI systems”.

However, this reference has completely disappeared in the definitively adopted version, since only the second paragraph (of the 14 June 2023 version) remains, which already stated that “*Providers and deployers of AI systems shall take measures to ensure 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 which the AI systems are to be used*”.

On the other hand, the list of those required to take training measures is fluctuating. Indeed, while Article 4 restrictively identifies the duty-holders of the training obligation as ‘*providers and deployers of AI systems*’, Recital 20 considers that AI proficiency concerns ‘*all relevant actors in the AI value chain*’, while Recital 91 only targets *high-risk AI system deployers*, which “*should ensure that the persons assigned to implement the instructions for use and human oversight as set out in this Regulation have the necessary competence, in particular an adequate level of AI literacy, training and authority to properly fulfil those tasks*”.

In any event, since training measures pertain to all relevant actors in the AI value chain, it is incumbent upon them to acquire the requisite knowledge to ensure appropriate compliance and correct implementation.

As outlined in Article 4, the beneficiaries of the AI training measures are defined as ‘*their staff and other persons dealing with the operation and use of AI systems on their behalf*’. This encompasses both the organisation’s employees and deployers.

It is noteworthy that the legislator has demonstrated a particular focus on the context of employment and worker protection. While, in principle, the AI Act considers AI systems employed in the context of employment, workforce management and access to self-employment to be high risk, Recital 20 expressly pertains to the matter of AI control in a professional capacity. Consequently, when the duty-holders of the measures are the beneficiaries’ employers, “*the wide implementation of AI literacy measures and the introduction of appropriate follow-up actions could contribute to improving working conditions and ultimately sustain the consolidation, and innovation path of trustworthy AI in the Union*”. But it is also important to consider the employment relationship itself. What will be the rights and obligations of employees affected by Article 4 measures? What are the consequences of termination of employment due to lack of AI training?

Will the employer's failure to comply with its training obligation result in an automatic obligation to pay compensation to the employee?

These sensitive issues show that workers' representatives in the company should be fully involved in AI workers' training.

3. Measures Relating To AI Literacy

Training shall be provided on the concepts necessary to make informed decisions about AI systems. Indeed, Article 4 requires providers and deployers to take measures to ensure a sufficient AI literacy level. Article 3(56) defines AI literacy as:

"Skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause".

It is regrettable that the legislator saw fit to give Article 4 the title 'AI literacy', as in the definition in Article 3 (56). AI literacy means the ability to manage the risks presented by AI through skills, knowledge and understanding. Organisations shall therefore train people considering their technical knowledge, experience, education and previous training, as well as the context in which the AI systems will be used and the people or groups that will use them.

Recital 20 in the AI Act is more detailed, as it states that this training shall cover "*the necessary notions to make informed decisions regarding AI systems*", and that:

"[t]hose notions may vary with regard to the relevant context and can include 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, and, in the case of affected persons, the knowledge necessary to understand how decisions taken with the assistance of AI will have an impact on them".

It is therefore considered that training can contribute to the protection of the health, safety and fundamental rights of individuals, who will be able to make informed decisions about AI. For this reason, Article 4 addresses all levels of AI system risk. These training measures therefore constitute a minimum and apply more broadly than for training relating to high-risk systems, for which training only comes into play if the risks cannot be eliminated.[11]

[11] See *supra* section 2.1.

It can thus be concluded that organisations will be considered as having properly trained the AI literacy measures' beneficiaries if the latter have received appropriate training on a regular basis, i.e., throughout the AI system life cycle.

Focus on AI literacy measures - AI literacy measures under the AI Act should be integrated into the continuous training offered by the organisation. The latter will also have to set up an internal structure that can adopt companies' internal AI literacy guidelines, for example.

Beneficiaries must have a clear vision of AI potential, a good understanding of its risks and dangers, and knowledge of how AI works. Training should not only have a technical purpose, but should also include other specialities, particularly legal and ethical, adapted to the applications deployed.

Beneficiaries should also be able to act, which includes knowing the methods and tools to limit human rights risks.

It will also have to expressly determine the procedures to be initiated within the organisation so as to deal with possible damages in the AI context. All measures adopted will take into account AI system functioning, will be adapted to the beneficiary's profile, and will take into consideration the impact on the persons concerned, i.e. the affected persons.

Cf. for a practical and comparative perspective: "AI Literacy living Repository" prepared by the European Commission, available at:

https://digital-strategy.ec.europa.eu/en/library/living-repository-foster-learning-and-exchange-ai-literacy

4. The Scope of Measures In Article 4

The matter of the scope of training measures is addressed from two complementary angles: the effects of the violation of Article 4 with regard to the AI Act (4.1.), and the remedies of the consequences of the breach in terms of the provider's or operator's civil liability (4.2.).

4.1 The effects of the violation of Article 4 with regard to the AI Act

We have already emphasised the absence of an administrative fine enacted by the European legislator in Article 99(4), on '*Sanctions*'.[12] However, one may wonder whether Member States could not introduce their own sanctions. Indeed, Article 99(1) provides that:

[12] Ibid.

"[i]n accordance with the terms and conditions laid down in this Regulation, Member States shall lay down the rules on penalties and other enforcement measures, which may also include warnings and non-monetary measures, applicable to infringements of this Regulation by operators, and shall take all measures necessary to ensure that they are properly and effectively implemented, thereby taking into account the guidelines issued by the Commission pursuant to Article 96. The penalties provided for shall be effective, proportionate and dissuasive. They shall take into account the interests of SMEs, including start-ups, and their economic viability".

The terms of this provision cast doubt on the possibility left to Member States to introduce provisions themselves, so as to formulate a possible sanction in case of violation of Article 4.

If we compare it with the GDPR,[13] we see this text was completely unambiguous, given that Article 83 on the general conditions for imposing administrative fines provided in paragraph 1 that "*Each supervisory authority shall ensure that the imposition of administrative fines pursuant to this Article in respect of infringements of this Regulation referred to in paragraphs 4, 5 and 6 shall in each individual case be effective, proportionate and dissuasive*". The text therefore took care to indicate that the violations sanctioned by administrative fines only concerned the paragraphs that listed the sanctions.

However, there is nothing of the sort in Article 99 of the AI Act. Even if there are common elements in the two texts in that sanctions are only imposed if they are '*effective, proportionate and dissuasive*', the AI Act does not link *'sanctions and other enforcement measures'* to the various violations of the regulation as the GDPR does. **Consequently, this leaves open the possibility for Member States, if not to provide for an administrative fine for the violation of Article 4 (which is considered unlikely), at least to perhaps issue warnings and pursue non-monetary measures.**

One may also ponder whether Article 4 infringement could potentially open up an alternative avenue of redress. Article 85, which pertains to the right to submit a complaint to a market surveillance authority, states that:

"[w]ithout prejudice to other administrative or judicial remedies, any natural or legal person having grounds to consider that there has been an infringement of the provisions of this Regulation may submit complaints to the relevant market surveillance authority".

[13] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), *OJEU* L 119, 4.5.2016, p. 1–88.

This recourse is very open in terms of the quality of action, since it is aimed at '*any natural or legal person*'. It also minimises the relevance in bringing proceedings, since it is sufficient for the person to have '*grounds to consider that there has been an infringement of the provisions*'. As it is not specified what kind of provisions may be infringed, it is conceivable that an infringement of Article 4 could give rise to a complaint to the competent market surveillance authority.

4.2 Remedying the consequences of the breach in terms of the provider's or operator's civil liability

If lack of AI control is proven and causes damage, this failure could constitute a fault justifying an action for liability under Member States' national law. As a result, the sanction of Article 4 would not be implemented in terms of compliance, as the infringement would not lead to an administrative fine – subject to the reservation mentioned above[14] – but to the provider's or deployer's civil liability.

Therefore, in the event of damage caused by AI systems, it might be appropriate for the judge to verify whether the organisation that has implemented training measures has allowed the beneficiaries to receive the required training. Thise AI literacy measure could be taken for all types of risk, including low- or zero-risk AI systems, where the AI system is the cause of the damage.

Conclusion

As of 2 February 2025, organisations have to apply Article 4. They must ensure that these measures' beneficiaries have a level of knowledge about AI that enables them to use AI in an informed manner. Organisations therefore have to implement not only technical measures, but at least legal and ethical training, to ensure responsible and informed use of AI.

However, Article 4 remains a rather ambiguous provision, with an unclear legal status. In this chapter, several clarifications have been proposed to help organisations adapt to this provision in the context of their compliance with the AI Act.

First, its position in the AI Act among the general measures makes it a separate text from the compliance requirements. Second, Article 4 is not accompanied by administrative fines, unless one assumes that the AI Act allows the Member State to impose its own penalties.

[14] See *supra* section 4.1.

Therefore, these elements suggest Article 4 on AI literacy is an ethical principle rather than a binding obligation. However, in the event of damage, nothing can prevent it from serving as the basis for civil liability action in the absence of AI training or in the event of inadequacy of the AI literacy measures proposed.

Bibliography

- High-Level Expert Group on AI, « Ethics guidelines for trustworthy AI », Report, European Commission, 08 April 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
- Veale, Michael, & Zuiderveen Borgesius, Frederik, "Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach", Computer Law Review International, 2021.
- Kaminski, Margot E., "The Right to Explanation, Explained", Berkeley Technology Law Journal, 2022, Vol. 37(1), pp. 1–63.
- Data Protection Working Party (WP29), Guidelines on Data Protection Impact Assessment (DPIA), WP 248 rev.01, October 2017.

IV - COMPLIANCE TOOLS AND PROCESSES

Chapter 9 - European Harmonised Standards under the EU AI Act

Olia Kanevskaia (Utrecht Unic.)

Introduction

This contribution discusses the role of technical standards in the AI Act and, in particular, the links between European and international AI standardisation. Standards harmonise expectations and practices across industries and provide predictability for market participants, thus enabling global trade. At the same time, standards have been essential tools for compliance with European legislation, and their role is becoming even more prominent in the European Digital Single Market.

To understand how and why standards are used in the AI Act, whether compliance with these standards is mandatory and which alternatives exist, one needs to know the legal rationale behind European standardisation and how it relates to the international rules and agreements pertaining to AI standards.

1. What is a technical standard?

According to the International Organization for Standardization (ISO), a standard is a technical document meant for repeated use that has been developed by a designated body in a consensus-based procedure.[1] Furthemore, standards should be based on science and experience and not on political considerations.[2]

There are different elements in this definition:

- Institutional dimension, meaning that a standard should be developed and approved by a body with recognised standardisation activities. The most well-known standards bodies are, at the global level, the ISO, the International Electrotechnical Commission (IEC) and the International Telecommunication Union (ITU) and, at the European level, the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunication Standards Institute (ETSI), jointly referred to as the European Standards Organisations (ESOs). The ESOs have different compositions; while CEN and CENELEC are

[1] ISO Guide 2:2004 Standardization and related activities — General vocabulary, *3.2.*

[2] Ibid, note 1.

comprised of national standards bodies of the EU and EFTA Member States and some neighbouring countries, ETSI also grants membership to private companies. National bodies, however, are often also comprised of companies, which arguably leaves ESOs dominated by private organisations.

However, there is no formal list of standards bodies that are deemed to have recognised standardisation activities, although some attributes of these bodies have been explained under international trade law (see below Section 2). Especially in the field of AI, there are many institutions developing informal rules and standards that are relevant for AI governance, including for instance, the Standards Association of the Institute of Electrical and Electronics Engineers (IEEE-SA).

- Procedural dimension, requiring a standard to be established by consensus of those participating in standards development processes. Each standards body has dedicated rules governing conditions for membership and participation in its standardisation committees. Individual participants collaborate in the technical committees of a standards body, and may represent either their employer or affiliation, or act in their personal capacity, depending on the operational rules (Baron and Kanevskaia, 2023).

While some institutions have voting requirements for adopting standards, in practice, consensus-based decision-making is more common (Bekkers and Lazaj, 2024, on the example of ETSI). The chair of a technical committee where a standard is developed plays a decisive role in establishing whether consensus has been reached.

Unlike laws, standards in principle do not have a binding force, meaning that compliance with them is voluntary. More often than not, standards are used as a tool to demonstrate compliance with legal requirements or, in certain cases, as a substitute for regulation where State-made laws are not available.

The rationale behind standards' voluntary nature is the following: standards are not developed in a democratic process by elected representatives of people, and thus should not be binding. This voluntary nature also legitimises the use of standards as a tool for regulatory compliance, allowing market players to use alternative means to demonstrate adherence to legal requirements.

Note - Standards may create legal effects in certain cases. For instance, in some jurisdictions, technical standards can be referenced in national laws and regulations as the only way to prove compliance with legal requirements; in the Netherlands, for instance, one of the standards that companies have to comply with is in order to pass an energy audit is the NEN-EN-ISO 50001:2018 standard for energy management systems.

Furthermore, even if the legislator allows alternative methods of compliance, these methods may not always be available in practice, leaving industry with no choice but to comply with the relevant standard. To illustrate, NEN 2580, a Dutch national standard for determining surface measurement is referenced in the Dutch building legislation as one of the possible options to comply with the legal requirement, yet in practice it is the only standard used in the Dutch construction industry.

Finally, companies may choose to comply with a particular standard for reputational and convenience reasons. Hence, even if voluntary, standards may still create legal consequences and affect the behaviour of different market actors.

2. Standards as barriers to and enablers of international trade

Standards have been noted to have a positive impact on countries' economic growth. They reduce production costs by enabling economies of scale and scope, promote exports and facilitate market access (Swann et al, 1996, Chen and Novy, 2012). They also bring benefits to consumers, ensuring safety and compatibility of products.

That said, standards can be used as "swords and shields" by creating trade barriers and disguising protectionist measures (e.g., Wirth, 2013). Given the central role of standards in regulating AI systems and technologies, as well as the increasing use of AI in international trade (WTO, 2024), it is important to understand the mechanics of international standardisation.

2.1. WTO legal framework

The World Trade Organization (WTO) is a Geneva-based and States-driven organisation that acts as a global arbiter of international trade. WTO functions on the basis of multilateral agreements signed by its members, which generally prohibit discrimination between how a state treats products and services from other contractual parties vis-à-vis domestic products or products imported from (an)other State(s), unless there are legitimate reasons that justify different treatment. Importantly, the WTO imposes on its members a number of transparency obligations and has a dispute settlement mechanism in place, which offers a forum to seek consultations and adjudication in the matter of trade-related disagreements among States.

The WTO rulebook covers three types of trade measures: 1) tariffs, 2) non-tariff trade barriers, and 3) measures related to the protection of intellectual property. Standards belong to non-tariffs barriers to trade, alongside with technical regulations and conformity assessment procedures.

However, not all trade-related aspects are regulated under the WTO agreements. For instance, despite the ongoing work programme, there is still no WTO agreement on e-commerce. Likewise, the WTO does not explicitly regulate trade in, and with, AI products and systems; rather, different WTO agreements are applicable to the different elements of AI regulation (WTO, 2024; Peng et al, 2021). This adds to the fragmentation of AI regulation that already exists due to the diverging national regulatory approaches (UNESCO, 2024).

2.2. TBT Agreement

Non-tariff barriers are regulated by the WTO under the Technical Barriers to Trade Agreement (TBT), which is binding upon all WTO Members. The TBT Agreement aims to strike a balance between avoiding unnecessary trade obstacles and allowing States to adopt trade restrictive rules in order to protect legitimate objectives,[3] as long as they do not constitute arbitrary or unjustifiable discrimination. Such objectives include the protection of health, safety, and environment, or requirements of national security. This list, however, is non-exhaustive, and a single TBT measure can cover multiple objectives. Importantly, the trade-restrictive measures must be necessary to achieve these objectives, meaning that less trade-restrictive alternatives are not available. [4]

The following provides an overview of the three TBT measures.

2.2.1 Technical regulations

Annex I of the TBT Agreement defines “technical regulations” as mandatory requirements establishing product characteristics or their related processes and production methods.[5] The TBT Agreement stipulates that States should base their technical regulations on the relevant international standards, or part thereof, unless these standards are ineffective or inappropriate to fulfill the legitimate objectives pursued by these measures,[6] for instance due to geographical or technological factors. Accordingly, there is a rebuttable presumption that technical regulations that are based on international standards do not create unnecessary trade barriers.[7] The TBT transparency obligations further require WTO Members to explain deviations of their technical regulations from international standards.[8]

[3] Agreement on Technical Barriers to Trade (adopted 15 April 1994, entered into force 1 January 1995) 1868 UNTS 120, (“TBT Agreement”), Article 2.2.

[4] Appellate Body Report, United States — Measures Concerning the Importation, Marketing and Sale of Tuna and Tuna Products WT/DS381/AB/R (US-Tuna II) (adopted 13 June 2012).

[5] TBT Agreement Annex I.1.

[6] TBT Agreement, Article 2.4.

[7] TBT Agreement, Article 2.5.

[8] TBT Agreement, Article 2.9.

2.2.2 International standards

The TBT Agreement makes an important distinction between technical regulations, which are mandatory, and standards, which are voluntary, requiring that neither be unnecessarily restrictive to trade.[9] The definition of a "standard" in Annex I of the TBT Agreement largely follows that of the ISO, with two exceptions:

1) TBT stipulates that standards are voluntary, while ISO/IEC Guide 2:2004 provides that standards can be both mandatory and voluntary;

2) TBT Agreement requires international standards to be based on consensus, but also covers non-consensus standards, while ISO clearly states that standards are based on consensus.[10]

The Agreement does not provide a definition of "international standards" apart from noting that such standards should be developed by an international body or system whose membership is open to the relevant bodies of all WTO Members,[11] presumably meaning the national standardisation committees. The TBT Agreement also does not provide a list of such international standards bodies. However, the Appellate Body, the highest WTO dispute settlement body, hasdclarified some criteria with which a standards body has to comply to be classified as such.

One of these requirements is that a standards body has to have "recognized" standards activities, meaning that that WTO members participate in its standards development processes and reference standards in their national regulations[12] (the latter requirement being somewhat circular since Members are obliged to base their technical regulations on international standards pursuant to 2.4 TBT). In this regard, an organisation can have "recognized" standardisation activity even if it has developed only one standard.[13]

Another condition for an organisation to be considered a developer of relevant international standards is of a procedural nature, and requires such standards bodies to comply with the six principles of the TBT Committee Decision: transparency, openness, impartiality and consensus, effectiveness and relevance, coherence and the development dimension.[14] These principles have been incorporated in many national

[9] See for instance TBT Agreement, Article 4.2 and Annex III.

[10] TBT Agreement, Annex I.2.

[11] TBT Agreement, TBT, Annex I.4.

[12] US-Tuna II.

[13] Ibid.

[14] TBT Committee, Decision, G/TBT/9 (13 November 2000).

legal frameworks on standardisation, as well as that of the EU,[15] and in many trade agreements.

Whether a standard is an international standard should thus be decided on a case-by-case basis through assessing institutional and procedural conditions of how the standard was developed. In the field of AI, initiatives such as UN AI Advisory Body Recommendations, UNESCO Recommendations on Ethics of AI and standards developed by the ISO/IEC JTC 1/SC 42 would most probably qualify as international standards, since these bodies are generally open to all WTO Members and have recognised standardisation activities. However, initiatives like IEEE AI Ethics Standards, while influential among industry players, are arguably not developed according to institutional and procedural conditions of the TBT Agreement; IEEE membership consists of entities and individuals, rather than national standards bodies or governmental agencies although, again, this will require an in-depth analysis.

Hence, classification of a standard as an "international standard," while not straightforward, is crucial in determining whether countries have a legal obligation to include this standard in their national laws or to explain any deviations from it.

2.2.3 Conformity assessment

Conformity assessment procedures are defined as procedures used to demonstrate compliance with the requirements of technical regulations or standards.[16] Similarly to these two instruments, conformity assessment procedures should not be discriminatory or create unnecessary trade barriers.[17] States likewise have an obligation to use existing relevant international guidelines or recommendations for conformity assessment, unless those are inappropriate,[18] and are required to comply with certain transparency obligations when adopting their national conformity assessment procedures.[19]

3. Standards supporting EU legislation

In the EU, standardisation is an important regulatory instrument that is rooted in product safety. EU standardisation is based in a public-private partnership where standards are used to support legislation and policy, and where there is a clear division of tasks between the European Commission [the legislator] and the ESOs [standards developers].

[15] Regulation 1025/2012/EU on European standardisation [2012] OJ L316/12, Rec.2

[16] TBT Agreement, Annex I.2.

[17] TBT Agreement, Article 5.1.

[18] TBT Agreement, Article 5.4.

[19] TBT Agreement, Articles 5.6 – 5.8.

3.1 New Approach and NLF

The EU standardisation policy, the "New Approach", was adopted in 1980s in response to the slow process of technical standards developed by the European Commission. In this regard, standards are necessary instruments to ensure harmonisation of technical requirements and support the completion of the EU Internal Market, enabling products made in one EU Member States to be legally sold in all EU countries.

The New Approach formula is as follows: the European Commission issues Directives and Regulations specifying essential safety requirements with which products must comply to be legally placed on the EU market and requests the three ESOs to develop *harmonised standards* to be used to demonstrate compliance with these essential requirements.[20] Once harmonised standards are developed, the Commission verifies, often with the help of Harmonised Standards (HAS) consultants, whether they are indeed compliant with the essential requirements and, upon positive assessments, approves these standards and publishes references to them in the EU Official Journal.[21] Market actors compliant with harmonised standards are presumed to comply with the essential requirements of the EU legislation.[22] Nevertheless, and as it currently stands in the EU legislation, harmonised standards are considered voluntary.[23]

The New Approach policy was updated in 2008 with the New Legislative Framework (NLF). This package of measures added the necessary elements of effective conformity assessment and market surveillance but also clarified the use of CE-marking.[24] The main principle, however, remained: the EU legislator developed the law, and ESOs developed voluntary standards.

Focus - The voluntary nature of standards has been challenged by the interpretation of the European Court of Justice (CJEU). In 2016, the CJEU held in the landmark case *James Elliott* that harmonised standards are part of EU law owing to their legal effects.[25]

This finding was repeated by the Court in the subsequent cases,[26] culminating in 2024 with the ruling in *PublicResourceOrg*.[27] In this judgment, the plaintiffs contested that

[20] Note that for AI standards, the request was directed to CEN and CENELEC only.

[21] Regulation 1025/2012, Article 10.

[22] *Ibid*, Rec. 5.

[23] *Ibid*, Article 2.

[24] See more at https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (accessed 3 July 2025).

[25] Case C-613/14 James Elliott Construction Ltd v Irish Asphalt Ltd ECLI:EU:C:2016:821, EU:C:2016:821.

[26] E.g. Anstar, Stichting Rookpreventie and Global Garden.

[27] Case C-588/21 P Public.Resource.Org and Right to Know v European Commission and Others ECLI:EU:C:2024:70.

harmonised standards developed by CEN/CENELEC were under the paywall and not freely accessible to public, while the ESOs argued that making their standards publicly available would breach their copyright and harm their commercial interests, putting into question their institutional survival. The CJEU found on appeal that there is an overriding public interest that justifies the disclosure of the requested harmoniszd standards under EU Regulation 1049/2001.

While CEN/CENELEC were compelled to make the requested standards available as a result of the *PublicResource.org* judgment, the case also posed new and unresolved questions regarding ESOs' copyrights over harmonised standards, their business model and the role of the European Commission in standardisation processes.

3.2 Harmonised standards in the AI Act

The AI Act imposes a number of *ex ante* obligations that providers and deployers of high-risk AI systems and general-purpose AI (GPAI) models have to follow in order for their products to be marketed in the EU. In this regard, Article 40 states that systems and models are presumed to be in conformity with the AI Act's requirements for high-risk AI systems (Chapter III Section 2) and obligations for provides of GPAI models (Chapter V, Sections 2 and 3) if they comply with harmonised standards. This provision follows the "New Approach" formula by leaving the concretisation of legal requirements to ESOs.

The current version of Article 40 was introduced in June 2023, and differed significantly from the initial version of the draft AI Act from April 2021. In particular, the updated Article clarified the requirements for the European Commission's standardisation request to ESOs, i.e., that it should 1) include reporting obligations and documentation of processes on improvement of AI systems' resource performance, 2) be prepared in consultation with the AI Board and relevant stakeholders, and 3) specify that harmonised standards need to be clear and consistent among themselves as well as with harmonised standards developed in other sectors.[28] This provision furthermore highlights ESOs' reporting obligation to the Commission regarding stakeholders' representation.[29]

Another additional obligation that made it into the final text of the AI Act is directed to standardisation participants, which should promote investment and innovation in AI, contribute to strengthening global standards cooperation and ensure balanced representation of interests and effective participation of all relevant stakeholders in AI standardisation.[30] Importantly, standards developers shoud take into account existing international standards on AI that are consistent with EU values, fundamental rights and

[28] AI Act, Article Art 40 (2).

[29] *Ibid* and Article Regulation 1025/2012, Article 24.

[30] AI Act, Article Art 40 (3).

interests.[31] While this to some extent reiterates the TBT obligation to base national technical regulations on the relevant international standards, the AI Act appears to condition the use of such standards on their conformity with EU values and interests.

Given that AI technologies touch upon such issues as fundamental rights and ethics, the inclusion of this provision in the AI Act is not surprising. In a similar vein, the EU's apprehensive approach to AI standards also appears from the fact that the European Commission does not use HAS consultants in its approval processes, and verifies these standards' compliance with the essential requirements by itself.

Furthermore, and unlike the TBT obligations, Article 40[3] seems to be directed not to Member States who develop binding regulations, but to ESOs. This does not take away the fact that harmonised standards should not be not unnecessarily restrictive to trade; however, and arguably, it may be considered as providing more leeway to ESOs who develop voluntary standards rather than mandatory technical regulations.

By comparison, EU national standards bodies have a "standstill obligation" to withdraw national standards that are not in conformity with European harmonised standards.[32] An equivalent obligation does not extent to national or regional standards that deviate from international standards, unless they are used as a basis for technical regulations, in which case it is up to the States whether or not such standards should be applied.

Note - The European Commission has yet another instument in its toolbox to determine the direction of technical standards. Article 41 of the AI Act provides that where the harmonisation request is not accepted by ESOs, or harmonised standards are not delivered in time, insufficiently address fundamental rights concerns or do not comply with the request, the European Commission may develop common specifications to be used instead of harmonised standards.[33] While common specifications are not new in the New Approach legislation, the question remains whether they will be adopted for the AI Act, who will develop them and whether they will guarantee the same level of technical expertise and consensus among all affected stakeholders.

3.3 Interplay between international and European standardisation in AI

European standards bodies have concluded a number of agreements with their international counterparts. For instance, CEN is linked to ISO by the Vienna Agreement,[34]

[31] Ibid.

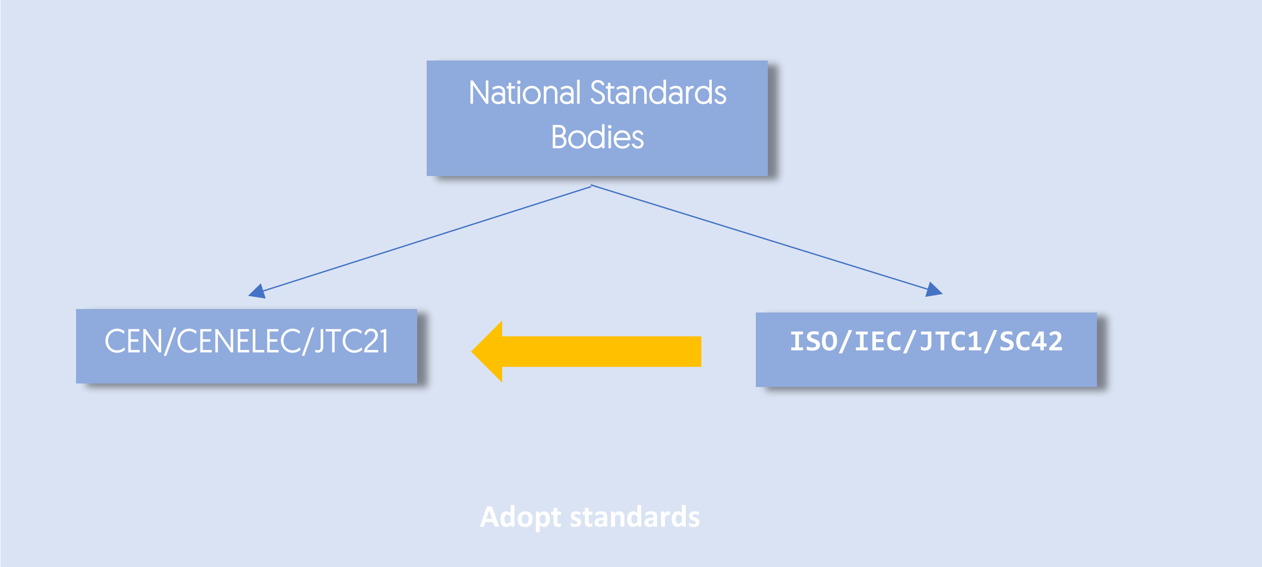
[32] Regulation 1025/2012, Article 3[6].

[33] AI Act, Article Art 40 [1].

[34] Agreement on Technical Cooperation between ISO and CEN [Vienna Agreement] [signed 27 June 1991].

while CENELEC is linked to IEC by the Frankfurt Agreement.[35] These agreements stipulate increasing cooperation but also coherence of international and European standards.

Furthermore, both international and European standards bodies consist of national committees. Hence, an overlap between standards participants, and the content of standards, is almost inevitable.



In principle, and according to WTO rules, the EU should implement international AI standards, i.e., those of ISO and IEC, when they develop their technical regulations, unless these standards are inefficient or inappropriate to achieve the legitimate objectives of the AI Act. In turn, the AI Act has a dual legal basis: the protection of fundamental rights[36] and supporting the internal market.[37] In particular, the former is likely to be used as a legitimate objective to deviate from international standards that, according to the EU, may fall short of protecting fundamental rights.

Example - The EU already deviates from international AI standards. For instance, ISO/IEC 42001:2023 on AI management systems for organisations was deemed unsuitable for managing the risks of developing high-risk AI systems and insufficient in terms of human oversight requirements. As such, it was considered inadequate for protecting fundamental rights within the meaning of the AI Act. Therefore, instead of adopting this

[35] Agreement on Technical Cooperation between IEC and CENELEC [Frankfurt Agreement] (signed 27 October 2016).

[36] Treaty of Functioning of the EU, Article 16.

[37] Ibid, Article 114.

standard as planned, CEN/CENELEC/JTC21 is developing a European standard for AI risk management systems.

Compliance with the TBT obligation to use international standards is also challenging due to the following:

- Firstly, it is unclear which standards should be considered "relevant international standards for AI," given the rise of private standard bodies developing globally recognised AI standards.
- Secondly, it is unclear to what extent the "protection of EU values" can constitute a legitimate objective to deviate from international standards.
- Thirdly, if EU harmonised standards are considered "law", which seems to be the path the CJEU is taking, they risk being qualified as "technical regulations" under the TBT Agreement, strengthening the obligation to base harmonised standards for high-risk AI systems on the relevant international AI standards.

The EU's concern with international standards for AI is not without reason. International standards bodies are open to foreign actors, whose ideas on ethics and fundamental rights protection may differ from those of the EU. At the same time, by deviating from international standards, the EU risks creating unnecesary trade barriers or even being accused of protectionist regulation.

Conclusion

Standards play a crucial role in ensuring compliance of high-risk AI systems and GPAI models with the requirements of the EU AI Act. However, AI standardisation is full of uncertainties, not least around the legal effects of standards, their development processes and the role of common specifications. While the EU's obligations under WTO law require technical regulations to be based on the relevant international standards, there is still a lack of clarity about which standards these are, by which standards bodies they are developed, and when, and to which extent, the EU may deviate from these requirements.

Due to the fundamental rights and ethics concerns triggered by AI, the balance between enabling international trade and safeguarding fundamental national interests is not easy to find. In the near future, some clarifications may be provided by the WTO TBT Committee, where the EU may be invited to discuss its policy on AI standardisation.

Bibliography

- Justus Baron and Olia Kanevskaia, ‘Global Rivalry over Leadership in ICT Standardization : SDO Governance amid Changing Patterns of Participation’ in Panagiotis Delimatsis, Stephanie Bijlmakers and M Konrad Borowicz (eds), *The Evolution of Transnational Rule-Makers through Crises (Cambridge University Press 2023) 287-309*
- Rudi Bekkers and Elona Lazaj, ‘Voting or Consensus ? An Empirical Study of Decision-Making in the European Standards Body ETSI’ (2024) 37(5) *Innovation : The European Journal of Social Science Research* 1336
- Natalie Chen and Dennis Novy, ‘On the Measurement of Trade Costs : Direct vs Indirect Approaches to Quantifying Standards and Technical Regulations’ (2012) 11(3) *World Trade Review* 401
- Panagiotis Delimatsis, ‘“Relevant International Standards” and “Recognised Standardization Bodies” under the TBT’ in Panagiotis Delimatsis (ed), *The Law, Economics and Politics of International Standardization* (Cambridge University Press 2015) 132
- Emmanuelle Ganne, Lauro Locks and Ankai Xu, ‘Trading with Intelligence : How AI Shapes and Is Shaped by International Trade’ (2024) available at https://www.wto.org/english/res_e/publications_e/trading_with_intelligence_e.htm
- Olia Kanevskaia, ‘ICT Standards Bodies and International Trade : What Role for the WTO ?’ (2022) 56(3) *Journal of World Trade*
- Shin-yi Peng, Ching-Fu Lin and Thomas Streinz (eds), *Artificial Intelligence and International Economic Law* (National Tsing Hua University and New York University School of Law (Cambridge University Press, 2021)
- Peter Swann, Paul Temple and Mark Shrumer, ‘Standards and Trade Performance : The UK Experience’ (1996) 106(438) *Economic Journal* 1297
- UNESCO, Consultation Paper on AI Regulation: Emerging Approaches Across the World (2024) CI/DIT/2024/CP/01.
- David A Wirth, ‘The International Organization for Standardization : Private Voluntary Standards as Swords and Shields’ in Geert Van Calster, Denise Prévost and Maria Garcia (eds), *Research Handbook on Environment, Health and the WTO* (Edward Elgar Publishing 2013) 139

Chapter 10 - Guidelines, Delegated Acts and Codes of Practice: A Mix of Practical Tools from the EU Commission and other Authorities

Céline Castet-Renard (Univ. Ottawa)

Introduction

While the AI Act is notably comprehensive - being a regulation rather than a directive and encompassing no fewer than 113 articles and 13 annexes - it still requires supplementation through a wide array of practical tools. These tools operationalise the Regulation's provisions and serve to introduce varying levels of normativity into the regulatory framework, ranging from soft law instruments such as guidelines to hard law instruments such as binding standards and technical specifications, as well as delegated acts.

These tools are rarely used independently. In practice, they usually function together, complementing and reinforcing each other to form a layered and adaptive regulatory ecosystem. This interaction enables flexibility and sometimes includes stakeholders, which is important in a field as dynamic and rapidly evolving as artificial intelligence (AI).

When considering whether the AI Act introduces any true innovations in the realm of norm-setting, one must recognise that the most significant and novel development lies in the process established for the creation of codes of practice. These codes are to be developed iteratively and collaboratively by diverse groups of stakeholders, including industry representatives, AI providers, rights-holders, technical experts, civil society organisations, academia and public authorities. This multi-stakeholder and iterative approach represents a meaningful shift toward participatory norm production, aiming to ensure transparency and that the rules governing AI remain broadly legitimate, and responsive to technological and societal change. This process stands out as a noteworthy normative innovation and warrants close attention as the process is still ongoing.

We will begin by examining the “Delegated Acts” to the European Commission (EC). These acts enable the EC to update or refine certain provisions without the need for a full legislative procedure, ensuring that the regulatory framework remains adaptable and responsive to technological developments. Following this, we will turn our attention to the “Guidelines” issued by the EC. Unlike Delegated Acts, these Guidelines are non-binding and serve primarily as interpretative tools. Their purpose is to provide practical clarification on how the AI Act should be understood and applied, assisting both regulators and market participants in achieving compliance. While they do not create new legal obligations, they play a crucial role in harmonising the implementation of the AI Act across Member States by offering a shared understanding of its requirements.

Finally, we will explore the “Code of Practice” on General Purpose AI Models (GPAIM), a particularly innovative governance mechanism actively promoted and facilitated by the EC. Its voluntary nature allows providers to commit to best practices and shared standards, fostering responsible development and deployment of general-purpose AI models. The collaborative drafting of the Code supported by independent experts reflects a new model of co-regulation, where public authorities and private actors work together with the help of experts to shape norms in a fast-evolving technological landscape. These three kinds of practical tools, including binding delegated acts, interpretative guidelines, and voluntary codes of practice, illustrate the EC’s multi-layered regulatory strategy, which combines hard law, soft law, and co-regulatory approaches to ensure both legal certainty and flexibility in the governance of AI.

1. Delegated Acts: Exercise of delegation by the European Commission

The procedure for exercising a delegation of power is outlined under Article 290 of the Treaty on the Functioning of the European Union[1]. This provision grants the European Commission (EC) the authority to adopt delegated acts as specified in Article 97 of the AI Act. The delegation remains valid for a period of five years and is automatically extended for successive identical periods unless either the European Parliament or the Council objects. Furthermore, the delegation of power can be revoked at any time by either the European Parliament or the Council.

The framework for these delegated acts is detailed in the Interinstitutional Agreement of 13 April 2016 on Better Law-Making, along with the Better Regulation Toolbox 2023[2].

[1] Article 290 TFEU, 1. A legislative act may delegate to the Commission the power to adopt non-legislative acts of general application to supplement or amend certain non-essential elements of the legislative act. The objectives, content, scope and duration of the delegation of power shall be explicitly defined in the legislative acts. The essential elements of an area shall be reserved for the legislative act and accordingly shall not be the subject of a delegation of power. 2. Legislative acts shall explicitly lay down the conditions to which the delegation is subject; these conditions may be as follows: (a) the European Parliament or the Council may decide to revoke the delegation; (b) the delegated act may enter into force only if no objection has been expressed by the European Parliament or the Council within a period set by the legislative act. For the purposes of (a) and (b), the European Parliament shall act by a majority of its component members, and the Council by a qualified majority. 3. The adjective "delegated" shall be inserted in the title of delegated acts.

[2] Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European Commission on Better Law-Making, OJ L 123, 12.5.2016, p. 1–14.

Prior to adopting a delegated act, the EC is required to consult with experts designated by each Member State. Upon adoption of a delegated act, the EC must simultaneously notify the European Parliament and the Council. Any delegated act adopted shall only enter into force if neither the European Parliament nor the Council expresses an objection within a period of three months following the notification of the act, or if both the European Parliament and the Council have informed the EC that they do not object.

The AI Act permits the EC to adopt delegated acts on various topics, such as modifying risk levels and amending annexes. These delegated acts allow the EC to respond quickly to changing conditions or new information. By doing so, the Commission ensures dynamic and effective adaptation of regulations while maintaining transparency and accountability to stakeholders and the public. Indeed, delegated acts are often subject to a consultation process whereby experts, concerned industries, and citizens can contribute before final decisions are made.

Focus - The EC has the authority to adopt delegated acts to modify the risk levels associated with AI systems in several instances. **Modifying risk levels may involve adjusting safety thresholds for certain AI systems or models.** Article 6[6] and Article 6[7] permit the modification of derogations that allow an AI system not to be classified as high-risk if it does not pose a significant risk of harm to health, safety, or fundamental rights, by adding or removing conditions. Another example is given in Article 43[6] which stipulates that high-risk AI systems, as referred to in points 2 to 8 of Annex III [internal control], are subject to the conformity assessment procedure outlined in Annex VII [external control]. Also, Article 51[3] allows for the amendment of thresholds concerning GPAIM classified with systemic risk.

The EC has also the authority to adopt delegated acts to amend annexes. Several examples illustrate this power. Article 7[1] and Article 7[3] allow the European Commission to amend Annex III by adding or modifying use-cases of high-risk AI systems or by removing them. In this example, amending Annex III means updating lists of regulated AI systems based on scientific advances or regulatory needs. Article 11[3] allows the amendment of Annex IV to ensure that the technical documentation provides all the information necessary to assess the compliance of the system. Article 43[5] is about amending Annexes VI and VII by updating them, considering technical progress [internal and external controls]. Article 47[5] enables the EC to amend Annex V by updating the content of the EU declaration of conformity. Article 52[4] allows for the amendment of Annex XIII by specifying and updating the criteria for the designation of GPAIM with systemic risk. Article 53[6] authorises the amendment of Annexes XI and XII in light of evolving technological developments to update the obligations of transparency for the authorities and providers of AI systems who intend to integrate the

GPAIM into their AI systems. Finally, Article 53(5) details the measurement and calculation methodologies to allow for comparable and verifiable documentation included in Annex XI on GPAI models' transparency obligations to the authorities.

Delegated acts constitute a means to ease the procedure of modifying directives or regulations. They are obviously not a new process, but they are particularly suited for framing a technology whose deployment in the market is still nascent. The power is conferred to the EC for a period of five years from 1st August 2024 (AI Act, Article 97(2)). The Commission shall draw up a report in respect of the delegation of power not later than nine months before the end of the five-year period.

According to Article 98(1), the Commission shall be assisted by a committee within the meaning of Regulation (EU) No 182/2011[3]. This Regulation lays down the rules and general principles governing the mechanisms which apply where a legally binding Union act identifies the need for uniform conditions of implementation and requires that the adoption of implementing acts by the Commission be subject to the control of Member States. More precisely, the Commission shall be assisted by a committee composed of representatives of the Member States.

2. Guidelines drafted by the EC

In addition to adopting delegated acts, the European Commission can issue guidelines as another practical tool. Importantly, there is no requirement for a specific provision in the AI Act for the Commission to decide to adopt them.

Note - These guidelines, outlined in Article 96(1) of the AI Act, are not mandatory and serve mainly to facilitate interpretation of the AI Act. They provide legal explanations and practical examples to help stakeholders understand and comply with the AI Act's requirements. They shall take due account of relevant harmonised standards and common specifications, highlighting the interaction between various practical tools.

Howerver, guidelines enacted by the EC do not constitute an authoritative interpretation, which is instead left to the Court of Justice of the European Union (CJEU).

As of 2 February 2025, the first rules under the AI Act have started to apply. These include the AI system definition, AI literacy, and a very limited number of prohibited AI use-cases that pose unacceptable risks in the EU, as outlined in the AI Act. The first two guidelines released by the EC on February 2025 are related to prohibited artificial intelligence

[3] Regulation (EU) No 182/2011 of the European Parliament and of the Council of 16 February 2011 laying down the rules and general principles concerning mechanisms for control by Member States of the Commission's exercise of implementing powers, *OJ L 55, 28.2.2011, p. 13–18.*

practices established by Regulation (EU) 2024/1689 (AI Act)[4] and the definition of an AI system[5].

Regarding the former, it should be borne in mind that the AI Act aims to promote innovation, while ensuring a high level of protection of health, safety and fundamental rights in the Union, including democracy and the rule of law. To this end, it classifies AI systems into four different risk categories, including prohibited, high-risk, transparency risk and minimal to no risk.

The guidelines on prohibited artificial intelligence practices specifically address practices such as harmful manipulation, social scoring, and real-time remote biometric identification, among others. They are designed to ensure the consistent, effective, and uniform application of the AI Act across the European Union. These Guidelines constitute a first interpretation with practical examples of the prohibitions in Article 5 AI Act. The EC will provide additional support to operators and authorities on how to understand the prohibitions and collect further practical use-cases on an ongoing basis with input from providers and deployers of AI systems, the AI Board and other relevant stakeholders[6]. The EC will review these Guidelines as soon as this is necessary in view of practical experience gained in the implementation of the prohibitions and the pace of technological, societal, and regulatory developments in this area. This also includes any relevant experience from market surveillance enforcement actions and interpretations given by the CJEU on the prohibitions and other provisions of the AI Act examined in these Guidelines. During such a review, the Commission may decide to withdraw or amend these Guidelines.

Note - By issuing guidelines on the AI system definition, the Commission aims to assist providers and other relevant persons in determining whether a software system constitutes an AI system to facilitate the effective application of the rules. The AI Act does not apply to all systems, but only to those systems that fulfil the definition of an 'AI system' within the meaning of Article 3(1) AI Act. The definition of an AI system is therefore key to understanding the scope of application of the AI Act. The guidelines on the AI system definition are designed to evolve over time and will be updated as necessary, in particular in light of practical experiences, new questions and use-cases that arise.

[4] Annex C(2025) 884 final, 4.2.2025.

[5] Annex C(2025) 924 final, 6.2.2025.

[6] European Commission, *Guidelines on Prohibited Artificial Intelligence Practices*, para. (433), available at Legalianet.

3. Code of Practice on General Purpose AI Models (GPAIM) (Chapter V – Art. 51 to 56 AI Act): A True Regulatory Innovation

Finally, the Code of Practice on General Purpose AI Models (GPAIM), as laid out in Chapter V of the AI Act, represents a significant regulatory innovation. The definition of General-Purpose AI Models (GPAIM), as delineated in Article 3(63) of the AI Act, refers to AI models, including those trained with a large amount of data using self-supervision at scale, that display significant generality and ability to competently perform a wide array of distinct tasks regardless of the way the model is placed on the market. These models can be integrated into various downstream systems or applications, excluding those utilised for research, development, or prototyping activities prior to their market introduction.

Focus - The obligations of providers of general-purpose AI models are comprehensively outlined. Certain responsibilities are imposed on providers of general-purpose AI models, as stipulated in Article 53(1) of the AI Act. These responsibilities include the following:

- Keep and maintain up-to-date technical documentation (Annex XI) for the oversight of AI Office and National Competent Authorities
- Make information (Annex XII) available to downstream providers who intend to integrate the GPAI model into their AI systems
- Put in place a policy to comply with Union law on copyright and related rights
- Draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office (but with regard for trade secrets and confidentiality).

Working Groups of 13 experts (Chairs and Vice-Chairs), such as WG 1 on Transparency and Copyright, draft the Code of Practice[7]. This provides detailed and practical guidelines for ensuring compliance with the AI Act. According to Article 53(4) of the AI Act, providers of general-purpose AI models may rely on codes of practice to demonstrate compliance with Article 53(1), until a harmonised standard is published. While European harmonised standards presume conformity, adhering to the Code of Practice does not. Providers of general-purpose AI models who do not adhere to an approved code of practice or do not comply with a European harmonised standard shall demonstrate alternative adequate means of compliance for assessment by the Commission.

[7] See https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice

Article 53(5) of the AI Act combines practical tools and empowers the Commission to adopt delegated acts in line with Article 97. The goal is to detail measurement and calculation methodologies for facilitating compliance with Annex XI. On another note, the codes of practice are also related to standards. The AI Office shall, as appropriate, encourage and facilitate the review and adaptation of the codes of practice, in light of emerging standards (AI Act, Art. 56(8)).

Article 56 of the AI Act outlines the code of practice and defines the role of these practical tools. It specifies that the AI Office shall promote and facilitate the creation of codes of practice (AI Act, Art. 56(1)). More specifically, Article 56(2) mandates that both the AI Office and the AI Board use the codes of practice to cover at least the obligations set forth in Articles 53 and 55 of the AI Act, which pertain to the responsibilities of providers of general-purpose AI models, regardless of systemic risks.

Moreover, Article 56(3) of the AI Act states that the AI Office may invite all providers of GPAIM, as well as relevant national competent authorities, to participate in the drawing-up of codes of practice. Civil society organisations, industry, academia and other relevant stakeholders, such as downstream providers and independent experts, may support this process. Consequently, an iterative process for drafting the code of practice has been created through multi-stakeholder engagement. Industry, academia and civil society contributed to the work on codes of practice for general-purpose artificial intelligence[8] by answering a public consultation, as well as a call for expression of interest (during summer 2024) to participate in the Code of Practice Plenary sessions for commenting on the four drafting versions of the code (from October 2024 to April 2025).

The drafting process is based on this open multi-stakeholder consultation, and parallel dedicated "workshops" for providers and Plenary divided into four Working Groups (WG) lead by (Vice-)Chairs experts[9], including: WG1 on Transparency and Copyright-related rules (2 sub-groups), WG2 on Risk identification and assessment measures for systemic risks, WG3 Risk mitigation measures for systemic risks, WG4 Internal risk management and governance for general-purpose AI model providers[10]. Dedicated workshops are also organised with the European Parliament Members and with the

[8] See https://digital-strategy.ec.europa.eu/en/news/industry-academia-and-civil-society-contribute-work-code-practice-general-purpose-artificial

[9] Meet the Chairs leading the development of the first General-Purpose AI Code of Practice: https://digital-strategy.ec.europa.eu/en/news/meet-chairs-leading-development-first-general-purpose-ai-code-practice. The Chairs act pro bono and have demonstrated expertise in relevant areas, ability to fulfil the role (time commitments and operational experience) and independence (no financial interest).

[10] See https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-codepractice#:~:text=The%20Code%20of%20Practice%20will,of%20Practice%20to%20demonstrate%20compliance

Member State representatives in the AI Board Steering Group. Around 1000 stakeholders, including EU Member State representatives and European and international observers, participate in the Code of Practice Working Groups and Provider Workshops, which is a very unique and new way to create practical and voluntary rules.

The third draft of the Code of Practice was released on March 11, 2025[11]. This proposal represents the last one as the Code will be finalised based on stakeholder feedback to it. Compared to the previous two drafts, this version of the Code features a more streamlined structure, with refined commitments and measures. Alongside the third draft, the (Vice-)Chairs have decided to also propose an executive summary and an interactive website[12] to facilitate stakeholders' feedback in the discussions in WGs and dedicated workshops. In addition, the Chairs have also invited civil society organisations and downstream industry to additional workshops to allow for even more targeted interactions at the end of the process.

In parallel to and independently of the Code, the AI Office has taken complementary actions regarding the template for an adequate public summary of the training data envisaged in Article 53(1)(d) of the AI Act. The AI Office outlined a preliminary approach for its possible content and structure[13].

Furthermore, the AI Office is dedicated to ensuring a holistic understanding of the AI Act rules for general-purpose AI, complementing the drawing-up of the Code[14]. Therefore, the AI Office has published guidelines clarifying the scope of the rules[15]. They are related to the definitions of general-purpose AI models, the placing of models on the market and providers, including clarification of responsibilities along the value chain, such as the extent to which rules apply to downstream actors modifying or fine-tuning a general-purpose AI model. The guidance also addresses the exemption for models provided under free and open-source licence, the effects of the AI Act on models placed on the

[11] See https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts

[12] See https://code-of-practice.ai/?section=summary

[13] See Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models, 24 July 2025: https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models

[14] See https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts

[15] See Guidelines for providers of general-purpose AI models, 18 July 2025, https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act

market before August 2025 and other elements of clarification for the future implementation of the rules for general-purpose AI models[16].

Focus - The final code of practice had to be ready at the latest by 2 May 2025[17] as a tool for general-purpose AI model providers to demonstrate compliance with the AI Act, incorporating state-of-the-art practices. It is intended to facilitate the proper application of the AI Act and detail the rules for providers of general-purpose AI models and general-purpose AI models with systemic risks.

These rules will apply 12 months after the entry into force of the AI Act on 1 August 2025. Providers should be able to rely on the Code of Practice to demonstrate compliance.

After the publication of the Code, the AI Office and the AI Board assessed its adequacy and published their assessment[18]. Moreover, Article 56(6) states that the Commission may decide to approve the Code of Practice and give it general validity within the Union by means of an implementing act. In this case, the AI Office may invite all providers of general-purpose AI models to adhere to the codes of practice (AI Act, Art. 56(7)). On the contrary, if, by 2 August 2025, a code of practice cannot be finalised, or if the AI Office deems it is inadequate, the Commission will provide common rules for the implementation of the relevant obligations provided for in Articles 53 and 55 by means of implementing acts (AI Act, Art. 56(8)).

Conclusion

In summary, the AI Act incorporates various practical tools to support and uphold the rule of law. While these tools are presented separately, they are intended to be used together, particularly codes of practice, guidelines, standards, delegated acts, and implementing acts. This combination forms a complex regulatory ecosystem. Therefore, the accumulation of obligations from multiple sources is likely to cause confusion among stakeholders.

The Code of Practice on general-purpose AI models stands out as the most significant tool and the only entirely new mechanism introduced. It involves a wide range of stakeholders in an ambitious participatory process that, while not perfect, represents a

[16] For a general overview, 'Sixth AI Pact webinar on the General-Purpose AI Models and Code of Practice', liableble at: https://www.youtube.com/watch?v=jyGlYo5rE-Y

[17] AI Act, art. 56(9). However, this deadline was not met: the final Code was only published on 10 July 2025, following delays caused by objections raised by certain stakeholders.

[18] See European Commission Opinion on the assessment of the General-Purpose AI Code of Practice and AI Board Adequacy Assessment of the Code of Practice, 1 August 2025, available at: https://digital-strategy.ec.europa.eu/en/library/commission-opinion-assessment-general-purpose-ai-code-practice

step toward greater transparency and accountability for European citizens. For the first time, such a practical code, integrated into an EU Act, has been drafted by independent experts[19] rather than the European Commission. This shift can be attributed to the complexity of the subject matter and the limited resources of the new AI Office. However, it also reflects a transfer of political decision-making from legislators to experts. Therefore, it may serve as a strategic move to mitigate potential backlash—whether from providers of general-purpose AI models who may resist signing the Code or from other stakeholders who might strongly criticise it. Given the challenges involved, some level of opposition is likely inevitable.

Ultimately, while it is too early to assess the effectiveness of these practical tools, they play a crucial role in facilitating compliance with the AI Act and demonstrating adherence to its provisions. Therefore, we strongly recommend following these tools, as they help to interpret and implement the complex requirements of the AI Act.

[19] https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practicepublished-written-independent-experts.

Chapter 11 - Fundamental Rights Impact Assessment (FRIA) under the AI Act

Marion Ho-Dac (Artois Univ.)

& Lamprini Xenou (Paris-Est Créteil Univ.)

Introduction

According to Article 27 of the AI Act, certain deployers[20] of high-risk AI systems[21] "shall perform an assessment of the impact on fundamental rights that the use of such system may produce". This provision reflects *ex ante* regulation based on a preventive approach. Deployers have to assess the potential risks on fundamental rights posed by high-risk AI systems prior to their deployment and, if needed, take relevant actions to prevent (or at least mitigate) those risks. This requirement builds on the FRIA (for "Fundamental Rights Impact Assessment") methodology and aims to prevent harmful AI applications from being put into service on the European market[22].

Focus - FRIA is partly rooted in the Human Rights-Based Approach (HRBA), especially as it evolved in the fields of international development policy and public sector governance. The HRBA led to the creation of Human Rights Impact Assessments (HRIA) as a tool used to predict and evaluate the human rights implications of laws, projects, or policies in key domains such as trade agreements or public health policies. By comparison and more broadly, in the context of technical standardisation, Impact Assessment is a widely used tool, particularly when there are significant societal challenges associated with ethical and fundamental rights considerations.

See, for instance, ISO/IEC 29134:2017 Information technology — Security techniques — Guidelines for privacy impact assessment, adapted from ISO/IEC 29100:2011 and, in AI context, ISO/IEC 42005:2025 Information technology — Artificial intelligence (AI) — AI system impact assessment. Cf. A. Mantelero, "Artificial Intelligence", Elgar Encyclopedia of Human Rights (2022) 163–171.

[20] Pursuant to Article 3 of the AI Act, deployer means a 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.

[21] Except for AI systems in the domain of critical infrastructure pursuant to Annex III, §2 of the AI Act.

[22] Cf. ISO/IEC 42005 Information technology — Artificial intelligence — AI system impact assessment.

FRIA should be understood as a tool for AI design and pre-deployment phase, in line with the "by-design approach" already adopted in the field of data protection[23]. Article 27 is also a perfect illustration of the compliance logic imposed on AI operators and followed at a general level by the AI Act as a product safety regulation. It is based on a mandatory examination that the AI operators concerned must carry out themselves, without the involvement of a third party, and which only requires notification to the national market surveillance authorities concerned[24].

1. General rationale of (Art. 27) FRIA under the AI Act

1.1 Interconnection between FRIA and Risk Management System under the AI Act

Thanks to this provision, the burden of risk management is shared between AI providers and deployers, according to the AI-based risks introduced into society and the respective power of AI operators to manage those risks. The AI Act makes a distinction between, on the one hand, the risk management system that the AI provider must put in place under Article 9 of the AI Act and, on the other hand, the FRIA under Article 27 imposed on certain deployers of high-risk AI systems in Annex III.

Some experts argue that Article 9 also contains a form of FRIA[25]. This is because the risks covered by Article 9 and, more generally, by the AI Act as a whole, in accordance with its risk-based approach, include risks to fundamental rights. In that respect, the identification and treatment of risks pursuant to Article 9 overlap in part with the identification and treatment of risks under Article 27[26].

However, Article 9 goes further in that it also covers risks to health and safety and requires the development of a risk management system, which is itself incorporated into the quality management system for assessing the conformity of the AI systems

[23] See DPIA under Art. 35 GDPR. In the EU jurisdiction, Article 35 of General Data Protection Regulation (GDPR) requires businesses and public organisations under certain conditions to carry out an assessment of the impact on the protection of personal data of their processing of such data (i.e. Data Protection Impact Assessment - DPIA). GDPR specifies that DPIAs are particularly required in cases *of systemic and extensive evaluations of personal data,* and where that evaluation is based on *automated processing*, including profiling, and on which decisions are based that produce legal effects, or similarly, significantly affect the natural person. Cf. ISO 27701 *privacy impact assessment* et EN 17529 *on data protection and privacy by design and by default.*

[24] Art. 27, §3, AI Act.

[25] In that respect, see A. Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template' (2024) 54 *Computer Law & Security Review.*

[26] See Recital 96: "[…] While performing this assessment, the deployer should take into account information relevant to a proper assessment of the impact, including but not limited to the information given by the provider of the high-risk AI system in the instructions for use".

concerned prior to their being placed on the market[27]. The same remark applies as regards GPAI models that pose a systemic risk – including negative effects on fundamental rights[28] – since the providers shall also conduct a risk assessment under Article 55 of the AI Act[29]. Ultimately, it is important that the risk analyses carried out by AI providers and deployers complement each other.

1.2 Main issues stemming from [Art. 27] FRIA implementation under the AI Act

Against this background, Article 27 of the AI Act raises several questions related to its concrete implementation. Since fundamental rights have *per se* a general and abstract nature, how are risks of AI systems for those rights to be assessed in each concrete case? Plus, since AI is a very quickly evolving technology, novel rights as well as new harmful impacts can arise later in time, although they did not exist yet at the moment of the design of the AI system.

First, fundamental rights are often abstract by nature. It is therefore not easy to identify the concrete rights and obligations they may create *in casu*. Hence, it requires a case-by-case analysis which is demanding and may also be burdensome for organisations. The latter will have to determine concretely how their AI-based activities might pose risks to such rights towards individuals or the society at large, and how these rights apply to and potentially impact the deployment of the AI system concerned. Second, AI is a multifaceted and transnational technology that evolves with very high speed. Technical complexity, including opacity of certain AI algorithms, make it difficult for the deployer to assess and detect potential risks to fundamental human rights, in particular when the AI value chain is scattered among several multi-located operators.

Ultimately, those practical challenges in implementing Article 27 will need to be addressed thanks to a robust FRIA methodology offering a readable mapping of potential negative impacts on fundamental rights in the AI context, as well as actionable tools such as metrics, thresholds or balancing tests.

Comparative Perspective - Experience developed in the field of data protection under Article 35 GDPR (i.e. DPIA) could support the FRIA process. Both DPIA and FRIA under Article 27 of the AI Act constitute mandatory obligations to be fulfilled before any innovative solution is implemented in the real world with the objective to prevent or, at least, mitigate risk for individuals and society. The main difference is, however, that DPIA only concerns one

[27] On Article 9 of the AI Act, see *Guide*, A. Favreau, p. 61.

[28] Art. 3 (65) of the AI Act for the definition of systemic risk.

[29] On Article 55 of the AI Act, see in this Guide, G. Bernard, p. 109.

fundamental right, namely the protection of personal data, while Article 27 FRIA targets any fundamental right. This should be a central point of attention for AI deployers.

Article 27 has a general scope which implies to identify any categories of rights which could be affected by the AI system in each scenario. It will especially depend on the type of AI system and on the context of deployment (e.g. education, justice, employment, health services).

Prior to that analysis, the legal source of the fundamental rights to be protected needs to be identify. Unfortunately, Article 27 does not provide for a clear answer. Based on the scope of the AI Act enshrined in the EU jurisdiction and the numerous references made to the EU Charter of Fundamental Rights, the latter instrument should be considered as the legal basis of reference to implement the Article 27 FRIA. By contrast, it is uncertain whether the European Convention on Human Rights and as well as the national constitutions of Member States or any other international Treaty, such as the Universal Declaration of Human Rights (UDHR) together with the International Covenant on Civil and Political Rights (CCPR), and the International Covenant on Economic, Social and Cultural Rights (CESCR), could also be used to conduct the FRIA. This lack of precision is problematic since stakeholders ignore the standard of protection of each fundamental right potentially at stake.

By comparison, the Framework Convention on AI, adopted under the auspices of the Council of Europe[30], is much more accurate on this matter. Under this instrument, States Parties should take measures on AI risk and impact assessment in the context of human rights and fundamental principles of democracy and the rule of law[31]. The Convention expressively refers to the rights and principles enshrined in the respective legal systems of States Parties to the Convention[32].

2. Scope of (Art. 27) FRIA

Article 27, §1, of the AI Act lays down specific criteria on the personal and material scope of the FRIA with a view to limiting its application in order to balance adequate protection of rights with innovation. Therefore, only some AI operators under certain use cases are covered by the FRIA requirement of Article 27.

[30] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Vilnius, 5.IX.2024, CETS 225 (hereafter "Framework Convention on AI").

[31] Art. 16 of the Framework Convention on AI.

[32] See §32 of the Explanatory Report of the Framework Convention on AI, listing the main legal sources of human rights law acquis globally.

2.1. FRIA Material Scope

First, not all high-risk AI systems under the meaning of Article 6 of the AI Act are included in the FRIA scope. Only the AI systems listed in Annex III are covered; *a contrario*, it means that AI systems covered by the Union harmonisation legislation under Annex I of the AI Act are excluded. Moreover, within the eight high-risk domains listed in Annex III of the AI Act, one is excluded, i.e. critical infrastructure. Accordingly, AI systems intended to be used as safety components in the management and operation of critical infrastructure as defined in point 2 of Annex III are not included in the FRIA scope.

2.2. FRIA Personal Scope

Second, two main categories of deployers are included in the scope or Article 27. On the one hand, "deployers that are bodies governed by public law, or are private entities providing public services" shall implement a FRIA in the substantive scenarios described above. This refers to organisations that use the system, either because they are part of the government or administration of an EU Member State, and to private companies doing work for the public sector and in the public interest, such as in the field of education, social services or administration of justice. On the other hand, "deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III" are also in charge of performing a FRIA for the use of such systems. It refers to the domain of essential private/public services, in particular in the field of financial services. Organisations deploying AI systems "to evaluate the creditworthiness of natural persons or establish their credit score"[33] and "for risk assessment and pricing in relation to natural persons in the case of life and health insurance"[34] are tasked with the implementation of a FRIA.

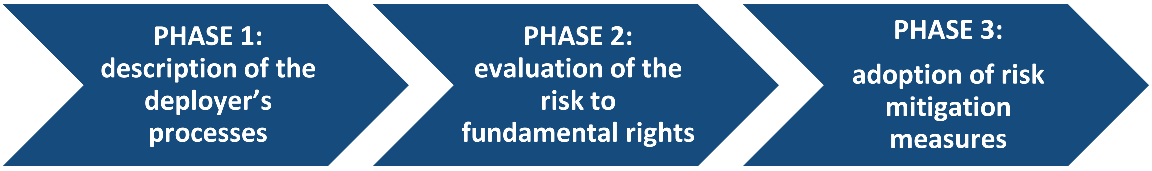
Ultimately, Article 27 mainly targets high-risk AI deployment in the public sector, vis-à-vis European citizens *lato sensu*, with two exceptions in the private sector (i.e. AI-based credit scoring and AI-based insurance pricing).

3. **The procedural structure of (Art. 27) FRIA**

The implementation of the FRIA has three main phases under the AI Act.

[33] Point 5 (b) of Annex III, AI Act.

[34] Point 5 (c) of Annex III, AI Act.


PHASE 1:
description of the deployer's processes
PHASE 2:
evaluation of the risk to fundamental rights
PHASE 3:
adoption of risk mitigation measures

3.1 Phase 1: description of the deployer's processes

The organisation shall provide for a detailed description of the envisioned use[s] of an AI system in the context of deployment and in accordance with the intended purpose of the system – as declared/mentioned by the AI provider, for instance in the notice of use. Moreover, the timing of the system deployment must be specified, i.e. "the period of time within which, and the frequency with which, each high-risk AI system is intended to be used"[35]. This temporal information is indeed crucial at the [next] stage of risk evaluation, in particular regarding the assessment of the level of risk to fundamental rights. The more frequent and prolonged the use of the system, the greater the likelihood that the [potential negative] impact on individual rights will materialise and could be significant.

3.2 Phase 2: evaluation of the risks to fundamental rights

The organisation shall then assess the risks to fundamental rights that might occur at different stages of the AI system's deployment. Article 27 of the AI Act does not provide for a list of fundamental rights that may be affected, nor for a cartography of potential risks to fundamental rights in the AI context. However, it refers to the instructions for use of the AI system [as established by the provider] which should contain informative data on potential risks of the system[36]. In particular, the instruction for use shall include information on "*any known or foreseeable circumstance, related to the use of the high-risk AI system [...] which may lead to risks to the health and safety or fundamental rights* [...]"[37].

Article 27 requests for deployers to identify, first, "the categories of natural persons and groups likely to be affected by [the AI system's] use in the specific context"[38] and, second, "the specific risks of harm likely to have an impact on the [said] categories of natural persons or groups of persons [...]"[39]. In doing so the article ensures that the

[35] Art. 27, §1, b], AI Act.
[36] Art. 13, §3, AI Act.
[37] Art. 13, §3, b] iii, AI Act.
[38] Art. 27, §1, c], AI Act.
[39] Art. 27, §1, d], AI Act.

system's real-world impact on people – especially in terms of fundamental rights – is properly assessed and anticipated. The same AI system may have different effects depending on how and where it is used. Therefore, it is crucial to implement a context-specific FRIA. It should also support appropriate mitigation measures in phase 3 (analysed below). Understanding which groups are at risk and what kind of harm is likely allows appropriate design of safeguards and more targeted human oversight measures.

In general, organisations are advised to take into account several factors, such as:

- Level of opacity of the AI system
- Vulnerability of certain categories of affected persons, like children
- Nature of the right(s) concerned (access to health system, education etc)

In practice, the FRIA model may assign a level of risk – low, medium, high – to each fundamental right that could be affected by the AI system. This level is based on two key factors: likelihood and severity, as suggested by the risk definition under the AI Act. The likelihood of occurrence indicates the probability that the risk to fundamental rights is realied. The severity refers to how serious the harm would be if it occurs. Based on the identified risks, a risk score may be calculated to reflect the overall level of potential harm to each fundamental right. This score combines the likelihood of the risk occurring with the seriousness of its impact. It should help classifying risks across different rights potentially affected by the AI system and taking mitigation measures to ensure the protection of AI subjects and their fundamental rights.

Ultimately, organisations should be able to explain why, despite potential implications, or even negative impact on fundamental rights, the deployment of the AI system is nevertheless a well-balanced choice.

3.3 Phase 3: adoption of risk mitigation measures

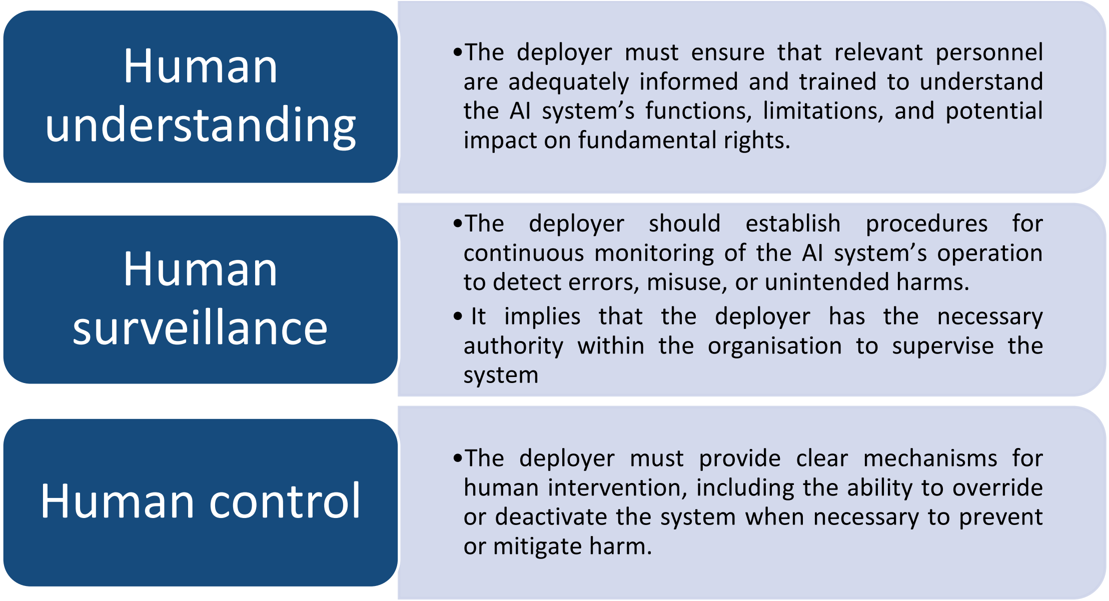
The AI Act mentions various organisational and/or technical measures to mitigate the fundamental rights risks identified.

On the one hand, it refers to human oversight measures, according to the instructions for use, as prepared by the AI provider. Deployers shall provide for a description of the implementation of such measures in order to limit the potential adverse impacts of the AI system[40]. The AI Act adopts a comprehensive and broad conception of human oversight; it consists of three main elements: human understanding of AI system, human

[40] Art. 27, §1, e), AI Act.

surveillance of the system, and human intervention/control in relation to the system[41]. Based on that conception, the deployer must clearly document the design and operational measures that ensure human oversight over the AI system, detailing how the organisation – as deployer –understands, monitors, and can intervene in its operation, with the objective of minimising risks to fundamental rights.

Finally, once the FRIA has been performed, "[…] the deployer shall notify the market surveillance authority of its results" based on the filled-out template to be developed by the AI Office (see below)[42]. This may increase the deployers' level of commitment[43].



Human understanding

- The deployer must ensure that relevant personnel are adequately informed and trained to understand the AI system's functions, limitations, and potential impact on fundamental rights.

Human surveillance

- The deployer should establish procedures for continuous monitoring of the AI system's operation to detect errors, misuse, or unintended harms.
- It implies that the deployer has the necessary authority within the organisation to supervise the system

Human control

- The deployer must provide clear mechanisms for human intervention, including the ability to override or deactivate the system when necessary to prevent or mitigate harm.

On the other hand, the FRIA shall include "measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanism"[44]. The deployer shall specify the remedial measures to be

[41] See M. Ho-Dac and B. Martinez, Human Oversight of Artificial Intelligence and Technical Standardisation (May 31, 2024). Available at SSRN: https://ssrn.com/abstract=5228774

[42] According to Article 27, §3, *in fine*, "in the case referred to in Article 46 §1, deployers may be exempt from that obligation to notify. It targets AI systems which, upon a duly justified request, have been authorised by a market surveillance authority of a Member Stated to be placed on its market, "[…] for exceptional reasons of public security or the protection of life and health of persons, environmental protection or the protection of key industrial and infrastructural assets" (art. 46, §1).

[43] In addition, any natural or legal person that has grounds to consider that there has been an infringement of this AI Act should be entitled to lodge a complaint to the relevant market surveillance authority (Art. 85, AI Act).

[44] Art. 27, §1, f), AI Act.

implemented in the event that the identified risks to fundamental rights materialise, such as notification obligations, corrective technical actions, audit and logging analysis, as well as review and redress procedures. As regards the organisation's internal governance, structural measures could be useful, such as designated bodies responsible for risk response, escalation protocols, and decision-making authority. The establishment of an accessible and effective complaint-handling mechanism is also key in allowing affected persons to seek redress.

Following the risk-based approach underpinning the AI Act, all these measures should be proportionate to the nature and severity of the risks and integrated into the deployer's overall compliance and accountability framework.

4. The formal structure of (Art. 27) FRIA

Article 27 does not provide for clear formal guidance on the implementation of FRIA, creating some uncertainties. However, it mentions that the AI Office shall publish a questionnaire to support AI deployers.

4.1 Uncertainties on FRIA formal structure

Three types of uncertainties arise from Article 27 as regards the formal nature of FRIA, its sanction in case of non-compliance by operators, and the place given to multi-stakeholder participation during the FRIA procedure.

Formal nature of FRIA. Reading Article 27, it seems that the FRIA model adopted by the AI Act is based on a checklist to be completed by AI deployers. A procedural approach is indeed necessary, because of the need to identify, assess and mitigate potential risks. However, when defining the FRIA methodology, the importance of the contextual dimension – of AI deployment – is central for the analysis and must be carefully considered. In this respect, a central question concerns the need to use an AI-based system rather than alternative possible solutions[45].

Therefore, it could be recommended to conduct the FRIA as follows: given the diversity of AI applications, the context of use and the rights and rightsholders potentially affected, a personalised approach is required, where experts design the most

[45] Cf. See UN Guidelines on human rights impact assessments of economic reforms adopted in 2018 by the United Nations: "17.1 An *ex-ante* human rights impact assessment is a structured process to review alternative policy options and analyse the impacts of proposed measures on human Rights".

appropriate and flexible model based, adapt the model to the specific use case and perform a detailed/contextual scenario-based analysis[46].

Sanction for non-respect of Article 27. The Act does not introduce specific administrative fines in the event of failure to comply with the obligation to carry out a FRIA by deployers[47]. It leaves it to the Member States to establish them in accordance with Article 99. This may weaken the effective protection of fundamental rights, as the sanction framework may be scattered within the EU, some States being more indulgent than others.

Limited participation of the stakeholders. Recital 96 of the AI Act states:

"Where appropriate, to collect relevant information necessary to perform the impact assessment, deployers of high-risk AI system, in particular when AI systems are used in the public sector, could involve relevant stakeholders, including the representatives of groups of persons likely to be affected by the AI system, independent experts, and civil society organisations in conducting such impact assessments and designing measures to be taken in the case of materialisation of the risks".

However, Article 27 of the AI Act does not contain a clear and mandatory participatory approach for the FRIA implementation. Recital 96 only provides for a "possible" involvement [*could involve*] "where appropriate", by deployers of relevant stakeholders, "including the representatives of groups of persons likely to be affected by the AI system, independent experts, and civil society organisations"[48] to collect relevant information necessary to perform the impact assessment.

The main added value of multistakeholder involvement in the FRIA process is certainly that it enables deployers to gather diverse, context-specific insights – particularly from affected communities – that are essential to accurately identifying and assessing potential impacts on certain fundamental rights. This approach could therefore be encouraged, on a voluntary basis, even if it is burdensome, or even costly to implement for the organisations.

[46] For an example of the implemention of FRIA methodology in the field of AI-based re-identifcation systems, see M. Ho-Dac & B. Martinez, Méthodologie du respect des valeurs de l'Union européenne *by design* par les systèmes d'intelligence artificielle – L'exemple des systèmes de réidentification de personnes ou d'objets, *Revue du Droit des Technologies de l'Information,* 2024/2, mars 2025, p. 67-89.

[47] A. Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template', *op. cit.*

[48] Ibid.

4.2 Certainty on FRIA formal structure

Pursuant to Article 27, §5, the AI Office will "[…] develop a template for a questionnaire, including through an automated tool, to facilitate deployers in complying with their obligations under this Article in a simplified manner". The Office will have to be very careful not to reduce the FRIA to an exercise based on a simple questionnaire; it would be a very narrow solution and an unfortunate orientation. It seems that the peculiar competences of the EU Agency of Fundamental Rights (FRA) are currently mobilised to support the AI Office in drafting the FRAI template. This is thus an encouraging move.

The questionnaire should assist deployers in fulfilling certain obligations of the FRIA, particularly in the planning and scoping the three phases (explained above), as well as some aspects of data collection during the assessment phase. However, such an approach cannot fully capture the contextual nature of the FRIA. A high-level of granularity of the questionnaire and all of the process would therefore be necessary.

Two complementary approaches can be mentioned at this formal stage of FRIA implementation.

On the one hand, future technical standards developed by standardisation bodies (in particular CEN-CENELEC JTC 21) could provide relevant structural elements and methodological tools for deployers. This will certainly be the case in the context of future standards on risk management by providers, as the risks of high-risk AI to fundamental rights must be included there and, consequently, identified, assessed and mitigated.

On the other hand, it is worth considering whether it would be appropriate to involve an independent third party in conducting the FRIA. Organisations could therefore opt for this approach in contexts where fundamental rights are particularly sensitive. At the same time, competent bodies should be certified for this purpose, as is the case for third-party conformity assessment for high-risk AI systems prior to their placing on the market.

Conclusion

FRIA under Article 27 constitute an interesting tool of the compliance method to safeguard fundament rights of affected persons on AI context. As a methodology for high-risk AI deployers, FRIA should be continually adapted to the technology, the context of use and the organisation itself deploying the system. It could also be part of future standards or normative deliverables in the standardisation context, developed to operationalise the quality management system, including the conformity assessment and the risk management system, of high-risk AI systems as regards fundamental rights.

Bibliography

- S. Bertaina, I. Biganzoli, R. Desiante, D. Fontanella, N. Inverardi, I. G. Penco, A. Claudio Cosentini, 'Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act', [2025] 56 Computer Law & Security Review
- J.H. Gerards, S. Kulk, A. Berlee, V.E. Breemen, Florianne Peters van Neijenhof, Getting the future right: Artificial intelligence and fundamental rights, Luxembourg, Fundamental Rights Agency (FRA), 2020
- H. Janssen, M. Seng Ah Lee and J. Singh, 'Practical fundamental rights impact assessments' [2022] 30(2) *International Journal of Law and Information Technology*
- Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template' [2024] 54 *Computer Law & Security Review*
- Martinez, M. Ho-Dac, Guide pratique du respect des valeurs de l'Union européenne et systèmes d'intelligence artificielle, *Université d'Artois*. 2024. ⟨hal-04665138⟩ (avalaible online)
- J. Salgado-Criado & C. Fernández-Aller, 'A Wide Human-Rights Approach to Artificial Intelligence Regulation in Europe', IEEE Technology and Society Magazine, vol. 40, n° 2, pp. 55-65, June 2021.
- ISO/IEC AWI 42005 Information technology — Artificial intelligence — AI system impact assessment.
- Fundamental Rights and Algorithms Impact Assessment (FRAIA), 31 July 2021.
- https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms
- The Ethics Certification Program for Autonomous and Intelligent System, 2021(IEEE CertifAIEdTM.

Chapter 12 - Europe's AI Regulatory Sandboxes: Germany, Spain, Norway, And Switzerland - Lessons for the Global South

Gaurav Sharma (International AI Policy and Advocacy Advisor)

Introduction

AI as a general-purpose technology has created shockwaves and its 'blackbox' nature has created fears within governance structures. Regulatory sandboxes are one way of arresting these fears and the EU AI Act has made mandatory[49] the introduction of regulatory sandboxes in every Member State. The AI wave has promoted sandboxes as a trusted approach; for example, the Singapore government monetary authority has introduced a FinTech regulatory sandbox framework[50]. Sandboxes are tools that allow innovators to test and experiment with new and innovative products, services, or processes under a controlled environment with defined timelines. This chapter provides an outlook on AI regulatory sandboxes' definition, how AI sandboxes are being envisioned in Europe, and lessons learnt through public exit reports. Four specific States' sandbox approaches are analysed: Germany, Norway, Spain and Switzerland. Norway and Switzerland are chosen as they are linked with the EU through membership in the Agreement on the European Economic Area (EEA) and are liable to comply with the EU AI Act. It shall also draw on lessons for the global south economies that have an established national AI strategy such as India, Brazil, Kenya, Indonesia. Ultimately, this part of the Guide is also intended to actively promote and pilot the use of AI regulatory sandboxes in social impact sectors, such as agriculture, education and urban development.

1. An AI regulatory sandbox: Why the need?

New age AI systems are a complex set of technologies and the emergence of Generative AI systems (hereinafter GenAI), such as ChatGPT, Gemini and others, has showcased the very large capacities of these systems to process enormously complex

[49] Article 57 of the AI Act.

[50] Monetary Authority of Singapore, FinTech Regulatory Sandbox, available at: https://www.mas.gov.sg/development/fintech/regulatory-sandbox

human datasets (text, audio, video) and their ability to infer knowledge just like in human conversations. The complication lies with regards to GenAI systems running on large language models (LLMs), which are difficult to interpret, explain and audit.

The risks associated with generative AI (GenAI) systems are multifaceted. First, data collected for one task can easily be repurposed for others, complicating the explanation and understanding of how GenAI operates. In addition, data governance remains a major concern, especially in public sector applications where digital trust is vital.

GenAI outputs can also lead to hallucinations and biases, with potentially serious societal impacts. For instance, the COMPAS AI system used for recidivism prediction in Florida led to more frequent inaccurate judgments for black defendants compared to white defendants. Such risks are hard to quantify and often remain hidden until detected, as seen with facial recognition technologies disproportionately affecting minority groups.

In summary, the complex and sometimes opaque risks of GenAI systems require vigilant assessment and management.

The impetus for establishing AI regulatory sandboxes is to promote the secure and trustworthy deployment of generative AI (GenAI) systems, particularly across social sector applications and public sector technologies. AI regulatory sandboxes play a crucial role in minimising data governance risks associated with GenAI, thereby protecting citizens' fundamental rights. By providing a controlled and transparent environment, AI regulatory sandboxes help create a trusted public framework for the responsible integration of GenAI systems in public sector technologies.

2. **The EU AI Act and the AI regulatory sandbox thinking**

In the EU AI Act, Article 3(55) defines AI regulatory sandboxes as “a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision”.

Note - Regulatory sandboxes in the EU assume two roles:

1) they foster business learning, i.e., the development and testing of innovations in a real-world environment;
2) they support regulatory learning, i.e., the formulation of experimental legal regimes to guide and support businesses in their innovative activities under the supervision of a regulatory authority[51].

[51] *Ibid.*

The legal regime of AI regulatory sandboxes is defined in the AI Act (Chapter IV, Articles 57 and 58). The EU adopts a nuanced approach to regulation and innovation in the emerging technology sector. Regulatory sandboxes represent a promising first step for testing solutions and frameworks that enable improved data access, foster cross-border technology exchange, and ultimately promote innovation.

The EU's robust sectoral regulatory infrastructure—including bodies such as the European Medicines Agency, the European Banking Authority (for the banking and financial services sector), as well as authorities in transport and energy—provides strong support for the creation and oversight of AI regulatory sandboxes. Additionally, the EU Council's conclusions on regulatory sandboxes and experimentation clauses demonstrate a strategic commitment to future-proofing regulation and responsibly addressing the disruptive challenges and opportunities posed by AI systems[52].The EU AI Act indeed highlights the development of AI regulatory sandboxes at both national and cross-border levels as a central mechanism for fostering the secure and trustworthy deployment of AI systems, particularly including safeguards around bias, transparency and social inclusion.

These sandboxes are designed as controlled frameworks—set up by national competent authorities—where AI solutions can be developed, trained, validated, and tested, including "in real world conditions," under regulatory supervision and based on a sandbox plan for a limited period. AI regulatory sandboxes are broadly envisioned in Articles 57 and 58 of the EU AI Act, which provide structured environments to foster AI innovation by enabling controlled experimentation and testing of novel AI technologies, products, and services during their development phase, prior to market placement. Within these sandboxes, the use of lawfully collected personal data is permitted in the public interest, subject to specific terms and conditions that ensure adequate protection of fundamental rights.

The EU AI Act reflects support for innovation through testing of AI applications in real world conditions with supervision and guidance by the competent national authorities such as the European Data Protection Supervisor[53]. The EU wants AI regulatory sandboxes to act as algorithmic testing environments for AI-based solutions in closed environments where the risks and challenges of the AI systems can be tested, vetted, and verified. Thus, AI sandboxes are viewed as secure and monitored testing sites that allow validation of AI's technological impact and help in further shaping AI regulations

[52] Council of the European Union, "Regulatory sandboxes and experimentation clauses as tools for an innovation-friendly, future-proof and resilient regulatory framework that masters disruptive challenges in the digital age", Conclusions adopted on 16 November 2020, document 13026/20.

[53] https://www.edps.europa.eu/artificial-intelligence/artificial-intelligence-act_en

and improving upon existing regulations. AI regulatory sandboxes are being looked upon to provide evidence for the pros and cons of the algorithmic output, and as a basis for annulling the solution altogether based on the risk.

3. Germany's Approach to AI Sandboxes: Real-world laboratory & experimental clauses

Germany welcomed the general approach documentation of the EU AI Act and, interpreting the statement by the former German Federal Minister Mr. Habeck, the EU AI Act fits well with the German start-up strategy, welcoming the fact that modern and innovation-friendly regulations for real laboratories have been introduced at the European level[54]. Germany released its updated national AI strategy in 2020[55] and emphasised testing in regulatory sandboxes for the transfer of innovation and for further development of legal frameworks to strengthen innovation capacity in AI. Germany's regulatory strategy is driven by the Federal Ministry for Economic Affairs and Climate Action (BMWK) and since 2019, the ministry has set up a regulatory sandbox coordinating office, called *Reallabore* translated to 'real-world laboratories' to implement and progress the Regulatory Sandboxes Strategy[56]. In November 2024, the Federal Cabinet approved the draft law to improve the framework conditions for testing innovations in real-world laboratories and to promote regulatory learning (Real-World Laboratories Act)[57], including AI-based technologies.

Germany's real-world laboratories are based on 'experimentation clauses', defined as 'a legal instrument that creates the necessary space to test innovations in the controlled environment of regulatory sandboxes'[58]. Reallabore is based on experimental clauses that present 'controlled exceptions' to technical legal requirements for testing and prohibitions for emerging technologies, such as AI, with the goal to seek and make testing legal requirements permanent. Germany updated its national AI strategy in

[54] Press Release – Key Technologies on EU Regulation on AI, December 2022, (translated from German to English). https://www.bmwk.de/Redaktion/DE/Pressemitteilungen/2022/12/20221206-zitat bundesminister-robert-habeck-eu-verordnung-zu-kunstlicher-intelligenz.html

[55] Artificial Intelligence Strategy of the German Federal Government, December 2020, https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf

[56] Regulatory Sandboxes – Testing Environments for Innovation and Regulation, http://www.bmwi.de/Redaktion/EN/Dossier/regulatory-sandboxes.html

[57] Real-world laboratories – test rooms for innovation and regulation, https://www.bmwk.de/Redaktion/DE/Dossier/reallabore-testraeume-fuer-innovation-und regulierung.html

[58] New flexibility for innovation. Guide for formulating experimentation clauses, December 2020, https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/guide-new-flexibility-for-innovation-en-web-bf.pdf?__blob=publicationFile&v=2

2020[59] and emphasised testing in regulatory sandboxes for the transfer of innovation and for further development of legal frameworks to strengthen innovation capacity in AI. An expert report, entitled 'Regulatory sandboxes as test spaces for innovation and regulation: Production of a guide for formulating experimentation clauses' and commissioned by BMWi [Federal Ministry for Economic Affairs and Energy], was released in December 2020, presenting a systematic and practice-oriented guide to help lawmakers from different legal fields to develop legally secure and pro-innovation experimentation clauses[60]. The German federal states are also looking at the AI policy at a federal level and at least one document on AI has been produced in each of the states, based on the respective competency sectors[61]. For example, mobility is the focus for the state of Baden-Wurttemberg's AI strategy and the state of Hessen is targeting the finance sector as part of its federal AI policy.

Experimentation clauses have been accepted as an agile and efficient regulatory instrument to balance the risks of innovation with adequate checks and balances and provide enough maneuvering for AI innovation to flourish. Experimentation clauses have been applied to the transport sector for 'autonomous driving' as adequate legislation and institutional capacity is present, thus providing adequate administrative understanding on the subject.

For example, the federal highway research institute [BASt] has identified 24 test fields, and 140 projects related to '*automated and connected driving*' to bring more transparency to relevant research on autonomous driving[62]. One such test field applied is the development of a decentralised data platform on which AI makes decisions for cooperative driving tasks to the vehicles[63]. The German federal states welcome the systemic inclusion of experimentation clauses in legislation to improve the framework for regulatory sandboxes for testing new ideas and solutions on a case-by-case intervention and based on the agility provision in the regulation.

[59] Artificial Intelligence Strategy of the German Federal Government, December 2020, https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf.

[60] Ibid.

[61] Artificial Intelligence made in X: The German AI policy landscape, blogpost by Laura Liebig, Alexander von Humboldt Institute for Internet and Society [HIIG.de], November 2022, https://www.hiig.de/en/german-ai-policy/

[62] Automated and connected driving: Testfeldmonitor makes projects and testfields transparent, November 2021, https://innovation-mobility.com/en/testfields-autonomous-driving-germany-testfeldmonitor/.

[63] Automated formation of rescue lanes in complex scenarios through intelligent networking [AORTA – German abbreviation], https://www.testfeldmonitor.de/Testfeldmonitoring/DE/Suche/Testfeldmonitoring_Detailansicht_Projekt.html;jsessionid=DE09BEDDED7B676A8C39A0CCC51D8BC4.live11294?cms_projektId=228

4. AI regulatory sandbox in Spain: Independent Supervisory Agency, AESIA

Spain kicked off its pilot regulatory sandbox in AI in June 2022 in cooperation with the EU Commission and, by year-end, had opened a call for organisations to participate in the AI sandbox focusing on high-risk AI and general-purpose AI applications across sectors. Spain is the first country in the European Union (EU) to set up a supervisory agency for AI, named the 'Spanish Agency for the Supervision of Artificial Intelligence (*AESIA*)'[64], whose governance is under the 'Ministry of Economic Affairs and Digital Transformation'. A state secretariat for digitalisation and artificial intelligence has also been established[65]. AESIA will supervise the creation, use and commercialisation of AI systems, especially those that might pose a threat to public safety or affect fundamental rights (such as the right to privacy).

Focus - The Spanish Agency for the Supervision of Artificial Intelligence (AESIA) is to act as an independent entity with the veto to sanction the use of potentially harmful AI systems and will be closely linked to the EU AI Act. AESIA will also engage in training, dissemination, and awareness activities for a responsible, sustainable, and reliable use of AI systems with a focus on high-risk AI systems mentioned in the Annex III (but not limited thereto) of the EU AI Act such as biometrics, critical infrastructure, education, law enforcement, vocational training, employment, worker management etc. Spain will create a "*national AI seal*", a type of certificate that accredits that the AI systems deployed in the country meet the requirements demanded by Europe.[66]

The pilot's programmes are looking to operationalise the requirements of the regulations, alongside other features, such as conformity assessments, post-market activities and human oversight[67]. The initiative is expected to create "*easy-to-follow, future-proof best practice guidelines*" [68], alongside an array of other practical explanation guides and materials to assist companies - particularly SMEs and start-ups - in compliance with the Spanish legal framework. The project is funded under the framework of the recovery and resilience facility of the Spanish Recovery plan, as part

[64] What to expect from Europe's first AI oversight agency, February 2023, https://algorithmwatch.org/en/what-to-expect-from-europes-first-ai-oversight-agency/.

[65] Ministry of economic affairs and Digital Transformation - Secretaries of State (mineco.gob.es).

[66] Ibid.

[67] First regulatory sandbox on Artificial Intelligence presented | Shaping Europe's digital future (europa.eu) (accessed on 14-01-2023).

[68] Spanish Government, Regulatory Sandbox pilot programme for AI systems (RD Sandbox), designed to operationalise AI Act requirements—including conformity assessments, post-market activities and human oversight—and to develop practical guidelines and explanatory materials for SMEs and start-ups.

of the Spanish National AI strategy. It is expected that the results of the pilot programme will create guidance on methods to control and monitor compliance that can be used by each individual Member State's national authorities and help in common drafting of documentation, cross-border sharing of lessons learnt and collaboration in post-market monitoring. The Spanish are looking to incorporate a continuous feedback loop for their pilot AI sandbox.

5. Norway and the AI sandbox: Data Protection focus

Norway wants to promote the development and implementation of ethical and responsible AI from a privacy perspective[69]. The Norwegian AI definition in its national AI strategy is, "Artificially intelligent systems perform actions, physical or digital, based on the interpretation and processing of structured or unstructured data, for the purpose of achieving a given goal. Some AI systems can also adapt by analyzing and taking into account how previous actions have affected their surroundings"[70]. Some AI systems can adapt their behavior by analysing how the environment is affected by their previous actions. The Norwegian data protection authority (DPA) is responsible for the AI regulatory sandboxes.

Norway is leveraging AI regulatory sandboxes to thoroughly address issues related to the use of personal data in AI applications. This approach is closely aligned with the requirements of the EU's General Data Protection Regulation (GDPR), ensuring that data protection and privacy remain central throughout AI development and deployment. The Norwegian Data Protection Authority (Datatilsynet) invites all organisations—ranging from startups to large public enterprises—facing challenges with personal data in AI to participate in the AI sandbox program. The sandbox structure is designed to provide benefits on three levels: supporting participating organisations, enhancing the expertise and oversight capacity of the data protection authority, and safeguarding the interests of society at large.

The Norwegian AI sandbox is operating with three main principles for responsible AI: lawfulness, ethics, and robustness. These three principles are based on the "Ethics

[69] Framework for the Norwegian Data Protection Authority regulatory sandbox for artificial intelligence, https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/rammeverk-for-den-regulatoriske-sandkassen/.

[70] National Strategy for Artificial Intelligence, https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf.

guidelines for trustworthy AI[71]" prepared by the European Commission's High-Level Expert Group on AI. The Norwegian DPA is responsible for monitoring compliance with data protection requirements relevant to the EU AI Act. In addition, the DPA holds supervisory authority under various sector-specific Norwegian laws, including the Police Databases Act, the Personal Health Data Filing System Act, the Health Research Act, the Health Records Act, and regulations under the Working Environment Act (notably those addressing video surveillance and email access). In fulfilling its mandate, the DPA provides guidance, issues recommendations, and collaborates with other authorities to address regulatory challenges arising at the intersection of AI and data protection.

Norway's AI sandbox is designed to assist participating organisations in complying with existing data protection regulations, and it explicitly welcomes all projects that address the use of personal data in AI. The Norwegian Data Protection Authority adopts an inclusive approach, deliberately avoiding exclusionary or experimental legal clauses that would require deviations from established data protection laws—a contrast to certain practices in countries like Germany. As a result, the sandbox does not accommodate projects that would necessitate changes or adjustments to existing legal frameworks. Instead, the focus remains on supporting responsible AI innovation strictly within the boundaries of current data protection legislation, ensuring legal certainty and alignment with both Norwegian and broader European data protection standards.

The AI sandbox in Norway has completed two years of operation in its beta phase and has published exit reports for each project, making the findings publicly available. The Norwegian government has now established the "Regulatory Sandbox"—known as "Sandkassa"—as a permanent entity with dedicated funding and a permanent staff, demonstrating a sustained commitment to responsible AI development.

6. AI regulatory sandbox Switzerland: Innovation and Economics

The Swiss approach to AI regulatory sandboxes is pioneered by the Zurich Canton[72]. It was initiated in March 2022 and is known by the name, '*Innovation Sandbox for AI*. The Zurich Canton's Department of Economic Development is leading the AI sandbox strategy with three strategic pillars; to foster innovation, engage society and promote Zurich as a regional AI hub. The sandbox environment is structured to function as a coordinating agency rather than as a supervisory authority. Its primary focus is on facilitating compliance through guidance and coordination, rather than direct

[71] Ethical Guidelines for trustworthy AI, April 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[72] https://www.zh.ch/en/wirtschaft-arbeit/wirtschaftsstandort/innovation-sandbox.html.

enforcement or regulatory oversight. The *Innovation Sandbox* provides a test environment for collaboration among the public administration, research, and the private sector - particularly start-ups and small and medium enterprises (SMEs) - with a goal to promote responsible innovation.

The *Innovation AI sandbox* orientation is aligned with the EU approach and supports responsible innovation for the development and testing of AI systems. The cantonal strategy of Zurich is horizontal in nature to incorporate the wider interests of all sectors and is driven to foster innovation, understand societal needs and promote AI literacy and acceptance, thereby making Zurich a key innovation location for AI in Europe. Currently there is a network of eight different cantons with specific domain requirements in AI. The first phase of the AI *Innovation Sandbox* was completed in March 2024, with findings published in the public domain. The focus is to learn and secure data privacy challenges and provide general guidance to startups and SMEs.

The *Innovation AI Sandbox* involves the various participations of the Zurich Department of Economic Development, Department of Statistics, State Chancellery, ETH AI Center, University of Zurich (Digital Society Initiative), Swiss ICT, and Centre for Information Technology Society and Law (ITSL). The Swiss are looking to use this '*Innovation AI Sandbox'* as a form of regulatory consulting with access to use-case-specific datasets, for a limited timeframe. A distinctive feature of the Swiss AI sandbox approach is the provision of regulatory consulting to companies applying for AI use-case testing. This measure facilitates direct communication between participating companies and Swiss authorities, enabling in-depth discussions about the proposed use case, necessary datasets, and optimal compliance strategies.

For example, regarding the use-case of solving the problem of smart parking in cities, a company named Parquery[73] requested for public camera imagery of cars. During the regulatory consultations, it was negotiated that only the edited images of the cars would be provided, as high-resolution images conflicted with personal data privacy through numberplates and were not needed to implement the intended AI solution. The most critical delivery of the Innovation AI Sandbox has been the "*exit reports*" which are publicly released and are helping to build citizen trust in the use of AI technologies in the public sector.

[73] Parquery, camera based smart parking, https://parquery.com/how-swiss-retail-store-profits-smart-parking-analytics/.

7. Europe’s AI Sandboxes and learnings for the Global South

The AI regulatory sandbox approach being developed in Europe offers a practical, evidence-based framework for fostering AI innovation, shaping policy, and supporting SMEs and startups. This model provides a viable and accessible pathway for organisations to develop and test AI solutions in collaboration with regulators, ultimately facilitating safer and more effective integration of AI into the market. The exit reports and lessons from Europe’s AI sandboxing experiments shall span across sectors and are intended to present an evidence-based use of AI systems.

The AI sandboxing approaches also showcase the need for high-quality knowledge resources, particularly for AI use in high-risk applications.

The Global South economies could take cues from Europe’s AI sandbox approaches and align in accordance with their respective national AI strategies and prioritise sectoral learning in the inclusion of GenAI systems based on need.

The following are three important lessons from Europe’s AI sandbox approaches for the Global South:

Focus on harms of high-risk AI systems: As the Global South faces unique socio-economic challenges such as poverty, hunger and climate change, the lessons from Spain’s sandbox approach could better explain, explore and urge caution regarding the impact of high-risk systems such as biometrics, facial recognition and personal data protection issue. The Global South could benefit by identifying the use of high-risk AI systems in social impact sectors and better align risk management strategies. Lessons on the use and safety of personal datasets, particularly for use by public administration agencies, could be mutually beneficial to the Global South and the EU, as the EU has a digital for development 2030 agenda to promote digitalisation in the Global South.

Use-case induced legislation: The added advantage of Europe’s AI sandboxes for the Global South would lie in customising AI solutions for specific use-cases and testing for socio-economic impact. The lessons from exit reports and completed projects in Europe could provide vital plug-ins in setting up evidence-based use-case mandates of AI systems and best practices for integrating existing sector specific legislations in social sectors such as healthcare and agriculture. For example, most of the data protection exit reports feed directly into the realm of ‘responsible AI’ and can help in better articulation of governance frameworks for setting up a regional or sub-regional AI regulatory authority in the Global South.

Alternatives to hard regulation: Europe’s AI regulatory sandboxes have demonstrated that innovation and regulation can be effectively balanced and tested within existing legal frameworks. Germany’s use of the “Experimental Clause” exemplifies how legislatures can be adapted based on practical evidence and the building of regulatory trust. These experimentation clauses offer an alternative to developing standalone AI regulations by enabling efficient governance through close inter-departmental cooperation and shared oversight of emerging technologies.

Conclusion

Europe's AI regulatory approaches are grounded in the rule of law, the protection of fundamental rights, and strong safeguards for data privacy. These principles form the foundation of the European strategy for fostering digital trust within society. The complexity of high-risk AI systems demands explanation and accountability, and Europe's varied AI regulatory sandboxes are a boon in this direction. Europe's AI regulatory sandboxes showcase the aim of developing use-case-specific sectoral explanations and present good validation and explainability avenues to foster digital trust in the use of high-risk AI systems in public sector technologies. As AI sandboxing in Europe is guided to help innovation and support small and medium scale enterprises (SMEs) and startups, the results are a litmus test closely watched by Global South economies that are much in need of AI-enabled technologies to solve the grand challenges of poverty, hunger, education and balance the responsible and trustworthy use of AI systems with limited and mostly stretched legal and regulatory institutions.

V- COMPLIANCE EVALUATION IN PRACTICE

Chapter 13 - Conformity Assessment Procedure

Guillaume Bernard, LNE Representative

– French Laboratoire national de métrologie et d'essais

Introduction

Along with the creation and publication of the AI Act, the European Standardisation Organisations (ESOs) CEN-CENELEC[1] have been tasked by the European Commission with defining and publishing the harmonised standards that will help clarify how to answer essential requirements of the AI Act. This is especially the case for the provisions contained in Chapter 3, Section 2 (Article 8 to Article 15) on the requirements of high-risk AI systems, where the content of the harmonised standards will be used both by AI providers and by the operators who will ensure the conformity of the systems (e.g. Notified Bodies, AI Safety Institutes, AI Office). This means that, currently, AI providers who want to prepare to obtain the "CE marking"[2] can only use the requirements provided in the articles of the AI Act, which at this stage lack sufficient detail for implementation. However, non-European resources that cover specific parts of the AI Act exist.

1. Existing tools for AI Act conformity

First, several international standards can already be used by AI providers and authorities to assess the conformity of AI systems.

For instance, regarding Article 9 on risk management system, the ISO-IEC 23894 standard (*Information Technology - Artificial Intelligence - Guidance on risk management)* has been published in 2023 and provides categories of risks — including technical, societal, legal, and ethical risks — that can be used by AI providers to identify, assess and mitigate the risks associated with the design, development, deployment and use of their AI systems. Likewise, ISO 42001 (*Information Technology - Artificial Intelligence – Management Systems)*, also published in 2023, provides requirements on the quality management systems of organisations that design, develop, deploy, or use AI systems.

[1] CEN-CENELEC, *European Committees for Standardization*, https://www.cencenelec.eu/. Art. 40 of the AI Act allows the European Commission to mandate CEN and CENELEC to develop harmonised standards.
[2] Art. 48 of the AI Act.

More recently, ISO 5259-5 (*Artificial Intelligence – Data quality for analytics and machine learning*) was published in 2025. It provides a comprehensive framework for managing data quality in the context of analytics and machine learning, helping organisations to ensure that the data used to train, test, and validate AI systems is accurate, relevant, complete, and fit for purpose — a key requirement for building trustworthy and compliant AI. It is not yet known if any of these standards will be adapted and adopted as part of the harmonised standards[3] and there are already some critics of the content of these documents. For instance, ISO/IEC 42001 establishes a general framework for an Artificial Intelligence Management System (AIMS), focusing on overarching AI governance, risk management, and organisational responsibilities. However, it does not set out detailed requirements for the internal quality management system (QMS) used during the development of high-risk AI systems, as explicitly required by Article 17 of the AI Act.

However, even if not adopted as European standards and used for conformity assessment, these standards can still be useful to an AI provider, as they provide technical frameworks and best practices that support the design and development of compliant AI systems.

Besides technical standards, methodologies and processes already exist in the scientific community that can be used to help in elucidating specific articles and requirements of the AI Act. The methodology traditionally used to evaluate the performance of AI systems - established through the early evaluation campaigns led by NIST[4] (National Institute of Standards and Technology) and widely adopted in benchmarking exercises - ensures traceability, transparency, and reproducibility of the evaluation protocol, which aligns with the requirements set out in the AI Act, particularly under its Article 15. Similarly, expertise in data analysis and the ability to assess data quality - such as evaluating representativeness or identifying bias – are helpful in meeting specific requirements under Article 10, including those related to bias detection. As with standards more generally, it remains uncertain whether and how these approaches and methodologies will be incorporated into the harmonised standards. However, it can reasonably be expected that there will be at least some alignment between the harmonised standards and these methodologies, given the expertise of those involved in their development.

[3] For a standard — even an ISO standard — to become a "harmonised standard" under the AI Act, it must be mandated by the European Commission through a formal standardisation request. Such requests are addressed exclusively to CEN, CENELEC and/or ETSI, which are the only three European Standardisation Organisations (ESOs) officially recognised by the EU pursuant to Regulation (EU) No 1025/2012.

[4] National Institute of Standards and Technology (NIST), *Face Recognition Vendor Test (FRVT)*, ongoing since 2000, available at: https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt.

In addition to international standards and existing methodological frameworks, some auditing procedures are already in place to support AI providers in aligning with the requirements of the AI Act. These procedures typically involve the evaluation of technical documentation, as described in Article 11, as well as interviews with development and management teams to assess compliance. Although the AI Act has not yet been applied to high-risk AI systems, certain organisations already offer conformity-related services. For instance, IPN in Portugal[5] conducts audits and assessments of AI systems in the healthcare domain, while LNE in France[6] provides certification services focusing on AI development processes. These early auditing initiatives contribute to structuring compliance practices and can serve as reference points for future formal assessments under the AI Act.

Structure of the chapter

Given this general context of existing conformity tools, this chapter aims to describe in more detail the LNE certification procedure, with the objective of helping the reader to understand how conformity with the AI Act can be achieved. In the second section, the procedure generally used to evaluate AI systems performance will be described (2). In the following section, both the approach to analysing the process used to develop an AI system and the auditing aspect, with a focus on the LNE certification, will be explained (3). Finally, to provide a broader perspective, the chapter will dedicate a section to various organisations and initiatives aimed at ensuring the safety and trustworthiness of the AI market (4), before making a few concluding remarks (5).

2. AI performance evaluation procedure

2. 1 General description

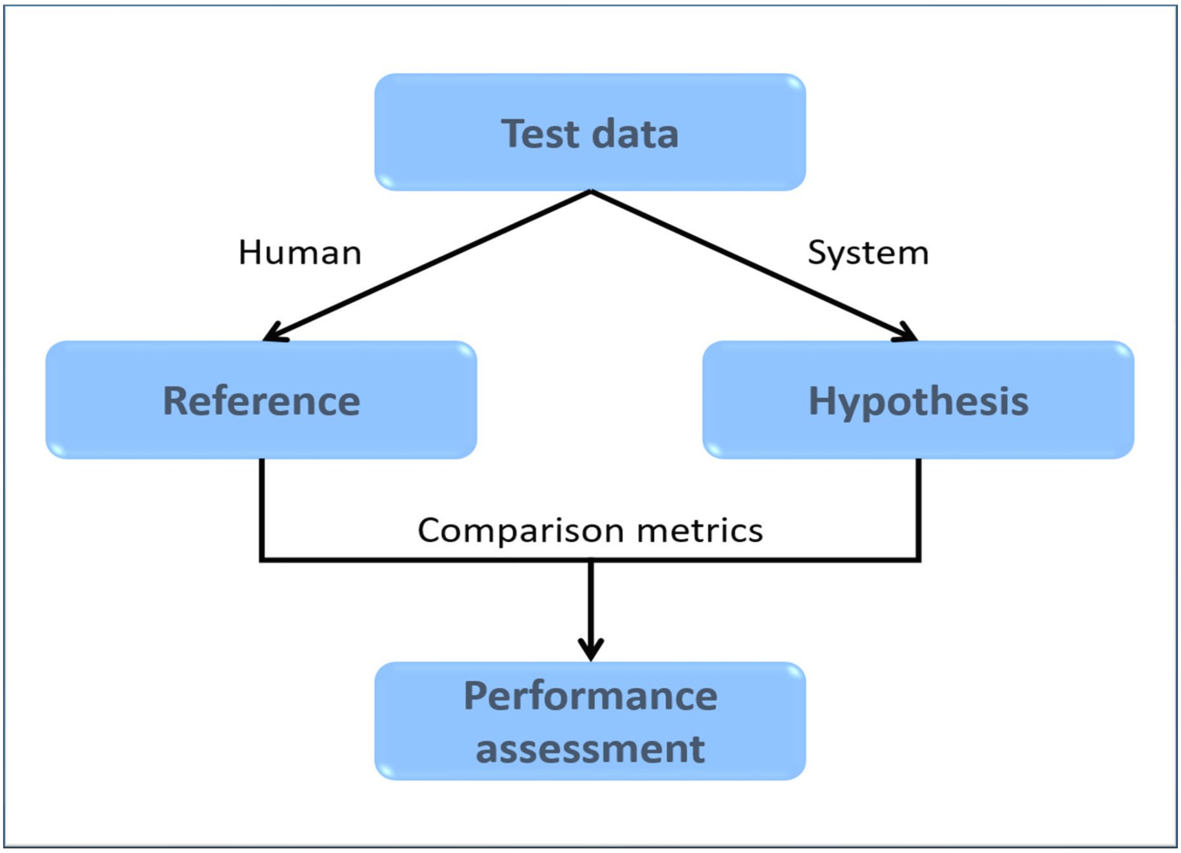
The evaluation of AI system performance is primarily linked to Articles 10 (Data and data governance) and 15 (Accuracy, robustness and cybersecurity) of the AI Act. Given the black-box nature of many AI systems, there is broad consensus within the scientific community on the need to focus on the system's inputs - such as the data processed by the AI - and outputs, such as predictions or classifications. The standard approach involves constructing an evaluation dataset aligned with the AI system's domain of application. Human annotators then provide the reference values (or “ground truth”) by

[5] Instituto Pedro Nunes (IPN), (Portugal) *https://www.ipn.pt/.*

[6] Laboratoire National de Métrologie et d'Essais (LNE), Artificial Intelligence – Thematic Offerings, available at: https://www.lne.fr/en/offers/thematic/articial-intelligence

labelling the data. The same dataset is processed by the AI system to generate the hypotheses.

For example, in the case of a machine translation system from Arabic to French, the evaluation dataset contains sentences in Arabic. The reference dataset includes human-generated French translations, while the hypothesis dataset consists of translations generated by the AI system. The reference and hypothesis outputs are then compared using evaluation metrics specific to the application domain. These metrics yield a score that enables evaluators to assess the performance of the AI system. A visual representation of this procedure is provided in the figure below.



Given this procedure, two key documents are typically produced and expected, particularly in light of regulatory requirements: the evaluation plan and the evaluation report. These documents ensure the reproducibility and traceability of the evaluation process, and are therefore essential for demonstrating compliance with Article 15 (Accuracy, robustness and cybersecurity) of the AI Act.

2. 1.1 Evaluation plan

The evaluation plan generally contains the following sections:

- **Task description**. This section outlines the functionality of the AI system, the type of technology used, and other relevant contextual information. It is a critical foundation for the subsequent sections, particularly for identifying influence factors and defining appropriate evaluation metrics.

- **Technical specifications.** This part describes the format of the inputs and outputs of the AI system, including their structure and type.

- **Influence factors.** This section identifies the input features that may affect the AI system's performance. Recognising these factors is crucial, as their presence in the dataset determines how the evaluation data should be constructed. Influence factors are typically identified during the risk analysis process, which links this section to Article 9 (Risk management system) of the AI Act.

For example, for an AI system that processes images from a camera, a relevant influence factor could be the level of visibility (e.g. clear vs. low-light conditions).

- **Data specifications.** This section describes how the data used in the evaluation was collected, including information on sampling methods, class distributions, annotation procedures, and overall data qualification. These elements relate directly to the requirements of Article 10 (Data and data governance) of the AI Act. The identification of influence factors plays a central role here, as it ensures the representativeness of the evaluation dataset and helps avoid bias.

- **Evaluation metrics.** This part describes the metrics used to assess how closely the AI system's outputs (hypotheses) match the expected results (reference data). What the system produces is called the *hypothesis* (also referred to as the prediction or output), while what is expected as the correct result is called the *reference* (or *ground truth*), typically provided by human annotators. It also includes a justification for the choice of metrics, depending on the nature of the evaluation task.

For example, in a machine translation task, the "BLEU score"[7] – as **technical performance indicator** – may be used to measure the similarity between the system-generated translation and the human reference translation; in a binary classification task, metrics such as accuracy, precision, or F1-score may be more appropriate[8].

- **Evaluation protocol.** This section provides a general overview of the testing procedure and describes the configuration used during the evaluation (e.g. system settings, test environment, or variations in the data). In addition to standard performance testing, it may also include robustness testing, for instance, by using synthetic data specifically created to simulate certain conditions, as detailed in the data specifications section.

- **Performance scores.** This section provides guidance on how to interpret the scores obtained through the evaluation metrics, particularly in relation to specific thresholds, if applicable.

2.1.2 Evaluation report

The evaluation report presents the results produced by the AI system using the selected evaluation metrics to compare the reference data with the system's outputs. While a global score is typically provided for each metric, it is particularly important to report performance scores broken down by influence factors. **This analysis makes it possible to assess how the system performs across different conditions and to detect potential biases in its behaviour.**

The results are also compared with the performance objectives defined during the system's design phase, in order to verify whether the system meets the expected thresholds. Additionally, the report ensures full traceability of the evaluation process and,

[7] Short for "Bilingual Evaluation Understudy", method primarily used in natural language processing (NLP), particularly for evaluating the quality of machine-generated text, such as in machine translation systems. See Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "Bleu: a Method for Automatic Evaluation of Machine Translation", *In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

[8] Dembinsky et al. (2025), Unifying VXAI: A Systematic Review and Framework for the Evaluation of Explainable AI, arXiv, June 2025.

if necessary, enables the evaluation to be reproduced in the future to confirm the consistency of the results.

In addition to the analysis of metric results, complementary statistical analyses may be provided to enhance the understanding of the AI system's behaviour, such as, for example, the Wilcoxon signed-rank test, analysis of variance (ANOVA), or other relevant significance tests.[9]

2.2 Specificities of Large Language Models (LLMs)

The procedure presented in this section can be applied to most AI systems and domains, with a few adjustments. Indeed, LLMs have specific challenges associated. For instance, these systems are often multi-task and their inputs are in the form of a prompt – a small sentence or paragraph – and not a structured format. Similarly, their outputs consist of free-form text, which poses challenges for evaluation, though such issues are also encountered in traditional NLP systems. This means that the approach for evaluating non-language AI systems needs to be adapted. Furthermore, evaluation challenges associated with LLMs also include hallucinations, LLM as a judge (i.e. performing evaluative, decision-making, or classification functions), multilingualism, etc. The evaluation of LLMs is currently the focus of intense research efforts, both within the scientific community and across numerous industrial and collaborative projects.

3. **Certification of AI development processes**

3.1 What is a certification?

It is first important to clarify what certification means in relation to other components of the regulatory landscape, particularly technical standards and regulation. As a brief reminder, a regulation is designed by public authorities (with a national or European scope, for instance) and imposes a set of requirements and prohibitions. A technical standard is created by standardisation organisations (eg. IEEE or ISO), is voluntary and can give a presumption of conformity with an associated regulation (in the case of harmonised standards)[10]. Finally, a certification is awarded by a third party (accredited certification organisation) and ensures that an element (organisational system, process,

[9] These statistical tests are used to assess whether the differences observed between groups or conditions are statistically significant, beyond what could be attributed to random variation. For example, the Wilcoxon signed-rank test is suitable for comparing paired results without assuming a normal distribution, while ANOVA is commonly used to evaluate the effect of multiple influence factors on performance scores.

[10] On AI technical standards, see also in this Guide *supra*, O. Kanevskaia, p. 104.

person, product or service) conforms with requirements stated in a reference document, for instance a technical standard.

3.2 Motivation for an AI certification on processes

Metrology in the field of AI remains a controversial topic. Unlike other domains that rely on well-established measurement standards - such as the *metre* in dimensional metrology or the *watt* in power measurement - no equivalent reference exists for AI systems. The performance of an AI system cannot be assessed against a fixed measurement standard, regardless of the task. Instead, performance depends on *multiple factors*, particularly the data used for training and evaluation.

These data-related parameters can be highly variable and difficult to characterise, making a traditional metrology-based approach ineffective. However, the quality of the final product can be assured through the analysis of its development process. In this context, the focus shifts from directly evaluating the final product to ensuring that it was properly designed and validated.

This process-based approach builds confidence in the AI system by enforcing good practices and quality control throughout development. While performance remains important, the objective is to demonstrate that the system has been rigorously evaluated and meets the goals defined during its design phase.

3.3 The LNE certification of AI processes

The LNE certification is a process certification published in 2021 and publicly available on LNE's website.[11] It covers four main categories of processes: conception, development, evaluation and monitoring of data-based AI systems, including all types of machine learning approaches. This means that knowledge-based AI, which are included in the AI Act, are not covered by this certification[12]. Hybrid approaches, including for instance both data- and knowledge-based AI methods are, however, accepted.

The certification was initially developed to address three distinct profiles:

- Developers, for whom it serves as a commercial branding tool and a third-party guarantee of quality or maturity level;

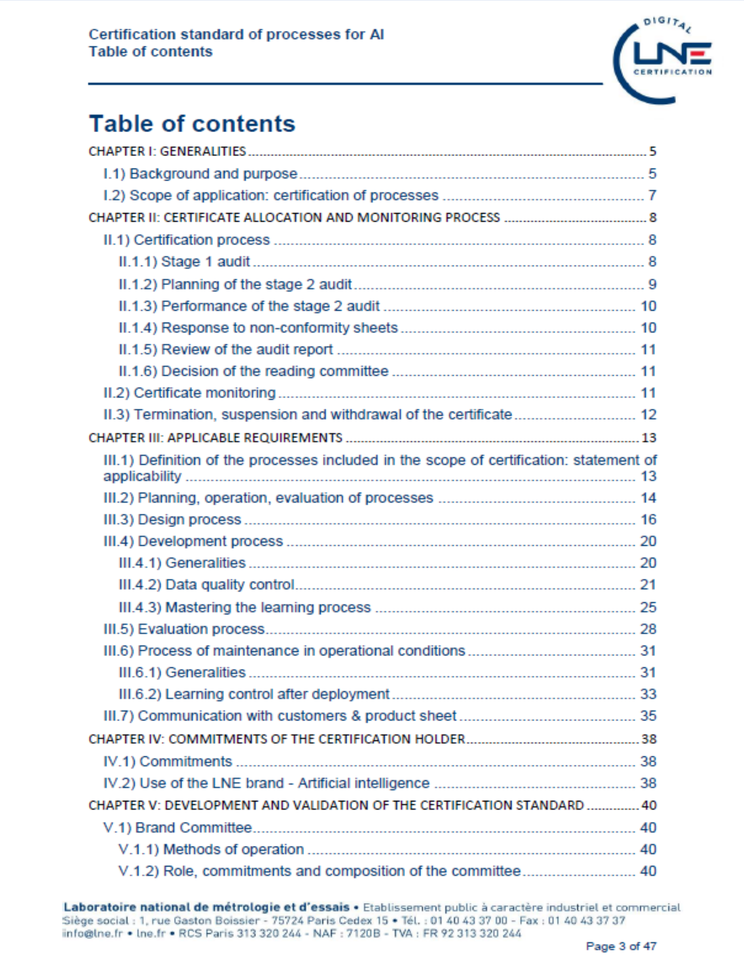
[11] https://www.lne.fr/en/services/certification

[12] Knowledge-based AI systems are typically built on explicitly encoded expert rules or structured knowledge representations (such as ontologies or logic-based inference engines), rather than on statistical models trained from data. As a result, they generally present a lower risk of opacity or unintended bias, since their decision logic is transparent and human-interpretable.

- **Decision-makers and prospective buyers**, who can use it as an objective basis for evaluating and comparing AI solutions;

- **Public authorities**, for whom it contributes to the broader acceptance of AI by reinforcing trust and promoting responsible development.

Since the adoption of the AI Act, the certification has also come to be seen as a tool to help AI system providers structure their organisation and demonstrate readiness for compliance.



Certification standard of processes for AI
Table of contents

Table of contents

Laboratoire national de métrologie et d'essais • Etablissement public à caractère industriel et commercial
Siège social : 1, rue Gaston Boissier - 75724 Paris Cedex 15 • Tél. : 01 40 43 37 00 - Fax : 01 40 43 37 37
info@lne.fr • lne.fr • RCS Paris 313 320 244 - NAF : 7120B - TVA : FR 92 313 320 244

Page 3 of 47

Figure 1. LNE certification - Table of contents (extract)

The certification was created through a working group composed of several actors in addition to the LNE:

- **Industrial groups** : Michelin, Orange, Thales
- **SMEs** : Arcure, Kickmaker, Scortex
- **Consulting companies** : Axionable, Capgemini Invent
- **Clusters** : IRT Railenium, Proxinnov, TOSIT

Seven meetings were planned from 2020 to 2021, with proof-of-concept audits with voluntary companies organised to test the first version of the certification. The final version was published in June 2021, with the first official audit conducted in September 2021.

Note - The certification contains around 150 requirements split between the four main categories of processes (conception, development, evaluation and monitoring) and is freely available at the following link:

https://www.lne.fr/fr/service/certification/certification-processus-ia

For each requirement, there are no imposed technical solutions but, rather, objectives to be reached (i.e. quality, control, monitoring), with examples to help the company understand the requirement. It is also necessary to document the processes and to know how to justify the decisions taken during the development of the AI system. This means that a strong focus is put on ensuring that the team members working on the project have access to the relevant information. The broad ecosystem (eg. customers, users, regulations, internal organisation) was considered when designing the certification, ensuring the content of the requirements will meet their needs. Finally, the content of the requirements and the general structure of the document were also defined using a risk-based approach, meaning that the certification has strong similarities with the AI Act.

An example of a conception requirement reads as follows:

"III.3.2. The AI functionality specifications (see III.3.1) must be made available to anyone involved in the design, development, evaluation or maintenance of the AI functionality.

For instance, AI functionality specifications may be made available to customers via the product sheet, available on the supplier's website, in customer documentation, etc." (p. 19, LNE AI certification standard)

All processes share common requirements: the input and output elements of the processes have to be documented, the resources needed to ensure the proper functioning of these processes must be clearly defined, the risks when deploying and using the AI functionality have to be considered, and the processes must be evaluated.

4. The auditing process

If an organisation applies for LNE AI certification, its request will be assessed to ensure that the minimum requirements are met. After this step, an auditing team composed of a quality process expert and an AI expert will meet the company's team, generally composed of project managers and more technical profiles, to challenge them on the requirements. The analysis will be conducted on a sample of AI projects depending on the size of the company. The assessment will be based both on the provided documentation during the audit and the answers from the AI team when interrogated on specific requirements. The expected documentation is quite similar to the list of documents from Annex IV of the AI Act.

During the audit, non-conformities (minor or major) can be identified by the auditing team depending on the documentation and the answers provided by the representatives of the company. Depending on the severity of the non-conformity, the company is given a different timeframe to address them; major non-conformities must be resolved within two weeks of the audit, while minor ones can be corrected by the time of the next renewal audit, one year later. An auditing report will be provided by the auditing team. This report will be analysed by additional LNE experts to determine whether the company can be granted the certification. If the certification is obtained by the company, an auditing cycle of three years will be initiated, during which two small audits and one medium audit will be conducted each year to ensure the company still follows the correct processes.

With regard to large language models (LLMs), it should be emphasised that this certification was developed prior to their rise in performance and widespread adoption. As a result, some of the current requirements may not fully address the specific challenges posed by LLM-based AI systems. The certification framework is currently being updated to incorporate requirements tailored to LLMs.

Note: This auditing process can be quite time-consuming and costly for a company. However, we believe that similar constraints will be expected for companies pursuing CE marking under the AI Act. This means that the LNE certification – or other existing certifications/standards – can be a serve as a good framework to prepare for the expected workload of the application of the AI Act. Even without undergoing the full auditing cycle, considering the content and structure of the LNE certification can still be highly relevant for companies aiming to achieve future conformity.

5. **Regulatory ecosystem**

5.1 Overview of the main actors and initiatives

In addition, in the wake of the publication of the AI Act, a number of institutions, projects or initiatives have also been created to help AI providers to achieve conformity. As a lot of these actions are still in their early phase, their objectives and perimeter of action may evolve with time.

The following institutions and projects are linked to AI Act conformity, some of which are directly mentioned in the text of the Regulation:

- Union Testing Facilities
- AI Regulatory Sandboxes
- Testing and Experimentation Facilities
- EU AI Office
- Artificial Intelligence Security Institutes

6. **Union Testing Facilities**

The Union Testing Facilities (UTFs) are public testing infrastructures established at the European level to support the implementation of the AI Act (Article 84). Their purpose is to support the Market Surveillance Authorities (MSAs), which cover high-risk AI systems and ensure the correct application of AI Act requirements.

UTFs are responsible for developing testing methodologies and providing platforms to assess whether AI systems deployed on the market comply with the Regulation. These facilities will also enable MSAs to apply those methodologies effectively in their supervision activities.

In addition, both the UTFs and the methodologies they develop will need to evolve over time to remain aligned with the state of the art. In this context, a pilot initiative was launched by the European Commission in 2025 to establish the first UTFs. The project, named NoLeFa-84, is available at: https://nolefa.eu/.

7. **AI regulatory sandboxes**

Regulatory sandboxes[13], defined in Article 57 of the AI Act, are controlled environments for innovation, facilitating the development, training, regulatory testing, and validation of

[13] On AI regulatory sandboxes, see also in this Guide *supra*, G. SHARMA, p. 139.

innovative AI systems before they are placed on the market. The aim is to help AI providers through the sandboxes to reach CE marking for the AI Act. While not directly equivalent to Union Testing Facilities (UTFs), they can be seen as complementary instruments, one supporting pre-market experimentation, the other focusing on post-market conformity assessment. Each Member State shall establish at least one AI regulatory sandbox by August 2026. A project has been launched in 2025, named EUSAiR, to design the framework of AI Regulatory Sandboxes [https://eusair-project.eu/].

8. Testing and Experimentation Facilities

The Testing and Experimentation Facilities (TEFs) project was launched in 2023 by the European Commission to bring a vertical/business approach to the problem of AI validation in several specific areas:

- Medical - TEF-Healthcare [https://tefhealth.eu/]
- Urban/Transport – Citcom.ai [https://citcomtef.eu/]
- Agrifood – AgrifoodTEF [https://agrifoodtef.eu/]
- Manufacturing – AI Matters [https://ai-matters.eu/]

Their main objective is to provide AI providers - in particular SMEs - with access to advanced infrastructure and expertise to test, evaluate, and fine-tune their AI systems in realistic conditions.

Beyond technical support, TEFs also play a strategic role in facilitating the launch of regulatory sandboxes within their respective sectors, by helping identify relevant use cases, preparing providers to operate in controlled environments, and ensuring alignment with the requirements of the AI Act.

9. EU AI Office

The EU AI Office, established under Article 64 of the AI Act, operates within the European Commission and is responsible for ensuring the consistent implementation of the regulation.[14] Its key tasks include supervising GPAI models, coordinating oversight and standardisation activities, and supporting innovation and international cooperation in the field of AI. The AI Office also coordinates a network of stakeholders in the implementation of the AI Act: Market Surveillance Authorities, UTFs, and the Network of Evaluators (external experts). Its work is structured within the New Legislative Framework,

[14] https://digital-strategy.ec.europa.eu/en/policies/ai-office

a simplified normative way for passing AI-related standards without going through the classical legislative process.[15]

10. Artificial Intelligence Safety Institutes

Artificial Intelligence Safety Institutes (AISIs) are national initiatives launched by several countries with the aim of developing public capabilities to assess and mitigate the risks posed by advanced AI systems. These institutes are designed to support national oversight, inform policy decisions, and contribute to global coordination efforts in AI safety. Several countries have already established their own institutes, including the United Kingdom, the United States, Japan, and Singapore. The international network of AI Safety Institutes was initiated during the AI Safety Summit held at Bletchley Park (United Kingdom) in November 2023. Since then, the initiative has been reinforced through a series of high-level global meetings, including the Seoul Summit (May 2024), the Paris Summit (February 2025), and a forthcoming summit to be hosted by India.

The AISIs cooperate at an international level on the development of shared evaluation methodologies and the exchange of safety-related information, which may then be adapted to national policies and priorities. [16]

Example for France

The French AISI, hosted in the INESIA (Institute for the evaluation and security of AI), was officially announced at the AI Action Summit 2025 in France.

INESIA is a grouping of four already existing public organisations: the National Institute for Research in Digital Science and Technology (Inria), the Center of Expertise for Digital Platform Regulation (PEReN), the French National Cybersecurity Agency (ANSSI) and the National Laboratory of Metrology and Testing (LNE).

The institute operates under the supervision of the General Secretariat for Defense and National Security and the Directorate General for Entreprises.

The main missions of the institute are:

- Systemic risk analysis in the field of national security
- Support for the implementation of AI regulationPerformance and reliability evaluation of AI models

[15] On that package of EU law measures that aim to improve market surveillance and boost the quality of conformity assessments, see the dedicated webpage: https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en

[16] On the international network of AI Safety Institutes and its mission statement, see : https://digital-strategy.ec.europa.eu/en/news/first-meeting-international-network-ai-safety-institutes

Conclusion

Among all existing methods to help AI providers to answer the requirements of AI Act, such as institutions or projects as AI regulatory sandboxes or the Testing and Experimentation Facilities (TEFs) there are many resources, either through a controlled environment or services.

The approach to evaluating the performance of AI systems is suited to answer requirements from Articles 10 and 15, and third-party certification such as the LNE AI Certification can help AI providers prepare their organisation for the application of the AI Act.

In addition, the European Commission and governments are promoting new initiatives to better regulate AI and ensure a controlled and secure environment, both for the citizens and AI providers, as illustrated by France's establishment of INESIA.

Chapter 14 - Compliance with the AI Act from a Corporate Perspective – Insights from an AI Expert

Axel Cypel

Introduction

When the first consultancy emails promising "AI Act readiness" began circulating, they signalled an emerging reality: compliance with the European Union Artificial Intelligence Act (AI Act) will soon be a business prerequisite rather than a theoretical curiosity. The Regulation, whose principal provisions take effect progressively from 2025[1], applies to any organisation that places an AI system on the EU market or whose system outputs are used within the Union, regardless of the organisation's domicile[2]. Much like the GDPR[3], the AI Act therefore functions extraterritorially and is expected to generate an extensive ecosystem of advisory, audit, and governance activities.

Focusing exclusively on the application of the AI Act within companies, this Chapter will attempt to bring some clarity as regarding important aspects of the Act, such as the diversity of transparency and documentation requirements contained therein and the concept and management of risk (low, medium, high). The purpose is to see whether the regulatory burden can be streamlined into broad, easily separable categories or, in other words, to engage in a kind of reverse-engineering of the legal text to do the opposite of what is done in AI: transform the implicit into explicit rules.

In this respect, this chapter aims to answer three practical questions every company must address (1), synthesise the Regulation's requirements into coherent operational domains (2), and highlight strategic implications, including interactions with the GDPR (3).

1. Three foundational questions for businesses confronted with the AI Act

Determining whether a company – regardless of its origin, i.e. whether it be European, American, or Chinese, etc. – is affected by the AI Act is relatively straightforward; if an AI

[1] Art. 113 of the AI Act.

[2] Art. 2, §1, of the AI Act.

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *OJ L 119, 4 May 2016, p. 1-88.*

system is made available or deployed in Europe by said company, the latter is covered by the Act. The global reach of the AI Act is clear from its Article 2 (1).

Given the implementation timeline of the AI Act under Article 113, the first thing for companies to do is to compile an inventory of AI systems within the organisation and, more specifically, of AI systems potentially used in Article 5 scenarios, i.e. prohibited AI practices such as systems deploying subliminal techniques, social scoring algorithms or predictive criminal justice based on profiling. In practice, the vast majority of organisations, whether SMEs or blue-chip companies, are unlikely to hold any such systems. In case of doubt, organisations should refer to the guidelines published by the AI Office, which specify use cases in which these prohibited AI practices may be characterised[4]. Furthermore, Article 111 limits the inventory's practical impact by confirming, in line with established legal principles, that the Regulation is not retroactive[5].

Once the Regulation is applicable – following a gradual timeline depending on its provisions – compliance converges around three analytical questions:

- System characterisation: Does the AI-based solution meet the AI Act definition of an AI system, and what output is it designed to deliver?
- Risk classification: Is the AI system prohibited, high-risk, subject to the specific transparency obligations under Article 50, or minimal-risk, i.e. excluded from the scope of the Act?
- Company role: Is the organisation acting as a provider, a deployer, or an end-user?

Each aspect should be examined separately.

1.1 *Is the system an AI system, and what does it produce?*

As with the GDPR, which applies only where personal data are processed, the AI Act applies solely when a system falls within its definition of an AI system. The definition adopted is broad and functional, encompassing everything done to date[6]. It embraces automation, autonomy, prediction and the generation of content through machine-learning techniques. Although purely deterministic, hard-coded routines without adaptive behaviour will generally fall outside the definition; anything based on machine

[4] Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), C(2025) 884 final, 4 February 2025.

[5] Article 111 specifies that compliance obligations apply solely to AI systems still in use once the Regulation enters into force; any system decommissioned beforehand falls outside its scope.

[6] Art. 3(1) of the AI Act and this guide see J. SENECHAL, p.14.

learning is considered AI. Accordingly, organisations should establish a comprehensive register of digital solutions to determine which qualify as AI.

Nonetheless, merely recognising that a system qualifies as AI remains insufficient; its intended purpose must likewise be clearly defined. General-purpose AI (GPAI), for instance, attracts specific obligations; where GPAI is deployed to generate synthetic media such as deepfakes, the deploying organisation must disclose unambiguously that the content is machine-generated[7]. It is therefore prudent to adopt a concise internal taxonomy that distinguishes *inter alia* between in-house adaptations of open-source models, commercially available off-the-shelf solutions and GPAI services, thereby ensuring that the correct compliance controls are applied to each category. This provides a coherent basis for consistent classification and the allocation of appropriate controls and responsibilities.

Organisations must therefore catalogue all digital solutions and determine which qualify as AI systems. They must also record the intended purpose of each system, because obligations differ for (i) general-purpose AI (GPAI), (ii) high-risk, purpose-specific AI, and (iii) minimal-risk or transparency-only use cases (e.g. deepfakes that require labelling).

1.2 *What is the associated risk level?*

Apart from the outright prohibitions under Article 5, the AI Act mainly focuses on 'high-risk' AI systems. This category is delineated explicitly in Article 6 of the Act and, by extension, in Annex III (supplemented by Annex I)[8]. For most commercial organisations, the pertinent sub-categories relate mainly to biometric identification and to AI deployed in matters related to education, vocational training or employment. The use cases for AI under Annex III are also abundant within government administrations (e.g. administration of justice, border control management, and law enforcement).

Responsibility for classification rests with the organisation itself; any assertion that a system falls outside the high-risk tier must be underpinned by cogent and up-to-date documentation[9].

Where a system is neither prohibited nor high-risk, the residual obligations are comparatively light and appear principally in Article 50, together with the somewhat

[7] Art. 50 of the AI Act and this guide see p. 23.

[8] On AI systems characterisation as high-risk, see. in this Guide, *supra*, p.24 and European Commission, Targeted stakeholder consultation on classification of AI systems as high-risk, 6 June–18 July 2025 (under way), launched to inform forthcoming Commission guidelines on high-risk AI systems classification and associated obligations.

[9] Art. 6 of the AI Act.

elliptical Article 4, which functions as a quasi-preamble[10]. Article 4 introduces an “AI-literacy” duty, creating a sort of obligation for organisations to ensure that staff members who develop, operate or are affected by AI systems receive appropriate training and awareness. By encouraging a culture of responsible AI use, this requirement justifies the allocation of [corporate] training budgets [within organisations] and, in return, strengthens an organisation’s overall compliance posture.

1.3 *What role does the company play?*

The distinction between the roles of provider and deployer is one of the main nuances introduced by the Regulation. This differentiation does not simplify matters; rather, the multiplication of roles disperses and reallocates responsibilities[11].

For companies, this is surely one of the major challenges in applying the Regulation; the same AI system – depending on its purpose, its users, any modifications, and its life-cycle stage – can cause the organisation’s role, and therefore its obligations, to shift. In practice, deployment often entails [technical] interventions that modify the system, thereby requalifying the organisation as a provider.

Article 25, titled “Responsibilities along the AI value chain”, confirms that liability may pass from one economic operator to another; this is indeed a well-established approach in European product safety law. This gives rise to a broader question: should AI systems be treated as products and thereby subjected to the product safety framework? The proposition is attractive; an AI model embedded into pure software or housed in a physical device is, functionally, a product. Yet contemporary AI is dominated by models trained on vast datasets using substantial computational resources, and these assets are typically owned by large technology companies. Allocating liability primarily to the deployer therefore eases enforcement but leaves the original producer with only partial accountability.

Numerous specific use cases illustrate the operational consequences. For instance, where an AI system is deployed to determine creditworthiness or to price life- or health-insurance policies, a Fundamental Rights Impact Analysis [FRIA] must be carried out and its findings reported to the national authority[12]. Article 27 assigns the impact analysis to the deployer, while the accompanying risk analysis remains the provider’s responsibility;

[10] On AI literacy under Article 4 of the AI Act, see in this Guide, N. Nevejans, p.101.

[11] If the aim were to shield the major American technology companies, it could hardly have been achieved more efficiently, even though the provider, beyond fulfilling the obligations imposed by this Regulation, is, as the manufacturer of the AI system, civilly liable under the 2024 Directive on product liability for defective products.

[12] On FRIA under Article 27 of the AI Act, see in this Guide, L. Xenou & M. Ho-Dac, p. 136.

isn't this two facets of the same exercise allocated to different actors? In that respect, Article 27 invites the deployer when performing its risk assessment to "… *tak[e] into account the information given by the provider pursuant to Article 13*"[13], i.e., transparency information on the high-risk AI system, including the instructions for use.

An EU-based subsidiary of a non-EU provider must ensure a Union representative is appointed[14]. Importers (often the position of European companies) must verify conformity documentation before placing a product on the market and must refuse to do so when substantive doubts arise[15]. This is a significant responsibility. How, in practice, can importers evaluate technical documentation when they have not participated in the development of the AI systems and, moreover how can they deal effectively, when doubts arise, with providers who wish to declare themselves compliant with the AI Act? This distribution of responsibilities, while aiming to safeguard end-users in the EU internal market, may generate significant uncertainties and risk for the businesses involved.

2. Rationalising the AI Act's requirements in establishing coherent operational domains

For operational purposes, the various requirements imposed by the AI Act can be categorised into three main areas. The first relates to governance, i.e. ensuring that a formal framework is in place to monitor AI-related activities within the company. The second involves documentation of everything that needs to be verified during the design of the AI system, from training-data management to the recording of design decisions and generated outputs. Finally, the third area covers the extensive set of administrative obligations.

2.1. Governance

Proof of an organisation's commitment to AI Act compliance may be demonstrated by establishing an *ad hoc* committee or a sub-committee, within the risk management department, in charge of overseeing governance and lending substance to the risk management system under Article 9[16]. Specific risks posed by AI systems must be identified, assessed, mitigated and formally accepted before the solution is released into production. These risks, whether technical or non-technical, include a wide range of legal issues such as those posed by generative AI – for instance, chatbots that inadvertently recommend a competitor's product or invent non-existent refund policies – as well as potential leaks of confidential or strategic data arising from uncontrolled use of such an AI-enhanced chatbot.

[13] Art. 27, §1, lit. d) of the AI Act.

[14] Art. 22 of the AI Act.

[15] Art. 23, §2 of the AI Act.

[16] On Risk Management under Article 9 of the AI Act, see in this Guide, A. Favreau, p. 61.

The dedicated *ad hoc* committee would also be responsible for ensuring that all other requirements are satisfied throughout the AI system's life-cycle. One of its initial tasks could be to inventory the AI applications in use across the organisation. It would likewise manage the resources allocated to meeting the aforementioned AI literacy obligation under Article 4. At this stage, integrating robust data-governance measures, familiar from the GDPR, would be essential to achieving genuine effectiveness.

2.2. Documentation

Here, the provisions on data governance[17], technical documentation[18], transparency[19] [Article 13], IT monitoring[20], security commitments[21] and human oversight[22] are best considered collectively. Rather than repeating each requirement verbatim, the following section offers critical observations from an operational perspective.

Data. No AI system can function without data, and the AI Act therefore attaches great importance to data quality[23]. From dataset management and lineage to bias identification, mitigation and – on the deployer's side – production-data monitoring, the Act's provisions are fully in line with established data-science practice. It is notable that the Regulation seeks to codify methods already familiar to AI professionals; one might reasonably speculate that leading firms in the tech sector have offered advice during its drafting.

Article 10[3] of the Act provides that "*training, validation and testing data sets shall be relevant, sufficiently representative and, to the best extent possible, free of errors and complete in view of the intended purpose*." While this language clearly aligns with current data-science terminology, the assumption that inductive AI systems can rely on *complete* datasets may be considered as problematic; indeed, this likely reflects a limited understanding of the technology's nature. In practice, reality cannot be captured exhaustively by data, nor is axiological neutrality attainable[24].

[17] Art. 10 of the AI Act.

[18] Art. 11 and 18 of the AI Act.

[19] Art. 13 of the AI Act.

[20] Art. 12 and 19 of the AI Act.

[21] Art. 15 of the AI Act.

[22] Art. 14 of the AI Act.

[23] On Data [quality] management under Article 12 of the AI Act, see in this Guide, J-M Vangyseghem, p. 77.

[24] The chapter does not address the question of bias but still highlights the inherent tension within the text concerning that matter. The learning of a model by induction is based on a sample of data, which in no way represents a complete picture of reality. On the one hand, this is by definition what a sample is [incomplete], and on the other hand, data are measurements of the world but not the world itself [incompleteness]. Thus, machine learning is an inherently biased form of modelling.

Record-keeping. High-risk AI systems must provide automatic, life-cycle event logging, a long-established requirement in information-technology practice.

Transparency. The Regulation's transparency obligations face two structural constraints[25]. First, meaningful disclosure is technically linked to the still-unresolved challenge of AI explainability. Secondly, Article 53 of the AI Act – in combination with the text-mining exemption of Directive (EU) 2019/790[26] – means that even GPAI models' providers are not obliged to reveal their full training datasets or demonstrate copyright clearance. In the absence of a robust theoretical framework and a mathematical theory that explains why deep-learning models perform so effectively, complete transparency remains aspirational rather than attainable.

Human oversight. Section 4(b) of that Article 14 requires that the persons tasked with human oversight be able "to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (i.e. automation bias)." The AI oversight staff must remain alert to automation bias and be able to interrupt a high-risk system via a reliable "stop function". This features a typical paradox: how can one be aware of an unconscious process? While such controls are straightforward for conventional machinery, their practical implementation in distributed, cloud-based services is considerably more complex. The provision further requires the involvement of at least two qualified individuals in the sensitive field of remote biometric identification.

Technical documentation. Annex IV of the AI Act fortunately supplies a template for technical documentation. The dossier for any high-risk AI system must contain comprehensive, descriptive information on the model, a detailed summary of the data used for development, performance metrics, arrangements for operation and maintenance, supervisory procedures and, ultimately, an EU declaration of conformity.

For design teams, meeting these specifications entails a significant organisational effort. The granularity demanded for data selection, data characteristics and engineering decisions is such that, in practice, it would almost be necessary for every data scientist to be supported by someone dedicated to taking notes on the progress of algorithmic development. Documentation should therefore be compiled progressively throughout the project; producing it retrospectively risks prioritising form over substance and may

[25] On transparency under Article 13 of the AI Act, see in this Guide, F. Guillaume, p. 80.

[26] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, Journal of Laws (Official Journal of the EU), L 130, 17 May 2019, pp. 92–125.

offer insufficient protection to end-users if its value is judged solely on the basis of the length of the report.

Instructions for use. The instructions for use must set out, *inter alia*, the identity and contact details of both provider and deployer, the salient characteristics of the AI system, foreseeable misuse risks, maintenance requirements and the measures enabling human oversight. Although the very term "instruction manual" may appear anachronistic for cutting-edge technology, the document is essential; by specifying the system's intended use, it delineates the provider's liability when the AI is employed for malicious, or prohibited, purposes.

2.3. Administration

All other obligations fall under the administrative procedure, characteristic of product safety regulation: the declaration of conformity[27], CE-marking[28], post-market monitoring[29], appointment of a representative[30], registration in the EU database[31] and incident reporting[32].

While an AI product designer may regard these tasks as burdensome, risk managers and counsel will see familiar territory. Consider a few illustrative projects:

- **Document-processing AI (computer vision for automatic document recognition and information extraction).** Fine-tuning an open-source model that contains personal data renders the organisation a *provider*, though the use case is not classified as high-risk.
- **AI-driven CV screening.** Recruitment falls squarely within Annex III; the system is therefore high-risk by default.
- **LLM-assisted email drafting.** Compliance hinges on receiving adequate technical documentation from the foundation-model provider, including open-source providers, where systemic risk thresholds apply. Should the organisation further fine-tune an open-source LLM, it becomes both the *provider* and *deployer* of a foundation model, with the corresponding obligations found under Articles 51 to 55[33]. Such requirements are manageable for large technology companies, but

[27] Art. 47 of the AI Act.

[28] Art. 48 of the AI Act.

[29] Art. 20 and of the AI Act.

[30] Art. 22 of the AI Act.

[31] Art. 71 of the AI Act.

[32] Art. 73 of the AI Act.

[33] On those provisions and GPAI models, see in this Guide, A. Latil, p.33.

considerably more onerous for smaller enterprises; in many cases, using foundation models "as-is" will be the pragmatic choice.

3. **Strategic implications of the AI Act compliance process**

From an objective perspective, the AI Act's documentation requirements largely codify recognised good practice: recording design decisions, validating data, calculating performance metrics, monitoring outputs and retaining the ability to switch off a malfunctioning system.

3.1. Comparison with the GDPR

As highlighted by many experts, the AI Act is first and foremost a *product* regulation[34]; unlike the GDPR, it does not *directly* confer new rights on individuals but imposes duties on the main actors in the AI value chain. Even so, several parallels emerge:

Impact assessments. The (above-mentioned) Fundamental Rights Impact Analysis (FRIA) for high-risk AI mirrors the GDPR's Data Protection Impact Assessment[35].

Incident notification. Providers and deployers must report serious incidents, echoing the GDPR breach-notification regime.

Representative. Non-EU entities must appoint an EU representative, as under the GDPR.

Privacy requirements. The AI Act adopts privacy-by-design and privacy-by-default principles[36]. However, Article 10[37] permits limited use of special-category data to mitigate bias, an "exception to the exception" that strains the GDPR's data-minimisation rule. Further tensions arise where training/validation/test datasets must be archived for audit, while the GDPR prescribes finite retention. Declaring a long retention period in the GDPR register may be necessary for high-risk projects; deleting data too soon would impede subsequent model work.

[34] See (inter alia) M. Almada & N. Petit, "The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights" (2025) 62 CMLR 85-120.

[35] Art. 35 of the GDPR.

[36] See Recital 69.

[37] [Art. 10] - 5: "To the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems in accordance with paragraph (2), points (f) and (g) of this Article, the providers of such systems may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons."

Right to explanation. Article 86 echoes GDPR Article 22 on automated decision-making, but narrows the safeguard; the AI Act grants only a right to an explanation, not a right to avoid automated processing. Although Recital 93 requires that individuals be informed, the "*Eliza effect"* demonstrates that disclosure alone does not eliminate cognitive or emotional impact[38].

3.2. The AI Act's risk-based approach

The AI Act adopts a risk-based approach that aligns with established risk management practice. Assessment criteria – such as proportionality, materiality, likelihood of occurrence and estimated impact – should be combined within a matrix to identify high-risk areas and the residual risks to be borne by the company. In most organisations, the main difficulty lies not in the analysis itself but in maintaining the necessary governance rhythm, typically through quarterly reviews. Even the complex evaluation of high-risk AI systems is subsumed within this method, long familiar to IT and audit functions.

Article 9 of the AI Act requires a formal risk management system, as mentioned above; accordingly, risks must be catalogued. Establishing a risk taxonomy, maintained by the Legal or Compliance department of the company and overseen by a dedicated AI-governance committee, should provide a sound foundation. The essential tools remain the same: accurate documentation and, where applicable, timely declarations to the supervisory authorities.

A practical recommendation is to draft a decision tree that directs teams to the relevant documentation requirements. The tree should begin by asking whether the system is high-risk, identify the organisation's role and then specify the associated obligations. Standardised templates – usable by internal and external auditors – will streamline subsequent reviews.

Risk-based frameworks do not require every compliance action to be implemented simultaneously; tasks may be prioritised, provided that an action plan and monitoring structure are put in place. The probability-*versus*-impact matrix supports proportionality, balancing on the one hand, the risks the company is prepared to accept and, on the other hand, the resources allocated, in line with the system's purpose and intended use.

[38] See A. Cai, "Eliza Re-examined: Relations between Humans and Robots", *The Gallatin Research Journal*, vol. 5, in Spring 2015, also available online at: https://confluence.gallatin.nyu.edu/sections/research/eliza-re-examined?utm_source=chatgpt.com

Although the AI Act endorses proportionality itself[39], reliance on this principle must not be misconstrued as licence for partial compliance.

Finally, the analysis must include the risk of inaction, i.e., the consequences of failing to address the AI Act's requirements and obligations. Reminding stakeholders of potential sanctions is often a useful motivator for completing the requisite documentation.

Conclusion

The law of the excluded middle asserts that a proposition is either true of false. Applied to the AI Act, one may say that this new EU regulation will either stand as a landmark, signalling an enlightened awareness of the societal transformation heralded by AI technologies, or that it will be sidelined as powerful technology companies continue to shape the AI normative governance agenda. Between these two outcomes no stable compromise appears possible; the prospect is binary, yet sufficient to sustain a measure of optimism.

[39] See Recital 26 of the AI Act: "In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be followed."

Chapter 15 - AI Act Compliance Action Plan in Practice - A Practical Testimony from “BioMérieux”

Yves Raisin (BioMérieux)

Introduction

This contribution provides a practice-based perspective on implementing the EU AI Act within bioMérieux, a global biotechnology company operating at the crossroads of health, diagnostics, and digital technologies. As both a data-driven company and a regulated healthcare actor, bioMérieux's journey towards AI Act compliance illustrates the interplay between legal requirements, technological constraints, and ethical commitments. The experience outlined here reflects the realities of operationalising compliance in a rapidly evolving regulatory and technological environment.

The AI Act introduces a risk-based framework that mandates proactive governance of AI systems across the EU. For bioMérieux, aligning with this framework means addressing multidisciplinary challenges, from explainability to data governance, and integrating compliance mechanisms into our existing quality and innovation culture. This testimony shares concrete lessons, governance structures, and strategic considerations relevant to other industry actors.

1. Key Compliance Challenges

1.1 Acceleration of Innovation vs. Regulatory Tempo

The development of large-scale models and advanced machine learning techniques is moving faster than the legislative process. Anticipating regulatory constraints while designing systems becomes essential, especially for applications in diagnostics or clinical decision support.

1.2 Multiplication of Regulatory Instruments

In addition to the AI Act, compliance requires alignment with a complex legal ecosystem: GDPR, IVDR, MDR, cybersecurity directives, and national ethical standards. Building a harmonised and scalable compliance framework is key to avoiding redundancy and gaps.

1.3 Explainability and Transparency

Some AI systems—especially black-box models—raise challenges in terms of interpretability. Ensuring traceability of datasets and algorithmic decisions is necessary, particularly for high-risk applications. Documentation must be intelligible to auditors and domain experts alike.

1.4 AI Literacy and Corporate Culture

Embedding AI literacy across departments (R&D, Legal, QA, Ethics) supports informed decision-making. Training is not just technical: it includes legal qualification of use cases, risk awareness, and ethical implications.

1.5 Data Security and Privacy

Given our role in processing sensitive medical and biological data, compliance with GDPR principles (lawfulness, purpose limitation, storage limitation, etc.) is reinforced by AI Act provisions on dataset quality and bias prevention.

1.6 Environmental Sustainability

AI Act Article 4(3) introduces sustainability as a compliance objective. For bioMérieux, this means assessing the environmental footprint of data processing and model training, in line with our ESG strategy.

1.7 IP and Transparency Tensions

Ensuring transparency while protecting proprietary algorithms and training datasets requires careful contractual drafting, including IP clauses and disclosure thresholds. These issues often arise in collaborative R&D or public-private partnerships.

1.8 Talent and Skills

Attracting and retaining profiles that combine AI engineering skills with regulatory literacy remains a challenge. Compliance is no longer a legal silo but a shared competency across teams.

2. **Governance and Deployment Methodology at BioMérieux**

At the heart of our compliance approach lies the implementation of an AI Management System (AIMS), designed in accordance with ISO/IEC 42001. This structured system enables risk-based oversight, continuous improvement, and full integration into bioMérieux's broader Quality Management System. The AIMS ensures that compliance obligations are not siloed but embedded into day-to-day operations.

To support this framework, a dedicated AI Governance Committee has been established. This multidisciplinary body, composed of representatives from Legal, Data Science, Ethics, Regulatory Affairs, and IT, convenes regularly to evaluate AI use cases, determine their risk classification, and assign internal responsibilities for documentation and monitoring.

All AI initiatives are subjected to a structured Use Case Mapping process. This internal classification mechanism maps each project to the AI Act's taxonomy—prohibited, high-risk, or limited-risk systems. This categorisation is operationalised through tailored compliance checklists and standardised templates to ensure consistency across departments.

The associated documentation efforts are role-based. Depending on whether bioMérieux is acting as a provider, deployer, or user, different obligations apply. Accordingly, we produce tailored documentation such as technical files, risk logs, FRIA (Fundamental Rights Impact Assessments), and data sheets to match regulatory expectations.

AI literacy is also a strategic priority. We deliver targeted, practical training modules—including case studies from within the company—to enhance legal and ethical awareness among technical and operational teams. These sessions demystify the AI Act's provisions and strengthen internal reflexes to detect and escalate potential non-compliant systems.

Lastly, the AIMS is designed to interface seamlessly with our existing regulatory frameworks. This includes integration with GDPR compliance tools (such as ROPA and DPIA registers), medical device certification files under MDR/IVDR, and cybersecurity protocols. This interoperability ensures legal coherence and operational efficiency across our compliance architecture.

In summary -

- Implementing an AI Management System (AIMS): Inspired by ISO/IEC 42001, a centralised AI governance mechanism ensures oversight and accountability.
- Setting Up an AI Governance Framework: Cross-functional committees (Legal, IT, R&D, Ethics) meet regularly to classify AI use cases, assess risk levels, and allocate responsibilities.
- Clear Internal Communication Channels: Compliance requirements are translated into operational language and integrated into existing quality systems.
- Training and Team Awareness: Targeted training sessions help demystify the AI Act and empower teams to identify risks, notably for high-risk or prohibited systems.
- Early Identification of Forbidden and High-Risk AI: A use-case mapping process, linked to the AI Act's Annex III, helps prioritise compliance resources.
- Documentation According to Actor Roles: Differentiated documentation is prepared based on whether the company acts as provider, deployer, or importer.

3. **Strategic Outlook: From Compliance to Competitive Advantage**

To fully embrace the AI Act is to view compliance not as a constraint, but as a strategic lever. At bioMérieux, we understand that regulatory alignment goes hand in hand with innovation, trust, and long-term competitiveness. This outlook is anchored in three core pillars:

Designing with Fundamental Rights in Mind. From the earliest stages of conception, AI systems must embed the rights to privacy, non-discrimination, and access to healthcare. These are not only legal mandates but also key to public trust and social responsibility. Our teams work to ensure that rights-based thinking is operationalised throughout the AI lifecycle, from data collection and model design to deployment and monitoring.

Leveraging Ethics and Trust as Innovation Catalysts. Ethical reflection is no longer an afterthought but a central driver of responsible innovation. By fostering internal debate and engaging external stakeholders—including patients, practitioners, and regulators—we create an environment where AI solutions can be both bold and credible. Trust becomes an asset, not a constraint.

Moving Toward Anticipatory Compliance. In a rapidly evolving legal landscape, waiting for enforcement is a losing strategy. We have adopted a mindset of continuous monitoring, iterative documentation, and proactive alignment with upcoming guidance and standards. This anticipatory posture allows us to adapt more swiftly, mitigate compliance risks, and remain resilient in the face of regulatory shifts.

Conclusion

The AI Act challenges healthcare actors to rethink how AI is designed, validated, and deployed. At bioMérieux, compliance has become a lever for reinforcing our trustworthiness and innovation capacity. Through structured governance, AI literacy, and integrated systems, we aim not only to meet legal obligations, but to shape a sustainable and ethical future for AI in diagnostics and global health.

AI Act Compliance Glossary

Chapter 1 - AI Systems & Models Taxonomy

AI system

A machine-based system designed to operate with autonomy and possibly adaptiveness, generating outputs that influence environments.

General-purpose AI model (GPAIM)

An AI model capable of performing a wide range of distinct tasks and trained on large datasets using self-supervised methods.

Taxonomy of AI systems

Structured classification of AI systems according to levels of risk: unacceptable, high, limited, and minimal, as defined by the AI Act.

Inference capability

Core functional trait of an AI system that allows it to generate outputs based on data inputs through reasoning or machine learning.

Adaptiveness

An AI system's ability to modify its behaviour or parameters based on new data post-deployment without explicit reprogramming.

Deployment context

The operational setting of an AI system, whether standalone or integrated into a product, influencing its regulatory treatment.

Chapter 2 - Regulating General-Purpose AI Models

Systemic risk

Risks with broad societal impact arising from high-impact capabilities of GPAI models, such as loss of control or harmful manipulation.

Floating-point operations (FLOPs)

A unit of computation used as a technical benchmark to assess model capacity and systemic risk under the AI Act.

Chapter 3 - AI Operators under the AI Act

Provider

Any entity that develops or has developed an AI system or GPAIM and places it on the market or into service under its name.

Deployer

A person or entity using an AI system under its authority, except for personal non-professional activities.

Substantial modification

Any change affecting the AI system's performance, intended purpose, or compliance status, triggering new obligations under Article 3(23).

Product manufacturer

An entity placing on the market a product embedding an AI system under its own name or trademark, considered a provider under specific conditions.

Chapter 4 - AI Literacy and Article 4

AI literacy

Skills, knowledge and understanding that allow providers, deployers and affected persons to make informed decisions about AI systems and understand associated risks.

Training obligation

Requirement for organisations to provide legal, ethical, and technical training throughout the AI system lifecycle, ensuring informed deployment and usage.

Chapter 5 - Risk Management System under the AI Act

Risk Management System (RMS)

A continuous, lifecycle-wide process for identifying, assessing, mitigating, and monitoring AI-related risks as defined in Article 9.

Residual risk

Risk that remains after mitigation measures have been applied, to be evaluated and documented under the AI Act.

Chapter 6 - Data Governance and Management Practices

Fundamental Rights Impact Assessment (FRIA)

A legal tool required by Article 27 for high-risk AI systems to assess their impact on fundamental rights, going beyond the GDPR's DPIA.

Cross-sectoral legislation

A regulatory approach acknowledging the AI Act's interplay with other legal regimes like GDPR, ensuring comprehensive governance.

Chapter 7 - Transparency under the AI Act

Transparency obligation

Obligation for AI operators to disclose key system characteristics, functionalities, and risks, especially for high-risk and transparency-risk systems.

Explainability

Capability of an AI system to provide understandable and meaningful insights into its functioning and outputs for stakeholders.

Chapter 8 - AI Literacy

Affected persons

Individuals impacted by the outputs or decisions of an AI system, entitled to receive AI literacy training under Article 4.

Regular training

Ongoing AI literacy measures that must be adapted to the technical profile and context of use throughout the system's lifecycle.

Chapter 9 - AI Regulatory Sandboxes

AI regulatory sandbox

A controlled framework set up by a competent authority offering providers the opportunity to develop, train, validate, and test AI systems in real-world conditions under regulatory supervision (AI Act, Art. 3(55), 57–58).

Experimentation clause

A legal instrument allowing temporary derogations from existing legal frameworks to enable the testing of innovative AI solutions under specific safeguards.

Chapter 10 - Codes of Conduct and Voluntary Measures

Code of practice (AI Act)

A voluntary, multi-stakeholder governance tool enabling GPAIM providers to demonstrate compliance with Articles 53–56 in the absence of harmonised standards.

Delegated act

A legal mechanism by which the European Commission may adopt non-legislative acts to supplement or amend certain non-essential elements of the AI Act.

Chapter 11 - Notified Bodies and Conformity Assessment

AI harmonised standards

Technical specifications developed by European Standardisation Organisations to support compliance with essential requirements of the AI Act

Conformity assessment

A structured procedure to demonstrate that high-risk AI systems comply with the AI Act, potentially involving internal checks, technical documentation review, and notified body audits.

AI Management System (AIMS)

An internal governance framework to ensure lifecycle-wide AI oversight and accountability mechanisms in an organisation (cf. ISO/IEC 42001)

Chapter 12 - Europe's AI regulatory sandboxes

AI Office

The EU-level administrative body responsible for supervising implementation of the AI Act, coordinating with national authorities, and facilitating the development of codes of practice and harmonised standards.

Market surveillance authority

National body designated to monitor, investigate, and enforce AI Act compliance through inspections, fines, or corrective measures.

Chapter 13 - Conformity Assessment Procedure

Systemic risk (expanded)

Risks arising from the capabilities of advanced GPAIMs (e.g., loss of control, cyber-offense, harmful manipulation), requiring EU-level oversight due to their transnational impact.

Exit report

A public document generated at the end of a sandbox cycle, summarising learnings, risk mitigations, and practical implications of tested AI use cases.

AI seal (Spain)

A national certification indicating that AI systems comply with the EU AI Act and national supervisory agency requirements.

Authors & Contributors

Axel Cypel

AI Business Strategist & AI Governance Expert

Axel Cypel is a senior executive specialising in AI project management, digital transformation, and responsible AI implementation. An engineer from Mines Paris – PSL, he has held leadership roles in risk management, data governance, and AI strategy within the banking sector. He is a frequent speaker on AI governance and projects, business strategies, and digital ethics, and the author of "Au cœur de l'intelligence artificielle" [De Boeck, 2020] as well as "Voyage au bout de l'IA" [De Boeck, 2023], laureate of the 2024 Prix Roberval.

Amélie Favreau

Full Professor, University Grenoble Alpes

Amélie Favreau is Full Professor of Law at the University Grenoble Alpes and Director of the INNOVACS Research Federation [CNRS/UGA]. Her research focuses on digital law, intellectual property, and the regulation of emerging technologies such as blockchain, quantum technologies, and AI.

Bernard Guillaume

AI Regulation Expert, French National Metrology and Testing Laboratory [LNE]

Bernard Guillaume is a technical expert in AI conformity assessment and standards at LNE. He leads the LEIA metrology lab, supporting AI transparency, robustness, and EU AI Act compliance. He collaborates with industry and regulators and frequently speaks at certification and testing forums across Europe.

Céline Castets-Renard

Full Professor, University of Ottawa

Céline Castets-Renard is a Full Professor at the University of Ottawa and Canada Research Chair in International and Comparative Law of AI. She also chairs the Research Chair on Accountable AI in a Global Context and serves as Vice Chair of the EU AI Office Working Group on Transparency and Copyright. Her research covers AI governance, data protection, platform regulation, and algorithmic accountability.

Olia Kanevskaia

Assistant Professor of European Economic Law & Technology, Utrecht University

Olia Kanevskaia is Assistant Professor at Utrecht University, affiliated with the Utrecht Centre for Regulation and Enforcement in Europe (RENFORCE). She holds a PhD in Law, an LL.M in International and European Public Law (cum laude), and an LL.B in International Law from Tilburg University. Her research focuses on EU economic law, governance of digital technologies, trade with disputed territories, and institutional legitimacy in ICT standardisation.

Prof. Florence Guillaume

Full Professor, University of Neuchâtel

Florence Guillaume is Professor of Civil Law, Private International Law, and Digital Law at the University of Neuchâtel since 2006. She founded the LexTech Institute (2020) and the CAS in Law & AI (2024). She is the former Dean of the Faculty of Law (2011–2014), she practiced law in Geneva and Zurich and served as Visiting Scholar at Stanford's CodeX and The Hague Conference on Private International Law.

Prof. Marion Ho-Dac

Professor of Private Law, University of Artois

Marion Ho-Dac is a professor of private law at the University of Artois [France], specialising in private international law and European Union [EU] Law. She is member of the editorial board of the *Revue Trimestrielle de Droit Européen* [Dalloz], and a co-editor of the EAPIL blog. One of her recent co-edited volumes explores the governance of Artificial Intelligence [AI] in the EU and its impact on consumer protection [Bruylant, 2023]. In 2024, Marion was appointed co-director of the 2025 Research Centre of the Hague Academy of International Law, focusing on AI and international law.

Dr. Arnaud Latil

Associate Professor [HDR], Sorbonne University

Arnaud Latil is Maître de Conférences [HDR] at Sorbonne University and researcher at SND [Sciences, Normes et Démocratie, UMR 8011]. Author of *Le droit du numérique : une approche par les risques* [Dalloz, 2023], he teaches digital law, cybersecurity, and EU AI regulation. He coordinates summer modules on AI's legal challenges at Sorbonne and contributes to EU policy debates on risk-based AI compliance.

Dr. Bernard Long

Postdoctoral Researcher, University of Artois

Bernard Long is a Postdoctoral Researcher in Private International Law at the University of Artois. His research focuses on the interplay between fundamental rights, technical standards, and legal compliance under the EU AI Act. He holds a PhD from University College Cork and actively contributes to academic and policy discussions on AI governance in Europe.

Nathalie Nevejans

Professor of Law & Ethics of Robotics and AI, University of Artois

Nathalie Nevejans is Professor of Private Law at the University of Artois and holds the Responsible AI Chair. She specialises in the legal and ethical dimensions of robotics and artificial intelligence. She leads research within the Centre "Droit, Éthique et Procédures" (EA 2471) and is regularly invited as an expert in academic and policy forums. Her notable publication includes the Treatise on Civil Robotics Law and Ethics (LEH, 2017), awarded the Francis Durieux Prize in 2019.

Marco Pasqua

Researcher, Catholic University of the Sacred Heart of Milan

Marco Pasqua is a legal scholar specialising in AI law, EU regulation, and anti-SLAPP measures. He holds a PhD in International and EU Law and lectures at the Catholic University of Milan and LIUC – Cattaneo University. A qualified lawyer in Italy, he co-chairs the EAPIL Young Research Network. His research focuses on private international law, cross-border AI regulation, anti-SLAPP frameworks, and soft law.

Cécile Pellegrini

Associate Professor, Lyon Catholic University (UCLy)

Cécile Pellegrini is Associate Professor of Private Law and Academic Director of the LLM in International Business Law at UCLy. She is a member of the CONFLUENCE research unit and associate researcher at CREDIP (Lyon 3). Her research focuses on digital contracts, AI regulation, private international law, European consumer law, and AI governance. She co-authored *Governance of Artificial Intelligence in the European Union* (Bruylant, 2023) and co-organised the "Achieving AI Act Compliance" workshop.

Yves Raisin

Global Data Protection Officer [DPO], bioMérieux

Yves Raisin is Global Data Protection Officer at bioMérieux. He oversees compliance with data protection and privacy regulations, including GDPR, and acts as the main liaison for data protection matters within the company. His work focuses on data governance, ethical use of health data, and ensuring transparency and accountability in processing personal information.

Béatrice Schütte

University Lecturer, University of Helsinki; Visiting researcher, University of Lapland

Béatrice Schütte holds a PhD in Law from Aarhus University. She conducts research on AI regulation, emotional AI, and civil liability frameworks. Her recent work includes a chapter on "Damage Caused by Emotional AI" [Springer]. She frequently contributes to conferences and publications on the legal and ethical aspects of AI.

Juliette Sénéchal

Professor of Private Law, University of Lille & Inria [Spirals Project]

Juliette Sénéchal is a specialist in digital law, AI ethics, and European private law. She holds a PhD in Law and has been seconded to Inria's Spirals project since 2023 for interdisciplinary research. She organises workshops on neuro-ethics, e-voting, and AI regulatory challenges, and frequently contributes to EU policy events and academic journals on AI governance.

Gaurav Sharma

International AI Policy and Advocacy Advisor

Gaurav Sharma is an International AI Strategy Consultant and Policy Advisor, with associations to Gates Foundation, Wadhwani AI, German Development Cooperation [GIZ]. He serves as Policy Advisory Fellow at the Centre for Responsible AI [CeRAI.in]. Gaurav is a former Humboldt fellow, and a tech-policy fellow. His career spans AI in the development sector, AI diplomacy and IT with expertise on AI governance, AI and global developmental agenda. Gaurav holds advanced degrees in international security, human rights law, and information technology.

Jean-Marc Van Gyseghem

Deputy Director & Scientific Coordinator, CRIDS, University of Namur — Senior Lecturer, University of Namur & UCLy — Attorney-at-Law, Brussels Bar

Jean-Marc Van Gyseghem is Deputy Director and Scientific Coordinator of the CRIDS (Research Centre on Information, Law & Society) at the University of Namur. He is Senior Lecturer at both the University of Namur and Lyon Catholic University (UCLy) and an attorney-at-law at the Brussels Bar. His research focuses on ICT law, data protection, digital health, and AI regulation. He directs advanced master programs and professional training in ICT law and is actively involved in EU-funded projects such as TEF-Health (Digital Europe Programme, grant No. 101100700).

Lamprini Xenou

Associate Professor, Université Paris-Est Créteil (UPEC)

Lamprini Xenou is Associate Professor of Public Law at UPEC and co-chair of the Marchés, Institutions, Libertés research group. A prize-winning PhD graduate from Paris II Panthéon-Assas, her research focuses on the direct effect of the EU Charter and AI-related public-law challenges. She regularly publishes on EU digital regulation and chaired panels including the 2025 “Achieving AI Act Compliance” workshop.