HOLISTIC EXAMINATION OF THE ARTIFICIAL INTELLIGENCE REGULATION (EU AI Act or AAI).

Essential aspects. Protected and vulnerable subjects

Carmen Muñoz García
Subproject 1.1. Mapping of AI Governance - Recommendations and Regulation
AI Governance (Project #1)" FOR GOOGLE.org
Delivered 14.07.2024

REGULATION (EU) 2024/1969
13 June 2024

CHAPTER I. General provisions Article 1.

1. *The objective of this Regulation is to improve the functioning of the internal market and to promote the adoption of human-centred and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety and fundamental rights enshrined in the Charter, including democracy, the rule of law and the protection of the , against the detrimental effects of AI systems (hereinafter 'systems') in the EU and to support innovation.*

2. *This Regulation establishes:*

(a) harmonised rules for the placing on the market, putting into service and use of AI systems in the EU;
(b) prohibitions on certain AI practices;
(c) specific requirements for high-risk AI systems and obligations for operators of such systems;
(d) harmonised transparency rules applicable to certain AI systems;
(e) harmonised standards for the placing on the market of general purpose AI models;
(f) rules on market monitoring, market surveillance, governance and enforcement;
(g) measures in support of innovation, with a particular focus on SMEs, including start-ups.

Article 2. Scope of application

Article 3. Definitions

Article 4. Literacy

PART 1

I. Key issues

1. The approach to take into account: studying the AI Act from a holistic perspective

The **EU AI Regulation, a** global regulatory milestone, **should be considered as a whole** for AI systems and for general-purpose AI models. As such, it should take into account that any AI tool developed and/or used in the EU, with some exceptions, is subject to the AI Act in its practical entirety. Hence, it considers that **its study should be approached from a holistic perspective** since any AI system or model subject to the Regulation is linked to the objectives predetermined by the EU co-legislators. The regulation as a whole contributes in a systematic or orderly manner to meeting them. Namely: (i) it is a uniform legal framework for the development and use of AI that promotes human-centred, ethically based AI, which, aimed at being safe and reliable, must guarantee consolidated rights and principles in the EU; (ii) furthermore, any action, implementation or development of systems in the internal market must be carried out in the knowledge that it is intended - by ensuring legal certainty - to facilitate investment and provide support for innovation (including SMEs and start-ups).

However, although **the study will be done in parts** - for purely systematic reasons, taking into account the progressive implementation of the Regulation - **this regulatory framework cannot be understood as a sum of AI systems and models**, with watertight regulation for each tool, with its own independent requirements and obligations. The evolution of this technology and the possibility of modifying its capabilities and risks determine that **any system**, given its properties, its self-learning, its malleability, the complexity and opacity of the tool, **is subject to meeting the EU's objectives** (all of the above). The regulation, in its entirety, is aimed at meeting these objectives, hence the need for an appropriate framework of market surveillance, monitoring, evaluation and control throughout the life of the tool. And in connection with all this, appropriate systems of governance, fines and sanctions in the field of public law should be foreseen. Also, a specific regulation of liability for damages to injured parties for the defects or negative consequences of AI tools which, in these cases, affect the sphere of the injured party.

2. Article 1. Objective of the Regulation

The objective set out in the first article, in the general provisions which are applicable to any AI system targeted by the Regulation, is the *raison d'être* of this regulation. It is to **improve the functioning of the** EU **internal market** and to promote the **adoption of human-centred and reliable AI**. To this end, it is necessary to ensure at the same time a high level of protection of health, safety and fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the detrimental effects of AI systems in the EU, as well as to encourage and support innovation (considering 176). The EU's aim to be a leader in this technology and, in addition, to preserve fundamental rights and consolidated values, seeks to build confidence in people to use this technology and thereby promote human welfare (explanatory memorandum to the EC proposal).

Broadly speaking, the Regulation seeks to strike a balance between fostering innovation in the internal market and the protection of individuals. To achieve these objectives, it creates a uniform legal framework, with harmonised rules for the placing on the EU market, putting into service and use of AI systems and AI models for general use, sets out

obligations and requirements for regulated entities in the EU and rules for market monitoring, market surveillance, governance and enforcement (sanctions, fines and civil liability). In addition to favouring innovation, also for SMEs and start-ups. If AI improves people's lives, society and the economy, it should benefit everyone. It is a tool for human beings and must ensure their safety and well-being (point 6). Hence the importance of a rigorous regulation that upholds ethical principles and EU values (considering 8, *in fine*). Consequently**, a level playing field and effective protection of the rights and freedoms of individuals must be ensured throughout the EU** (cf. point 21).

3. Article 2. Scope of application and derogations

To whom does the Regulation apply, who are the regulated entities, to which systems does it apply, and which systems are excluded from this regulatory framework? This is what Article 2 deals with in its entirety, on the one hand to include in its scope of application, on the other hand the enumeration is aimed at excluding certain systems or AI models.

AI systems and AI models falling within the scope of the Regulation are all those listed in Article 2(1) of the Regulation. Thus, it **shall apply to** (verbatim transcription):

a) suppliers placing on the market or putting into service AI systems or placing on the market AI models in general use in the EU, irrespective of whether those suppliers are established or located in the EU or in a third country;
b) those responsible for the deployment of AI systems that are established or located in the EU;
c) providers and deployers of AI systems that are established or located in a third country, where the output results generated by the AI are used in the EU;
d) importers and distributors of AI systems;
e) manufacturers of products that place an AI system on the market or put it into service together with their product and under their own name or trademark;
f) authorised representatives of suppliers not established in the EU;
g) affected persons who are located in the EU.

In addition, according to Article 2(2) of the EU AI Act, it shall apply: "to AI systems classified as high-risk AI systems in accordance with Article 6(1) and relating to products covered by EU harmonisation legislation listed in Section B of Annex I, only Articles 6(1), 102 to 109 and 112 shall apply. Article 57 shall apply only to the extent that the requirements for high-risk AI systems under this Regulation have been integrated into those EU harmonisation legislative acts.

In addition, the scope of application has an impact on the rest of the world, since any third party outside the EU that intends to operate in European commerce will be subject to the EU AI Act. Hence, its application extends to anyone operating with AI systems or models in our market. This is the so-called "Brussels effect".

They are **excluded from** the application of the Regulation (Art. 2(3) et seq. of the EU AI Act):

- Those areas that fall outside the EUD. Among others, the competences of the Member States (MS) in the field of national security (Art. 2(3) of the EU AI Act);
- AI, whether or not placed on the market, which is used exclusively for military, defence or national security purposes, irrespective of the type of entity carrying out these activities (Art. 2(3) of the EU AI Act);
- To public authorities of non-EU countries and international organisations where such authorities or organisations use AI systems in the framework of international agreements or cooperation for the purposes of law enforcement and judicial cooperation with the EU or with one or more Member States, provided that such third country or international organisation offers sufficient guarantees with respect to the protection of fundamental rights and freedoms of individuals (Art. 2(4) of the EU AI Act);
- It shall not affect the application of the provisions on the liability of intermediary service providers set out in Chapter II of Regulation (EU) 2022/2065 (Art. 2(5) of the EU AI Act);
- AI systems or models developed and put into service for the sole purpose of scientific research and development Art. 2(6) of the EU AI Act);
- The Regulation shall not affect the protection afforded by the Data Protection, Privacy and Confidentiality of Communications Regulations (Art. 2(7) of the EU AI Act), nor the rules on consumer protection and product safety (Art. 2(9) of the EU AI Act);
- It does not apply to any research, testing or development activities relating to AI systems or AI models prior to their placing on the market or putting into service, although they must be carried out in accordance with the EUTL. As an exception to this derogation, and subject to the Regulation, testing under real conditions (Art. 2(8) of the EU AI Act);
- It does not apply to natural persons using AI systems as part of a personal, non-professional activity (Art. 2(10) of the EU AI Act);
- The Regulation shall not prevent the EU or the Member States from providing more favourable rules for workers as regards the protection of their rights in respect of the use of AI systems by employers (Art. 2(11) of the EU AI Act);
- It does not apply to AI systems disclosed under free and open source licences, as long as they are not placed on the market or put into service as high-risk AI systems or as AI systems falling within the scope of Article 5 or Article 50 (Art. 2(12) of the EU AI Act);

However, given the above and in view of rapid technological developments, the European Commission is expected to assess, at the latest 5 years after the entry into force - from August 2024 - whether it is necessary to modify this scope of application (considering 174).

4. Article 3. Definitions

The definitions provided by the Regulation are for the purposes of application, and although initially there were many fewer, both in the Commission's proposal (April 2021) and in the European Parliament's amendments (June 2023), it has been decisive to establish the concept of artificial intelligence system (**aligned with that proposed by the OECD** in December 2023), and that of general-purpose AI model. For the time being, let us bring up generic definitions that will be of great use to us in advancing this regulatory framework, the one provided for an AI system is intended to be sufficiently broad and flexible to adapt to future technological evolutions:

(1) "AI system" means a machine-based system that is designed to operate with varying levels of autonomy and that can exhibit post-deployment adaptability, and that, for explicit or implicit purposes, infers from the input information it receives how to generate output results, such as predictions, content, recommendations or decisions, which may influence physical or virtual environments (Art. 3(1) of the EU AI Act);

In clear cohesion with the previous one, although with the intention of distancing it from the one contained in the first paragraph, the last definitions provided by Article 3 include ″general purpose AI model", making it clear that this tool is NOT an AI system on its own:

(63) "General-purpose AI model: an AI model, also one trained on a large volume of data using self-monitoring, that exhibits a considerable degree of meaningful generality and is capable of competently performing a wide variety of different tasks, regardless of how the model is introduced to 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 prior to market introduction.

The above definition is followed by others that are clearly cohesive with it, and which determine the enormous relevance that this model represents for the European co-legislators:

(64) "High-impact capabilities: capabilities that match or exceed the capabilities demonstrated by the most advanced general-purpose AI models.

(65) "Systemic risk" means a risk specific to high-impact capabilities general-purpose AI models, which have a significant impact on the EU market due to their scope or actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights or society as a whole, which may propagate on a large scale along the entire value chain".

(66) "General purpose AI system" means an AI system that is based on a general purpose AI model and that can serve a variety of purposes, both for direct use and for integration into other AI systems.

(67) "Floating point operation" means any mathematical operation or task involving floating point numbers, which are a subset of the real numbers normally represented on computers by an integer of fixed precision raised by the integer exponent of a fixed base".

5. Article 4. Literacy

This article, in a single paragraph, closes the general provisions of the text. Certainly, the final content of this article has come as a surprise to all of us who have followed the pre-legislative proceedings preceding the final text of the AI Regulation. Previously, the European Parliament's amendments (June 2023) **introduced in this article**

two relevant aspects, and if one was **AI literacy**, it was preceded by a first paragraph dealing with the **general principles applicable to all AI systems.** Well, these, in general provisions, would have been of obligatory application to all these tools, and yet they have been removed as a rule and moved to the recitals of the text for the purposes of being taken into account (principles that are also aligned with the OECD), and which in general terms are as follows:

- Human intervention and surveillance;
- Technical soundness and safety;
- Privacy and data governance;
- Transparency;
- Diversity, non-discrimination and equity;
- Social and environmental welfare.

As far as literacy is concerned, there is no doubt that meeting the obligations set out in the Regulation, and the objectives envisaged by the co-legislators, requires that the persons in charge of ensuring that AI systems and models have an adequate level of literacy, that they have the necessary support and authority, and that they are backed up by national and European authorities. However, in the context of the Regulation, this AI literacy, which should provide all relevant actors in the AI value chain with the necessary knowledge to ensure adequate compliance and proper implementation, obliges the European Commission to expand this claim. This is not only about literacy, but also, and in connection with the above, about generating public awareness and understanding of the benefits, risks, safeguards, rights and obligations in relation to the use of AI systems. In doing so, the EC will be supported by the European AI Committee. In addition, in cooperation with relevant stakeholders, the Commission and Member States should facilitate the development of voluntary codes of conduct to promote AI literacy among those involved in the development, operation and use of AI (considering 20).

Its definition is provided in Article 3(56):

> AI literacy' means the skills, knowledge and understanding that enable providers, deployers and others concerned, taking into account their respective rights and obligations in the context of this Regulation, to carry out an informed deployment of AI systems and to become aware of the opportunities and risks posed by AI, as well as the harm it may cause;

Undoubtedly, **the literacy process** (training and understanding of AI) **reaches any recipient of AI**, not only those who are obliged under the EU AI Act - the **operators -** but also **those who are not** (we will refer to both below), since all of us who make use of AI are exposed to the risks of the systems and models of this technology. Let us not forget what has already been said about **ensuring a level playing field and effective protection of the rights and freedoms of individuals throughout the EU** (considering 21). If it is essential that the regulatory framework provides mechanisms to minimise risks for individuals, special attention is needed for those who, due to circumstances of any types, find themselves in a vulnerable situation.

6. Parties subject to and protected by the Regulation

Having defined the above, Article 3(8), together with other definitions, includes under a common denomination all subjects that, by developing, implementing or using AI systems in the EU market, are subject to essential principles of the EUD, to multiple obligations and requirements, and therefore also to responsibilities. It is the operator:

> "Operator: a supplier, manufacturer of the product, deployer, authorised representative, importer or distributor.

And although each of these subjects is defined individually in the preceding paragraphs of Article 3, in paragraphs 3 to 7 (both inclusive), the term includes any actor involved in the AI value chain. However, its enumeration in general terms makes it possible, and at the same time difficult, to find out which subject contributes value to the tool and which subject is potentially liable for the damage it causes.

Nonetheless, among all these subjects, the EU AI Act will establish that for general-purpose AI models, it will be the providers of these models who must comply with the obligations set out in the text, regardless of whether these models are integrated in an AI system or form part of an AI system. However, it should be noted that, in Article 25, for "high risk" AI systems *any distributor, importer, deployer or third party* shall be considered a supplier in certain circumstances: (a) when he places his name or trademark on a high-risk AI system already placed on the market or put into service; (b) when he substantially modifies a high-risk AI system which continues to merit such a qualification according to Article 6; and (c) if the system or model in question, placed on the market or put into service, becomes a high-risk AI system.

Without a doubt, establishing the obligated parties is not a trivial issue; it will serve to determine who should be liable throughout the life cycle of the system or model when it causes damage or harm. At this point, establishing the subjective imputation or attribution of any harmful event will be key, as will be, beforehand, establishing which is the event -by action or omission- that deserves the role of cause in the production of any damage. But this issue, of great importance, requires separate study.

Once the obligated subjects have been delimited (and thus, who the possible responsible parties are), it is necessary to know **who are the subjects to be protected by the Artificial Intelligence Regulation**. To this end, let us bear in mind that the regulation aims to generate security and reliability in the development of AI through rigorous rules, governance mechanisms and systems of sanctions and responsibilities **to protect the rights of natural persons**. Moreover, the regulatory approach **based on the potential risk that the systems can produce for people** - health, security, fundamental rights, and even to democracies or the rule of law - determined the classification of systems into four types: (i) "unacceptable risk"; (ii) "high risk"; (iii) "limited or medium risk"; and (iv) "low or minimal risk" and **put the focus on protecting people**. This confirms who are real beneficiaries of the protection sought by this regulatory framework.

The Regulation, which **aims to protect natural persons in all cases** (although it cannot be ruled out that it also protects legal persons who are not operators), **does require greater guarantees - or so it claims - for those who are considered vulnerable**. There are mechanisms that appear to be sufficient to protect them with greater guarantees, but

there is still a long way to go to fully implement the instruments that protect the weaker party (governance, sanctions, etc.).

Having said that, **to which vulnerable persons does the EU AI Act refer?** Vulnerable persons are only mentioned in 6 recitals (hereinafter, considering) and 4 articles (hereinafter, art.) of the Regulation. Why is this the case? It makes clear, from its initial proposal by the European Commission (21 April 2021), that it is a priority objective of the Regulation: (i) to achieve technological leadership, and (ii) to ensure that AI is developed in an ethical and trustworthy manner, in line with European values, principles and fundamental rights. The rights and values enshrined in European law are preserved and AI must be human-centred.

I am transferring here the recitals of the text and the specific mention in the rules:

- (considering 60): AI systems used in migration, asylum and border control management, as they may affect persons in vulnerable situations, must be accurate, non-discriminatory and transparent, in order to ensure that the fundamental rights of the persons concerned are respected, and in particular their right to free movement, to non-discrimination, to personal privacy and protection of personal data, to international protection and to good administration.
- (considering 67): in terms of data quality and access to them, it must be ensured that the high-risk AI system does not become a source of discrimination. Data governance requirements should ensure that there is no bias, that existing discrimination is not perpetuated and amplified, and in particular with regard to persons belonging to certain vulnerable groups.
- (considering 93): those responsible for the deployment of a high-risk AI system can, by virtue of their position, detect potential risks not detected before by having a better knowledge of the context of use and of vulnerable persons or groups.
- (considering 132): Certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception, regardless of whether they are high-risk or not. They should therefore be subject to specific transparency obligations... In applying this obligation, the characteristics of natural persons who are vulnerable due to age or disability shall be taken into account.
- (considering 141): It is important to reduce risks and to allow monitoring by the competent authorities and therefore require potential suppliers to submit to the supervisory authority a plan of the test under real conditions... and require additional safeguards for persons belonging to certain vulnerable groups.
- (considering 165): Providers of non-high risk AI systems are encouraged to create codes of conduct. In addition, ALL providers and, where appropriate, developers of ALL systems - whether high-risk or not - and AI models are encouraged to apply, on a voluntary basis, additional requirements such as the EU Ethical Guidelines (2019) for Trustworthy AI,... including consideration of vulnerable persons and accessibility for persons with disabilities....
- (Art. 9.9): on "Risk management system" in high-risk systems: ...where the risk management system is implemented ... the high-risk AI system is likely to adversely affect persons under the age of 18 and, where appropriate, other vulnerable groups.
- (Art. 60): on "Testing of high-risk AI systems under real conditions outside controlled AI test spaces": suppliers (also potential suppliers) may test under real conditions when, among others, the following conditions are fulfilled

, that persons belonging to vulnerable groups due to age or disability are adequately protected.

- (Art. 79.2): on "Procedure at national level for systems presenting a risk" (to the health, safety or fundamental rights of individuals). Particular attention should be paid to AI systems that present a risk to vulnerable groups.
- (Art. 95.2, e): "Codes of conduct and guidelines": the AI Office and the Member States shall encourage and facilitate the development of codes of conduct, the voluntary application of codes of conduct for compliance by non-high risk systems with the requirements of high-risk systems, and shall take into account, among other elements, the assessment and prevention of harm from AI systems to vulnerable persons and groups.

7. Study plan in line with the progressive implementation of the Regulation

In view of the above, it is the purpose of this team to study and analyse whether the rules proposed by the Regulation (AI systems and models, governance, sanctions, monitoring and surveillance) are sufficient to protect both natural persons and those who might fit in the context of vulnerable persons or groups.

For the optimal analysis, and leaving aside the systems prohibited by being expelled from the European EU and its regulation, we will start with those rules that will be initially applicable after the entry into force of the regulation (twenty days after its publication in the OJEU of 12 July 2024). From there, we will move on to the study and analysis of the rules that will be applied later:

(1) **General rule**: after the entry into force of the text, the Regulation will apply after two years, i.e. from 2 August 2026 (Article 113 EU AI Act).
(2) **Exceptions, numerous and very relevant**:
    - After 6 months (from 2 February 2025):
        - Chapters I (General Provisions) and II (Prohibited Systems), according to Article 113.
        - It is also intended, from the outset, to "promote AI literacy" (considering 20), the implementation of codes of conduct (of a voluntary nature), and which aim at the adoption by operators of non-high risk AI systems of the specific requirements set out in Chapter III, Section 2 for high-risk systems;
    - After 9 months, i.e. from 2 May 2025: the codes of practice for general-purpose AI models (voluntary or otherwise) involve the harmonised adoption, on a proposal from the European AI Bureau, of the obligations laid down for these models, according to Article 56(9);
    - One year after entry into force, as of 2 August 2025, they will apply:
        - Chapter III, Section 4: notifying authorities and notified bodies
        - Chapter V: General-purpose AI models (GPAI models)
        - Chapter VII: governance;
        - Chapter XII: Penalties, with the exception of Article 101 on fines for suppliers of GPAI models;
        - Article 78 (Chapter IX), on confidentiality;
    - At 24 months, in penalties, Article 101;

- At 36 months, Article 6(1) on high-risk AI schemes and the corresponding obligations in the Regulation.

Hence, in this first year we will look at general provisions, codes of conduct, codes of good practice and models of GPAIs. In the final year (first semester), the study will focus on high-risk systems, governance and sanctions. However, other relevant legislative and non-legislative initiatives promoted or adopted by non-EU states or international institutions will be monitored. The aim is to identify whether other initiatives provide more and better protection for natural persons and, in particular, for vulnerable persons or groups. For all of this, there will be no shortage of real case studies that highlight the value of the exhaustive and protective regulation promoted by the EU. The aim is: (i) to analyse the regulation of AI and its impact on all the Member States; and (ii) to identify whether regulatory rigour, governance, systems and surveillance mechanisms are capable of protecting people and the vulnerable to the extent intended.