HOLISTIC REVIEW OF ARTIFICIAL INTELLIGENCE REGULATION

General-purpose AI models

Contents

Note: if you know how to use the table of contents function, the title formats are automatic.

Under no circumstances should you format any type of title, as it is very easy to make them uniform later.

It is important that each group has someone who manages the table of contents index to submit the final version.

Document format:

- Include the table of contents on the first page. If it is long, include the cover page and table of contents on the second page.
- Document title: Arial, 25, bold.
- Subtitle: Arial, 25.
- Level 1 title: Arial, 16, bold.
- Level 2 title: Arial, 14, bold.
- Secondary title: Arial, 12, bold and italic.
- Text font size: Arial, 12.
- The authorship and references to the Project will be unified.

1. Regulation of AI systems. A horizontal regulatory approach

The European legislator is promoting the regulation of artificial intelligence systems from a risk-based approach, depending on whether the tool is more likely to cause harm to "health, safety, and fundamental rights enshrined in the Charter of Fundamental Rights of the European Union (hereinafter "the Charter"), including democracy, the rule of law, and the protection of the environment" (recital 1 of the IA Regulation).

The threat or greater or lesser danger that any of these systems pose to individuals and established values or principles led the European Commission (proposal for a Regulation on artificial intelligence, April 21, 2021) to classify them into systems: (i) "unacceptable risk"; (ii) "high risk"; (iii) "limited or medium risk"; and (iv) "low or minimal risk." This regulation, which is rigorous and proportional to the risk posed by the tool, left general-purpose AI models (GPAI models) outside the scope of specific regulation. These models, which are more recent, were not included in the initial proposal for the Artificial Intelligence Regulation.

In this vein, the horizontal regulatory approach makes it possible to classify systems according to the risks that AI poses to **protected goods and rights of the individual**: mainly "health, safety, and fundamental rights." If they pose a threat (prohibited) or produce a greater than desirable risk (high-risk or limited-risk systems) throughout the tool's life cycle, the mandatory nature of the regulation and governance systems encourage immediate responses to prevent harm to individuals.

This classification will determine that, in view of the potential risks to individuals, the Union's firm intention is to prevent or minimize the risks of these systems. To this end, mandatory, binding rules are established, with obligations for the operator and rigorous technical requirements, not only *ex ante*, but also *ex post* from the moment any AI system is put into circulation.

2. Following the horizontal regulatory approach, general-purpose AI models burst into the scene

With the emergence of ChatGPT in November 2022, large generative artificial intelligence models showed within a few months that, with their extensive

capabilities to create content (multimodal and multifunctional), even if they were not trained to do so, they could also create risks for individuals and violate fundamental rights, health, safety, democracies, the rule of law, or the environment.

During this period, the proposal to regulate artificial intelligence in Europe continued to move forward, although European legislators could not ignore this tool, which they initially called "foundational models," although they were ultimately referred to in the Regulation as "GPAI models." If technology is at the service of human beings, to provide them with benefits and optimize results, there is no place for regulation that exposed individuals to the risks that large language models represent per se.

This category, and its place in the Regulation, following the emergence---and socialization---of Generative AI, led to the initiative of European co-legislators to insert specific regulation for GPAI models. Recently discovered and in the experimental phase throughout 2023, they have been subject to definitive regulation in the European Artificial Intelligence Regulation (June 13, 2024).

Initially, general-purpose models are regulated outside and separately from systems such as GPAI models, which may be simple and subject to fewer obligations and requirements than those referred to as "systemic risk" models. Basically, they are inserted into GPAI systems, and for them, the **transparency** required in Article 50 of the RIA for "non-high-risk systems" becomes the central axis of their development and operation.

Their widespread and generalized use, available to anyone, regardless of their knowledge of the tool or their intentions in using it, has highlighted that they can potentially cause harm, not only to their users, but also to third parties who are not using the tool.

Therefore, even if AI models **in their initial phase are integrated into medium-risk systems**, there is nothing to prevent them from eventually being incorporated into high-risk systems. Undoubtedly, whether it affects any of the protected assets or rights is key, but it will depend not only on whether it relates to critical areas, such as health, but also on whether the potential risk is related to the function that the tool performs in the specific area.

Let us imagine, on the one hand, a system for assigning medical shifts or a tool that assigns appointments according to pre-established criteria. Even though these operate in the field of health, they do not put it at risk.

2.1 Classification

In Article 3 of the AIAD, after multiple definitions relevant to the uniform application of the regulation in the European Union, the legislator, in the last of the contributions it provides for the purposes of the Regulation, defines what is meant for the purposes of this regulation by "general-purpose AI models" and what "general-purpose AI systems" are:

> (63) "General-purpose AI model: an AI model, including one trained with a large volume of data using large-scale self-supervision, that exhibits a significant degree of meaningful generality and is capable of competently performing a wide variety of different tasks, regardless of how the model is placed on the market, and which can be integrated into various downstream systems or applications, except for AI models used for research, development, or prototyping activities prior to their placing on the market."
>
> (66) "General-purpose AI system: an AI system based on a general-purpose AI model and capable of serving a variety of purposes, both for direct use and for integration into other AI systems."

Both definitions were incorporated following the political agreement of the European co-legislators (December 2023), and in them, the dependence of the latter on AI models is clear.

Initially, and currently, as far as we know, conversational *chatbots*, "large generative AI models are a typical example of a general-purpose AI model"[1].

Regarding these models, the EU legislator has not only proceeded to incorporate them in the final stages of political negotiation (trilogues September-December 2023) but has also classified GPAI models in the final draft according to whether they pose a "systemic risk." When are we dealing with these higher-risk models?

> (i) Those that have high-impact capabilities, according to Article 3(64), assessed in accordance with the parameters currently set out in Article 51(2) of the RIA. This is the case when the cumulative amount of computation used for training, measured in floating point operations, exceeds 10^25.
>
> (ii) And, in addition, they present a specific risk with significant repercussions on the Union 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 spread on a large scale throughout the entire value chain of the AI model, as stated in Article 3, paragraph 65).

Specific regulations can be found in **Title V of the AI Act**, which contains four sections with only six articles, from 51 to 56. These include the **classification of the model according to risk**, which will be relevant because, depending on the risk, the provider will be required to comply with more obligations, and the model will be required to meet a greater number of requirements.

2.2 Legal obligations and requirements

The designation of models as "systemic risk" determined that the former, which lack "probability," should be referred to as "simple" general-purpose AI models[2]. Let us address the effects of giving one designation or another to the model[3].

2.2.1 Obligations and technical requirements for general-purpose AI models or "simple GPAI"

These are not systems; their essential functional characteristics are their generality and ability to perform a wide range of different tasks. They can be introduced onto the market and marketed in various forms, and can be modified, refined, and even transformed[4]. However, they do not have the capabilities of "systemic risk" AI models, but they may undergo changes that could alter their initial designation, including being integrated into high-risk systems or being prohibited. They are covered in Title V, **Section 2,** under the heading "Obligations of providers of general-purpose AI models," specifically in Articles 53 and 54:

(i) Article 53. Obligations of providers of general-purpose AI models

Article 53 specifies the obligations for **all providers,** regardless of the model they incorporate or market in the Union. In this case, the following obligations are established as mandatory and general:

a) Preparation and maintenance of up-to-date **technical documentation of the model, including information on the training, testing, and evaluation process**, in accordance with Annex XI. This documentation must be made available, upon request, to the European AI Office and the competent national authorities (Article 53(1)(a).

b) They shall also make the documentation **available to downstream providers of AI systems who intend to integrate the** general-purpose AI **model** into their AI systems. This documentation shall, in a clear and understandable manner, enable the capabilities and limitations to be known (Article 53(1)(b)). In addition to being obliged **to inform and comply with current legislation on intellectual and industrial property,** confidential business information, or trade secrets in accordance with European and national law (Article 53(1)(c)

We note that providers of "simple" GPAI models that **are disclosed under a free and open-source license** allowing access, use, modification, and distribution of the model, and whose parameters, including weights, information on the model's architecture, and information on the model's use, are made available to the public, are exempt from these obligations. However, we can state that **this exception does not apply** in the case of models "with systemic risk" (Article 53(2)), nor when there is any type of consideration or monetization of the model (recital 103).

(ii) Article 54. Authorized representatives of providers of general-purpose AI models

In addition to the above, when dealing with any supplier from a third country (not a member of the European Union), this operator has the prior obligation to designate an authorized representative before introducing a general-purpose AI model. Both the supplier and the authorized representative are the subjects who, according to the Regulation, are responsible for the application of the rule. Therefore, they must be adequately informed about the AI system or model that the operator wishes to place on the market. It is the operator who must provide the authorities with all the necessary information about the model, including documentation, and prove that the tool meets the requirements of the European standard. In short, they must be able to be held accountable when required[5].

(iii) Technical requirements of Annexes XI and XII of the AI Act

For all GPAI models, without distinction, technical requirements must be met, and documentation must be completed with clear and accurate information.

In this regard, the third-to-last and second-to-last annexes of the Regulation specify what this documentation is, considering, in addition, that a subsequent supplier may be involved[6]:

a) **Annex XI,** "Technical documentation referred to in Article 53(1)(a) -- technical documentation for suppliers of general-purpose AI models." Two aspects of this documentation should be clarified: first, the general requirement contained in this annex, specifically in section one, which applies to all suppliers, whether they supply "simple" or "systemic risk" GPAI models; the second aspect to be taken into account is that such technical documentation relating to the model must be made available to the AI Office and to the national authorities with competence in this area;

b) **Annex XII** "Transparency information referred to in Article 53, paragraph 1(b) -- technical documentation from suppliers of general-purpose AI models for downstream suppliers who integrate the model into their AI system," and which must contain a minimum amount of information (in addition to a general description of the model, it must contain a more detailed description of the elements of the model and its development process, including other requirements). This documentation and information must be provided by the initial "supplier" (Article 3(3)) to the "downstream supplier" (Article 3(68)).

2.2.2 Obligations and technical requirements for general-purpose AI models "with systemic risk"

In the **first section of Title V (Articles 51 to 56 of the IAIR),** from the beginning of the regulation of "general-purpose AI models," the first two articles provide a unique regulation for models "with systemic risk." These are *classification rules,* with criteria for determining when the tool may pose more serious risks to individuals, society, or the economy. To this end, the European legislator determines when it will be designated as such and establishes the procedure for doing so. Articles 51 and 52 will be supplemented by Article 55 to establish the obligations required by providers, and Annex XIII concerning technical requirements:

(i) Article 51. *Rules for classifying general-purpose AI models as general-purpose AI models with systemic risk*.

In this Article 51, based on the generic definition of general-purpose AI models (Article 3(63) of the IA Regulation), and taking into account that the tool's high-impact capabilities (Article 3, paragraph 64 of the IAIR)[7] affect its classification as "General-purpose AI models with systemic risk" (Article 3, paragraph 65 of the AI Act)[8], essential criteria to be taken into account for this classification are established. On the one hand, **an objective criterion:** (**i**) for **models with high-impact capabilities** *assessed using appropriate technical tools and methodologies, such as indicators and benchmarks,* (**ii**) and**, in addition, whose capabilities have a significant impact on the internal 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.*

The designation as models with systemic risk will determine a more stringent regulatory regime in terms of the tool's technical obligations and requirements. However, the designation as models with systemic risk can be achieved in two ways: first, **via a rebuttable** *presumption,* in that it admits evidence to the contrary, and second**, via imposition by the Commission**.

(ii) Article 52. *Procedure.*

It is inevitable that, when an IPR model meets the objective requirement referred to in Article 51(1)(a) -- in terms of having *high-impact capabilities*, or *an equivalent impact* considering the criteria set out in Annex XIII -- the provider must notify the Commission. Thus, the **obligation to notify the Commission** "without delay and, in any case, within two weeks of meeting that requirement or of becoming aware that it will be met" that it has exceeded or will exceed the threshold is established. This determines, at first glance, the presumption that the model must bear a new designation, that of a model "with systemic risk."

If the provider fails to notify: **the Commission may decide** to classify the model as "systemic risk" and may also require compliance with the obligations imposed by the Regulation on providers of general-purpose AI models through various measures, including the **imposition of administrative fines**.

(iii) Article 55. Obligations of providers of general-purpose AI models with systemic risk.

The third section includes a single article regulating GPAI models with systemic risk. On the one hand, the first paragraph of Article 55, paragraph 1, refers us to the obligations established for any GPAI model in Article 53[9], so that the documentation and information obligations must also be fulfilled by providers of GPAI models "with systemic risk." In addition, other **additional obligations** are added**,** as specified in the same Article 55, paragraph 1 of the RIA:

a) Perform **model assessment in accordance with standardized protocols and tools**, which involves documenting that adversarial simulation tests have been carried out in order to reduce potential systemic risks.

b) **Assess and take the necessary measures to reduce the potential systemic risks** that may arise from the development, introduction, or use of the model in the internal market.

c) Monitor, **document, and report serious incidents and corrective measures to resolve them**, both to the AI Office and to the competent national authority. In short, be vigilant and report serious incidents without delay.

d) Take the necessary measures to ensure optimal **cybersecurity** protection for the model.

(iv) Technical requirements of Annex XIII of the AI Act

In addition to all the obligations for general-purpose AI models, there are those specific to this tool, which presents a systemic risk. The same applies to the technical requirements set for these models. Thus, in the last annex to the Regulation, specific criteria are added to those common to Annexes XI and XII (the latter from the initial supplier to the subsequent supplier):

c) **Annex XIII** "Criteria for the classification of general-purpose AI models with systemic risk referred to in Article 51." This is done in order to determine whether a model has capabilities with a high impact or an impact equivalent to those set out in Article 51(1)(a) of the IAIR. For this classification, the Commission will take into account (and I transcribe, almost verbatim, the criteria or indicators included in the aforementioned annex):

a. Number of model parameters
b. Quality or size of the data set used, measured, for example, through crypto tokens.
c. The amount of computation used to train the model, measured in floating point operations or indicated by a combination of other variables, such as the estimated cost of training, the estimated time required, or the estimated energy consumption for such training.
d. Input and output modalities of the model, whether same modality or multimodal (e.g., text to image), and 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. Benchmarks and evaluations of the model's capabilities, also considering the number of tasks without additional training, adaptability to learn new and different tasks, degree of autonomy and scalability, and tools to which it has access.

f. The impact on the internal market, considering its scope. For this purpose, the parameter is set that the model must be made available to at least 10,000 registered professional users in the EU.

g. Number of registered end users.

These, apart from the common technical requirements, are **part of the design and development process** of any AI model, in that the clear and transparent documentation and information that must accompany it contains all the parameters and indicators that will help to understand the model's capabilities and the potential risks it generates.

3. Codes of good practice

Article 56 of the Regulation, the last of Chapter VI, in the **fourth and final section**[10], in order to facilitate compliance and standardization of the obligations to be fulfilled by the provider of this tool, whether it is a simple GPAI model or one with systemic risk, promotes the creation of *codes of good practice* to help implement and comply with the duties and requirements demanded by the standard.

It is an essential tool for compliance with the obligations set out in the preceding articles and, given its relevance, the Commission may, by means of implementing acts, approve a code of good practice and grant it general validity in the EU (recital 117 of the AI Act)[11].

Any provider of general-purpose AI models, regardless of the risk involved, may use codes of good practice under the terms of Article 56. This practice, in the absence of a harmonized European standard, favors those who implement it, as it implies compliance with the obligations imposed by the Regulation according to the model in question, whether those contained in Article 53(1) or Article 55 of the Regulation.

For the time being, the supplier has the power to adhere to codes of good practice which, it should be noted, must include at least the obligations set out in Articles 53 and 55, depending on the model of AI for general use in question.

Compliance with these obligations is unquestionable, in one way or another, as they are ultimately considered a fundamental tool for compliance with the obligations imposed by the AIAR and must faithfully reflect that they comply with the purposes of the regulation.

However, Article 53(4) (for simple general-purpose AI models) and Article 55(2) (for models with systemic risk) refer to Article 56 of the IAO and offer **other alternatives** to the provider, insofar as it is up to them to prove that they comply with the obligations by other suitable means that will be submitted to the Commission for evaluation.

If the regulatory framework for GPAI models is to be applied one year after the entry into force of the RIA, in August 2025, the Code of Good Practice must be finalized by May 2, 2025, at the latest (Article 56, paragraph 9 of the RIA). These will be common rules for adopting the obligations set out in the Regulation, which are already being developed through consultations within the European Commission.

References

[1] Recital 99 of the AIA

[2] MUÑOZ GARCÍA, C. (2024) "General-purpose AI models and limited-risk and minimal-risk AI systems," in The European Regulation on Artificial Intelligence, edited by BARRIO ANDRÉS, M. Tirant lo Blanch, Valencia, 1st edition

[3] See the publication that summarizes the obligations and requirements set forth in the AIA, according to the model: MUÑOZ GARCÍA, C. (2024) "Taxonomy of general-purpose AI models. Probability of generating high-impact risks and the need to identify them," in Artificial Intelligence and Tort Law, edited by MORENO MARTÍNEZ, J.A. and FEMENIA LÓPEZ, P. J. Dykinson, Madrid, 1st edition (in press).

[4] Recital 97 of the AI Regulation.

[5] MUÑOZ GARCÍA, C. "Article 54 of the Artificial Intelligence Regulation," in Barrio Andrés, M, (coord.), Comments on the Artificial Intelligence Regulation, Editorial La Ley, Madrid, 2024 (available in October 2024)

[6] In Article 3, paragraph 68), the last of the definitions provided by the AIA, a "downstream provider" is defined as a provider of an AI system, including a general-purpose AI system, that integrates an AI model, regardless of whether the AI model is provided by the provider itself and is vertically integrated or provided by another entity under contractual relationships.

[7] The Regulation determines when AI models are considered tools with systemic risk when they have "high-impact capabilities: capabilities that match or exceed the capabilities shown by the most advanced general-purpose AI models" (Article 3, paragraph 64 of the AI Act).

[8] The Regulation classifies a model as having "systemic risk" when there is "a specific risk from the high-impact capabilities of general-purpose AI models, which have a significant impact on the Union 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 can spread on a large scale throughout the value chain."

[9] It will apply to any GPAI model, regardless of its capabilities and risks.

[10] MUÑOZ GARCÍA, C. "Article 56 of the Artificial Intelligence Regulation," Barrio Andrés, M, (coord.), Comments on the Artificial Intelligence Regulation, op. cit.

[11] The European Commission has opened a consultation period with a view to drawing up a Code of Practice covering the specific obligations contained in the Regulation for these models. The consultation period will be carried out with the aim of establishing rules relating to the required transparency, compliance with the obligations of Chapter V of the AIA, copyright and related rights, the identification of risks and assessment measures for systemic risks, as well as measures to mitigate them and internal risk

management and governance policies. About the consultation https://artificialintelligenceact.eu/es/introduction-to-codes-of-practice/