Published August 2024

Obligations for general-purpose AI models

By Phillip Lee and Uzma Chaudhry

If you were to read the European Commission's original AI Act proposal, published in April 2021, you would find it conspicuously devoid of references to general-purpose AI. With the benefit of hindsight, this might seem like a surprising omission. Yet, outside of the world of AI experts, few people had ever heard of general-purpose AI at the time the proposal was published.

Fast-forward to a little over one year later, OpenAI released ChatGPT to an unsuspecting public in November 2022, wowing them with its human-like, if sometimes unreliable, responses to their prompts. It quickly went viral, reportedly reaching 100 million users in just two months and becoming the fastest adopted consumer app of all time.

As a result, terms like large language models, generative AI and general-purpose AI began to enter the consciousness of European legislators, if not exactly the public consciousness. Clearly, the AI Act would need to regulate general-purpose AI, but how?

This was not an easy question to answer. The proposed law worked by placing AI systems into prohibited, high and low risk buckets to decide which rules to apply. However, by its very nature, general-purpose AI could be implemented across an unimaginably wide range of use cases that spanned the entire risk spectrum. The risks arising in any given scenario would necessarily depend on context, making it impossible to place general-purpose AI into a single risk bucket.

Consequently, Europe's legislators ultimately proposed an entirely new chapter of the AI Act dedicated specifically to regulating general-purpose AI models: Chapter V.

Distinguishing AI models from AI systems

As identified in Part 1 of this series, the difference between AI models and AI systems is critical.

This is because Chapter V sets out rules that address the use of general-purpose AI models. While the AI Act also defines the concept of a general-purpose AI system as a system based on a general-purpose AI model. This term is simply a subset of the broader concept of an AI system, and general-purpose AI systems are not addressed within Chapter V's rules.

Further, by specifying rules for general-purpose AI models, Chapter V takes a different regulatory approach from the one taken generally throughout the AI Act, which instead regulates AI systems, of which general-purpose AI systems are just one type. The rules applicable to an AI system, including any general-purpose AI systems, will be determined by whether they are prohibited, high or low risk.

This distinction is not accidental. According to Recital 97, "the notion of general-purpose AI models should be clearly defined and set apart from the notion of AI systems to enable legal certainty."

Article 3(63) of the act defines a general-purpose AI model as "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."

Therefore, to fully understand this definition, it is necessary first to understand what an AI model is and how it is different from an AI system.

The act does not define the concept of an AI model, but IBM helpfully explains "an AI model is a program that has been trained on a set of data to recognize certain patterns or make certain decisions without further human intervention." Recital 97 of the AI Act notes "AI models are essential components of AI systems" but "they do not constitute AI systems on their own." This is because "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."

An AI model can therefore be thought of as the program that powers the intelligence of an AI system, but it cannot be used on a stand-alone basis. Accordingly, an AI model must first be integrated with other software and/or hardware components, so users have a means to access and interact with the AI model via a user interface, such as using a dialogue box to submit prompts. The set of hardware and software components that integrate, and enable users to interact with, one or more AI models collectively comprise the AI system. For example, in very generalized terms, an autonomous vehicle can be thought of as an AI system that integrates multiple AI models to enable it to steer the vehicle, manage fuel consumption, apply brakes and so on.

What is a general-purpose AI model?

In general, the AI Act applies to AI systems, not AI models. As explained above, a general-purpose AI model:

- Is an AI model, not an AI system, although it may be integrated into an AI system.
- Is trained with a large amount of data using self-supervision at scale. For example, ChatGPT 3 was reportedly trained on at least 570 gigabytes of data or about 300 billion words.
- Displays significant generality and is capable of competently performing a wide range of distinct tasks.

However, the act only regulates AI models that are placed on the EU market. "AI models that are used for research, development or prototyping activities before they are placed on the market" are excluded from the definition of a general purpose-AI model under Article 3(63) and from the scope of the act under Article 2(8).

Types of general-purpose AI models covered by the act

Chapter V distinguishes between general-purpose AI models with and without systemic risk. This distinction reflects the need to have stricter regulatory controls for general-purpose AI models with systemic risk due to their potential for significant harmful effects if not closely regulated.

To this end, under Article 3(65) of the AI Act, systemic risk is defined as "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."

At first glance, this definition appears circular. A general-purpose AI model with systemic risk is one presenting risks that would have significant impact and are "specific to the high-impact capabilities of general-purpose AI models." However, the definition hints at the types of concerns AI Act legislators believe general-purpose AI could present, namely "negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale."

As to what these "negative effects ... propagated at scale" could include, Recital 110 lists "major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content."

It continues that these might result in "chemical, biological, radiological, and nuclear risks ... offensive cyber capabilities ... the capacity to control physical systems and interfere with critical infrastructure; risks from models of making copies of themselves or 'self-replicating' or training other models ... harmful bias and discrimination ... the facilitation of disinformation or harming privacy with threats to democratic values and human rights."

How to identify a general-purpose AI model with systemic risk

For the purposes of the AI Act, there are two ways for a general-purpose AI model to be deemed to present a systemic risk.

First, under Article 51(1-2), the general-purpose AI model must have "high impact capabilities," as evaluated by "appropriate technical tools and methodologies, including indicators and benchmarks."

For these purposes, a general-purpose AI model is presumed to have high impact capabilities if the cumulative amount of computation used for training is greater than 10^{25} floating point operations.

To put this in human terms, according to some estimates, the computational power of the human brain is approximately in the order of 10^{16} to 10^{17} floating point operations. However, this is a crude and imprecise comparison for all sorts of reasons, not least that, while considerably slower than a computer, the brain is capable of much greater parallel processing at much lower levels of energy consumption. Nevertheless, it does provide a simple way for nonengineers to picture the type of computing power concerned.

Second, a general-purpose AI model can be determined to have high impact capabilities by the European Commission, which it can do either on its own initiative or following a qualified alert from the Scientific Panel of Independent Experts created under Articles 51(1)(b), 68 and 90 of the act. In reaching such a determination, the Commission must have regard to certain criteria set out in Annex XIII.

The Commission must publish a list of general-purpose AI models with systemic risk per Article 52(6) and can adopt delegated legislation to amend and supplement the thresholds, benchmarks and indicators that determine what qualify as high impact capabilities under Article 51(3) to keep pace with evolving technological developments.

Obligations for providers of all general-purpose AI models

Providers of general-purpose AI models with or without systemic risk must comply with the obligations set out in Article 53 and Article 54 of the AI Act. These primarily address technical documentation requirements, the provision of transparency information to providers of AI systems that integrate the general-purpose AI models, compliance with EU copyright rules and the need for non-EU model providers to appoint an EU representative.

Providers of general-purpose AI models without systemic risk have fewer obligations than those with systemic risk. For that reason, while providers of general-purpose AI models without systemic risk only need to comply with Articles 53 and 54, providers of models with systemic risk have additional compliance responsibilities under Article 55.

Obligations that apply to all providers of general-purpose AI models, with or without systemic risk, include the following:

- → Prepare and maintain technical documentation about the general-purpose AI model, including its training and testing process and evaluation results, containing the mandatory information set out in Annex XI, listed in the Annex section below. The European Commission's AI Office and national competent authorities can require the general-purpose AI model provider to provide this documentation on request. See also Article 91(1).

- → Make certain information and documentation available to providers of AI systems that integrate the general-purpose AI model so they have a good understanding of the capabilities and limitations of the model and can comply with their own obligations under the AI Act. This must include the mandatory information set out in Annex XII, listed in Table 2.

- → Put a policy in place to comply with EU copyright and related rights rules. This should include a means to identify and comply, through state-of-the-art technologies, with any reservation of rights expressed by rights holders.

- Prepare and make publicly available a detailed summary of the general-purpose AI model's training content using a template provided by the AI Office that is not yet available as of the date of this article. This latter requirement has raised eyebrows among providers of general-purpose AI models over concerns that it may force them to reveal trade secrets about their training content.

The first two points above do not apply to providers of open-source general-purpose AI models unless they have systemic risk, provided these models can be used and adapted without restriction and that information about their parameters, including weights, model architecture and model usage are made publicly available.

In addition, and with more than a passing nod toward EU representative requirements under the EU General Data Protection Regulation, non-EU providers of general-purpose AI models must additionally appoint an authorized representative in the EU per Article 54(1). This appointment must be via a written mandate that authorizes the representative to:

- Verify that the general-purpose AI model provider has prepared the required technical documentation and otherwise fulfilled its obligations under Article 53, as described above, and Article 55, if it provides a general-purpose AI model with systemic risk, as described below.
- Keep a copy of the general-purpose AI model provider's required technical documentation for a period of 10 years after the model is placed on the market, so it is available to the European Commission's AI Office and national competent authorities, in addition to its contact details.
- Provide the AI Office with the compliance information and documentation necessary to demonstrate the general-purpose AI model provider's compliance upon request.
- Cooperate with the AI Office and competent authorities upon request in any action they take in relation to the general-purpose AI model, including when it is integrated into AI systems available in the EU.

Once again, this requirement does not ordinarily apply to providers of open-source general-purpose models, unless those models have systemic risk.

Obligations of providers of general-purpose AI models with systemic risks

As already noted, providers of general-purpose AI models with systemic risk are subject to additional obligations under Article 55 of the AI Act. In addition to the rules already described above, they must also:

- Perform model evaluation in accordance with standardized protocols and tools reflecting the state of the art, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating the systemic risks described above.
- Assess and mitigate possible systemic risks at an EU level, including their sources, that may stem from the development, sale or use of general-purpose AI models with systemic risk.

- → Keep track of, document and report relevant information about serious incidents without undue delay to the AI Office, and to national competent authorities as appropriate, including possible corrective measures.

- → Ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model.

Regarding the requirement to document and report relevant information about serious incidents, a key question is how this requirement will be operationalized in practice, and further guidance would be welcomed in this respect. However, it is clear that this requirement is distinct from the requirement for high-risk AI systems' providers and deployers to report serious incidents under Article 26(5) and Article 73.

Codes of practice for general-purpose AI

To demonstrate their compliance and pending the EU's adoption of harmonized standards for general-purpose AI, pursuant to Article 40, providers of general-purpose AI models, with or without systemic risk, can adhere to codes of practice which are expected to be drawn up and finalized by the AI Office within nine months after the AI Act enters into force. This would follow consultation with the AI Board and national competent authorities, as well as industry, academic and civil society stakeholders under Article 56.

The European AI Office launched a consultation for a first Code of Practice for general-purpose AI models 30 July 2024.

When does this take effect?

The AI Act's rules for general-purpose AI model providers come into effect in two phases under Articles 111(3) and 113.

Providers of older general-purpose AI models placed on the EU market before 2 Aug. 2025 have up to three years from the act's entry into force to comply, i.e. until 2 Aug. 2027. However, providers of newer general-purpose models, that is, all other general-purpose AI model providers, have up to 12 months after the act enters into force to come into compliance, i.e. until 2 Aug. 2025.

Practical steps for general-purpose AI

Any organization using general-purpose AI will need to ask itself the following questions and implement compliance measures accordingly:

- → Is the general-purpose AI in question a general-purpose AI model to which Chapter V applies or, instead, a general-purpose AI system that must then be categorized as prohibited, high or low risk to determine which rules apply under the AI Act?

- → Is the organization in question the provider of the general-purpose AI model? Chapter V applies only to providers of general-purpose AI models.

- → Does the general-purpose AI model present systemic risk? If not, it will be subject only to the rules in Articles 53 and 54. If so, it will be subject to additional rules in Article 55.

- → Has the AI Office produced any applicable codes of practice yet under Article 56?

If so, consider alignment with these as a means of demonstrating compliance with the AI Act.

- Is the general-purpose AI model provider established outside the EU? If so, it must appoint an authorized representative in the EU in accordance with Article 54.
- Are you a provider of an older or newer general-purpose AI model for the purposes of Articles 111(3) and 113? This will determine when the AI Act's rules apply to you, and when you need to come into compliance.

Annex

Mandatory information to be included in technical documentation for general-purpose AI models	
WITH OR WITHOUT SYSTEMIC RISK	WITH SYSTEMIC RISK
A general description of the general-purpose AI model, including: → The tasks that the model is intended to perform and the type and nature of AI systems in which it can be integrated. → The acceptable use policies applicable. → The date of release and methods of distribution. → The architecture and number of parameters. → The modality, such as text or image, and format of inputs and outputs. → The license. A detailed description of the elements of the model referred to above and relevant information of the process for the development, including the following elements: → The technical means required to integrate the general-purpose AI model in AI systems, such as instructions for use, infrastructure and tools. → The design specifications of the model and training process, including training methodologies and techniques. The key design choices include the rationale and assumptions, what the model is designed to optimize, and the relevance of the different parameters. → Information on the data used for training, testing and validation, when applicable, including the type and provenance of data and curation methodologies, such as cleaning and filtering; the number of data points, their scope and main characteristics; how the data was obtained and selected; as well as all other measures to detect the unsuitability of data sources and methods to detect identifiable biases, when applicable. → The computational resources used to train the model, such as the number of floating point operations, training time and other relevant details related to the training → The known or estimated energy consumption of the model. When the energy consumption of the model is unknown, the energy consumption may be based on information about computational resources used.	→ A detailed description of the evaluation strategies, including evaluation results, based on available public evaluation protocols and tools or other evaluation methodologies. Evaluation strategies shall include evaluation criteria, metrics and methods for identifying limitations. → A detailed description, when applicable, of the measures implemented to conduct internal and/or external adversarial testing, such as red teaming and model adaptations, including alignment and fine tuning. → When applicable, a detailed system architecture description that explains how software components build or feed into each other and integrate into the overall processing.