Published July 2024

Understanding and assessing risk

By Eduardo Ustaran and Uzma Chaudhry

A defining feature of the EU AI Act that has stood out since the European Commission's first proposal in 2021 is the now largely favored "risk-based approach." However, this is not the first time the approach has been featured in EU regulation. For example, the EU General Data Protection Regulation requires safeguards to be implemented according to the level of risk associated with data processing activities. Similarly, the AI Act places obligations on operators depending on the risk category of their AI use. The goal is to mitigate the risk of AI while promoting innovation to reap the benefits of this transformative technology.

The reason behind this model of regulation is an implicit acknowledgment that technology, such as AI, can be beneficial or risky depending on its uses. By placing risk regulation central to the new law, legislators have sought to craft a legislation that does not regulate a particular technology but what we make of the technology through its use. This is even more relevant in the context of AI, given its role as an emerging technology that can, and will, be deployed for almost unlimited applications from the very trivial to the existential.

This article provides insights on how risk is defined and addressed in the AI Act and unpacks the risk-based approach through a breakdown of the definitions and classification criteria for each risk category identified under the new law.

Understanding risk

Understanding around risk within the AI Act is established through two significant definitions.

Risk is defined under Article 3(2) of the AI Act as "the combination of the probability of an occurrence of harm and the severity of that harm."

"Product presenting a risk" is mentioned in Article 79 of the AI Act and is defined under Article 3(19) of Regulation (EU) 2019/1020 on market surveillance and product compliance. At its core, the AI Act is aimed at promoting the uptake of human-centric and trustworthy AI, while ensuring a high level of protection of health, safety and fundamental rights within the EU.

As such, to the extent that AI is seen as a product, a product presenting a risk under Regulation EU 2019/1020 is defined as one that has "the potential to affect adversely health and safety of persons in general, health and safety in the workplace, protection of consumers, the environment, public security, and other public interests protected by applicable Union harmonisation legislation to a degree which goes beyond that is considered reasonable and acceptable in relation to its intended purpose or under the normal or reasonably foreseeable conditions of use of the product concerned. This includes duration of use and the product's putting into service, installation, and maintenance requirements.

Assessing risk

Essentially, the AI Act identifies different types of AI, in accordance with the different levels of risk they present, as follows:

- → Prohibited AI systems, which by their nature present an unacceptable level of risk.
- → High-risk AI systems.
- → AI systems with transparency risks.
- → General-purpose AI models.
- → General-purpose AI models with systemic risk.
- → Other types of AI that do not fall within the above categories.

Prohibited AI systems

Although the AI Act does not define this risk category, Article 5 exhaustively lists examples of unacceptable AI practices that can threaten the rights of individuals located in the EU.

Prohibited AI systems include systems that deploy subliminal techniques, social scoring systems, predictive policing based solely on profiling or personal characteristics, systems used for untargeted scraping of facial images from the internet or CCTV footage for creating or expanding facial recognition databases, emotion recognition systems in the workplace and schools, and biometric categorization systems for deducing or inferring protected characteristics.

Although real-time remote biometric identification systems in publicly accessible spaces for law enforcement have also been banned, the AI Act provides exceptions for their use. For example, exemptions will be granted to law enforcement if they are deployed for:

- → Conducting targeted searches for victims of abduction, human trafficking and sexual exploitation and searches for missing persons.
- → Preventing specific, substantial and imminent threats to life or safety of natural persons, or to prevent the threat of a genuine and foreseeable threat of a terrorist attack.
- → Identifying and locating a person suspected of committing a crime for the purpose of conducting a criminal investigation, or prosecuting or executing offenses mentioned in Annex II, such as terrorism, trafficking, sexual exploitation, child pornography, murder or illicit trade, among others.

High-risk AI systems

High-risk AI systems were the original focus of the proposed regulation, and although different types of AI have been included as part of the legislative process, they remain at the core of

the regulatory framework. The AI Act sets forth criteria to identify whether an AI system does or does not classify as high risk. However, at the outset, it is useful to note an AI system will be classified as high risk depending on the specific purpose for which it is used. Therefore, careful assessment and understanding of the relevant use cases is essential to determine whether a given AI system qualifies as high risk.

For an AI system to be classified as high risk under the AI Act, it will either meet both conditions set forth in Article 6(1) or be listed as high risk under Annex III.

According to Article 6(1) if the "system is intended to be used as a safety component of a product, or the AI system itself is a product, covered by the Union harmonisation legislation listed in Annex I" and is "required to undergo a third-party conformity assessment, with a view to the placing on the market or putting into service of that product," then it meets the classification criteria of being a high-risk AI system.

Alternatively, if the system does not fall within the criteria set out in Article 6(1), then it may be listed as high risk under Annex III, pursuant to Article 6(2). Use cases mentioned under Annex III include biometric systems; critical infrastructure; education and vocational training; employment, workers management and self-employment; access to enjoyment of essential private and public services and benefits; law enforcement; migration, asylum and border control management; administration of justice and democratic processes.

However, if an AI system is listed under Annex III but meets any of the conditions set forth in Article 6(3), then it will not be considered high risk, provided it is not used to profile people. This includes systems that are intended to:

- → Perform a narrow procedural task.
- → Improve the result of a previously completed human activity.
- → Detect decision-making patterns or deviations from prior decision-making patterns.
- → Perform a preparatory task for an assessment relevant for the use cases listed under Annex III.

AI systems with transparency risks

This category, governed by Article 50 of the AI Act, sets forth obligations for both providers and deployers of certain AI systems. According to the Articles 50(1) and 50(2), which set forth obligations for providers, AI systems with transparency risks include AI systems that interact directly with natural persons and AI systems that generate synthetic audio, image, video or text content.

Additionally, according to Articles 50(3) and 50(4), which set forth obligations for deployers, this category includes emotion recognition systems, biometric categorization systems, AI systems that generate or manipulate images, and audio or video content constituting a deepfake AI system that generates or manipulates text published with the purpose of informing the public on matters of public interest.

Based on the above examples, and according to the interpretative guidance of Recital 132 it appears the specific transparency requirements

capture systems that may or may not raise high risks. If they raise high risks, transparency obligations will have to be fulfilled without prejudice to the transparency obligations listed for high-risk AI systems. However, as the AI Act enters its phased implementation, this also creates significant uncertainties about what should be regarded as low, high or unacceptable risk. This may raise complexities that affect both those who are meant to comply with the law and those who are meant to implement and enforce it.

General-purpose AI models and those with systemic risk

Chapter V of the AI Act lays down a legal framework for two types of general-purpose AI: general-purpose AI models and general-purpose AI models with systemic risk.

It is important to note, while the AI Act provides a legal framework for regulation of AI "systems," Chapter V departs from that approach and lays down a framework for the regulation of general-purpose AI "models."

A general-purpose AI system is defined in the AI Act as a system based on a general-purpose AI model with the capability to serve a variety of purposes that can either be used directly or integrated in other AI systems.

A general-purpose AI model is therefore part of the broader technical architecture that underpins a general-purpose AI system and is defined as an AI model "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." However, this definition does not cover AI models that are used before release on the market for research, development and prototyping activities.

The AI Act also establishes a subset of general-purpose AI with systemic risk, which refers to a general-purpose AI model that "has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks," or "based on a decision of the Commission, ex officio or following a qualified alert from the scientific panel" has those same high impact capabilities with "regard to the criteria set out in Annex XIII." The AI Act then goes on to say a general-purpose AI model "shall be presumed to have high impact capabilities ... when the cumulative amount of computation used for its training measured in floating point operations is greater than 10^{25}." Crucially, the concept of systemic risk is not assessed by a use case but by the computing power of the relevant AI model.

Conclusion

The AI Act represents the most sophisticated example of the so-called risk-based approach to European regulation. The degree of granularity with which it classifies the various levels of risk potentially created by AI technology is one of its defining factors and is a key contributor to its complexity. As AI technology evolves and the AI Act becomes effective in practice, determining which obligations apply to a specific type of AI will be almost as challenging as deploying the necessary measures for compliance with the act.