CDT EUROPE

A Series on the EU AI Act

Pt. 1 – An Overview

March 2024

Authored by
***Laura Lazaro Cabrera**, Counsel and Director of the Equity and Data Programme, CDT Europe*
***Iverna McGowan**, Director, CDT Europe*

The AI Act was politically agreed to by the Committee of Permanent Representatives on 2 February 2024 and endorsed by MEPs in the Internal Market and Consumer Protection (IMCO) committee and the Civil Liberties, Justice, and Home Affairs (LIBE) committee on 13 February 2024. Today, on 13 March 2024, the Act was approved by Parliament in plenary.

In this explainer, we dive into the key features of the AI Act, ranging from the artificial intelligence (AI) systems that it seeks to regulate and the obligations created for actors throughout the AI supply chain, to the oversight bodies and mechanisms created to monitor compliance. This paper is in two parts, the first looking at the regulatory approach of the AI Act, including its hierarchy of risks and corresponding mitigation measures. The second part looks specifically at the governance structures and oversight and enforcement mechanisms.

Whilst we touch upon some of the issues relating to human rights and democracy, the main purpose of this piece is to provide an overview of what the Act provides for. Our next publications will look at specific areas of human rights risks.

An Overview of the Risk-Based Regulatory Approach

The AI Act in its current form regulates AI in four key ways, by:

1. Prohibiting specific types of AI systems;
2. Categorising some types of AI as high-risk and imposing additional obligations on them;
3. Defining and setting out obligations for general-purpose AI (GPAI) models, including GPAI with systemic risks; and
4. Setting out measures for dealing with AI systems presenting risks at the national level.

Prohibited AI Systems

The AI Act prohibits eight types of AI. Although, as we detail in this and future explainers, in some cases it is questionable whether we should refer to a prohibition at all given the very broad exceptions that the text allows for.

The prohibitions can be broadly split into two categories: AI systems prohibited based on their actual or likely effects, and AI systems prohibited based on their function, regardless of consequences. In the first category, thresholds are generally high: two prohibitions require a showing of a risk of significant harm, while another requires detrimental or unfavourable treatment. Neither of these concepts are actually defined by the AI Act.

Prohibited AI systems based on their effects	**Manipulation in decision-making.** AI systems deploying subliminal, purposefully manipulative or deceptive techniques, with the objective or effect of materially distorting behaviour by impairing a person's ability to make an informed decision, causing that person to make a decision that they would not have otherwise taken in a manner that causes or is likely to cause them or others *significant* harm. See Article 5(1)(a).
	Vulnerability exploitation. AI systems exploiting persons' vulnerabilities because of their age, disability or specific social or economic situation, with the objective or effect of "materially distorting" the behaviour of a person in a way that is reasonably likely to cause them or someone else *significant* harm. See Article 5(1)(b).
	Social scoring. AI systems evaluating or classifying individuals based on social behaviour or known, inferred or predicted personal or personality characteristics if the resulting "social score" results in detrimental or unfavourable treatment that is i) unjustified or disproportionate and/or ii) unrelated to the contexts in which the data was originally generated or collected. This section can be read as requiring *a showing of detrimental or unfavourable treatment*. See Article 5(1)(c).

Prohibited AI systems based on their function	**Predictive policing.** AI systems making risk assessments to assess or predict risk of offending, based *solely* on the profiling of a natural person or assessing their personality traits and characteristics – this excludes instances where the AI supports a human assessment which is based on objective and verifiable facts. See Article 5(1)(d).
	Image scraping. AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage. See Article 5(1)(e).
	Emotion recognition. AI that infers emotions of a natural person in the areas of workplace and education settings unless the AI is intended to be put in place for medical or safety reasons. See Article 5(1)(f).
	Biometric categorisation to generate sensitive data. AI systems carrying out the biometric categorisation of individuals based on their biometric data to deduce/infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation – i.e. AI using biometric categorisation to deduce special category data within the meaning of GDPR. See Article 5(1)(g).
	Real time facial recognition. The deployment of real-time biometric identification for law enforcement purposes in publicly accessible spaces unless exceptions apply. These exceptions are broad and include: i) finding victims, ii) preventing a specific, substantial and imminent threat to human life or threat of terrorist attack which is present genuine and foreseeable; or iii) localising/identifying a person suspected of crime, if that crime is in Annex IIA and punishable by custodial sentence or detention order of 4y+. See Article 5(1)(h).

These prohibitions are set to apply six months after the entry into force of the AI Act. Violation of these prohibitions may result in a fine of up to €35M or, if the offender is a company, up to 7% of its total worldwide annual turnover for the preceding financial year, whichever is higher. For EU institutions and bodies engaging in prohibited AI practices, a fine of up to €1.5M may be imposed.

High Risk Systems: A Two-Part Assessment

The AI Act adopts a risk-based approach in its regulation of AI systems. As CDT has previously commented, the AI Act is not entirely clear in its categorisation of risk, at times oscillating between context, technology and human rights. This can also be seen in the way that the Act does not point to a single definition of high-risk, but adopts a flexible approach.

There are only two cases where AI can be conclusively designated as high-risk. First, if an AI acts as a safety component for a product – or is itself a product – and the product is subject to a conformity assessment pursuant to existing legislation. This means that an AI system will be categorised as high risk if the product that the AI supports (or the AI itself), due to its nature, is governed by specific legislation other than the AI Act which requires an assessment of its compliance with applicable standards and technical regulations. Examples of AI-enabled safety technologies could include fall detection or site inspection software deployed in the construction sector to prevent worker injuries. Second, an AI system will be conclusively deemed high risk if it falls within one of the areas identified in Annex III *and* it carries out profiling activities: in other words, if it carries out the automated processing of personal data in one of the areas covered by Annex III to evaluate certain personal aspects relating to an individual, in particular to analyse or predict aspects concerning that person's behaviour, interests or personal preferences among others.

Annex III to the Act lists the areas where AI deployment would be considered high risk, and is subject to modification by the Commission.

Currently, the following areas are designated as high-risk:

- Biometrics
- Critical infrastructure
- Education and vocational training
- Employment, workers management and access to self-employment
- Access to and enjoyment of essential private/public services and benefits
- Law enforcement
- Migration, asylum and border control management
- Administration of justice and democratic processes

However, an AI system being deployed in any of the above areas does not, in the absence of profiling, conclusively establish that it is high risk. Irrespective of whether an AI system is listed in Annex III, an AI system will not be considered high risk if it does not pose a significant risk of harm to the health, safety or fundamental rights of people, including by not materially influencing the outcome of decision making (Article 6 (2a)). The Act goes on to list examples of AI falling short of the high risk threshold, including systems intended to i) perform a narrow procedural task; ii) improve the result of a previously completed human activity; iii) detect decision-making patterns or deviations thereof; iv) or to perform a preparatory task.

The ultimate assessment of whether or not an AI system is high risk is left to the discretion of the AI provider. If an AI provider considers that their AI system is not high risk despite falling within one of the high risk areas, all it must do is to document this assessment (Article 6 (2b)) and provide it to the relevant authorities, but only upon request. Separately, a market surveillance authority can carry out an evaluation of an AI system when it has sufficient reasons to consider that the system is high risk and has not been categorised as such (Article 80(1)).

A Tiered Obligation System

The AI Act creates tiered obligations for the various actors involved in the launch of an AI system, depending on their role in the supply chain. If an AI system is high risk within the meaning of the Act, obligations apply to providers, deployers, distributors and importers.

Entity	Definition	Obligations
Providers	Person or entity that develops or has an AI system or GPAI and places it on the market or puts it into service under their name or trademark	Providers are responsible for ensuring compliance with requirements of AI system, including risk-management, quality control in relation to datasets used to develop the AI (Article 10), preparation of technical documentation prior to placing on the market (Article 11), record-keeping (Article 12), maintenance of accuracy, robustness and cybersecurity (Article 15) and human oversight during the lifecycle of the AI (Article 14). Crucially, providers are in charge of ensuring the AI goes through a conformity assessment, obtaining a declaration of conformity, and registering the AI in a public EU database (Article 51).
Deployers	A person or entity using an AI system under its authority	Deployers must monitor the operation of the AI and keep logs (Article 29). Additional obligations apply if the deployers control the AI system generally, if they control the input data, or if they're financial institutions or employers. Deployers also have the obligation to carry out a FRIA if they are a public body or a private body carrying out public functions, and must inform people that they are being subject to the use of a high-risk AI system if this is the case.
Distributors	Person or entity that makes an AI system available on the EU market	Distributors must ensure that the provider has undertaken all relevant steps as above – if they have any reason to suspect that the AI is not in conformity, they must not place the AI system on the market.
Importers	Any person or entity in the EU that places on the market an AI system with the name or trademark of a person or entity established outside the EU	Importers must ensure that the provider has undertaken all relevant steps as outlined above and that they have appointed a representative in the territory of the EU (Article 26).

Failure by any of these entities to fulfill their obligations related to high-risk AI systems shall be subject to a fine up to €15M or, if the offender is an entity carrying out an economic activity, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher. Different fine levels apply to non-compliant EU institutions or bodies, which can be fined up to €750,000 for any non-compliance with the AI Act other than in relation to prohibited practices.

AI Models

The Act clearly distinguishes between an AI model and an AI system, and considers that every AI system contains an AI model as a component, but an AI model on its own does not amount to an AI system. The Act regulates general purpose AI *models*, with additional obligations for general purpose AI models that pose systemic risks.

General Purpose AI Models

General purpose AI (GPAI) models are broadly defined under the Act as models that display significant generality and are capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market.

Where GPAI models are concerned, the only entity which is subject to obligations under the AI Act is the provider. The obligations are markedly reduced in comparison to high risk AI systems: upon placing a GPAI model on the market, a provider must prepare technical documentation and provide information for downstream parties (Article 52c). However, these obligations do not apply to providers of GPAI models which are "open source", i.e. made publicly accessible under a free and open license allowing for access/usage/modification and distribution, whose parameters are publicly available. All providers of GPAI models must put in place a policy outlining compliance with copyright law and publish a detailed summary of the content used for training of the model.

Obligations for GPAI models will apply within 12 months from the date of entry into force of the AI Act.

General Purpose AI Models Presenting a "Systemic Risk"

A systemic risk will exist if a GPAI model has high-impact capabilities, evaluated on the basis of appropriate technical tools and methodologies, or significant impact on the internal market due to its reach.

The Act recognises as high-impact capabilities those that match or exceed the capabilities recorded in the most advanced GPAI models (Article 2(64)). While the GPAI models meeting this definition will change over time as technology continues to develop, the AI Act sets a baseline by creating a rebuttable presumption that a GPAI model will pose a systemic risk if floating point operations (FLOPs) are over 10^25 – a number that the Commission reserves the right to adjust in the future. Providers can rebut this presumption, but in any event must notify the Commission that their GPAI model has crossed the threshold no less than two weeks after it does so (Article 52(1)).

Separately, the Commission has the power to make individual decisions designating a GPAI model as having systemic risk (Article 52 (4)) based on the criteria in Annex IXc. Some of the criteria considered by the Commission include the quality or size of the training data set, the number of business and end users, its input and output modalities, its degree of autonomy and scalability, and the tools it has access to.

If a GPAI model is deemed to pose a systemic risk, it is still allowed to operate subject to special additional obligations being imposed on the provider (Article 55), including assessment and mitigation of risk, the monitoring of serious incidents, and ensuring an adequate level of cybersecurity. GPAI models with systemic risk are not exempt from any obligations even if the models are publicly accessible (Article 53 (2)).

The same level of penalties apply as for high-risk AI: failure to comply with obligations shall be subject to a fine not exceeding €15M or, if the offender is a company, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher.

AI Systems "Presenting a Risk at National Level"

There is one fallback provision in the AI Act to deal with systems that are not classified as high-risk or GPAI models: AI systems presenting a risk at national level (Article 79). This provision mirrors pre-existing legislation on product safety and applies when an AI system has the potential to affect adversely health and safety or fundamental rights of persons, or public interests, to a degree which goes beyond that considered reasonable and acceptable in relation to its intended purpose or under the normal or reasonably foreseeable conditions of use of the product concerned.

Where an AI system presents such a risk, the Act provides for a market surveillance authority (MSA) to request for corrective actions to be undertaken by the relevant operator – that is, any entity involved with the AI regardless of whether it is the provider, deployer, importer or distributor – to bring the AI system into compliance, to withdraw the AI system from the market, or to recall it within a period of no more than 15 days.

If the AI system presenting a risk also happens to be "high risk" within the meaning of the AI Act, tiered obligations apply on a range of entities involved with the system, with obligations being triggered as soon as the systemic risk materialises.

Governance and Oversight Bodies

Oversight Bodies

The AI Act sets up monitoring of AI systems at two levels: national and regional.

1. National

Compliance with the bulk of the obligations contained in the AI Act is largely overseen by member state authorities or bodies. The Act designates two relevant authorities at member-state level, generally referred to in the text as "national competent authorities": notifying authorities and market surveillance authorities.

Notifying authorities, rather than monitoring AI systems and AI providers directly, monitor and formally authorise conformity assessment bodies, i.e. the bodies responsible for carrying out the conformity assessment which certifies a high risk AI system is in conformity with the Act's requirements. At least one notifying authority is needed per member state. (Article 28). Because a conformity assessment is only needed for high-risk AI systems, notifying authorities only play a role when it comes to this subset of AI. They play no role in relation to general-purpose AI or AI presenting a risk at the national level.

Market surveillance authorities (MSAs) are the default oversight entity for all types of AI, including high-risk. They pre-date the AI Act – the concept was actually borrowed from a pre-existing EU law on product safety and regulation. There is no limit to the number of market surveillance bodies that any member state may have, and a constellation of these authorities already exist across Europe. There must be at least one per member state, and its nature can be flexible – except in the case where a high-risk AI system is deployed in a law enforcement context, in which case the nature of the authority is prescribed by the Act. Regardless of the number of market surveillance authorities in any given member state, only one market surveillance authority per state must be designated as the point of contact for the purposes of enforcement of the AI Act (Article 70). The fact that the Act largely leaves the choice of MSAs to member states is significant, as it potentially allows for differences in expertise and approach between MSAs across member states.

MSAs play a key role in ensuring compliance with the AI Act, and have a range of powers to enable them to do so – including exempting AI from restrictions under the Act, for example by authorising a high-risk AI system while dispensing with the need for a conformity assessment (Article 46) or authorising testing in the real world outside regulatory sandboxes (Article 60(4)). Additionally, MSAs are the primary entity to whom reporting is addressed under the Act. They must be notified, alongside any data protection agency, of the use of a real-time biometric identification system (Article 5 (4)) and post biometric identification system (Article 26 (10)). They are also the recipient of any fundamental rights impact assessment (FRIA) undertaken pursuant to Article 27. A market surveillance authority can "double" as a notified body (i.e. an approved conformity assessment body) where a high-risk AI system is intended to be put into service by law enforcement, immigration or asylum authorities as well as EU institutions, bodies or agencies (Article 43 (1)). Importantly, market surveillance authorities are the entity who receives notifications of serious incidents pursuant to Article 73(1).

MSAs can make findings of non-compliance with the Act under Article 83, pursuant to which they can require the provider to put an end to non-compliance within a fixed period of time, and subsequently take measures to restrict or prohibit the AI system (Article 83(2)).

2. Regional

In addition to the member state level authorities, the AI Act reserves various follow-up activities for various European regional bodies, some old and some new. While the European Commission is largely responsible for the bulk of these implementation activities, the AI Act creates roles for four additional entities, whose input pursuant to the Act can be categorised as either voluntary or mandatory. Of these four entities, one pre-exists the Act (the AI Office); two are created by the Act itself (the European AI Board and the advisory forum); and another one remains to be created by the Commission by way of implementing act (the scientific panel of experts). Only the AI Office and the Commission are mandated by the Act to take steps relating to its implementation, while the input by the Board, the advisory forum and the scientific panel of experts, while detailed, is largely voluntary.

The AI Office and the European Commission

The AI Office was created separately in a Commission decision from January 2024 to implement the AI Act. Its role under the Act is to set standards for evaluating the capabilities of and monitoring the application of rules to GPAI models and systems. It is empowered to create Codes of Practice, and monitor possible infringements of the AI Act by these models and systems. The Commission decision further tasks the AI Office to coordinate enforcement action where the Digital Markets Act or Digital Services Act are also engaged.

The Act however reserves the most prominent role for the Commission. The tasks allocated to the Commission range from the production of codes of practice and guidelines to the issuance of standardisation requests and the imposition of fines for GPAI. Most notably, the Commission is tasked with adopting delegated acts (i.e non-legislative acts supplementing the AI Act) in relation to the following:

- Amending the criteria for a derogation from the “high-risk” classification (Article 6 (6)).
- Amending Annex III by adding, modifying or removing use cases of high-risk AI systems (Article 7).
- Amending the technical documentation requirements for high-risk AI contained in Annex IV (Article 11(3)) and for GPAI contained in Annex IXa (Article 53(6)).
- Amending the conformity assessment procedures set out in Annex VI and VII (Article 43(5)).
- Amending the thresholds for systemic risk GPAIs (Article 52 (4)).

Satellite Bodies

The Act also envisages a role for a range of bodies that can best be described as satellite bodies to the extent that the AI Act does not compel them to take specific action. Instead, their inputs are either voluntary or conditional upon another entity making a direct request for their expertise.

Entity	Supports	Can issue opinions or recommendations on its own	Description
The European AI Board	The AI Office	Yes	Composed of member state representatives, it has an advisory role to the AI office and is empowered to issue guidance of its own initiative or at the request of the Commission on any matter pertaining to the AI Act and its application. It also plays a role coordinating market surveillance authorities.
The Advisory Forum	The Board and the Commission	No	Members are appointed from a cross-section of civil society, academia, and private sector. Permanent members of the advisory forum named by the Act include CEN, CENELEC, ETSI and FRA (Article 58a). The forum can only prepare guidance at the request of either of the Board or the Commission.
The scientific panel of independent experts	AI Office and national market surveillance authorities upon request	Yes	The Act instructs the Commission to establish a scientific panel by way of delegated act. The panel is to be composed of experts in the AI field, and its primary role is to evaluate the capabilities of and provide advice on GPAI, alerting the Office of any possible systemic risks.

Oversight Mechanisms

The AI Act sets out mandatory documentation procedures for both high risk AI systems and GPAI models. Not every documentation process has a related reporting obligation – in fact, only the fundamental rights impact assessment result needs to be communicated to a relevant member state authority. Neither the conformity assessment for high risk AI nor the technical documentation for a GPAI model need to be reported to any authority as the Act considers it sufficient for the providers to produce and keep those documents.

Procedures for High Risk AI

- **Conformity assessments.** The Act requires providers of high risk AI to seek conformity assessments to ensure they comply with the requirements laid out by the Act. The entity carrying out the conformity assessment depends on the area the high-risk AI is being deployed in. If it is in connection to biometrics, a notified body must undertake the assessment; if it is in connection to any other area identified as high-risk in Annex III, the provider must undertake the assessment itself. Once the assessment result is obtained, there is no requirement for the assessment to be reviewed or approved by any authorities. Instead, the Act asks providers to provide a competent authority with the relevant documentation upon "reasoned request".

- **Fundamental Rights Impact Assessment (FRIA).** Deployers of high-risk AI which are bodies governed by public law, or private operators providing public services, or operators specifically deploying credit-scoring AI or risk assessment-pricing AI in the field of life and health insurance, must carry out a fundamental rights impact assessment prior to placing the AI system in the market and notify the market surveillance authority of the result. However, once that FRIA is undertaken, there is nothing in the AI Act: i) preventing a deployer from launching a high-risk AI system – even if it performed poorly in the assessment; ii) requiring the market surveillance authority to "approve" the FRIA prior to deployment and iii) creating explicit power for market surveillance authorities to halt the deployment of a high-risk AI system if they "fail" the FRIA. The AI Office is tasked by the Act to develop a questionnaire to facilitate the FRIA (Article 27).

Procedures for GPAI

- **Technical documentation.** Providers of GPAI models must provide relevant technical documentation, including the GPAI training and testing process and the results of its evaluation, which shall contain, at a minimum, the elements set out in Annex IXa for the purpose of providing it, upon request, to the AI Office and the national competent authorities. However, an exception to this requirement applies if the GPAI model is open source.

In addition, ad hoc procedures exist to enable entities to alert the authorities upon a specific risk materialising. Where a high risk AI system causes a serious incident, providers must report to the market surveillance authorities; where it becomes apparent that a GPAI model poses or will pose a systemic risk, providers must notify the Commission within a two-week period.

Conclusion

The AI Act is the first regulation of its kind. The resulting text is ambitious, lengthy and complex. It is nonetheless crucial for members of the public and civil society to achieve a robust understanding of this regulation: its impacts will be wide-ranging, and its successes – as well as its failings – will inform future AI legislation.

The Commission faces a busy few months as it concludes the hiring process for the newly established AI Office and prepares to issue guidelines on the practical implementation of the Act. Close coordination with experts and civil society will be crucial to ensure that the interpretation and application of the AI Act going forward ensures its effectiveness and is consistent with the Act's own articulated goals: protecting fundamental rights, democracy and the rule of law.

Find more from the CDT Europe team on the EU AI Act at *cdt.org*.

*The **Center for Democracy & Technology (CDT)** is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.*