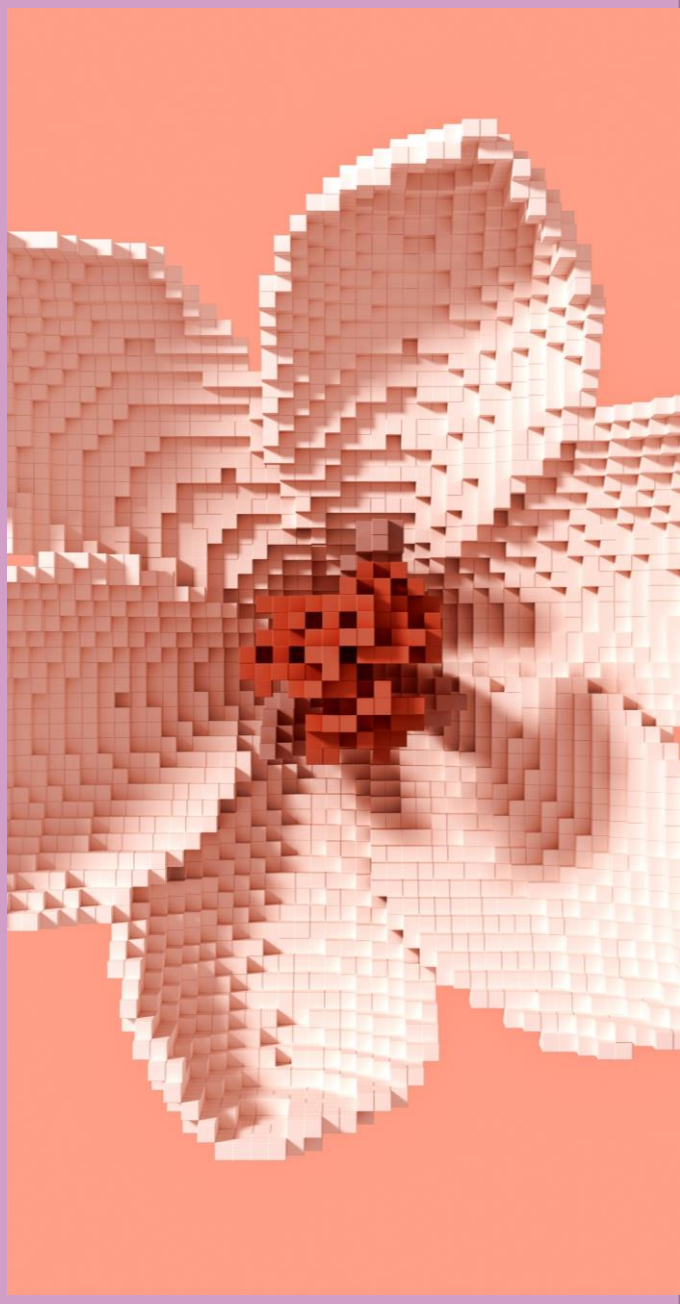
Bird & Bird

EU reaches deal on AI Act

Key insights from the political agreement

December 2023



Contents

The legislative process for the Artificial Intelligence Act ("AI Act") began in April 2021 and reached an important milestone during the final political trilogue from 6 December to 8 December 2023. The negotiations in this extraordinary marathon session resolved several highly contentious issues, some of which had just arisen in recent months. In this article, we provide an overview of the key elements of this deal and a first assessment of its implications.

1 Preliminary remarks

1.1 Premise

The information available and this initial assessment should be treated sceptically due to the following two factors:

- At present we do neither have a new consolidated text of the AI Act nor an official statement that conclusively addresses the recent agreement. This article is based on various reports, some first hand, some from secondary sources, of the terms agreed. It is important to treat this information with caution, as some of these reports even give conflicting details.
- Further developments are expected in the coming weeks, in particular the translation of the political agreement into a final text. Several technical trilogue negotiations are scheduled to take place between now and February 2024. These negotiations and the subsequent the work of lawyer-linguists are expected to fine-tune the key points recently agreed at the political level and may lead to further amendments of certain provisions (see Section 3 for more details).

1.2 Our approach

This article shall reflect what is currently known about the results of the final political trilogue. Recognising the significant demand for information from clients, we have analysed and consolidated information from a range of sources. We have then compared these findings with the European Commission's first draft of April 2021 (hereinafter **"Commission's First Draft"**) and the European Parliament's negotiating position of June 2023 (hereinafter **"EP's Negotiating Position"**).[1]

This comparative analysis allowed us to identify two distinct categories of news:

- Items that were eventually agreed in a way that is identical (or very similar) to the provisions of the previous drafts of the AI Act. These items have already been extensively covered over the past months.
- Genuinely new items introduced in the final stages of negotiations, such as the tiered approach to regulating foundation models.

This article is aimed at those who already have a basic understanding of the AI Act and have followed its legislative journey in a cursory way. In particular, it is intended for those who, after the final political trilogue, asked themselves: "*What exactly has been agreed? What is really new?*".

[1] All references to articles refer to the last full text version of the AI Act available, namely the EP's Negotiating Position. See our Reading Version of that position to conveniently look up the articles.

2 Contents of the latest agreement

2.1 *Which AI systems will be covered by the AI Act?*

The negotiating parties adopted a revised version of the OECD's definition of AI:

> "*An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that [can] influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.*"

This definition marks a departure from the definition of the EP's Negotiating Position. It reintroduces the term "content" similar to what was seen in the Commission's First Draft, and thereby unequivocally encompasses generative AI systems. There had been an opinion among right holders that generative AI was potentially not covered by the AI Act, as the definition of the EP's Negotiating Position did not explicitly mention "content" as potential output.

However, this revised definition is notably broad, to the extent that it could arguably apply more to software in general than to AI systems specifically. For instance, to pick up a popular example, even a basic Excel function could potentially fall under this definition. It remains to be seen whether the final text will diverge from the agreement or if the forthcoming recitals will at least provide a more precise delineation for AI.

2.2 *What are the obligations for all AI systems?*

To our knowledge, the negotiations have not led to any changes in the obligations for all AI systems. As in the EP's Negotiating Position, all AI systems must comply with the following six basic principles:

- **Human agency and oversight:** AI systems should serve people, respect human dignity and autonomy, and allow for human control and oversight.
- **Technical robustness and safety:** AI systems should be designed to minimize harm, be resilient to problems and resistant to misuse by malicious parties.
- **Privacy and data governance:** AI systems should comply with privacy and data protection rules and process data of high quality and integrity.
- **Transparency:** AI systems should be traceable and explainable, making clear to users that they are interacting with AI and informing them of its capabilities, limitations, and their rights.
- **Diversity, non-discrimination and fairness:** AI systems should promote diversity, equal access, gender equality, cultural diversity, and avoid biases and discrimination.
- **Social and environmental well-being:** AI systems should be sustainable, environmentally friendly, and beneficial to all humans, while assessing long-term impacts on individuals.

2.3 How will the AI Act regulate high-risk AI systems?

High-risk classification & obligations

The recent trilogue has led to an agreement on the classification of specific high-risk areas, initially proposed by both the European Commission and Parliament. These areas include education, employment, critical infrastructure, public services, law enforcement, border control, and the administration of justice. This agreement underscores a shared commitment to closely monitoring and regulating AI applications in sectors with significant societal impact. The agreement also, to a large extent, retains the obligations for high-risk systems outlined in previous drafts, e.g., conducting conformity assessments, integration of quality and risk management systems, registration, post-market monitoring by surveillance authorities.

Filter system for high-risk classification

A notable enhancement is the introduction of a new filter system, that came up for the first time in the negotiations in October 2023, designed to capture only genuine high-risk applications. Even if a system is generally covered by the catalogue of high-risk systems, it loses its classification if one of the following conditions applies:

- The AI system is intended to perform a "narrow procedural task" only (e.g., transforming unstructured into structured data).
- The AI system is meant to review or improve the result of a previously completed human activity (i.e., merely providing an additional layer to human activity).
- The AI system is purely intended to detect decision-making patterns or deviations from prior decision-making patterns (e.g., to flag potential inconsistencies or anomalies).
- The AI model is used to perform only preparatory tasks to an assessment relevant to the critical use cases (e.g., file handling).

Fundamental rights impact assessment

There is a confirmation of the obligation, introduced in the EP's Negotiating Position (Article 29a), for certain deployers of high-risk systems to conduct a Fundamental Rights Impact Assessment. This requirement extends to public bodies and private entities that provide essential public services, such as hospitals, schools, banks, and insurance companies, when deploying high-risk systems. This obligation aims to ensure that the deployment of such systems is consistent with fundamental rights, reinforcing the need for a cautious and responsible approach to integrating AI into critical societal functions. The assessment shall include: a description of the deployer's processes for using the high-risk AI system or the intended period and frequency of use.

Right to lodge complaints

A further significant new development is the provision for citizens to lodge complaints about AI systems and obtain explanations for decisions made by high-risk AI systems that affect their rights. This measure enhances transparency and accountability, providing a mechanism for individuals to understand and challenge AI-driven decisions. It is unclear to what extent this right to information must also include the Fundamental Rights Impact Assessment. For companies, the exact scope of this information will be crucial in assessing the amount of substantiation required – similar to the initial uncertainty regarding the exact scope of the right of access according to Art. 15 GDPR.

2.4 What AI systems will be prohibited under the AI Act?

Agreement on prohibited systems

The recent trilogue resulted in an agreement on several bans on AI systems, reflecting elements of both the Commission's First Draft and the EP's Negotiating Position. Agreed prohibitions include:

- **Manipulative techniques (Article 5.1.a):** Techniques designed to manipulate users into behaviour that may cause physical or psychological harm.
- **Systems that exploit vulnerabilities (Art. 5.1.b):** Systems that target vulnerable groups, such as children or the disabled, in order to exploit their vulnerabilities.
- **Categorization based on sensitive characteristics (Art. 5.1.ba):** Systems that categorize individuals based on sensitive characteristics such as race, political opinions or religious beliefs.
- **Social scoring (Art. 5.1.c):** Systems that score individuals for social behavior or trustworthiness based on non-transparent criteria.
- **Predictive policing (Article 5.1.da):** Systems designed to solely assess a person's risk of committing future crimes based on an AI's prediction.
- **Databases based on bulk scraping of facial images (Article 5.1.db):** Prohibition of AI systems that create or expand facial recognition databases through untargeted scraping of faces from the internet or CCTV footage.

Updates to prohibited systems

Amendments to prohibitions were agreed as follows:

- **Emotion recognition (Art. 5.1.dc):** Now limited to prohibition in the workplace and educational environment, with a new exemption for medical and safety reasons (e.g., monitoring the tiredness levels of a pilot). EP's Negotiating Position contained a ban in law enforcement and border management that has been apparently dropped.
- **Real-time remote biometric identification (Art. 5.1.d):** Prohibited except for three law enforcement exceptions: law enforcement activities related to 16 specified crimes (e.g., trafficking in human beings and murder), locating missing victims of certain crimes and the prevention of terror attacks. These exceptions were to some extent present in Commission's First Draft and contested in the EP's Negotiating Position. The current agreement additionally introduces an extensive catalogue of safeguards to avoid misuse (e.g., prior authorisation by a judicial or independent administrative authority).
- **Ex-post remote biometric identification (Art. 5.1.dd):** Prohibited except for several law enforcement exceptions, where it is strictly necessary based on national legislation and requires prior authorization from an independent authority. The Commission will monitor potential abuses.

The recently discussed export ban, which would have prevented EU-based companies from selling *prohibited systems* abroad, was given up.

2.5 How is General Purpose AI regulated in the AI Act?

Based on the information currently available, it appears that the AI Act will now exclusively focus on General Purpose AI (hereinafter **"GPAI"**), also covering what was previously considered to be Foundation Models in the EP's Negotiating Position. For the sake of clarity, this article will assume this equivalence and refer only to GPAI.

Definition of GPAI

The definition of GPAI shall expressly cover large generative AI Models and was therefore amended as follows:

> "*'general-purpose AI model' means an AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is released on the market and that can be integrated into a variety of downstream systems or applications.*"

Minimum requirements

An agreement was reached on the establishment of minimum requirements for all GPAI, as proposed in the EP's Negotiating Position. These standards include technical documentation, information to downstream providers, transparency measures (such as watermarking) and compliance with copyright rules. The latter will be discussed in more detail in the section on copyright below (see Section 2.7). The recent agreement on minimum requirements for all GPAIs appears to be less stringent than the regulation in the EP's Negotiating Position. For example, the previous draft included provisions to ensure that content generation does not violate EU law and respects citizens' fundamental rights (see Section 2.8 for potential consequences of this).

New tier: systemic risk GPAI

An important new development is the increased requirements for so-called **"Systemic Risk GPAI"**. The threshold for classifying a model as posing **"Systemic Risk"** is set at a computational power used for training exceeding 10^{25} FLOPs[2]. This threshold currently includes only the largest of the large language models, such as OpenAI's GPT-4 and possibly Google's Gemini – thus (intentionally?) favouring the smaller European models (e.g., from Aleph Alpha and Mistral). The current agreement already provides for the possibility for the Commission to adapt this threshold to the state-of-the-art or add further criteria besides the computational power (such as the number of users or the degree of autonomy of the model).

[2] In computing, floating point operations per second (FLOPS) is a measure of computer performance.

Requirements for systemic risk GPAI

The new requirements for systemic risk GPAI include:

- **Model evaluation:** Perform model evaluation according to standardised protocols and tools.
- **Risk assessment:** Assess and mitigate possible systemic risks that may stem from the development, placing on the market, putting into service or use of the models.
- **Red teaming:** Necessity to conduct and document adversarial testing to mitigate Systemic Risks.
- **Cybersecurity:** Maintaining an adequate level of cybersecurity for the AI model and its physical infrastructure.
- **Incident reporting:** Obligation to report serious incidents directly to the European Commission.
- **Energy consumption:** Providers are required to track, document, and report on the known or estimated energy consumption of the model.

Even though the exact contours of these requirements have not yet been finalised, it should be noted that they represent a significant tightening compared to the originally envisaged single-tier regulation of GPAI in the EP's Negotiating Position.

The Commission will have the authority to approve codes of practice with general validity within the EU. Providers of GPAI models with Systemic Risk may rely on those codes of practice to demonstrate compliance with the above-mentioned obligations, until a harmonised standard is established.

2.6 Is the AI Act applicable to open-source models?

The treatment of open-source AI models in the AI Act has evolved in the latest agreement. The EP's Negotiating Position already included an exemption for open-source models. However, this exemption was limited in scope, as all open-source foundation models (now GPAI) were subsequently brought back within the scope through a counter-exemption.

The newly established regulation concerning open-source AI systems appears to be more expansive esp. regarding GPAI: Only those open source GPAI systems classified as **"Systemic Risk GPAI"** will covered by the AI Act (i.e., exceeding the threshold of 10^{25} FLOPs). This will potentially exclude smaller GPAIs – a demand that has grown louder during the previous negotiations.

2.7 Is copyright covered by the AI Act?

The latest agreement on the AI Act contains copyright provisions for providers of GPAI models. A provision already included in the EP's Negotiation Position is the **"Training Content Summary"**. Providers must make a **"sufficiently detailed"** summary of the content used for training the AI model publicly available. This is intended to create transparency about the training data used. However, not every single training data should be listed. Instead, leaked recitals suggest that the summary should contain the main catalogues – private or public – and a narrative explanation about the data used. Templates for such lists may allegedly be published by the AI Office.

New is the requirement to create a copyright policy. Providers of GPAI should ensure that they comply with EU copyright law – in particular to observe the reservation of rights holders against text and data mining, which is laid down in Art. 4 Directive (EU) 2019/790. Thus, the leaked recitals explicitly recognize this reservation for text and data mining also for *generative* AI (which was partially controversial). It remains to be seen whether such a copyright policy must be flanked with organizational (e.g., licensing guidelines) and technical requirements (such as state-of-the-art filter technologies) – which is mentioned in some sources. Should such technical and organizational measures become necessary, it remains to be seen how they will interact with European copyright law, such as Directive (EU) 2019/790, the case law of the European Court of Justice, and also with the content moderation mechanisms of the Digital Services Act (**"DSA"**). However, from other sources (not related to the AI Act) it emerges, the EU Commission is planning to review the EU copyright framework in 2024 and this assessment will also be very significant for questions about the interplay with the AI Act. On the substance, the leaks suggest that the AI Act may require a certain level of prevention and handling of copyright issues from providers of GPAI models.

2.8 Does the AI Act require content moderation? How does it relate to the DSA?

As hinted in *section 2.5 and 2.7*, the latest agreement does not contain a requirement for general content moderation for providers of GPAIs (e.g., to prevent hate speech). Thus, deployers cannot fall back on a statutory minimum framework for content moderation in all GPAI models but must look at the policies of the individual GPAI models. This may be important for deployers if the use of the GPAI models needs to comply with certain corporate communication policies or if the potential risks for internal or external use cases are to be assessed.

So far, the leaks have also not indicated whether the regulations on content moderation for host providers and online platforms under Art. 16 et seq. of the DSA will be extended to generative AI. For example, infringing content due to gaps in safeguards could be reported (e.g., if certain prompt hacking generates output containing hate speech), which the GPAI provider could then have fixed centrally. Such a reporting could also be done via trusted flaggers who specialise in finding such gaps. Without an extension to these rules of the DSA, their applicability would need to be assessed on a case-by-case basis for the individual GPAIs (platforms).

Furthermore, leaked recitals reveal that, if a very large online platform or very large search engine (Tier 4 of the DSA, heavily regulated), integrate a system or a model into their service regulated under the DSA, the system or model is also subject to the risk management framework under the AI Act. It remains to be seen in the final text to what extent the fulfilment of obligations under the DSA can also be used to satisfy obligations under the AI Act.

2.9 *What are the penalties for violating the AI Act?*

The penalty structure in the AI Act has undergone revisions compared to the EP's Negotiating Position, resulting in a reduction of the fines. The current penalties are tiered as follows:

- **Prohibited systems & non-compliance with data requirements:** Up to 7% of the company's annual worldwide turnover or EUR 35 million. A slight reduction from the previously proposed 7% or EUR 40 million in the EP's Negotiating Position.
- **Obligations for system and model providers:** Up to 3% of the company's annual worldwide turnover or EUR 15 million. Similar to the penalties proposed in the EP's Negotiating Position.
- **Failure to provide accurate information:** Up to 1.5% of the company's annual worldwide turnover or EUR 7.5 million. Compared to the EUR 5 million previously proposed in the EP's Negotiating Position and EUR 10 million in the Commission's First Draft.

For each of the above categories of infringement, the threshold will be the lower of the two amounts for SMEs and the higher for other companies.

2.10 *When will the AI Act be applicable?*

The implementation timeline for the AI Act is structured to allow different provisions to come into force at varying intervals after the legislation enters into force. The overarching framework of the AI Act will be applicable **24 months** following its entry into force. However, specific provisions will have different timelines:

- **Prohibited systems**: 6 months after the AI Act's entry into force.
- **Requirements for GPAI:** 12 months after the AI Act's entry into force.
- **Some requirements for high-risk systems**: 36 months after the AI Act's entry into force.

This staggered timeline means that some companies, particularly those dealing with potentially prohibited systems, will need to achieve compliance with the AI Act as early as 2024. Given the significant lead time typically required to change internal business processes, it is imperative that organizations begin to address the requirements of the AI Act promptly.

3 Outlook and timeline

As mentioned in the *Premise Chapter*, the next phase in the legislative procedure of the AI Act involves further technical trilogue negotiations followed by the work of lawyer-linguists.

In more detail:

- A series of technical negotiations will be needed to translate the political agreement into a final legal text. These negotiations will last until January 2024. A consolidated text will presumably be available by the end of January 2024 at the earliest.
- The provisional text will need to be adopted by the COREPER (Committee of Permanent Representatives) and by the Internal Market and Civil Liberties Committees in the European Parliament - potentially in late January or early February 2024.
- The text will then be revised by lawyer-linguists for several weeks before being formally adopted by the plenary of the Parliament and then by the Council at ministerial level - both expected in April 2024.
- The following publication in the EU's Official Journal could take place in May or June 2024, with the AI Act's entry into force 20 days later - possibly by the end of the second quarter of 2024. The periods set out in Section 2.10 will commence on that date.

We will continue to monitor developments in relation to the AI Act closely. As more information becomes available, we will update our analysis to reflect these changes.

Once the final text of the AI Act is available, we will also publish an updated version of our easy-to-read version of the AI Act. This updated version will retain all the conveniences of the previous edition, such as a clickable table of contents, to ensure that it remains a practical and accessible resource for understanding the provisions and implications of the Act.

4 Contact

Oliver Belitz

Senior Associate

Frankfurt, Germany
+4969742226148
oliver.belitz@twobirds.com

Dr. Simon Hembt

Senior Associate

Frankfurt, Germany
+4969742226203
simon.hembt@twobirds.com

Bird & Bird maintains of the largest pan European IP practices, giving the group a strong international reach. The practice continues to advise on a number of high-value cross-border disputes and transactional matters, with an increasing amount of matters involving digital assets, rights and artificial intelligence.

Legal 500, 2023

Thank you

twobirds.com

Abu Dhabi • Amsterdam • Beijing • Bratislava • Brussels • Budapest • Casablanca • Copenhagen • Dubai • Dublin • Dusseldorf • Frankfurt • The Hague • Hamburg • Helsinki • Hong Kong • London • Luxembourg • Lyon • Madrid • Milan • Munich • Paris • Prague • Rome • San Francisco • Shanghai • Shenzhen • Singapore • Stockholm • Sydney • Warsaw