

Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world

Following 3-day 'marathon' talks, the Council presidency and the European Parliament's negotiators have reached a provisional agreement on the proposal on harmonised rules on artificial intelligence (AI), the so-called **artificial intelligence act**. The draft regulation aims to ensure that AI systems placed on the European market and used in the EU are **safe** and respect **fundamental rights** and EU values. This landmark proposal also aims to stimulate investment and innovation on AI in Europe.

This is a historical achievement, and a huge milestone towards the future! Today's agreement effectively addresses a global challenge in a fast-evolving technological environment on a key area for the future of our societies and economies. And in this endeavour, we managed to keep an extremely delicate balance: boosting innovation and uptake of artificial intelligence across Europe whilst fully respecting the fundamental rights of our citizens.

Carme Artigas, Spanish secretary of state for digitalisation and artificial intelligence

The AI act is a **flagship** legislative initiative with the potential to foster the development and uptake of safe and trustworthy AI across the EU's single market by both private and public actors. The main idea is to regulate AI based on the latter's capacity to cause harm to society following a '**risk-based**' approach: **the higher the risk, the stricter the rules**. As the first legislative proposal of its kind in the world, it can set a **global standard** for AI regulation in other jurisdictions, just as the GDPR has done, thus promoting the European approach to tech regulation in the world stage.

The main elements of the provisional agreement

Compared to the initial Commission proposal, the main new elements of the provisional agreement can be summarised as follows:

- rules on **high-impact general-purpose AI models** that can cause systemic risk in the future, as well as on high-risk **AI systems**
- a revised system of **governance** with some enforcement powers at EU level
- extension of the list of **prohibitions** but with the possibility to use **remote biometric identification** by law enforcement authorities in public spaces, subject to safeguards
- better protection of rights through the obligation for deployers of high-risk AI systems to conduct a **fundamental rights impact assessment** prior to putting an AI system into use.

In more concrete terms, the provisional agreement covers the following aspects:

Definitions and scope

To ensure that the **definition** of an AI system provides sufficiently clear criteria for distinguishing AI from simpler software systems, the compromise agreement aligns the definition with the approach proposed by the OECD.

The provisional agreement also clarifies that the regulation does not apply to areas outside the scope of EU law and should not, in any case, affect member states' competences in **national security** or any entity entrusted with tasks in this area. Furthermore, the AI act will not apply to systems which are used exclusively for **military** or **defence** purposes. Similarly, the agreement provides that the regulation would not apply to AI systems used for the sole purpose of **research and innovation**, or for people using AI for non-professional reasons.

Classification of AI systems as high-risk and prohibited AI practices

The compromise agreement provides for a **horizontal layer of protection**, including a high-risk classification, to ensure that AI systems that are not likely to cause serious fundamental rights violations or other significant risks are not captured. AI systems

presenting only **limited risk** would be subject to very light **transparency obligations**, for example disclosing that the content was AI-generated so users can make informed decisions on further use.

A wide range of **high-risk** AI systems would be authorised, but subject to a set of requirements and obligations to gain access to the EU market. These requirements have been clarified and adjusted by the co-legislators in such a way that they are more **technically feasible** and **less burdensome** for stakeholders to comply with, for example as regards the quality of data, or in relation to the technical documentation that should be drawn up by SMEs to demonstrate that their high-risk AI systems comply with the requirements.

Since AI systems are developed and distributed through complex value chains, the compromise agreement includes changes clarifying the **allocation of responsibilities** and **roles of the various actors** in those chains, in particular providers and users of AI systems. It also clarifies the relationship between responsibilities under the AI Act and responsibilities that already exist under other legislation, such as the relevant EU data protection or sectorial legislation.

For some uses of AI, risk is deemed **unacceptable** and, therefore, these systems will be banned from the EU. The provisional agreement bans, for example, cognitive **behavioural manipulation**, the untargeted **scrapping** of facial images from the internet or CCTV footage, **emotion recognition** in the workplace and educational institutions, **social scoring**, **biometric categorisation** to infer sensitive data, such as sexual orientation or religious beliefs, and some cases of **predictive policing** for individuals.

Law enforcement exceptions

Considering the specificities of **law enforcement authorities** and the need to preserve their ability to use AI in their vital work, several changes to the Commission proposal were agreed relating to the use of AI systems for law enforcement purposes. Subject to appropriate **safeguards**, these changes are meant to reflect the need to respect the confidentiality of sensitive operational data in relation to their activities. For example, an emergency procedure was introduced allowing law enforcement agencies to deploy a high-risk AI tool that has not passed the **conformity assessment** procedure in case of urgency. However, a specific mechanism has been also introduced to ensure that **fundamental rights** will be sufficiently protected against any potential misuses of AI systems.

Moreover, as regards the use of real-time **remote biometric identification** systems in publicly accessible spaces, the provisional agreement clarifies the objectives where such use is strictly necessary for law enforcement purposes and for which law enforcement authorities should therefore be exceptionally allowed to use such systems. The compromise agreement provides for **additional safeguards** and limits these exceptions to cases of victims of certain crimes, prevention of genuine, present, or foreseeable threats, such as terrorist attacks, and searches for people suspected of the most serious crimes.

General purpose AI systems and foundation models

New provisions have been added to take into account situations where AI systems can be used for many different purposes (**general purpose AI**), and where general-purpose AI technology is subsequently integrated into another high-risk system. The provisional agreement also addresses the specific cases of general-purpose AI (GPAI) systems.

Specific rules have been also agreed for **foundation models**, large systems capable to competently perform a wide range of distinctive tasks, such as generating video, text, images, conversing in lateral language, computing, or generating computer code. The provisional agreement provides that foundation models must comply with specific **transparency obligations** before they are placed in the market. A stricter regime was introduced for **'high impact'** foundation models. These are foundation models trained with large amount of data and with advanced complexity, capabilities, and performance well above the average, which can disseminate systemic risks along the value chain.

A new governance architecture

Following the new rules on GPAI models and the obvious need for their enforcement at EU level, an **AI Office** within the Commission is set up tasked to oversee these most advanced AI models, contribute to fostering standards and testing practices, and enforce the common rules in all member states. A **scientific panel of independent experts** will advise the AI Office about GPAI models, by contributing to the development of methodologies for evaluating the capabilities of foundation models, advising on the designation and the emergence of high impact foundation models, and monitoring possible material safety risks related to foundation models.

The **AI Board**, which would comprise member states' representatives, will remain as a coordination platform and an advisory body to the Commission and will give an important role to Member States on the implementation of the regulation, including the design of codes of practice for foundation models. Finally, an **advisory forum** for stakeholders, such as industry representatives, SMEs, start-ups, civil society, and academia, will be set up to provide technical expertise to the AI Board.

Penalties

The fines for violations of the AI act were set as a percentage of the offending company's global annual turnover in the previous financial year or a predetermined amount, whichever is higher. This would be €35 million or 7% for violations of the banned AI applications, €15 million or 3% for violations of the AI act's obligations and €7,5 million or 1,5% for the supply of incorrect information. However, the provisional agreement provides for **more proportionate caps** on administrative fines for SMEs and start-ups in case of infringements of the provisions of the AI act.

The compromise agreement also makes clear that a natural or legal person may make a complaint to the relevant **market surveillance authority** concerning non-compliance with the AI act and may expect that such a complaint will be handled in line with the dedicated procedures of that authority.

Transparency and protection of fundamental rights

The provisional agreement provides for a **fundamental rights impact assessment** before a high-risk AI system is put in the market by its deployers. The provisional agreement also provides for increased **transparency** regarding the use of high-risk AI systems. Notably, some provisions of the Commission proposal have been amended to indicate that certain users of a high-risk AI system that are public entities will also be obliged to register in the **EU database** for high-risk AI systems. Moreover, newly added provisions put emphasis on an obligation for users of an **emotion recognition system** to inform natural persons when they are being exposed to such a system.

Measures in support of innovation

With a view to creating a legal framework that is more innovation-friendly and to promoting evidence-based regulatory learning, the provisions concerning **measures in support of innovation** have been substantially modified compared to the Commission proposal.

Notably, it has been clarified that AI **regulatory sandboxes**, which are supposed to establish a controlled environment for the development, testing and validation of innovative AI systems, should also allow for testing of innovative AI systems in real world conditions. Furthermore, new provisions have been added allowing **testing** of AI systems in **real world conditions**, under specific conditions and safeguards. To alleviate the administrative burden for smaller companies, the provisional agreement includes a list of actions to be undertaken to support such operators and provides for some limited and clearly specified derogations.

Entry into force

The provisional agreement provides that the AI act should apply **two years** after its entry into force, with some exceptions for specific provisions.

Next steps

Following today's provisional agreement, work will continue at technical level in the coming weeks to finalise the details of the new regulation. The presidency will submit the compromise text to the member states' representatives (Coreper) for endorsement once this work has been concluded.

The entire text will need to be confirmed by both institutions and undergo legal-linguistic revision before formal adoption by the co-legislators.

Background information

The Commission proposal, presented in April 2021, is a key element of the EU's policy to foster the development and uptake across the single market of safe and lawful AI that respects fundamental rights.

The proposal follows a risk-based approach and lays down a uniform, horizontal legal framework for AI that aims to ensure legal certainty. The draft regulation aims to promote investment and innovation in AI, enhance governance and effective enforcement of existing law on fundamental rights and safety, and facilitate the development of a single market for AI applications. It goes hand in hand with other initiatives, including the coordinated plan on artificial intelligence which aims to accelerate investment in AI in Europe. On 6 December 2022, the Council reached an agreement for a general approach (negotiating mandate) on this file and entered interinstitutional talks with the European Parliament ('trilogues') in mid-June 2023.

Press office - General Secretariat of the Council

Rue de la Loi 175 - B-1048 BRUSSELS - Tel.: +32 (0)2 281 6319

press@consilium.europa.eu - www.consilium.europa.eu/press