AI Act

The AI Act (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689) (Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence)is the first-ever comprehensive legal framework on AI worldwide. The aim of the rules is to foster trustworthy AI in Europe.

The AI Act sets out a clear set of risk-based rules for AI developers and deployers regarding specific uses of AI. The AI Act is part of a wider package of policy measures to support the development of trustworthy AI, which also includes the AI Innovation Package (https://ec.europa.eu/commission/presscorner/detail/en/ip_24_383), the launch of AI Factories (https://digital-strategy.ec.europa.eu/en/policies/ai-factories) and the Coordinated Plan on AI (https://digital-strategy.ec.europa.eu/en/policies/plan-ai). Together, these measures guarantee safety, fundamental rights and human-centric AI, and strengthen uptake, investment and innovation in AI across the EU.

To facilitate the transition to the new regulatory framework, the Commission has launched the AI Pact (https://digital-strategy.ec.europa.eu/en/policies/ai-pact), a voluntary initiative that seeks to support the future implementation, engage with stakeholders and invite AI providers and deployers from Europe and beyond to comply with the key obligations of the AI Act ahead of time.

Why do we need rules on AI?

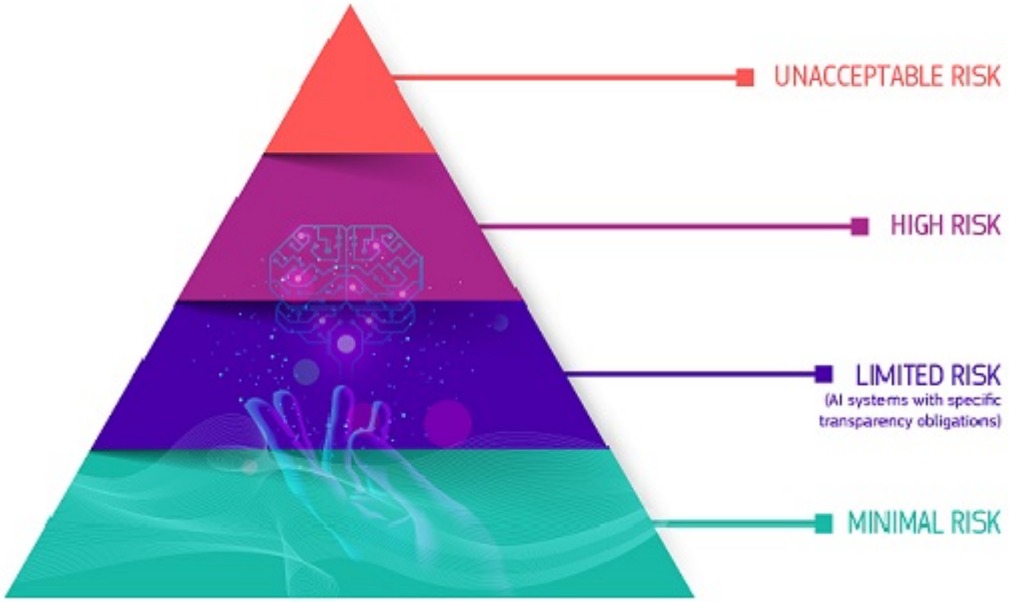
The AI Act ensures that Europeans can trust what AI has to offer. While most AI systems pose limited to no risk and can contribute to solving many societal challenges, certain AI systems create risks that we must address to avoid undesirable outcomes.

For example, it is often not possible to find out why an AI system has made a decision or prediction and taken a particular action. So, it may become difficult to assess whether someone has been unfairly disadvantaged, such as in a hiring decision or in an application for a public benefit scheme.

Although existing legislation provides some protection, it is insufficient to address the specific challenges AI systems may bring.

A risk-based approach

The AI Act defines 4 levels of risk for AI systems:



Unacceptable risk

All AI systems considered a clear threat to the safety, livelihoods and rights of people are banned. The **AI Act prohibits eight practices**, namely:

1. harmful AI-based manipulation and deception
2. harmful AI-based exploitation of vulnerabilities
3. social scoring
4. Individual criminal offence risk assessment or prediction
5. untargeted scraping of the internet or CCTV material to create or expand facial recognition databases
6. emotion recognition in workplaces and education institutions
7. biometric categorisation to deduce certain protected characteristics
8. real-time remote biometric identification for law enforcement purposes in publicly accessible spaces

High risk

AI use cases that can pose serious risks to health, safety or fundamental rights are classified as high-risk. These **high-risk** use-cases include:

- AI safety components in critical infrastructures (e.g. transport), the failure of which could put the life and health of citizens at risk
- AI solutions used in education institutions, that may determine the access to education and course of someone's professional life (e.g. scoring of exams)
- AI-based safety components of products (e.g. AI application in robot-assisted surgery)
- AI tools for employment, management of workers and access to self-employment (e.g. CV-sorting software for recruitment)
- Certain AI use-cases utilised to give access to essential private and public services (e.g. credit scoring denying citizens opportunity to obtain a loan)
- AI systems used for remote biometric identification, emotion recognition and biometric categorisation (e.g AI system to retroactively identify a shoplifter)
- AI use-cases in law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence)
- AI use-cases in migration, asylum and border control management (e.g. automated examination of visa applications)
- AI solutions used in the administration of justice and democratic processes (e.g. AI solutions to prepare court rulings)

High-risk AI systems are subject to **strict obligations** before they can be put on the market:

- adequate risk assessment and mitigation systems
- high-quality of the datasets feeding the system to minimise risks of discriminatory outcomes
- logging of activity to ensure traceability of results
- detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance

- clear and adequate information to the deployer
- appropriate human oversight measures
- high level of robustness, cybersecurity and accuracy

Transparency risk

This refers to the risks associated with a need for transparency around the use of AI. The AI Act introduces specific disclosure obligations to ensure that humans are informed when necessary to preserve trust. For instance, when using AI systems such as chatbots, humans should be made aware that they are interacting with a machine so they can take an informed decision.

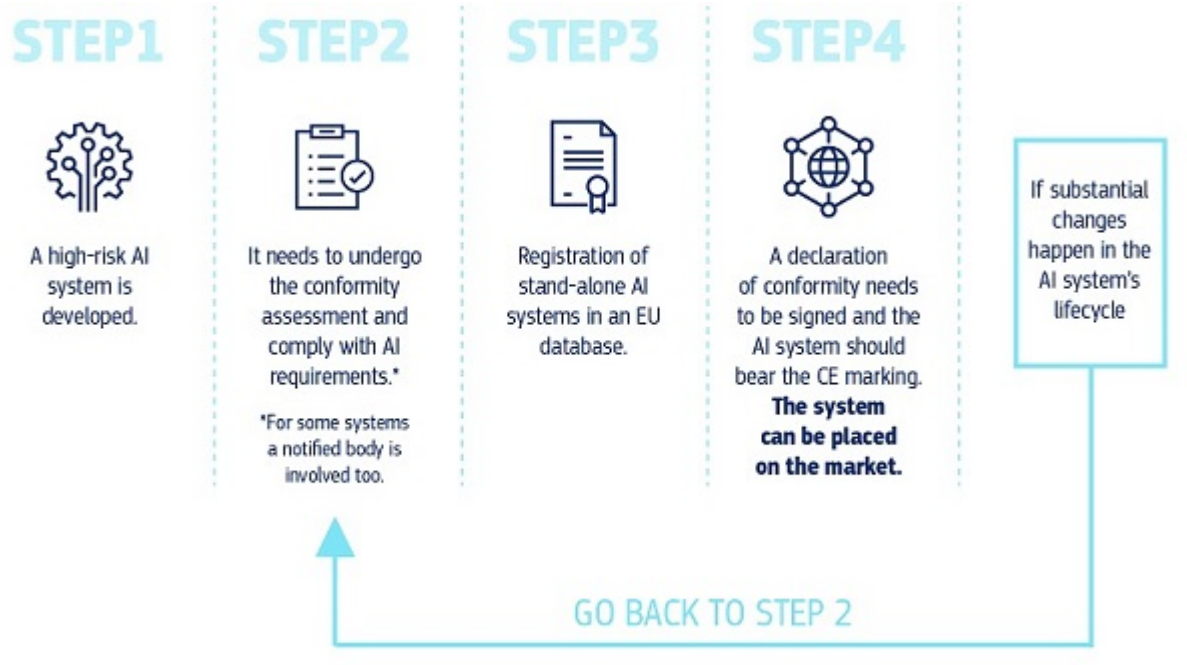
Moreover, providers of generative AI have to ensure that AI-generated content is identifiable. On top of that, certain AI-generated content should be clearly and visibly labelled, namely deep fakes and text published with the purpose to inform the public on matters of public interest.

Minimal or no risk

The AI Act does not introduce rules for AI that is deemed minimal or no risk. The vast majority of AI systems currently used in the EU fall into this category. This includes applications such as AI-enabled video games or spam filters.

How does it all work in practice for providers of high-risk AI systems?

How does it all work in practice for providers of high-risk AI systems?



Once an AI system is on the market, authorities are in charge of market surveillance, deployers ensure human oversight and monitoring, and providers have a post-market monitoring system in place. Providers and deployers will also report serious incidents and malfunctioning.

A solution for the trustworthy use of large AI models

General-purpose AI models can perform a wide range of tasks and are becoming the basis for many AI systems in the EU. Some of these models could carry systemic risks if they are very capable or widely used. To ensure safe and trustworthy AI, the AI Act puts in place rules for providers of such models. This includes transparency and copyright-related rules. For models that may carry systemic risks, providers should assess and mitigate these risks.

The AI Act rules on general-purpose AI will become effective in August 2025. The AI Office is facilitating the drawing-up of a Code of Practice to detail out these rules. The Code should represent a central tool for providers to demonstrate compliance with the AI Act, incorporating state-of-the-art practices.

Governance and implementation

The European AI Office (https://digital-strategy.ec.europa.eu/en/policies/ai-office) and authorities of the Member States are responsible for implementing, supervising and enforcing the AI Act. The AI Board, the Scientific Panel and the Advisory Forum steer and advise the AI Act's governance. Find out more details about the Governance and enforcement of the AI Act (https://digital-strategy.ec.europa.eu/en/policies/ai-act-governance-and-enforcement).

Next steps

The AI Act entered into force on 1 August 2024, and will be fully applicable 2 years later on 2 August 2026, with some exceptions:

- prohibitions and AI literacy obligations entered into application from 2 February 2025
- the governance rules and the obligations for general-purpose AI models become applicable on 2 August 2025
- the rules for high-risk AI systems - embedded into regulated products - have an extended transition period until 2 August 2027
