> *'An AI system is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and 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'.*

The definition is not intended to cover simpler traditional software systems or programming approaches, and the Commission has been tasked to develop **guidelines** on its application.

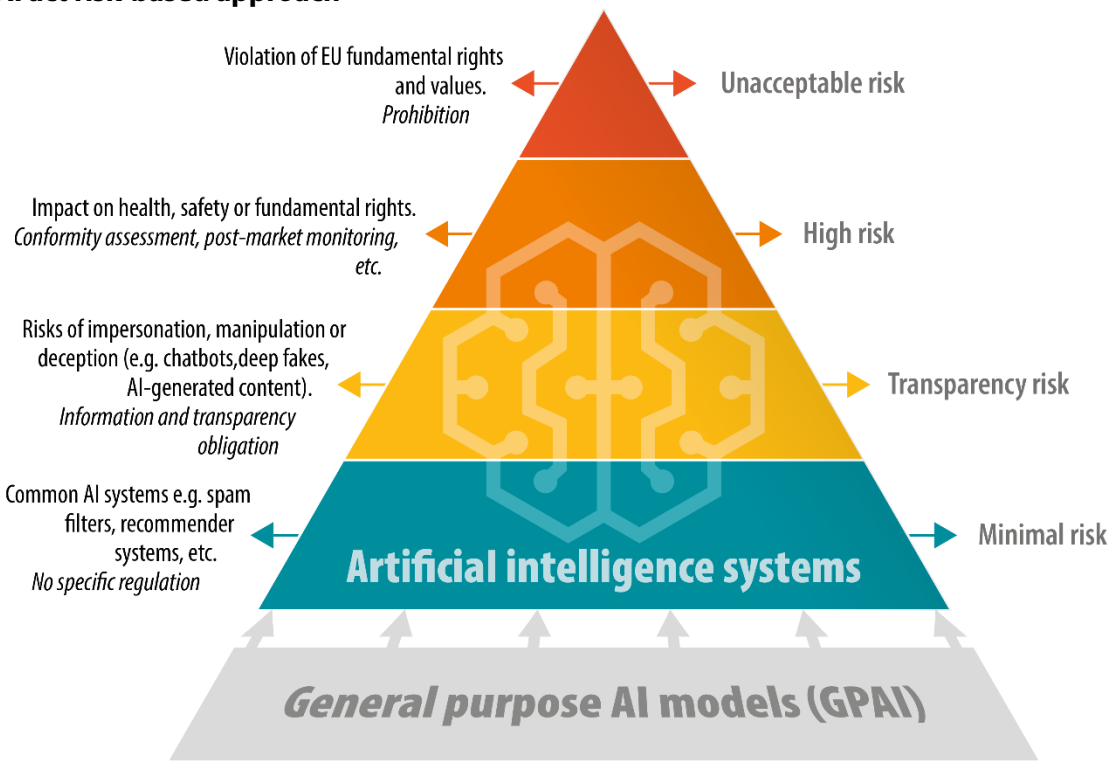
The act also contains a definition of **general purpose artificial intelligence (GPAI) models** 'that are trained with a large amount of data using self-supervision at scale', that display 'significant generality' and are 'capable to competently perform a wide range of distinct tasks' and 'can be integrated into a variety of downstream systems or applications'. Furthermore, the AI act defines **general-purpose AI systems** as systems based on a GPAI model, which have the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

Scope of application

The AI act applies primarily to providers and deployers putting AI systems and GPAI models into service or placing on the EU market and who have their place of establishment or who are located in the EU, as well as to deployers or providers of AI systems that are established in a third country, when the output produced by their systems is used in the EU.[12] However, AI systems placed on the market, put into service, or used by public and private entities for **military, defence or national security** purposes, are excluded from the scope. Similarly, the AI act will not apply to AI systems and models, including their output, which are specifically developed and put into service for the sole purpose of **scientific research and development**. Furthermore, as matter of principle, the regulation does not apply prior to the systems and models being put into service or placed on the market (sandboxing rules may apply in this case).

Risk-based approach

EU AI act risk-based approach



Data source: European Commission

The final agreement maintains the risk-based approach proposed by the Commission and classifies AI systems into several risk categories, with different degrees of regulation applying.

- **Prohibited AI practices**. The final text prohibits a wider range of AI practices as originally proposed by the Commission because of their harmful impact:
  - AI systems using subliminal or manipulative or deceptive techniques to distort people's or a group of people's behaviour and impair informed decision-making, leading to significant harm;
  - AI systems exploiting vulnerabilities due to age, disability, or social or economic situations, causing significant harm;
  - Biometric categorisation systems inferring race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation (except for lawful labelling or filtering in law-enforcement purposes);
  - AI systems evaluating or classifying individuals or groups based on social behaviour or personal characteristics, leading to detrimental or disproportionate treatment in unrelated contexts or unjustified or disproportionate to their behaviour;
  - 'Real-time' remote biometric identification in public spaces for law enforcement (except for specific necessary objectives such as searching for victims of abduction, sexual exploitation or missing persons, preventing certain substantial and imminent threats to safety, or identifying suspects in serious crimes);
  - AI systems assessing the risk of individuals committing criminal offences based solely on profiling or personality traits and characteristics (except when supporting human assessments based on objective, verifiable facts linked to a criminal activity);
  - AI systems creating or expanding facial recognition databases through untargeted scraping from the internet or CCTV footage;
  - AI systems inferring emotions in workplaces or educational institutions, except for medical or safety reasons.
- **High-risk AI systems**. The AI act identifies a number of use cases in which AI systems are to be considered high risk because they can potentially create an adverse impact on people's health, safety or their fundamental rights.
  - The **risk classification** is based on the intended purpose of the AI system. The function performed by the AI system and the specific purpose and modalities for which the system is used are key to determine if an AI system is high-risk or not. High-risk AI systems can be safety components of products covered by **sectoral EU law** (e.g. medical devices) or AI systems that, as a matter of principle, are considered to be high-risk when they are used in **specific areas** listed in an annex.[13] The Commission is tasked with maintaining an EU database for the high-risk AI systems listed in this annex.
  - A new test has been enshrined at the Parliament's request (**'filter provision'**), according to which AI systems will not be considered high-risk if they do not pose a significant risk of harm to the health, safety or fundamental rights of natural persons.[14] However, an AI system will always be considered high-risk if the AI system performs profiling of natural persons.
  - Providers of such high-risk AI systems will have to run a **conformity assessment procedure** before their products can be sold and used in the EU. They will need to comply with a range of requirements including for testing, data training and cybersecurity and, in some cases, will have to conduct a fundamental rights impact assessment to ensure their systems comply with EU law. The conformity assessment should be carried out either based on

internal control (self-assessment) or with the involvement of a notified body (e.g. biometrics). Compliance with European harmonised standards to be developed will grant high-risk AI systems providers a **presumption of conformity**. After such AI systems are placed in the market, providers must implement post-market monitoring and take corrective actions if necessary.

- **Transparency risk.** Certain AI systems intended to interact with natural persons or to generate content may pose specific risks of impersonation or deception, irrespective of whether they qualify as high-risk AI systems or not. Such systems are subject to information and transparency requirements. Users must be made aware that they interact with chatbots. Deployers of AI systems that generate or manipulate image, audio or video content (i.e. **deep fakes),** must disclose that the content has been artificially generated or manipulated except in very limited cases (e.g. when it is used to prevent criminal offences). Providers of AI systems that generate large quantities of **synthetic content** must implement sufficiently reliable, interoperable, effective and robust techniques and methods (such as watermarks) to enable marking and detection that the output has been generated or manipulated by an AI system and not a human. Employers who deploy **AI systems in the workplace** must inform the workers and their representatives.
- **Minimal risks.** Systems presenting minimal risk for people (e.g. spam filters) will not be subject to further obligations beyond currently applicable legislation (e.g., GDPR).
- **General-purpose AI (GPAI).** The regulation provides specific rules for general-purpose AI models and for general-purpose AI models that pose systemic risks.
  - **GPAI system transparency requirements**. All GPAI models will have to draw up and maintain up-to-date technical documentation and make information and documentation available to downstream providers of AI systems. All providers of GPAI models have to put a policy in place to respect Union **copyright law**, including through state-of-the-art technologies (e.g. watermarking), to carry out lawful text-and-data mining exceptions as envisaged under the Copyright Directive. Furthermore, GPAIs must draw up and make publicly available a sufficiently **detailed summary of the content used in training the GPAI models** according to a template provided by the AI Office.[15] Finally, if located outside the EU, they will have to appoint a **representative** in the EU. However, AI models made accessible under a **free and open source** will be exempt from some of the obligations (i.e. disclosure of technical documentation) given they have, in principle, positive effects on research, innovation and competition.[16]
  - **Systemic-risk GPAI obligations**. GPAI models with **'high-impact capabilities'** could pose a systemic risk and have a significant impact on the internal market, due to their reach and their actual or reasonably foreseeable negative effects (on public health, safety, public security, fundamental rights, or the society as a whole). GPAI providers must therefore notify the European Commission if their model is trained using a **total computing power** exceeding 10^25 FLOPs (i.e. floating-point operations per second). When this threshold is met, the presumption will be that the model is a GPAI model posing systemic risks.[17] In addition to the requirements on transparency and copyright protection falling on all GPAI models, providers of systemic-risk GPAI models are required to **constantly assess and mitigate the risks** they pose and to ensure cybersecurity protection. That requires, inter alia, keeping track of, documenting and reporting serious incidents (e.g. violations of fundamental rights) and implementing corrective measures.
  - **Codes of practice and presumption of conformity**. GPAI model providers will be able to rely on codes of practice to demonstrate compliance with the

obligations set under the act. By means of implementing acts, the Commission may decide to approve a code of practice and give it a general validity within the EU, or alternatively, provide common rules for implementing the relevant obligations. Compliance with a European harmonised standard grants GPAI providers the presumption of conformity. Providers of GPAI models with systemic risks who do not adhere to an approved code of practice will be required to demonstrate adequate alternative means of compliance.

Sandboxing and real-world testing

The measures to support investment in AI systems have been strengthened. National authorities must establish at least one AI regulatory sandbox at national level to facilitate the development and testing of innovative AI systems under strict regulatory oversight.[18] Such **regulatory sandboxes** provide for a controlled environment that fosters innovation and facilitates the development, training, testing and validation of innovative AI systems for a limited time before their placement on the market or entry into service. The AI regulatory sandbox must enable, where appropriate, testing of AI systems in real-world conditions outside of a laboratory for a limited period (subject to compliance with EU data protection law rules and principles). Furthermore, to accelerate the development and placing on the market of high-risk AI systems, providers or prospective providers of such systems may also test them in **real-world conditions** – even without participating in an AI regulatory sandbox – if they respect some guarantees and conditions (e.g. ask for specific consent, submit their real-world testing plan to the market surveillance authority).

Enforcement and institutional setting

The implementation of the act will be the responsibility of a number of national and EU-level actors. Member States must establish or designate at least one market surveillance authority and at least one notifying authority to ensure the application and implementation of the act. **Heavy fines** will fall on non-compliant entities.[19] At EU level, a range of actors including the Commission, the AI Board, the AI Office, the EU standardisation bodies (CEN and CENELEC) and an advisory forum and scientific panel of independent experts will support the implementation of the act. The **EU AI Office** was established to provide advice on the implementation of the new rules, in particular as regards GPAI models and to develop codes of practice to support the proper application of the AI act.

'Entry into force' timelines

Prohibited systems have to be phased out within **six months** after the act enters into force. The provisions concerning GPAI and penalties will apply **12 months** after the act enters into force, and those concerning high-risk AI systems apply **24 months** after entry into force (36 months after entry into force for AI systems covered by existing EU product legislation). The codes of practice envisaged must be ready, at the latest, nine months after the AI act enters into force. The implementation of the AI act requires a number of steps to be taken. In the coming months, the Commission is expected to issue various **implementing, delegated and guidelines** related to the act[20] and to oversee the **standardisation process** required for implementing the obligations.[21]

Policy debate latest issues. Academics have raised a number of questions as regards the final text of the AI act and the implementation challenges lying ahead. Hacker welcomes the final AI act text but stresses, inter alia: that alignment with existing sectoral regulation is incomplete (which results in unnecessary and highly detrimental red tape); compliance costs will be substantial, especially for SMEs developing narrow AI models; the threshold of 10^25 FLOPs for a default categorisation of systemic risk models is too high; and calls for European supervision and monitoring of remote biometric identification to avoid the risk that some Member States circumvent the rules enshrined in the AI act.[22] Kutterer argues the AI act's implementation will require a robust taxonomy setting out the correlation of risk classification and model capabilities and assessing the developments of open sources models.[23] Helberger and others call for the AI act to be complemented by an additional set of exercisable rights to protect citizens from AI-generated harm, with additional legislation to