BRIEFING
EU Legislation in Progress

European Parliament

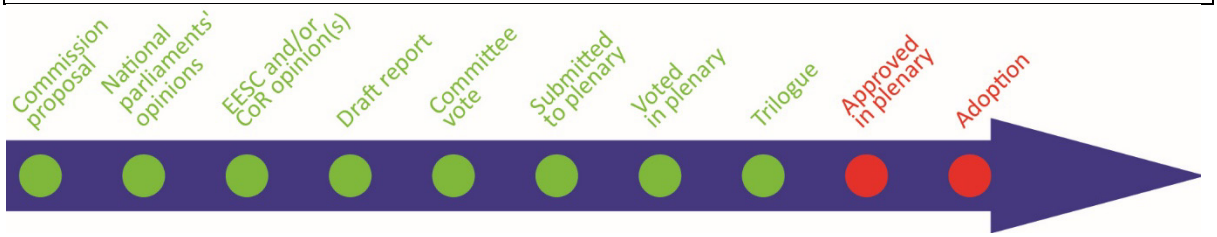
Artificial intelligence act

OVERVIEW

European Union lawmakers reached a political agreement on the draft artificial intelligence (AI) act in December 2023. Proposed by the European Commission in April 2021, the draft AI act, the first binding worldwide horizontal regulation on AI, sets a common framework for the use and supply of AI systems in the EU. It offers a classification for AI systems with different requirements and obligations tailored on a 'risk-based approach'. Some AI systems presenting 'unacceptable' risks are prohibited. A wide range of 'high-risk' AI systems that can have a detrimental impact on people's health, safety or on their fundamental rights are authorised, but subject to a set of requirements and obligations to gain access to the EU market. AI systems posing limited risks because of their lack of transparency will be subject to information and transparency requirements, while AI systems presenting only minimal risk for people will not be subject to further obligations. The regulation also provides specific rules for general purpose AI (GPAI) models and lays down more stringent requirements for GPAI models with 'high-impact capabilities' that could pose a systemic risk and have a significant impact on the internal market.

The provisional agreement has been endorsed by the Committee of Permanent Representatives of EU Member States and by Parliament's two lead committees. Parliament's plenary vote on the final agreement is scheduled for the March plenary session. The AI act must also be endorsed by Council and published in the EU's Official Journal before entering into force.

Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts		
Committees responsible:	Internal Market and Consumer Protection (IMCO) and Civil Liberties, Justice and Home Affairs (LIBE) (jointly under Rule 58)	COM(2021)206 21.4.2021 2021/0106(COD)
Rapporteurs:	Brando Benifei (S&D, Italy) and Dragoş Tudorache (Renew, Romania)	
Shadow rapporteurs:	Deirdre Clune, Axel Voss (EPP); Petar Vitanov (S&D); Svenja Hahn, (Renew); Sergey Lagodinsky, Kim Van Sparrentak (Greens/EFA); Rob Rooken, Kosma Złotowski (ECR); Jean-Lin Lacapelle, Jaak Madison (ID); Cornelia Ernst, Kateřina Konecna (The Left)	Ordinary legislative procedure (COD) (Parliament and Council on equal footing – formerly 'co-decision')
Next steps expected:	Final first-reading vote in plenary	



EPRS | European Parliamentary Research Service

Author: Tambiama Madiega
Members' Research Service
PE 698.792 – March 2024

EN

Introduction

AI technologies are expected to bring a wide array of **economic and societal benefits** to a wide range of sectors, including environment and health, the public sector, finance, mobility, home affairs and agriculture. They are particularly useful for improving prediction, for optimising operations and resource allocation, and for personalising services.[1] However, the implications of AI systems for **fundamental rights** protected under the EU Charter of Fundamental Rights, as well as the **safety risks** for users when AI technologies are embedded in products and services, are raising concern. Most notably, AI systems may jeopardise fundamental rights such as the right to non-discrimination, freedom of expression, human dignity, personal data protection and privacy.[2]

Given the fast development of these technologies, in recent years AI regulation has become a central policy question in the European Union (EU). Policy-makers pledged to develop a **'human-centric' approach to AI** to ensure that Europeans can benefit from new technologies developed and functioning according to the EU's values and principles. In its 2020 White Paper on Artificial Intelligence, the European Commission committed to **promote the uptake of AI** and **address the risks associated** with certain uses of this new technology. After having initially adopted a **soft-law approach** with the publication of its non-binding 2019 Ethics Guidelines for Trustworthy AI and Policy and investment recommendations, the European Commission shifted towards a **legislative approach**, calling for the adoption of harmonised rules for the development, placing on the market and use of AI systems.

Leading the EU-level debate, the Parliament called on the Commission to assess the impact of AI and to draft an EU framework for AI, in its wide-ranging 2017 recommendations on civil law rules on robotics. In 2020 and 2021, Parliament adopted a number of non-legislative resolutions calling for EU action,[3] as well as two legislative resolutions asking the Commission to establish a legal framework of ethical principles for the development, deployment and use of AI, robotics and related technologies in the Union and harmonising the legal framework for civil liability claims and imposition of a regime of strict liability on operators of high-risk AI systems.

In the past, the Council has repeatedly called for the adoption of common AI rules, including in 2017 and 2019. In 2020, the Council called upon the Commission to put forward concrete proposals that take existing legislation into account and follow a risk-based, proportionate and, if necessary, regulatory approach.

The Commission launched a broad public consultation in 2020 and published an Impact Assessment of the regulation on artificial intelligence, a supporting study and a draft proposal, which received feedback from stakeholders.[4] In its impact assessment, the Commission identified several problems raised by the development and use of AI systems, due to their **specific characteristics**, namely: (i) opacity (limited ability of the human mind to understand how certain AI systems operate), (ii) complexity, (iii) continuous adaptation and unpredictability, (iv) autonomous behaviour, and (v) functional dependence on data and on the quality of data.

AI regulatory approach in the world. An increasing number of countries worldwide are designing and implementing AI governance legislation and policies. While the United States of America (USA) had initially taken a lenient approach towards AI, calls for regulation have recently been mounting. The White House has released the Blueprint for an AI Bill of Rights, a set of guidelines to protect the rights of the American public in the age of AI and President Joe Biden signed an executive order on AI in 2023. The Cyberspace Administration of China issued some guidelines on generative AI services, while the UK has announced a pro-innovation approach to AI regulation, which largely regulates AI via existing laws. At international level, the Organisation for Economic Co-operation and Development (OECD) adopted some non-binding Principles on AI, in 2019, UNESCO embraced a set of Recommendations on the Ethics of AI in 2021, the G7 agreed some International Guiding Principles on Artificial Intelligence in 2023 and the Council of Europe is currently finalising an international convention on AI. Furthermore, in the context of the newly established EU-US tech partnership (the Trade and Technology Council), the EU and the USA are seeking to develop a mutual understanding on the principles underpinning trustworthy and responsible AI.

The changes the proposal would bring

The draft AI act was designed as a **horizontal EU legislative instrument** applicable to all AI systems placed on the market or used in the Union, based on Article 114 and Article 16 of the Treaty on the Functioning of the European Union (TFEU) following the logic of the new legislative framework (NLF), i.e. the EU's approach to ensuring a range of products comply with the applicable legislation when they are placed on the EU market through conformity assessments and the use of CE marking.

The Commission proposed enshrining in EU law a legal definition of 'AI system' referring to a range of software-based technologies using specific techniques and approaches ('**machine learning**', '**logic and knowledge-based**' systems, and '**statistical**' approaches) that could be complemented through the adoption of **delegated acts** to factor in technological developments.

The Commission also proposed to adopt a **risk-based approach** whereby legal intervention was tailored to concrete level of risk. Four categories were identified.

First, the draft act proposed to explicitly ban the following **harmful AI practices** that are considered to be a clear threat to people's safety, livelihoods and rights, because of the 'unacceptable risk' they create:

- AI systems that deploy harmful manipulative 'subliminal techniques';
- AI systems that exploit specific vulnerable groups (physical or mental disability);
- AI systems used by public authorities, or on their behalf, for social scoring purposes;
- 'Real-time' remote biometric identification systems in publicly accessible spaces for law enforcement purposes, except in a limited number of cases.[5]

Second, the draft act proposed to regulate **high-risk AI systems** that create adverse impact on people's safety or their fundamental rights. The draft text distinguished between two categories of high-risk AI systems.

- Systems used as a safety component of a product or falling under EU health and safety harmonisation legislation (e.g. toys, aviation, cars, medical devices, lifts).
- Systems deployed in eight specific areas specified in Annex (e.g. law enforcement)**,** which the Commission could update as necessary through delegated acts.

Such high-risk AI systems would have to comply with a range of requirements particularly on risk management, testing, technical robustness, data training and data governance, transparency, human oversight, and cybersecurity before being placed on the market or put into service. AI systems that conform to new harmonised EU standards would benefit from a presumption of conformity with the draft AI act requirements.

Third, AI systems presenting **limited risk**, such as systems that interacts with humans (i.e. chatbots), emotion recognition systems, biometric categorisation systems, and AI systems that generate or manipulate image, audio or video content (i.e. deepfakes), would be subject to a limited set of transparency obligations.

Finally, all other AI systems presenting only **low or minimal risk** could be developed and used in the EU without conforming to any additional legal obligations. However, the proposed AI act envisaged the creation of codes of conduct to encourage providers of non-high-risk AI systems to apply the mandatory requirements for high-risk AI systems voluntarily.

The proposal required Member States to designate one or more competent authorities, including a **national supervisory authority**, which would be tasked with supervising the application and implementation of the regulation, and proposed to establish a **European Artificial Intelligence Board** (composed of representatives from the Member States and the Commission) at EU level. National **market surveillance authorities** would be responsible for assessing operators' compliance with the obligations and requirements for high-risk AI systems. Administrative **fines** of

varying scales (up to €30 million or 6 % of the total worldwide annual turnover), depending on the severity of the infringement, were set as sanctions for non-compliance with the AI act.

Some measures were tailored to foster investments. The Commission proposed that Member States, or the European Data Protection Supervisor, could establish a **regulatory sandbox**, i.e. a controlled environment that facilitates the development, testing and validation of innovative AI systems (for a limited period of time) before they are put on the market. Sandboxing would enable participants to use personal data to foster AI innovation, without prejudice to the GDPR requirements. Other proposed measures were tailored specifically to small-scale providers and **start-ups**.

Advisory committees

The European Economic and Social Committee and the European Committee of the Regions adopted their opinions in 2021 and in 2022, respectively.

National parliaments

The deadline for the submission of reasoned opinions on the grounds of subsidiarity was 2 September 2021. Contributions were received from the Czech Chamber of Deputies and the Czech Senate, the Portuguese Parliament, the Polish Senate and the German Bundesrat.

Stakeholder views[6]

Definitions were a contentious point of discussion among stakeholders. The Big Data Value Association, an industry-driven international not–for-profit organisation, stressed that the definition of AI systems was quite broad and would cover far more than what is subjectively understood as AI, including the simplest search, sorting and routing algorithms, which would consequently be subject to new rules. Furthermore, they asked for clarification of how components of larger AI systems (such as pre-trained AI components from other manufacturers or components not released separately), should be treated. AmCham, the American Chamber of Commerce in the EU, suggested avoiding over-regulation by adopting a narrower definition of AI systems, focusing strictly on high-risk AI applications (and not extended to AI applications that are not high-risk, or software in general).

While they generally welcomed the proposed AI act's **risk-based approach**, some stakeholders supported wider prohibition and regulation of AI systems. Civil rights organisations called for a ban on indiscriminate or arbitrarily targeted use of biometrics in public or publicly accessible spaces, and for restrictions on the uses of AI systems, including for border control and predictive policing.

The European Enterprises Alliance stressed that there was general uncertainty about the roles and responsibilities of the different actors in the AI value chain (developers, providers, and users of AI systems). This was particularly challenging for companies providing general purpose application programming interfaces or **open-source AI models** that are not specifically intended for high-risk AI systems but are nevertheless used by third parties in a manner that could be considered high-risk. They also called for 'high-risk' to be redefined, based on the measurable harm and potential impact. AlgorithmWatch underlined that the applicability of specific rules should not depend on the type of technology, but on the impact it has on individuals and society. They called for the new rules to be defined according to the impact of the AI systems and recommend that every operator should conduct an impact assessment that assesses the system's risk levels on a case-by-case basis. Climate Change AI called for climate change mitigation and adaptation to be taken into account in the classification rules for high-risk AI systems and impose environmental protection requirements.

The European Consumer Organisation, BEUC, stressed that the proposal required substantial improvement to guarantee **consumer protection**. The organisation argued that the proposal should have a broader scope and impose basic principles and obligations (e.g. on fairness, accountability and transparency) upon all AI systems, as well as prohibiting more comprehensively harmful practices (such as private entities' use of social scoring and of remote biometric

identification systems in public spaces). Furthermore, consumers should be granted a strong set of rights, effective remedies and redress mechanisms, including collective redress.

There were opposing views on the impact of the proposed regulation on **investment**. A study by the Centre for Data Innovation (representing large online platforms) highlighted that the compliance costs incurred under the proposed AI act would likely provoke a chilling effect on investment in AI in Europe, and could particularly deter small and medium-sized enterprises (SMEs) from developing high-risk AI systems. According to the study, the AI act would cost the European economy €31 billion over the next five years and reduce AI investments by almost 20 %. However, such estimates of the compliance costs were challenged by the experts from the Centre for European Policy Studies, as well as by other economists. The European Digital SME Alliance warned against overly stringent conformity requirements, and asked for effective SME representation in the standards-setting procedures and for mandatory sandboxes in all EU Member States.

Academic and other views

While generally supporting the Commission's proposal, critics called for amendments, including revising the 'AI systems' definition, ensuring a better allocation of responsibility, strengthening enforcement mechanisms and fostering democratic participation.[7] Among the main issues were:

AI systems definition

The legal definition of 'AI systems' contained in the proposed AI act has been heavily criticised. Smuha and others warned the definition lacks clarity and may lead to legal uncertainty, especially for some systems that would not qualify as AI systems under the draft text, while their use may have an adverse impact on fundamental rights.[8] To address this issue, the authors proposed to **broaden the scope of the legislation** to include explicitly all computational systems used in the identified high-risk domains, regardless of whether they are considered to be AI. Ebers and others consider that the scope of 'AI systems' was overly broad, which may lead to **legal uncertainty** for developers, operators, and users of AI systems and ultimately to over-regulation.[9] They called on EU law-makers to exempt AI systems developed and used for **research purposes** and **open-source software** (OSS) from regulation. Other commentators questioned whether the proposed definition of 'AI systems' is truly **technology neutral** as it refers primarily to 'software', omitting potential future AI developments.

Risk-based approach

Academics also called for amendments, warning that the risk-based approach proposed by the Commission would not ensure a high level of protection of fundamental rights. Smuha and others argued that the proposal does not always accurately recognise the wrongs and harms associated with different kinds of AI systems and therefore does not appropriately allocate responsibility. Among other things, they recommended adding a procedure that enables the Commission to **broaden the list of prohibited AI systems**, and proposed banning existing manipulative AI systems (e.g. deepfakes), social scoring and some biometrics. Ebers and others called for a **more detailed classification of risks** to facilitate industry self-assessment and support, as well as **prohibiting more AI systems** (e.g. biometrics), including in the context of **private use**. Furthermore, some highlighted that the draft legislation did not address **systemic sustainability risks** created by AI, especially in the area of climate and environmental protection.[10]

One of the major concerns raised was that the rules on prohibited and high-risk practices might prove ineffective in practice, because the risk assessment was proposed to be left to provider **self-assessment**. Veale and Zuiderveen Borgesius warned that most providers can arbitrarily classify most high-risk systems as adhering to the rules using self-assessment procedures alone. Smuha and others recommended exploring whether certain high-risk systems would not benefit from a conformity assessment carried out by an **independent entity** prior to their deployment.

Biometrics regulation. A study commissioned by the European Parliament recommended, inter alia, to empower the Commission to adapt the list of prohibited AI practices periodically, under the supervision of the European Parliament, and the adoption of a more comprehensive list of 'restricted AI applications' (comprising real-time remote biometric identification without limitation for law enforcement purposes). Regulation of facial recognition technologies (FRTs) was one of the most contentious issues.[11] The European Data Protection Supervisor (EDPS) and the European Data Protection Board (EDPB) called for a general ban on any uses of AI for the automated recognition of human features in publicly accessible spaces.

Governance structure and enforcement and redress mechanisms

Ebers et al. stressed that the AI act **lacks effective enforcement structures**, as the Commission proposed to leave the preliminary risk assessment, including the qualification as high-risk, to the providers' self-assessment. They also raised concerns about the excessive delegation of regulatory power to private European standardisation organisations (ESOs), due to the lack of democratic oversight, the impossibility for stakeholders (civil society organisations, consumer associations) to influence the development of standards, and the lack of judicial means to control them once they have been adopted. Instead, they recommended that the AI act codify a set of legally binding requirements for high-risk AI systems, which ESOs may specify through harmonised standards.

Commentators regretted a crucial gap in the AI act – the lack of provision provide for **individual enforcement rights**. Ebers and others stressed that individuals affected by AI systems and civil rights organisations have no **right to complain** to market surveillance authorities or to sue a provider or user for failure to comply with the requirements. Similarly, Veale and Zuiderveen Borgesius warned that, while some provisions of the draft legislation aim to impose obligations on AI systems users, **no mechanism for complaint or judicial redress** was available to them. Smuha and others recommended amending the proposal to include, inter alia, an **explicit right of redress for individuals** and **rights of consultation and participation for EU citizens** regarding the decision to amend the list of high-risk systems in Annex III.

It has also been stressed that the text proposed by the Commission **lacked proper coordination** mechanisms between authorities, in particular concerning **cross-border infringement**. Furthermore, guidance would be desirable on how to ensure compliance with transparency and information requirements, while simultaneously **protecting intellectual property rights and trade secrets**, not least to avoid diverging practices in the Member States.

Legislative process

The **Council** adopted its common position in December 2022. In **Parliament**, the file was assigned jointly (under Rule 58) to the Committee on Internal Market and Consumer Protection (IMCO) and the Committee on Civil Liberties, Justice and Home Affairs (LIBE), with Brando Benifei (S&D, Italy) and Dragoş Tudorache, Renew, Romania) appointed as rapporteurs. Parliament adopted its negotiating position (499 votes in favour, 28 against and 93 abstentions) in June 2023, with substantial amendments to the Commission's text. Following protracted negotiations, the Council and the European Parliament reached a provisional agreement on the AI act on 9 December 2023. The European Parliament's LIBE and IMCO committees endorsed the final text in a joint vote on 13 February 2024, with an overwhelming majority (71 votes in favour, 8 votes against and 7 abstentions). The European Parliament will now vote on the final agreement on the AI act at the March 2024 plenary session, before it is endorsed by Council ands published in the EU's Official Journal. The main points of the EU AI rules are:

Definitions

The AI act enshrines in EU law a definition of **AI systems** aligned with the revised definition agreed by the OECD:

> *'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.

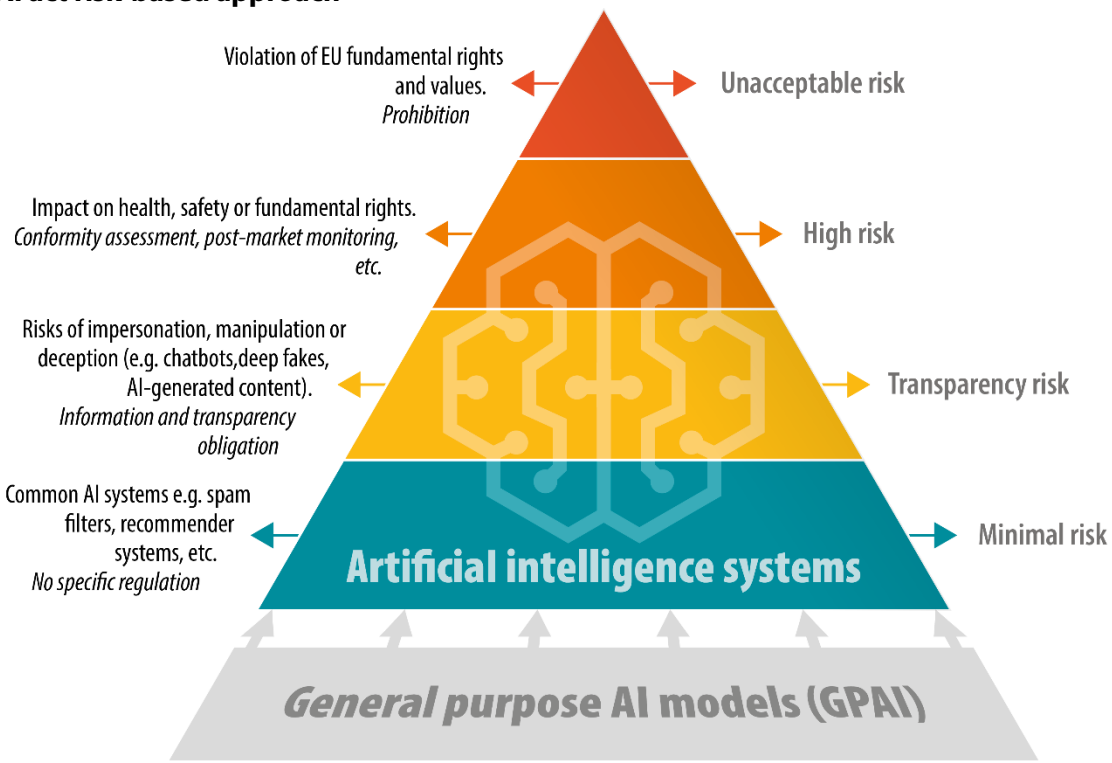
The act also contains a definition of **general purpose artificial intelligence (GPAI) models** 'that are trained with a large amount of data using self-supervision at scale', that display 'significant generality' and are 'capable to competently perform a wide range of distinct tasks' and 'can be integrated 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

control the potential environmental impact of training AI models and protect worker's rights and to define further a set of requirements that research organisations must comply with to benefit from the research exemption.[24] Also, some argue that the AI act does not go far enough in preventing and/or mitigating the specific risks associated with chatbots. Timely standardisation will be key to ensuring adequate implementation of the AI act, for instance, to ensure the robustness of high-risk AI systems and the watermarking of AI-generated content while, in the meantime, the EU is fostering the adoption of voluntary codes of conduct and of an AI Pact to mitigate the potential downsides of generative AI. Some academics warn that that the standardisation and codification processes might not include representative groups of stakeholders and risks privileging regulated parties. Ensuring international harmonisation of AI governance has become a key topic for policymakers. More cooperation on aligning AI governance between the EU and the USA is seen as crucial for AI's democratic governance.[25] Key questions such as setting a common terminology and addressing dual-use and military AI applications have been raised in this respect. Finally, generative AI is seen as a disruptive technology that will likely mean amending EU laws and regulation, including in intellectual property rights, privacy and data protection and cybersecurity.

EUROPEAN PARLIAMENT SUPPORTING ANALYSIS

AI Repository, EPRS, STOA Centre for Artificial Intelligence (C4AI), October 2023.

Biometric Recognition and Behavioural Detection, Policy Department for Citizens' Rights and Constitutional Affairs, August 2021.

Dalli H., Artificial Intelligence Act: Initial Appraisal of the European Commission Impact Assessment, EPRS, July 2021.

Dumbrava C., Artificial intelligence at EU borders: Overview of applications and key issues, EPRS, July 2021.

Madiega T. A. and Mildebrath H. A., Regulating facial recognition in the EU, EPRS, September 2021.

Madiega T., Artificial intelligence act and regulatory sandboxes, EPRS, March 2022.

Madiega T., General-purpose artificial intelligence, EPRS, March 2023.

Madiega T., Generative AI and watermarking, EPRS, December 2023.

OTHER SOURCES

Artificial Intelligence Act, European Parliament, Legislative Observatory (OEIL).

Novelli C. et al., Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity, 2024.

Hacker P., Comments on the final trilogue version of the AI act, 2024.

ENDNOTES

1 See European Commission, Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) 2021/0106 (COD), Explanatory memorandum.

2 See for instance, High-Level Expert Group, Ethics Guidelines for Trustworthy AI, 2019.

3 See, inter alia, Recommendations on intellectual property, criminal law, education, culture and audiovisual areas, and regarding civil and military AI uses.

4 For an overview see H. Dalli, Artificial intelligence act, above.

5 It was proposed to allow FRTs (i) for targeted search for potential victims of crime, including missing children; (ii) to prevent a specific, substantial and imminent threat to the life or physical safety of persons or of a terrorist attack; and (iii) for the detection, localisation, identification or prosecution of a perpetrator or individual suspected of a criminal offence referred to in the European Arrest Warrant Framework Decision.

6 This section aims to provide a flavour of the debate and is not intended to be an exhaustive account of all different views on the proposal. Additional information can be found in publications listed under 'supporting analysis'.

7 For an in-depth analysis of the proposals and recommendations for amendments see N. Smuha et al., How the EU can achieve legally trustworthy AI: A response to the European Commission's proposal for an artificial intelligence act, Elsevier, August 2021; M. Ebers, and others, The European Commission's proposal for an artificial intelligence act—A critical assessment by members of the Robotics and AI Law Society (RAILS), J 4, no 4: 589-603, October 2021.

8 N. Smuha, et al., above, at pp. 14-15.;M. Veale and F. Zuiderveen Borgesius., Demystifying the draft EU AI act, 22(4) *Computer Law Review International*, July 2021.

9 See M. Ebers and others, above.

10 See V. Galaz and others, Artificial intelligence, systemic risks, and sustainability, Vol 67, *Technology in Society*, 2021.

[11] For an overview, see T. Madiega and H. Mildebrath, Regulating facial recognition in the EU, 2021.

[12] The act applies to private organisations as well as to public authorities.

[13] The Annex refers to AI systems used in areas of critical infrastructures (e.g. road traffic), education and vocational training, employment worker management and access to self-employment, access to essential private and public services and benefits (e.g., creditworthiness evaluation), law enforcement, border control, administration of justice and democratic processes, biometric identification, categorisation and emotion recognition systems (outside the prohibited categories).

[14] An AI system will not be considered as high-risk if one or more of the following criteria are fulfilled: (i) the AI system is intended to perform a narrow procedural task; (ii) the AI system is intended to improve the result of a previously completed human activity; (iii) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review; or (iv) the AI system is intended to perform a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III.

[15] Established by European Commission decision in January 2024 the AI Office enters into force in February 2024.

[16] Furthermore, open-source models must comply with the AI act when they are integrated into prohibited AI practices or into high-risk systems and when they are considered to present systemic risk.

[17] FLOPs, or Floating-Point Operations Per Second, measure a computer's processing speed. The threshold should be adjusted over time to reflect technological and industrial changes. Moreover, the Commission is entitled to take individual decisions designating a GPAI model posing systemic risk if it is found that such model has capabilities or impact equivalent to those captured by the FLOP threshold on the basis of an overall assessment of criteria (e.g. quality or size of the training data set, number of business and end users, degree of autonomy and scalability). In the USA, President Biden's AI executive order set 10^26 FLOPs as the threshold for AI models that need to be reported to the government with details of their training, capabilities and security.

[18] Additional AI regulatory sandboxes at regional or local levels or jointly with other Member States' competent authorities may also be established. The European Data Protection Supervisor may also establish an AI regulatory sandbox for the EU institutions, bodies and agencies.

[19] For instance, up to €35 million or 7 % of the total worldwide annual turnover of the preceding financial year (whichever is higher) for infringements on prohibited practices or non-compliance related to requirements on data.

[20] Implementing acts must be adopted by the Commission to establish common specifications for requirements for high-risk systems, to approve codes of practice on generated or manipulated content and to specify common rules for implementation if such codes of practice are deemed not adequate. Delegated acts will need to be adopted to identify conditions for AI systems to not be considered high-risk and to specify and update criteria of GPAI posing systemic risk, inter alia. The AI Office will have to draw up the codes of practice for GPAI providers.

[21] The Commission mandated the European Standardisation Organisations (CEN-CENELEC) to deliver a series of European standards to implement the AI act by January 2025.

[22] See P. Hacker, Comments on the final trilogue version of the AI act, 2024.

[23] See C. Kutterer, Regulating foundation models in the AI act: from "high" to "systemic" risk, 2024.

[24] See N. Helberger and others, The Amsterdam Paper: Recommendations for the technical finalisation of the regulation of GPAI in the AI act, 2024. See also, P. Chavez, An AI challenge: Balancing open and closed systems, 2023.

[25] See A. Engler, The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment, 2023.

DISCLAIMER AND COPYRIGHT

eprs@ep.europa.eu (contact)
www.eprs.ep.parl.union.eu (intranet)
www.europarl.europa.eu/thinktank (internet)
http://epthinktank.eu (blog)

Third edition. 'EU Legislation in Progress' briefings are updated at key stages of the legislative procedure.