AI regulation introduced in Europe – setting the way forward for the rest of the world or slowing down the adoption of new tech?

by Ron Moscona

The EU AI Act, which was passed by the European Parliament on 13 March and is set to become law later this year, will probably be the world's first legislation to introduce a general regulatory framework for artificial intelligence systems. The European Union is known to pioneer the responsible regulation of emerging technology, and the AI Act is no exception: it intends to govern the development, deployment and use of AI systems in order to protect against risk and uphold health, safety and fundamental rights, while balancing technological development and competition. If successful in achieving that balance, it will likely inform legislative standards around the world. The framework comes at a time where artificial intelligence is increasingly becoming more easily and readily deployable in a range of sectors and products, presently with little or no limitations to its usages.

Overview

The Act will cover AI systems that are deployed anywhere in the European Union, defining an AI system broadly as: "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 framework classifies AI systems according to the risk they present and sets out categories ranging from "unacceptable risk" to "minimal risk", each subject to varying degrees of regulation. The majority of the Act focuses on requirements for developers of "high risk" systems, including regulatory oversight and conformity assessments, data quality and traceability, robustness and accuracy, human oversight, and transparency – but there are also obligations for commercial users, importers, distributors and product manufacturers. Further, the Act prohibits certain systems that are deemed to present unacceptable risk. Rules are also laid down for general-purpose AI models, being those systems that are widely integrated into a range of products without a specific purpose.

Related People

- Ron Moscona

Related Capabilities

- Artificial Intelligence
- Technology
- Europe
- Privacy & Social Media

Prohibited systems

The Act bans, with very limited exceptions, AI systems which pose unacceptable threats to the safety, livelihood and rights of EU citizens, including systems designed to:

- Manipulate or influence people's behaviour subconsciously;
- Exploit vulnerabilities of specific groups in order to distort their behaviour;
- Conduct social scoring (i.e. categorising people based on their behaviour, race, political opinions, sexual orientation, etc.);
- Enable facial recognition using web-scraping;
- Conduct real-time biometric identification, except in very limited law enforcement situations.

Additionally, systems designed to do anything else which is already against the law are prohibited.

High-risk systems

Systems considered high-risk are those with significant potential to cause harm or affect safety or fundamental rights. Specifically, AI systems used in certain sectors are designated as high-risk, including systems used in medical devices, lifts, machinery, aviation, automotive and transportation, education, border control, and the management of critical infrastructure.

Of course, products within many of these sectors are already regulated products in the European Union and subject to safety standards and conformity assessment requirements. The Act designates as "high risk" any AI system used as part of a regulated product falling within any of the list of existing product regulations set out in Annex II of the Act. In addition, Article 6(1) designates as "high risk" any system intended to be used as a safety component or as a safety product, if such product or system is subject to a legal requirement in the EU to undergo a third-party conformity assessment.

Further, under Article 6(2) of the Act, regardless of the particular product or system in which an AI model might be deployed, if the AI system is developed for use within certain sectors listed in Annex III of the Act it will be considered "high-risk". These listed sectors include education and vocational training, biometrics, law enforcement, employment and critical infrastructure. Any use of an AI system in one of the listed areas is deemed "high-risk" regardless of whether the system is part of a regulated product subject to EU safety standards. For example, in the employment sector, AI systems might be used to filter out prospective applicants; in the education sector, a system might be deployed to assist with deciding on admissions. Such systems will be deemed "high-risk" under the Act.

Where a system is designated as high-risk, most obligations fall on the "provider", i.e. the developer, of that system. Providers must conduct a Fundamental Rights Impact Assessment before placing their system on the market, and also must continuously assess risk after that. The assessment will evaluate the risks posed by the system and how these will be mitigated. If such a provider believes their system is not high-risk

despite being designated as such by the Act, it must conduct and register an assessment to prove this before making the system available, in order to take advantage of the exemption for systems where there is no “significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision-marking” (Article 6(2a)).

Providers of high-risk systems, as part of their ongoing risk assessment obligations, will need to integrate processes and documentation to minimise risks, and must ensure that the data used to train and test the system are governed in accordance with that assessment and that cybersecurity is maintained. Further, providers must put together technical documentation to demonstrate a system’s compliance with the Act and must keep records and logs of safety incidents. Such records must be provided to deployers of the system, but also to competent authorities if requested as part of a provider’s co-operation obligations under Article 23. A key concept of the Act is human oversight: high-risk systems must be developed in such a way to allow them to be overseen by natural persons when deployed, allowing users to intervene and override decisions made by AI algorithms. Transparency obligations dictate that users be informed when interacting with AI systems.

Conformity assessment procedures and regulatory oversight

To guarantee compliance with the standards and requirements laid out in the Act, before putting a high-risk system on the market, the providers will have to issue a declaration of conformity and apply a CE mark to the system. To do so, they will have to follow conformity assessment procedures, either in accordance with existing regulations for Annex II or Article 6(1) regulated products (such as medical devices, machinery, or automotives) or in accordance with the conformity assessment procedures laid out in the Act for AI systems used in non-regulated products within the designated high-risk sectors (as listed in Annex III).

The conformity assessment procedure will in many cases involve notified bodies which will have the role of assessing the high-risk system’s conformity with the rules and standards laid out in the legislation. Notified bodies have a key role in product safety conformity assessment and certification in the EU and are designated to perform that role by the member states. However, the providers, as well as importers, distributors and deployers, of the AI systems will have the ultimate legal responsibility to ensure compliance with the regulatory requirements.

Compliance obligations do not fall upon the heads of providers alone. Further down the supply chain, deployers of the AI system (being commercial users who exploit a system in their own product, rather than end-users who interact with the system) need to take measures to ensure that they use a high-risk system in accordance with the instructions from the provider. Importers of high-risk systems are also responsible for ensuring that a system conforms with the Act, and distributors must verify that a system bears the CE conformity marking before distributing it.

Consequences for non-compliance can be serious: for example, breaches of certain

high-risk system obligations can result in a fine of the higher of €35M and 7% of turnover, with small and medium size enterprises ("SMEs") and start-ups being subject to a fine which is the lower of the two figures.

General purpose systems

General-purpose AI (GPAI) models are those with no specific purpose and which instead generate output or content in response to a prompt from the user. Such models are frequently integrated into downstream apps and systems due to their range of uses and operate via extensive training using a huge amount of data (e.g. images or text). The obvious examples that spring to mind are tools like ChatGPT, or models used to create AI images or deep-fakes.

Such models, regardless of whether they form high-risk systems or are subject to other requirements under the Act, have separate obligations. Given the models do not have specific intended purposes, risk assessments are difficult due to the huge range of potential uses. Providers of such models instead need to evaluate generic “systemic risks” that might apply in most usages, such as the generation of hate speech or misinformation, assisting or enabling fraud, and so forth. Models presenting systemic risks are the more powerful GPAIs, and will be subject to further obligations, including performing evaluations to identify and mitigate such risks, keep track of incidents and corrective measures taken, and maintain adequate cybersecurity.

Obligations for all GPAI providers include publishing summaries of the content used to train the model and on the testing process, and to make detailed documentation on this available to downstream deployers of the model. Other notable obligations include:

- Labelling content as AI generated: Under Article 52, providers of systems, whether GPAI or not, which generate image, video, text or audio content, need to mark such content as AI-generated. The effects of this will mean that AI-generated images (such as the Willy’s Chocolate Experience advert) or videos (such as deep-fakes, which the obligation expressly applies to) need to be labelled as such. Additionally, AI-generated public interest articles or other media text also need to be labelled unless they have undergone human review.

- Compliance with copyright laws: Article 52C requires providers of GPAI models to put in place a policy to respect EU copyright laws. The Copyright in the Digital Single Market legislation (EU Directive 2019/790) requires EU member states to provide for an exception to copyright protection for the purpose of enabling data mining of published and lawfully accessible works. The exception opens up a wealth of content resources for training AI models, free from copyright restrictions, which can be highly valuable for developers. But publishers can reserve their rights if they wish to prevent their content being used for data mining. As part of the copyright protection policy that providers of GPAI models are required to adopt, technical means would have to be put in place to ensure that systems that scrape data from content resources for the purpose of training AI models can automatically identify and respect such reservations of rights. These rules are particularly relevant given the recent wave of copyright infringement lawsuits against GPAI developers relating largely to the use of published content for training AI models.

Other parts of the Act: Low-risk systems and balancing innovation

AI systems not prohibited or designated as high-risk, and which are not GPAI, are subject to little or no regulation under the Act. Systems which are not high risk but which still generate certain AI content such as chatbots are considered limited risk, and developers must still adhere to some obligations, including transparency to ensure users are notified that content they interact with is AI-generated (see above). Otherwise, systems presenting minimal or no risk are unregulated.

Certain members of the industry have expressed concern that over-regulation will stifle the development of beneficial technology or will present a disproportionate burden on businesses. Specific parts of the Act attempt to address effects on technological advancement, especially for SMEs, by encouraging responsible development of AI systems. Article 53 requires “regulatory sandboxes” in each member state, to foster AI innovation in a controlled setting under the supervision of authorities. SMEs and start-ups are to be prioritised under this provision.

Implementation and consequences

Following the Act being approved by the European Council and published in the EU’s Official Journal (likely to occur around May), it will become law and various parts and obligations will enter into force over a period of six months to three years following that. In addition to the role of notified bodies in relation to conformity assessments, the Act establishes the AI Office, part of the European Commission, to implement and enforce the obligations, and dictates that the Office will within 18 months publish guidelines to provide providers and deployers with practical examples of the various categories of risk and detailing actionable steps in order to assist them in complying with their obligations.

Impact on industry and markets

The rapid development of AI and the significant implications that the technology could soon have on people’s lives have raised great concerns not just from commentators and academics but also from governments and many industry leaders. The risks are broad, varied and very real. It is no surprise that authorities are keen to address and regulate the potential threats the technology could introduce to people’s rights and safety.

The EU AI Act does not set out to exhaustively identify or evaluate each and every potential risk. Rather, whilst identifying some obvious areas of concern, its main function is to put the onus on developers of AI systems and on other organisations that wish to put on the market products and services powered by AI, to identify and weigh the risks and to address them in the framework of a regulatory system that mandates oversight, transparency and accountability.

The European regulatory approach will be successful if it can ensure that AI technologies and AI powered products and services placed on the market in Europe are designed to service the requirements of consumers and other users whilst avoiding serious harm. To achieve this, a significant investment will have to be made by EU member states to build

the necessary regulatory capabilities and to guarantee a high level of enforcement. Robust, highly professional and efficient regulatory and enforcement mechanisms can promote safety and quality without stunting safe and responsible innovation. If those systems can be put in place quickly enough and if they are equipped with the necessary resources, the European regulatory framework can achieve the task of protecting consumers in Europe and it can help to nudge the global AI industry in the right direction. It could also potentially provide a blueprint for regulations to be introduced in other parts of the world.

But Europe alone cannot shape a global industry (particularly one that has so far been led by major players elsewhere, particularly in the US and China). Other major countries will have to play their parts as well, whether by adopting similar regulatory processes as the new EU framework or through alternative approaches, to ensure that AI technology is used to make the world a better place for people, whilst eliminating or minimising the potential threats.