Margaret Hartnett
Co-founder, Progressio AI Ltd
State of Open: The UK in 2024
Phase One: AI and Open Innovation

Margaret Hartnett
Co-founder, Progressio AI Ltd

The EU's AI Act is the first comprehensive regulation targeting AI, focusing on fundamental rights and safety risks in AI development and deployment within the EU. It aims to ensure responsible AI use, promote innovation, and prevent fragmentation of the single market. The Act categorises AI systems into risk levels, with high-risk systems facing extensive evaluation and monitoring. It mandates transparency for limited risk systems and exempts minimal risk ones. Recent negotiations introduced tiers for general purpose AI, with varying obligations.

Fines for violations depend on the severity and type of infringement. Persistent non-compliance may lead to restrictions or withdrawal of high risk AI systems from the EU market. The final text is yet to be published, leaving some details unknown, such as the exact definition of AI systems and classifications for high risk systems. Margaret Hartnett takes us through what we need to know.

Thought Leadership: The International View

While several EU laws (e.g., the General Data Protection Regulation (GDPR)) already apply to AI applications, the AI Act is the EU's first comprehensive horizontal, cross-sectoral regulation focusing on AI. The AI Act addresses fundamental rights and safety risks stemming from the development, deployment, and utilisation of AI systems within the EU. The primary goals of the AI Act are to ensure the responsible and ethical use of AI technologies while fostering innovation and competitiveness in the EU. Another objective is to avoid fragmentation of the EU single market by setting harmonised rules on the development and placing on the market of 'lawful, safe and trustworthy AI systems' thereby ensuring legal certainty for all actors in the AI supply chain.

In essence, the AI Act regulates entry to the EU single market. Companies and state authorities that provide or deploy AI systems in the EU must comply with the rules set out in the AI Act. The AI Act also has extraterritorial effect, because it will apply whenever an AI-based system is used in the EU, regardless of where the provider or operator is based – or whenever an output of such a system is used within the EU, regardless of where the AI system itself is based. However, the AI Act will not apply to AI systems "which are used exclusively for military or defence purposes" or to "AI systems used for the sole purpose of research and innovation".

The AI Act adopts a risk-based approach, categorising AI systems into different risk levels based on their potential impact on fundamental rights, health and safety, and societal well-being. This classification includes four categories of risk ("unacceptable", "high", "limited" and "minimal"), plus one additional category for general-purpose AI ("GPAI").

AI applications deemed to represent unacceptable risks are banned. These include:

- biometric categorisation systems that use sensitive characteristics (e.g. political, religious, philosophical beliefs, sexual orientation, race);
- untargeted scraping of facial images from the Internet or CCTV footage to create facial recognition databases;
- emotion recognition in the workplace and educational institutions;
- social scoring based on social behaviour or personal characteristics;
- manipulation of human behaviour to circumvent free will;
- exploiting the vulnerabilities of people (due to their age, disability, social or economic situation);
- certain applications of predictive policing; and
- some uses of "real-time" biometric systems in publicly accessible spaces by law enforcement.

AI systems deemed to be high risk are required to undergo extensive evaluation before being introduced to the market and ongoing monitoring throughout their operational life cycle. Specifically, high-risk AI systems must comply with comprehensive obligations regarding risk mitigation, data governance, detailed documentation, human oversight, transparency and provision of information to users, robustness, accuracy, and cybersecurity. Such AI systems may also be required to undergo fundamental rights impact assessments.

High-risk AI systems will also be subject to conformity assessments to evaluate their compliance with the Act. Conformity assessments may be done by self-assessment or third parties (i.e. a notifying body appointed by EU member states under the AI Act). Notifying bodies may also carry out audits to check whether a conformity assessment is carried out properly.

A final agreed list of high-risk AI system categories is not yet available. However, while changes may be expected to specific details, the broad application areas covered by the original draft text of the AI Act are likely to remain, namely those associated with critical sectors, such as healthcare, education, employment and recruitment, critical infrastructure, access to public and private services (including credit-scoring), law enforcement, border control and administration of justice.

AI applications classified as being limited-risk, such as chatbots, certain emotion recognition and biometric categorization systems and systems for generating deep fakes are only subject to transparency obligations. These include informing users that they are interacting with an AI system; and marking synthetic audio, video, text and images content as being artificially generated or manipulated for users and in a machine-readable format.

AI systems representing minimal risks are not regulated. Instead, stakeholders are encouraged to build codes of conduct.

In recent trilogue negotiations, an amended tiered approach was agreed for obligations of GPAI systems/models. The first tier applies to all GPAI models. It requires providers to adhere to transparency requirements by drawing up technical documentation (unless the GPAI models are in the R&D phase or they are open source); to comply with EU copyright law; and to provide detailed summaries about the content used for training.

The second tier applies to GPAI models with systemic risk. These GPAI models are subject to more stringent obligations including conducting model evaluations; assessing and mitigating systemic risks; conducting adversarial testing; reporting serious incidents; ensuring cybersecurity and reporting on their energy efficiency". GPAI models with systemic risk may comply with the AI Act by adhering to codes of practice, until harmonised EU standards are published.

Fines for violations of the AI Act will depend on the type of AI system, size of company and se- verity of infringement and will range from:

- 7.5 million euros or 1.5% of a company's total worldwide annual turnover (whichever is higher) for the supply of incorrect information; to
- 15 million euros or 3% of a company's total worldwide annual turnover (whichever is higher) for violations of the AI Act's obligations; to
- 35 million euros or 7% of a company's total worldwide annual turnover (whichever is higher) for violations of the banned AI applications.

In cases of persistent non-compliance, the high-risk AI systems may be restricted or withdrawn from the EU market.

The final text of the AI Act has not yet been published and adopted by the Council and the EU Parliament. However, it may be published in early 2024 and a leaked version of the text was available on 22 January25. Until the official text, certain details remain unknown, for example, the precise definition of "AI systems" and the final list of classifications for high risk AI systems. We also look forward to greater clarity regarding the obligations of developers and deployers of open source.

First published by OpenUK in 2024 as part of
State of Open: The UK in 2024
Phase Three "Open Source and Market Shaping"