What will the EU’s proposed act to regulate AI mean for consumers?

Publication Date: 2024-03-14

Author: Dan Milmo

Section: Technology

Tags: Artificial intelligence (AI), European Union, explainers

Article URL: https://www.theguardian.com/technology/2024/mar/14/what-will-eu-proposed-regulation-ai-mean-consumers

The European Union’s proposed AI law was endorsed by the European parliament on Wednesday, and is a milestone in regulating the technology. The vote is an important step towards introducing the legislation. It is now expected to be rubber stamped by a council of ministers, becoming law within weeks. However, the act will come into force in stages, with a cascade of deadlines for compliance over the next three years. “Users will be able to trust that the AI tools they have access to have been carefully vetted and are safe to use,” said Guillaume Couneson, a partner at the law firm Linklaters. “This is similar to users of banking apps being able to trust that the bank has taken stringent security measures to enable them to use the apps safely.” The bill matters outside the EU because Brussels is an influential tech regulator, as shown by GDPR’s impact on the management of people’s data. The AI act could do the same. “Many other countries will be watching what happens in the EU following the adoption of the AI act. The EU approach will likely only be copied if it is shown to work,” Couneson added. How does the bill define AI? A basic definition of AI is a computer system that carries out tasks you would normally associate with human levels of intelligence, such as writing an essay or drawing a picture. The act itself has a more detailed take, describing the AI technology it regulates as a “machine-based system designed to operate with varying levels of autonomy”, which obviously covers tools like ChatGPT. This system may show “adaptiveness after deployment” – ie it learns on the job – and infers from the inputs it receives “how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments”. This definition covers chatbots, but also AI tools that, for instance, sift through job applications. As detailed below, the legislation bans systems that pose an “unacceptable risk”, but it exempts AI tools designed for military, defence or national security use, issues that alarm many tech safety advocates. It also does not apply to systems designed for use in scientific research and innovation. “We fear that the exemptions for national security in the AI Act provide member states with a carte blanche to bypass crucial AI regulations and create a high risk of abuse,” said Kilian Vieth-Ditlmann, deputy head of policy at German non-profit organisation Algorithmwatch, which campaigns for responsible AI use. How does the bill tackle the risks posed by AI? Certain systems will be prohibited. These include systems that seek to manipulate people to cause harm; “social scoring” systems that classify people based on social behaviour or personality, like the one in Rongcheng, China, where the city rated aspects of residents’ behaviour; Minority Report-style attempts at predictive policing; monitoring people’s emotions at work or in schools; “biometric categorisation” systems that sift people based on their biometric data (retina scans, facial recognition, fingerprints) to infer things such as race, sexual orientation, political opinions or religious beliefs; and compiling facial recognition databases through scraping facial images from the internet or CCTV. Exemptions for law enforcement Facial recognition has been a contentious factor in the legislation. The use of real-time biometric identification systems – which covers facial recognition technology on live crowds – is banned, but allowed for law enforcement in a number of circumstances. Law enforcement can use such technology to find a missing person or prevent a terror attack, but they will need approval from authorities – although in exceptional circumstances it can be deployed without prior approval. What about systems that are risky but not banned? The act has a special category for "high risk" systems that will be legal but closely observed. Included are systems used in critical infrastructure, like water, gas and electricity, or those deployed in areas like education, employment, healthcare and banking. Certain law enforcement, justice and border control systems will also be covered. For instance, a system used in deciding whether someone is admitted to an educational institution, or whether they get a job, will be deemed high-risk. The act requires these tools to be accurate, subject to risk assessments, have human oversight, and also have their usage logged. EU citizens can also ask for explanations about decisions made by these AI systems that have affected them. What about generative AI? Generative AI – the term for systems that produce plausible text, image, video and audio from simple prompts – is covered by provisions for what the act calls "general-purpose" AI systems. There will be a two-tiered approach. Under the first tier, all model developers will need to comply with EU copyright law and provide detailed summaries of the content used to train the model. It is unclear how already-trained models will be able to comply, and some are already under legal pressure. The New York Times is suing OpenAI and Getty Images is suing StabilityAI, alleging copyright infringement. Open-source models, which are freely available to the public, unlike "closed" models like ChatGPT's GPT-4, will be exempt from the copyright requirement. A tougher tier is reserved for models that pose a "systemic risk" – based on an assessment of their more human-like "intelligence" – and is expected to include chatbots and image generators. The measures for this tier include reporting serious incidents caused by the models, such as death or breach of fundamental rights, and conducting "adversarial testing", where experts attempt to bypass a model's safeguards. What does it mean for deepfakes? People, companies or public bodies that issue deepfakes have to disclose whether the content has been artificially generated or manipulated. If it is done for "evidently" artistic, creative or satirical work, it still needs to be flagged, but in an "appropriate manner that does not hamper the display or enjoyment of the work". Text produced by chatbots that informs the public "on matters of public interest" needs to be flagged as AI-made, but not where it has undergone a process of human review or editorial control – which exempts content that has had human oversight. Developers of AI systems also need to ensure that their output can be detected as AI-made, by watermarking or otherwise flagging the material. What do AI and tech companies think? The bill has received a mixed response. The largest tech companies are publicly supportive of the legislation in principle, while wary of the specifics. Amazon said it was committed to collaborating with the EU "to support the safe, secure and responsible development of AI technology", but Mark Zuckerberg's Meta warned against overregulation. "It is critical we don't lose sight of AI's huge potential to foster European innovation and enable competition, and openness is key here," the company's head of EU affairs said. In private, responses have been more critical. One senior figure at a US company warned that the EU had set a limit for the computing power used to train AI models that is much lower than similar proposals in the US. Models trained with more power than 10 to the power of 25 "flops", a measure of computing power, will be hit with burdensome requirements to prove they don't create system risks. This could prompt European companies to simply up stakes and move west to avoid EU restrictions. What are the punishments under the act? Fines will range from €7.5m or 1.5% of a company's total worldwide turnover – whichever is higher – for giving incorrect information to regulators, to €15m or 3% of worldwide turnover for breaching certain provisions of the act, such as transparency obligations, to €35m, or 7% of turnover, for deploying or developing banned AI tools. There will be more proportionate fines for smaller companies and startups. The obligations will come into effect after 12 months, so at some point next year, once the act becomes law, prohibition of certain categories comes into force after six months. Providers and deployers of high-risk systems have three years to comply. There will also be a new European AI office that will set standards and be the main oversight body for GPAI models.