Engineering 38 (2024) 5–7

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

News & Highlights

European Union Issues World's First Comprehensive Regulations for Artificial Intelligence

Chris Palmer

Senior Technology Writer

In March 2024, European Union (EU) lawmakers passed the world's first comprehensive set of regulations governing the use of artificial intelligence (AI) [1]. The EU's AI Act, two and a half years in the making, was initially drawn up as a landmark bill to reduce harm in areas in which AI was thought to pose the biggest risks to people, such as in health care, education, and security, as well as banning uses that pose "unacceptable risks," including manipulation of human behavior and evaluation of individuals' trustworthiness based on personal characteristics. According to the regulations, which will go into effect in stages over the next two years, "high-risk" AI systems will require risk-mitigation strategies, high-quality data sets, transparency, better documentation, and human supervision. The most common current AI uses, such as augmenting recommendation engines and email spam filters, will see far less oversight.

The new regulations have arrived none too soon for many experts. "If you think about pretty much any product, toothpaste for example, it has to go through a ton of testing and regulatory approval, from its ingredients to the way it is marketed," said Chirag Shah, professor of information science and co-director of the Center for Responsibility in AI Systems and Experiences at the University of Washington (Seattle, WA, USA). "For something like AI that has so much potential to change lives in both good and bad ways to have no accountability has been very troubling."

Nearly one year after the technology firm OpenAI (San Francisco, CA, USA) released its ChatGPT chatbot, companies are fiercely competing to develop ever-more powerful generative artificial intelligence (GAI) systems. By producing text, images, videos, and computer programs in response to human queries, GAI systems can make information more accessible and accelerate technological development (Fig. 1). Yet they also are associated with substantial risk for harmful outcomes [2].

GAI systems could flood the internet with misinformation and "deepfakes"—videos of computer-generated faces and voices that can be indistinguishable from those of real individuals (Fig. 2). In addition, the widespread use of commercial "black box" AI tools—those that cannot explain how their results are produced—have been shown to introduce biases and inaccuracies that diminish the validity of public and scientific knowledge while seeming genuine and authoritative [3,4]. "We cannot effectively govern AI without more data about how it is made because it is a black box, and it is becoming even more opaque," said Susan Aaronson, professor of international affairs at the George Washington University (Washington, DC, USA) and co-principal investigator of the US National Science Foundation's Institute for Trustworthy AI Institute for Law and Society.

In the long run, AI's potential downsides could erode people's trust in politicians, the media, other institutions, and each other. "We are at this crossroads with AI," said Lee Tiedrich, professor of law at Duke University (Durham, NC, USA) and a member of the Global Partnership on Artificial Intelligence, a Paris, France-based think tank focused on the responsible use of AI. There are many potential benefits for social good and national security, Tiedrich said, noting projections estimating that AI could enhance the global economy by more than 15 trillion USD annually [5]. But, he said, "If AI is not developed and deployed responsibly, it could cause much harm."

AI has quickly and widely penetrated the world's digital ecosystem, with more than 1.7 billion people using ChatGPT in its first year of release [6]. The chatbot and others like it are powered by

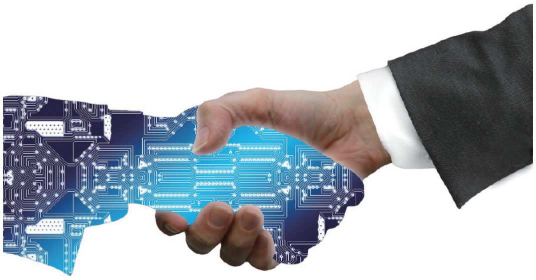

Fig. 1. Powerful new AI tools hold great promise for helping humanity solve many of its most formidable challenges. But used incorrectly—or maliciously—they also have great potential to cause substantial harm. This risk led EU legislators in March 2024 to pass the AI Act, the world's first comprehensive set of regulations governing the use of AI. Credit: Gerd Altman/Pixabay (CC0).

https://doi.org/10.1016/j.eng.2024.06.006

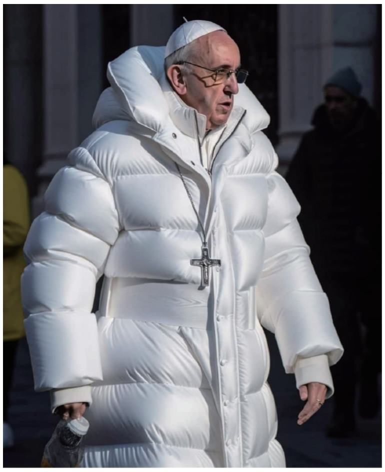

Fig. 2. This image of Pope Francis in a puffy winter coat demonstrates the ability of powerful GAI systems to produce misleading and false information. Many people believed this AI-generated work to be an actual photograph and it went viral, shared and viewed by millions across social media. Note that the image has been deemed in the public domain (not subject to any copyright) because "it is the work of a computer algorithm or artificial intelligence and does not contain sufficient human authorship to support a copyright claim." Credit: Wikimedia Commons (public domain).

so-called foundation models, which generate responses by scraping data from across the internet. However, according to the Foundation Model Transparency Index created and maintained by the Center for Research on Foundation Models at Stanford University (Palo Alto, CA, USA), none of the ten biggest foundation models in 2023 received a passing grade for transparency. The highest scoring model was Meta's Llama 2, which met only 54 out of the 100 different metrics of transparency assessed by the Index [7].

While technology companies may claim that they can police themselves, this is probably too much to expect. Although nearly all AI companies have in-house "responsible AI" teams whose goal is to reduce the potential for harm, such groups are often the first to see cuts during layoffs or during the rush to be first to market, Shah said. Technology companies can also change their AI ethics policies at any time. Most notably, OpenAI, for example, started off as an "open" AI research laboratory that promised to freely collaborate with other institutions by making its patents and research available to anyone. But in March 2023, when it introduced its latest generation GPT-4 chatbot [2], a powerful and multimodal version that can analyze text, audio, and images, the company followed the lead of many other AI startups, moving to withhold all information about how its models are trained and the hardware they are built on to protect its competitive advantage [7].

To meet its harm-reduction goals, the EU's AI Act establishes a risk-based framework for regulating AI products and applications. The regulation imposes legally binding rules requiring technology companies to notify people when they are interacting with a chatbot or with biometric categorization or emotion recognition systems; it also requires companies to label AI-generated media with digital watermarks [8].

All organizations that offer essential services, such as insurance and banking, are required to conduct assessments regarding the impact of their AI systems on people's fundamental rights [9]. In addition, companies must comply with EU copyright law, better document how their code works, and share more information about their models' training data. The latter stipulation may, however, cause problems for some companies. "Sharing training data may create tension because some of the data could be proprietary or include personal information, opening companies up to legal action," Tiedrich said. According to the new law, though, companies that do not share this information could have their products removed from the marketplace.

There are extra provisions for the most powerful AI models, which are currently determined by the computing power needed to train them. For example, technology companies must share how secure and energy-efficient their AI models are [9]. But how these additional rules might apply to models such as GPT-4 or Google's Gemini remains unclear, because the companies have yet to disclose how much computing power was used to train their models [10]. However, as AI technology continues to evolve, the EU will likely change how the regulations measure the "power" of AI systems [9].

The EU will probably need to eventually change other aspects of the act as well, Aaronson said. "My problem with the AI Act is that it will be quickly out of date," she said. For example, she said, early drafts of the rules did not mention the term "foundation model." The rules only began specifically referring to those models in 2023, right around when GAI systems first debuted to great fanfare.

Notably, the AI Act outright bans cognitive behavioral manipulation, the indiscriminate scraping of facial images from the internet or closed-circuit television surveillance footage, social scoring, and the use of biometric categorization systems to infer race, sexual orientation, and political and religious beliefs [11]. Technology companies now have two years to implement all the rules, although the bans on unacceptable uses will apply after six months, and companies developing foundation models will have to comply with the law within one year [9].

Some worry that the new rules could slow progress. "Bigger companies can afford lots of lawyers to help set up oversight structures, but some smaller entrepreneurial companies may not have the resources to comply," Tiedrich said. "The concern is that the amount of regulation could have an adverse effect on innovation."

For now, the AI Act does not apply to military and defense uses of AI. However, European police forces may only use biometric identification systems in public places if they first get approval from a human court official, and only for 16 specific crimes, including terrorism, sexual exploitation of children, drug trafficking, and "exceptional circumstances relating to public security" [9].

To keep track of all this, the AI Act sets up a new European AI Office for coordinating compliance, implementation, and enforcement. Fines for noncompliance are steep: from 1.5% to 7% of a firm's global sales, depending on the severity of the offense and the size of the company [11]. The office will also solicit citizen complaints about AI systems as well as respond to public requests for explanations for how AI systems generated their answers [8]. However, determining who is at fault when a model is out of compliance or how a model comes to certain conclusions may be tricky, Shah said. "Assigning blame to an AI system for producing discriminatory results, for example, may pose challenges because there is no specific code in the system saying, 'Do not approve a home loan for this type or that type of person,' " Shah said. "Rather, the system likely has learned this on its own from training data generated from many sources. That is not the fault of the developers. Likewise, the learning is unsupervised, meaning you cannot really control how weights are assigned and how the models are tuned. There will likely be many cases where it just will not be possible to explain how the model made a particular decision."

Much like the EU's other recent pioneering forays into the regulation of technology, including right to repair [12], device-charging standards [13], and the General Data Protection Regulation (GDPR) that safeguards personal privacy preferences in personal data handling and processing [14], the AI Act could become a global standard. The Act will really impact how AI systems are developed because it affects so many companies, Tiedrich said. "I would not be surprised if companies create baseline product specifications that meet all regulations globally—that is more likely than having different products for regions with different levels of regulation."

Still, other countries are continuing to explore implementing their own AI regulations. While some US congressional hearings on AI have focused on the possibility of creating a new federal regulatory agency for AI, US President Joe Biden in October 2023 signed an executive order distributing responsibility for AI governance among several federal agencies, tasking each with overseeing aspects of AI considered within their purview [15]. For example, the National Institute of Standards and Technology will develop digital watermarking systems and roll out "red-team testing," in which virtuous actors try to misuse a system to assess its security, to help to ensure the security and trustworthiness of powerful AI systems. The Biden administration further recommended in March 2024 that federal agencies increase their use of AI systems, but with greater transparency of these uses to those affected by them [16]. In November 2023, the United Kingdom hosted an AI safety summit but currently has no plans to legislate like the EU [3]. Other regulatory actions are being considered in Africa [17] and in countries including Brazil, Canada, and Japan [15].

In August 2023, the Cyberspace Administration of China announced that it will require developers of GAI systems to prevent the spread of misinformation or content that challenges Chinese socialist values [18]. The Chinese government also requires companies to register their algorithms and disclose information about training data and performance [3], and the country is currently developing a broader set of regulations and testing methods for enforcement [18].

The full impact of the newly approved EU regulations remains to be seen, although just drawing people's attention to the very real potential for harm is a good start. "One thing I would like to see come out of the AI Act is consumers becoming more aware," said Shah, citing the EU's GDPR as an example. "Every time I visit Europe, I have to consent to all the cookies, which makes me aware that I am being tracked and they are using my information," he said. "Awareness of what AI could be up to may have a similarly positive impact."

References

[1] Fung B. EU approves landmark AI law, leapfrogging US to regulate critical but worrying new technology [Internet]. Atlanta: CNN; 2024 Mar 13 [cited 2024 Apr 6]. Available from: https://www.cnn.com/2024/03/13/tech/ai-european-union/index.html.

[2] Mackenzie D. Surprising advances in generative artificial intelligence prompt amazement—and worries. Engineering 2023;25:9–11.

[3] Bockting CL, van Dis EAM, van Rooij R, Zuidema W, Bollen J. Living guidelines for generative AI—why scientists must oversee its use. Nature 2023;622:693–6.

[4] Leslie M. Artificial intelligence could revolutionize science—if we can trust it. Engineering 2024;35:4–6.

[5] Rao AS, Verweij G. Sizing the prize. What's the real value of AI for your business and how can you capitalise? London: PricewaterhouseCoopers International Limited; 2017.

[6] DeVan C. On ChatGPT's one-year anniversary, it has more than 1.7 billion users—here's what it may do next [Internet]. Englewood Cliffs: CNBC LLC; 2023 Nov 30 [cited 2024 Apr 19]. Available from: https://www.cnbc.com/2023/11/30/chatgpts-one-year-anniversary-how-the-viral-ai-chatbot-has-changed.html.

[7] Strickland E. Top AI shops fail transparency test [Internet]. New York City: IEEE Spectrum; 2023 Oct 22 [cited 2024 Apr 6]. Available from: https://spectrum.ieee.org/ai-ethics.

[8] Heikkilä M. The AI Act is done. Here's what will (and won't) change [Internet]. Cambridge: MIT Technology Review; 2024 Mar 19 [cited 2024 Apr 6]. Available from: https://www.technologyreview.com/2024/03/19/1089919/the-ai-act-is-done-heres-what-will-and-wont-change/.

[9] Heikkilä M. Five things you need to know about the EU's new AI Act [Internet]. Cambridge: MIT Technology Review; 2023 Dec 11 [cited 2024 Apr 6]. Available from: https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act.

[10] Leslie M. Artificial intelligence, like cryptocurrency, eats energy—lots of it. Engineering 2024;32:7–9.

[11] Chee FY, Coulter M, Mukherjee S. Europe agrees landmark AI regulation deal [Internet]. London: Reuters; 2023 Dec 12 [cited 2024 Apr 6]. Available from: https://www.reuters.com/technology/stalled-eu-ai-act-talks-set-resume-2023-12-08/.

[12] O'Neill S. European Union puts teeth in right to repair. Engineering 2021;7(9):1197–8.

[13] Palmer C. European Union legislates charging port standard. Engineering 2023;23:7–9.

[14] Burgess M. What is GDPR? The summary guide to GDPR compliance in the UK [Internet]. San Francisco: Wired; 2020 Mar 24 [cited 2024 Apr 6]. Available from: https://www.wired.com/story/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018/.

[15] Strickland E. What you need to know about Biden's sweeping AI order [Internet]. New York City: IEEE Spectrum; 2023 Oct 30 [cited 2024 Apr 6]. Available from: https://spectrum.ieee.org/biden-ai-executive-order.

[16] Hoover A. The White House puts new guardrails on government use of AI [Internet]. San Francisco: Wired; 2024 Mar 28 [cited 2024 Apr 6]. Available from: https://www.wired.com/story/white-house-new-guardrails-government-use-of-ai/.

[17] Tsanni A. Africa's push to regulate AI starts now [Internet]. Cambridge: MIT Technology Review; 2024 Mar 15 [cited 2024 Apr 6]. Available from: https://www.technologyreview.com/2024/03/15/1089844/africa-ai-artificial-intelligence-regulation-au-policy.

[18] Yang Z. Four things to know about China's new AI rules in 2024 [Internet]. Cambridge: MIT Technology Review; 2024 Jan 17 [cited 2024 Apr 6]. Available from: https://www.technologyreview.com/2024/01/17/1086704/china-ai-regulation-changes-2024/.