Top tech firms commit to AI safeguards amid fears over pace of change

Publication Date: 2023-07-21

Author: Kari Paul

Section: Technology

Tags: Artificial intelligence (AI), US politics, news

Article URL: https://www.theguardian.com/technology/2023/jul/21/ai-ethics-guidelines-google-meta-amazon

Top players in the development of artificial intelligence, including Amazon, Google, Meta, Microsoft and OpenAI, have agreed to new safeguards for the fast-moving technology, Joe Biden announced on Friday. Among the guidelines brokered by the Biden administration are watermarks for AI content to make it easier to identify and third-party testing of the technology that will try to spot dangerous flaws. Speaking at the White House, Biden said the companies' commitment were “real and concrete” and will help “develop safe, secure and trustworthy” technologies that benefit society and uphold values. “Americans are seeing how advanced artificial intelligence and the pace of innovation have the power to disrupt jobs in industries,” he said. “These commitments are a promising step that we have a lot more work to do together.” The president said AI brings “incredible opportunities”, as well as risks to society and economy. The agreement, he said, would underscore three fundamental principles – safety, security and trust. The White House said seven US companies had agreed to the voluntary commitments, which are meant to ensure their AI products are safe before they release them. The announcement comes as critics charge AI's breakneck expansion threatens to allow real damage to occur before laws catch up. The voluntary commitments are not legally binding, but may create a stopgap while more comprehensive action is developed. A surge of commercial investment in generative AI tools that can write convincingly human-like text and churn out new images and other media has brought public fascination as well as concern about their ability to trick people and spread disinformation, among other dangers. The tech companies agreed to eight measures: Using watermarking on audio and visual content to help identify content generated by AI. Allowing independent experts to try to push models into bad behavior – a process known as “red-teaming”. Sharing trust and safety information with the government and other companies. Investing in cybersecurity measures. Encouraging third parties to uncover security vulnerabilities. Reporting societal risks such as inappropriate uses and bias. Prioritizing research on AI's societal risks. Using the most cutting-edge AI systems, known as frontier models, to solve society's greatest problems. The voluntary commitments are meant to be an immediate way of addressing risks ahead of a longer-term push to get Congress to pass laws regulating the technology. Some advocates for AI regulations said Biden's move is a start but more needs to be done to hold the companies and their products accountable. “History would indicate that many tech companies do not actually walk the walk on a voluntary pledge to act responsibly and support strong regulations,” said a statement from James Steyer, founder and CEO of the non-profit Common Sense Media. The guidelines, as detailed at a high level in a fact sheet the White House released, some critics have argued, do not go far enough in addressing concerns over the way AI could impact society and give the administration little to no remedies for enforcement if the companies do not abide by them. “We need a much more wide-ranging public deliberation and that’s going to bring up issues that companies almost certainly won’t voluntarily commit to because it would lead to substantively different results, ones that may more directly impact their business models,” said Amba Kak, the executive director of research group the AI Now Institute. “A closed-door deliberation with corporate actors resulting in voluntary safeguards isn’t enough,” Kak said. “What this list covers is a set of problems that are comfortable to business as usual, but we also need to be looking at what’s not on the list – things like competition concerns, discriminatory impacts of these systems. The companies have said they’ll ‘research’ privacy and bias, but we already have robust bodies of research on both – what we need is accountability.” Voluntary guidelines amount to little more than self-regulation, said Caitriona Fitzgerald, the deputy director at the non-profit research group, the Electronic Privacy Information Center (Epic). A similar approach was taken with social media platforms, she said, and it didn’t work. “It’s internal compliance checking and it’s similar to what we’ve seen in the FTC consent orders from the past when they required Facebook to do internal privacy impact assessments and they just became a box-checking exercise.” The Senate majority leader, Chuck Schumer, has said he will introduce legislation to regulate AI. He has held a number of briefings with government officials to educate senators about an issue that’s attracted bipartisan interest. A number of technology executives have called for regulation, and several went to the White House in May to speak with Biden, vice-president Kamala Harris and other officials. Senator Mark Warner said the guidelines released on Friday are a start but that “we need more than industry commitments”. “While we often hear AI vendors talk about their commitment to security and safety, we have repeatedly seen the expedited release of products that are exploitable, prone to generating unreliable outputs, and susceptible to misuse,” Warner said in a statement. But some experts and upstart competitors worry that the type of regulation being floated could be a boon for deep-pocketed first-movers led by OpenAI, Google and Microsoft, as smaller players are elbowed out by the high cost of making their AI systems known as large language models adhere to regulatory strictures. The software trade group BSA, which includes Microsoft as a member, said on Friday that it welcomed the Biden administration’s efforts to set rules for high-risk AI systems. “Enterprise software companies look forward to working with the administration and Congress to enact legislation that addresses the risks associated with artificial intelligence and promote its benefits,” the group said in a statement. Several countries have been looking at ways to regulate AI, including European Union lawmakers who have been negotiating sweeping AI rules for the 27-country bloc. The details of the European legislation are still being hashed out, but the EU AI Act contains robust regulations that would create significant consumer protections against the overreach, privacy violations and biases of certain types of high-risk AI models. Meanwhile conversations in the US remain in the early stages. Fitzgerald, of Epic, said while the voluntary guidelines are just one in a series of guidelines the White House has released on AI, she worries it might cause Congress to slow down their push to create regulations. “We need the rules of the road before it gets too big to regulate,” she said. The UN secretary general, António Guterres, recently said the United Nations was “the ideal place” to adopt global standards and appointed a board that will report back on options for global AI governance by the end of the year. The United Nations chief also said he welcomed calls from some countries for the creation of a new UN body to support global efforts to govern AI, inspired by such models as the International Atomic Energy Agency or the Intergovernmental Panel on Climate Change. The White House said on Friday that it had already consulted on the voluntary commitments with a number of countries. Associated Press contributed to this story