Biden hails ‘bold action’ of US government with order on safe use of AI

Publication Date: 2023-10-30

Author: Dan Milmo

Section: Technology

Tags: Artificial intelligence (AI), Biden administration, Computing, Joe Biden, news

Article URL: https://www.theguardian.com/technology/2023/oct/30/biden-orders-tech-firms-to-share-ai-safety-test-results-with-us-government

Joe Biden signed an executive order on Monday that he called the most “significant” action that any government has taken on the safe deployment of AI. “We’re going to see more technological change in the next 10, maybe next five years than we’ve seen in the last 50 years,” the US president said at a press conference. “AI is all around us. Much of it is making our lives better … but in some case AI is making life worse.” Kamala Harris, the vice-president and the US’s representative at the global AI safety summit in the UK this week, said that the government has a “a moral, ethical and societal duty to make sure AI is adopted and advanced in a way that protects the public from potential harm and ensures that everyone is able to enjoy its benefits”. Harris emphasised the administration’s belief that the US is the global leader in AI and that the order should serve as a model for international action. “It is American companies that lead the world in AI innovation,” Harris said. “It is America that can catalyse global action and build global consensus in a way that no other country can. And under President Joe Biden that will continue to lead on AI.” While Biden said the executive order “represents bold action”, Congress still needs to act quickly to ensure the safe deployment and development of AI. Under the order, tech companies will be required to share test results for their artificial intelligence systems with the US government before they are released. The government will also set stringent testing guidelines. “As we advance this agenda at home, the administration will work with allies and partners abroad on a strong international framework to govern the development and use of AI,” said the order. The AI directives issued by the White House include: • Companies developing AI models that pose a threat to national security, economic security or health and safety must share their safety test results with the government. • The government will set guidelines for so-called red-team testing, where assessors emulate rogue actors in their test procedures. • Official guidance on watermarking AI-made content will be issued to address risk of harm from fraud and deepfakes. • New standards for biological synthesis screening – to identify potentially harmful gene sequences and compounds – will be developed to mitigate the threat of AI systems helping to create bioweapons. The White House chief of staff, Jeff Zients, said Biden had given his staff a directive to move with urgency on the AI issue. “We can’t move at a normal government pace,” Zients said Biden told him. “We have to move as fast, if not faster than the technology itself.” The White House said the sharing of test results for powerful models would “ensure AI systems are safe, secure and trustworthy before companies make them public”. Under the provisions on AI-made deepfakes, the US Department of Commerce will issue guidance to label and watermark AI-generated content to help differentiate between authentic interactions and those generated by software. Referring to the watermarking plans, the order stated: “Federal agencies will use these tools to make it easy for Americans to know that the communications they receive from their government are authentic – and set an example for the private sector and governments around the world.” The order also covers areas such as privacy, civil rights, consumer protections and workers’ rights. Civil liberties and digital rights group have largely lauded the executive order as a good first step. Alexandra Reeve Givens, the chief executive of the Center for Democracy and Technology, a non-profit digital rights group, said it marks a milestone that shows the entire government will support “the responsible development and governance of AI”. “It’s notable to see the administration focusing on both the emergent risks of sophisticated foundation models and the many ways in which AI systems are already impacting people’s rights – a crucial approach that responds to the many concerns raised by public interest experts and advocates,” Givens said in a statement. But the success of the order lies in how enforceable and actionable the directives are, Givens said. “We urge the administration to move quickly to meet relevant deadlines, and to ensure that any guidance or mandates issued under the EO are sufficiently detailed and actionable to have their intended effect.” Caitriona Fitzgerald, the deputy director of the Electronic Privacy Information Center (Epic) said the privacy protections in the order are particularly necessary given the lack of federal protections in the US. “While Epic continues to call on Congress to pass a comprehensive privacy law that limits the mass data collection that fuels harmful uses of technology, this executive order is a significant step towards establishing the necessary fairness, accountability, and transparency guardrails to protect people from discrimination and inequality facilitated by AI systems,” Fitzgerald said in a statement. However, some groups focused on surveillance are not as optimistic. Albert Fox Cahn, the Surveillance Tech Oversight Project, said the approach taken in the order would enable further AI abuses. For one, he said, the White House order relies on AI auditing techniques that “can be easily gamed by companies and agencies”. “The worst forms of AI, like facial recognition, don’t need guidelines, they need a complete ban,” Fox Cahn said. “Many forms of AI simply should not be allowed on the market. And many of these proposals are simply regulatory theater, allowing abusive AI to stay on the market.” According to a White House official, the to-do lists within the order will be implemented and fulfilled over the range of 90 to 365 days, with the safety and security items facing the earliest deadlines. Elsewhere in the order, a national security memorandum will direct the US military and intelligence community on how to use AI safely and ethically. It also calls on Congress to pass legislation protecting Americans’ data privacy. Federal agencies will develop guidelines for evaluating privacy-preserving techniques in AI systems. Concerns around bias are addressed with an order to provide guidance to landlords, federal benefits programmes and federal contractors to prevent AI algorithms from exacerbating discrimination. A key immediate concern about AI systems is that they inadvertently repeat underlying biases in the datasets they are trained upon. Best practice will also be developed on using AI in the justice system, in areas such as sentencing, predictive policing and parole. The threat of disruption in the jobs market is addressed with an order to develop best practices for mitigating the harms from job displacement, by providing to “prevent employers from undercompensating workers, evaluating job applications unfairly, or impinging on workers’ ability to organise”. Government agencies will also be issued guidance on using AI including standards to protect rights and safety. The Federal Trade Commission, a competition watchdog, will be encouraged to use its powers if there are any distortions in the AI market. In a nod to efforts to regulate AI around the world including discussions at this week’s safety summit, the White House said it would also accelerate the development of AI standards with international partners. Also on Monday, the G7 group of nations published a code of conduct for organisations developing advanced AI systems. These included watermarking AI-made content, testing models externally and prioritise using AI to address challenges such as the climate crisis and global health problems.