Tech firms sign ‘reasonable precautions’ to stop AI-generated election chaos

Publication Date: 2024-02-16

Author: Unknown

Section: Technology

Tags: Artificial intelligence (AI), Amazon, Meta, Google, Microsoft, OpenAI, TikTok, news

Article URL: https://www.theguardian.com/technology/2024/feb/16/tech-companies-precautions-ai-election

Major technology companies signed a pact Friday to voluntarily adopt “reasonable precautions” to prevent artificial intelligence tools from being used to disrupt democratic elections around the world. Executives from Adobe, Amazon, Google, IBM, Meta, Microsoft, OpenAI and TikTok gathered at the Munich Security Conference to announce a new framework for how they will respond to AI-generated deepfakes that deliberately trick voters. Twelve other companies – including Elon Musk’s X – are signing on to the accord. “Everybody recognizes that no one tech company, no one government, no one civil society organization is able to deal with the advent of this technology and its possible nefarious use on their own,” said Nick Clegg, president of global affairs for Meta, the parent company of Facebook and Instagram, in an interview before the summit. The accord is largely symbolic, but targets increasingly realistic AI-generated images, audio and video “that deceptively fake or alter the appearance, voice or actions of political candidates, election officials and other key stakeholders in a democratic election, or that provide false information to voters about when, where and how they can lawfully vote”. The companies aren’t committing to ban or remove deepfakes. Instead, the accord outlines methods they will use to try to detect and label deceptive AI content when it is created or distributed on their platforms. It notes the companies will share best practices with each other and provide “swift and proportionate responses” when that content starts to spread. The vagueness of the commitments and lack of any binding requirements likely helped win over a diverse swath of companies, but disappointed advocates who were looking for stronger assurances. “The language isn’t quite as strong as one might have expected,” said Rachel Orey, the senior associate director of the Elections Project at the Bipartisan Policy Center. “I think we should give credit where credit is due, and acknowledge that the companies do have a vested interest in their tools not being used to undermine free and fair elections. That said, it is voluntary, and we’ll be keeping an eye on whether they follow through.” Clegg said each company “quite rightly has its own set of content policies”. “This is not attempting to try to impose a straitjacket on everybody,” he said. “And in any event, no one in the industry thinks that you can deal with a whole new technological paradigm by sweeping things under the rug and trying to play Whac-a-Mole and finding everything that you think may mislead someone.” Several political leaders from Europe and the US joined Friday’s announcement. Vera Jourová, the European Commission vice-president, said while such an agreement can’t be comprehensive, “it contains very impactful and positive elements”. She also urged fellow politicians to take responsibility to not use AI tools deceptively and warned that AI-fueled disinformation could bring about “the end of democracy, not only in the EU member states”. The agreement at the German city’s annual security meeting comes as more than 50 countries are due to hold national elections in 2024. Bangladesh, Taiwan, Pakistan and most recently Indonesia have already done so. Attempts at AI-generated election

interference have already begun, such as when AI robocalls that mimicked the US president Joe Biden's voice tried to discourage people from voting in New Hampshire's primary election last month. Just days before Slovakia's elections in November, AI-generated audio recordings impersonated a candidate discussing plans to raise beer prices and rig the election. Fact-checkers scrambled to identify them as false as they spread across social media. Politicians also have experimented with the technology, from using AI chatbots to communicate with voters to adding AI-generated images to ads. The accord calls on platforms to "pay attention to context and in particular to safeguarding educational, documentary, artistic, satirical and political expression". It said the companies will focus on transparency with users about their policies and work to educate the public about how they can avoid falling for AI fakes. Most companies have previously said they're putting safeguards on their own generative AI tools that can manipulate images and sound, while also working to identify and label AI-generated content so that social media users know whether what they're seeing is real. But most of those proposed solutions haven't yet rolled out and the companies have faced pressure to do more. That pressure is heightened in the US, where Congress has yet to pass laws regulating AI in politics, leaving companies to largely govern themselves. The Federal Communications Commission recently confirmed that AI-generated audio clips in robocalls are against the law, but that doesn't cover audio deepfakes when they circulate on social media or in campaign advertisements. Many social media companies have policies in place to deter deceptive posts about electoral processes – AI-generated or not. Meta says it removes misinformation about "the dates, locations, times and methods for voting, voter registration or census participation" as well as other false posts meant to interfere with someone's civic participation. Jeff Allen, a co-founder of the Integrity Institute and a former Facebook data scientist, said the accord seems like a "positive step" but he'd still like to see social media companies taking other actions to combat misinformation, such as building content-recommendation systems that don't prioritize engagement above all else. Lisa Gilbert, the executive vice-president of the advocacy group Public Citizen, argued Friday that the accord is "not enough" and AI companies should "hold back technology" such as hyper-realistic text-to-video generators "until there are substantial and adequate safeguards in place to help us avert many potential problems". In addition to the companies that helped broker Friday's agreement, other signatories include chatbot developers Anthropic and Inflection AI; voice-clone startup ElevenLabs; chip designer Arm Holdings; security companies McAfee and TrendMicro; and Stability AI, known for making the image-generator Stable Diffusion. Notably absent is another popular AI image-generator, Midjourney. The San Francisco-based startup didn't immediately respond to a request for comment Friday. The inclusion of X – not mentioned in an earlier announcement about the pending accord – was one of the surprises of Friday's agreement. Musk sharply curtailed content-moderation teams after taking over the former Twitter and has described himself as a "free-speech absolutist". In a statement Friday, the X CEO Linda Yaccarino said "every citizen and company has a responsibility to safeguard free and fair elections". "X is dedicated to playing its part, collaborating with peers to combat AI threats while also protecting free speech and maximizing transparency," she said.