Tech companies sign accord to combat AI-generated election trickery

BY MATT O'BRIEN AND ALI SWENSON
Updated 8:42 AM AKST, February 16, 2024
Share

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.

Tech executives from Adobe, Amazon, Google, IBM, Meta, Microsoft, OpenAI and TikTok gathered at the Munich Security Conference to announce a new voluntary framework for how they will respond to AI-generated deepfakes that deliberately trick voters. Twelve other companies — including Elon Musk's X — are also 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 ahead of 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 may disappoint pro-democracy activists and watchdogs looking for stronger assurances.

"The language isn't quite as strong as one might have expected," said Rachel Orey, 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 whack-a-mole and finding everything that you think may mislead someone."

The agreement at the German city's annual security meeting comes as more than 50 countries are due to hold national elections in 2024. Some have already done so, including Bangladesh, Taiwan, Pakistan, and most recently Indonesia.

Attempts at AI-generated election interference have already begun, such as when AI robocalls that mimicked U.S. 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 liberal candidate discussing plans to raise beer prices and rig the election. Fact-checkers scrambled to identify them as false, but they were already widely shared as real across social media.

Politicians and campaign committees also have experimented with the technology, from using AI chatbots to communicate with voters to adding AI-generated images to ads.

Ahead of Indonesia's election, the leader of a political party shared a video cloning the face and voice of the deceased dictator Suharto. The post on X disclosed the video was generated by AI, but some online critics called it a misuse of AI tools to intimidate and sway voters.

Friday's accord said in responding to AI-generated deepfakes, platforms "will pay attention to context and in particular to safeguarding educational, documentary, artistic, satirical, and political expression."

It said the companies will focus on transparency to users about their policies on deceptive AI election content and work to educate the public about how they can avoid falling for AI fakes.

Many of the 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 if what they're seeing is real. But most of those proposed solutions haven't yet rolled out and the companies have faced pressure from regulators and others to do more.

That pressure is heightened in the U.S., where Congress has yet to pass laws regulating AI in politics, leaving AI companies to largely govern themselves. In the absence of federal legislation, many states are considering ways to put guardrails around the use of AI, in elections and other applications.

The Federal Communications Commission recently confirmed 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.

Misinformation experts warn that while AI deepfakes are especially worrisome for their potential to fly under the radar and influence voters this year, cheaper and simpler forms of misinformation remain a major threat. The accord noted this too, acknowledging that "traditional manipulations ("cheapfakes") can be used for similar purposes."

Many social media companies already have policies in place to deter deceptive posts about electoral processes — AI-generated or not. For example, 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, co-founder of the Integrity Institute and a former data scientist at Facebook, said the accord seems like a "positive step" but he'd still like to see social media companies taking other basic actions to combat misinformation, such as building content recommendation systems that don't prioritize engagement above all else.

In addition to the major platforms 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 from the accord is another popular AI image-generator, Midjourney. The San Francisco-based startup didn't immediately return a request for comment Friday.

The inclusion of X — not mentioned in an earlier announcement about the pending accord — was one of the biggest 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."

But in a statement Friday, 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.

___

The Associated Press receives support from several private foundations to enhance its explanatory coverage of elections and democracy. See more about AP's democracy initiative here. The AP is solely responsible for all content.