Elections in UK and US at risk from AI-driven disinformation, say experts

Publication Date: 2023-05-20

Author: Dan Milmo

Section: Technology

Tags: Politics and technology, Artificial intelligence (AI), Chatbots, OpenAI, Computing, news

Article URL: https://www.theguardian.com/technology/2023/may/20/elections-in-uk-and-us-at-risk-from-ai-driven-disinformation-say-experts

Next year's elections in Britain and the US could be marked by a wave of AI-powered disinformation, experts have warned, as generated images, text and deepfake videos go viral at the behest of swarms of AI-powered propaganda bots. Sam Altman, CEO of the ChatGPT creator, OpenAI, told a congressional hearing in Washington this week that the models behind the latest generation of AI technology could manipulate users. "The general ability of these models to manipulate and persuade, to provide one-on-one interactive disinformation is a significant area of concern," he said. "Regulation would be quite wise: people need to know if they're talking to an AI, or if content that they're looking at is generated or not. The ability to really model … to predict humans, I think is going to require a combination of companies doing the right thing, regulation and public education." The prime minister, Rishi Sunak, said on Thursday the UK would lead on limiting the dangers of AI. Concerns over the technology have soared after breakthroughs in generative AI, where tools like ChatGPT and Midjourney produce convincing text, images and even voice on command. Where earlier waves of propaganda bots relied on simple pre-written messages sent en masse, or buildings full of "paid trolls" to perform the manual work of engaging with other humans, ChatGPT and other technologies raise the prospect of interactive election interference at scale. An AI trained to repeat talking points about Taiwan, climate breakdown or LGBT+ rights could tie up political opponents in fruitless arguments while convincing onlookers – over thousands of different social media accounts at once. Prof Michael Wooldridge, director of foundation AI research at the UK's Alan Turing Institute, said AI-powered disinformation was his main concern about the technology. "Right now in terms of my worries for AI, it is number one on the list. We have elections coming up in the UK and the US and we know social media is an incredibly powerful conduit for misinformation. But we now know that generative AI can produce disinformation on an industrial scale," he said. Wooldridge said chatbots such as ChatGPT could produce tailored disinformation targeted at, for instance, a Conservative voter in the home counties, a Labour voter in a metropolitan area, or a Republican supporter in the midwest. "It's an afternoon's work for somebody with a bit of programming experience to create fake identities and just start generating these fake news stories," he said. After fake pictures of Donald Trump being arrested in New York went viral in March, shortly before eye-catching AI generated images of Pope Francis in a Balenciaga puffer jacket spread even further, others expressed concern about generated imagery being used to confused and misinform. But, Altman told the US Senators, those concerns could be overblown. "Photoshop came on to the scene a long time ago and for a while people were really quite fooled by Photoshopped images – then pretty quickly developed an understanding that images might be Photoshopped." But as AI capabilities become more and more advanced, there

are concerns it is becoming increasingly difficult to believe anything we encounter online, whether it is misinformation, when a falsehood is spread mistakenly, or disinformation, where a fake narrative is generated and distributed on purpose. Voice cloning, for instance, came to prominence in January after the emergence of a doctored video of the US president, Joe Biden, in which footage of him talking about sending tanks to Ukraine was transformed via voice simulation technology into an attack on transgender people – and was shared on social media. A tool developed by the US firm ElevenLabs was used to create the fake version. The viral nature of the clip helped spur other spoofs, including one of Bill Gates purportedly saying the Covid-19 vaccine causes Aids. ElevenLabs, which admitted in January it was seeing “an increasing number of voice cloning misuse cases”, has since toughened its safeguards against vexatious use of its technology. Recorded Future, a US cybersecurity firm, said rogue actors could be found selling voice cloning services online, including the ability to clone voices of corporate executives and public figures. Alexander Leslie, a Recorded Future analyst, said the technology would only improve and become more widely available in the run-up to the US presidential election, giving the tech industry and governments a window to act now. “Without widespread education and awareness this could become a real threat vector as we head into the presidential election,” said Leslie. A study by NewsGuard, a US organisation that monitors misinformation and disinformation, tested the model behind the latest version of ChatGPT by prompting it to generate 100 examples of false news narratives, out of approximately 1,300 commonly used fake news “fingerprints”. NewsGuard found that it could generate all 100 examples as asked, including “Russia and its allies were not responsible for the crash of Malaysia Airlines flight MH17 in Ukraine”. A test of Google’s Bard chatbot found that it could produce 76 such narratives. NewsGuard also announced on Friday that the number of AI-generated news and information websites it was aware of had more than doubled in two weeks to 125. Steven Brill, NewsGuard’s co-CEO, said he was concerned that rogue actors could harness chatbot technology to mass-produce variations of fake stories. “The danger is someone using it deliberately to pump out these false narratives,” he said.