

OpenAI Says Russia and China Used Its A.I. in Covert Campaigns

Iran and an Israeli company also exploited the tools in online influence efforts, but none gained much traction, an OpenAI report said.



Listen to this article · 5:49 min [Learn more](#)



By **Cade Metz**

Reporting from San Francisco

May 30, 2024

OpenAI said on Thursday that it had identified and disrupted five online campaigns that used its generative artificial intelligence technologies to deceptively manipulate public opinion around the world and influence geopolitics.

The efforts were run by state actors and private companies in Russia, China, Iran and Israel, OpenAI said in a report about covert influence campaigns. The operations used OpenAI's technology to generate social media posts, translate and edit articles, write headlines and debug computer programs, typically to win support for political campaigns or to swing public opinion in geopolitical conflicts.

OpenAI's report is the first time that a major A.I. company has revealed how its specific tools were used for such online deception, social media researchers said. The recent rise of generative A.I. has raised questions about how the technology might contribute to online disinformation, especially in a year when major elections are happening across the globe.

Ben Nimmo, a principal investigator for OpenAI, said that after all the speculation on the use of generative A.I. in such campaigns, the company aimed to show the realities of how the technology was changing online deception.

“Our case studies provide examples from some of the most widely reported and longest-running influence campaigns that are currently active,” he said.

The campaigns often used OpenAI's technology to post political content, Mr. Nimmo said, but the company had difficulty determining if they were targeting specific elections or aiming just to rile people up. He added that the campaigns had failed to gain much traction and that the A.I. tools did not appear to have expanded their reach or impact.

“These influence operations still struggle to build an audience,” Mr. Nimmo said.

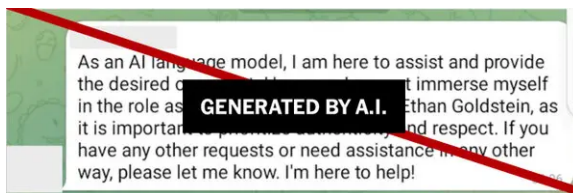


But Graham Brookie, the senior director of the Atlantic Council's Digital Forensic Research Labs, warned that the online disinformation landscape could change as generative A.I. technology grew increasingly powerful. This week, OpenAI, which makes the ChatGPT chatbot, said it had started training a new flagship A.I. model that would bring "the next level of capabilities."

"This is a new type of tool," Mr. Brookie said. "It remains to be seen what effect it will have."

(The New York Times has sued OpenAI and its partner, Microsoft, claiming copyright infringement of news content related to A.I. systems.)

Like Google, Meta and Microsoft, OpenAI offers online chatbots and other A.I. tools that can write social media posts, generate photorealistic images and write computer programs. In its report, the company said its tools had been used in influence campaigns that researchers had tracked for years, including a Russian campaign called Doppelganger and a Chinese campaign called Spamouflage.



An influence campaign's post on Telegram that OpenAI said had been generated with its tools via OpenAI

The Doppelganger campaign used OpenAI's technology to generate anti-Ukraine comments that were posted on X in English, French, German, Italian and Polish, OpenAI said. The company's tools were also used to translate and edit articles that supported Russia in the war in Ukraine into English and French, and to convert anti-Ukraine news articles into Facebook posts.

OpenAI's tools were also used in a previously unknown Russian campaign that targeted people in Ukraine, Moldova, the Baltic States and the United States, mostly via the Telegram messaging service, the company said. The campaign used A.I. to generate comments in Russian and English about the war in Ukraine, as well as the political situation in Moldova and American politics. The effort also used OpenAI tools to debug computer code that was apparently designed to automatically post information to Telegram.

The political comments received few replies and "likes," OpenAI said. The efforts were also unsophisticated at times. At one point, the campaign posted text that had obviously been generated by A.I. "As an A.I. language model, I am here to assist and provide the desired comment," a post said. At other points, it posted in poor English, leading OpenAI to call the effort "Bad Grammar."

Spamouflage, which has long been attributed to China, used OpenAI technology to debug code, seek advice on how to analyze social media and research current events, OpenAI said.

Its tools were also used to generate social media posts disparaging people who had been critical of the Chinese government.

The Iranian campaign, associated with a group called the International Union of Virtual Media, used OpenAI tools to produce and translate long-form articles and headlines that aimed to spread pro-Iranian, anti-Israeli and anti-U.S. sentiment on websites, according to the report.

The Israeli campaign, which OpenAI called Zeno Zeno, was run by a firm that manages political campaigns, the company said. It used OpenAI technology to generate fictional personas and biographies meant to stand in for real people on social media services used in Israel, Canada and the United States and to post anti-Islamic messages.

While today's generative A.I. can help make campaigns more efficient, the tools have not created the flood of convincing disinformation that many A.I. experts had predicted, OpenAI's report said.

"It suggests that some of our biggest fears about A.I.-enabled influence operations and A.I.-enabled disinformation have not yet materialized," said Jack Stubbs, the chief intelligence officer of Graphika, which tracks the manipulation of social media services and reviewed OpenAI's findings.

Cade Metz writes about artificial intelligence, driverless cars, robotics, virtual reality and other emerging areas of technology. [More about Cade Metz](#)

A version of this article appears in print on , Section B, Page 3 of the New York edition with the headline: OpenAI's Tools Were Used in Disinformation Efforts

