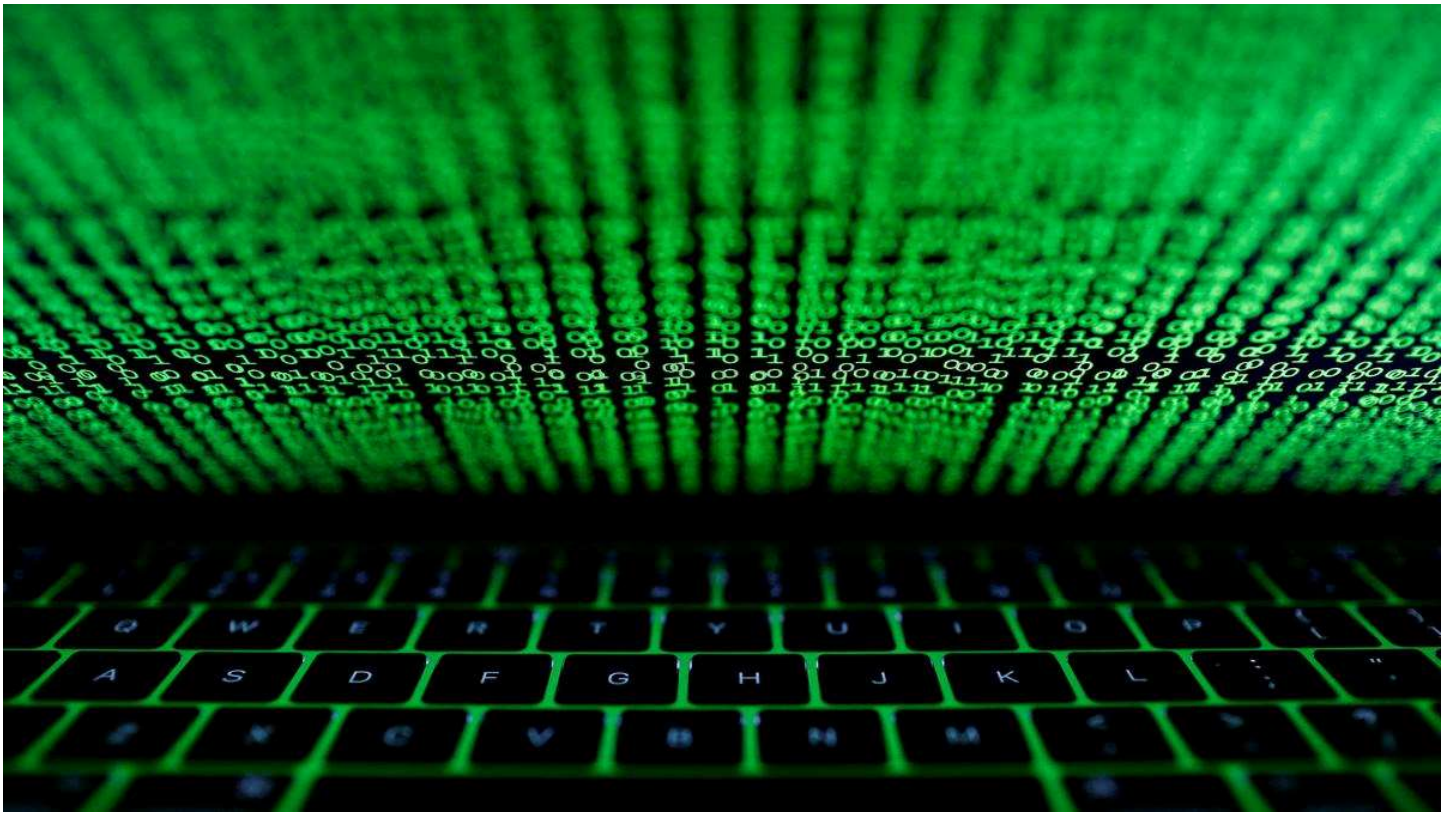
How an Iranian group used ChatGPT to influence U.S. presidential election

The Microsoft-backed company behind ChatGPT claimed that it disrupted a "covert" Iranian influence operation targeting the U.S. presidential election. How was ChatGPT used by these operatives?

Updated - August 17, 2024 06:07 pm IST Published - August 17, 2024 04:59 pm IST

SAHANA VENUGOPAL, POULOMI CHATTERJEE



FILE PHOTO: OpenAI said it banned ChatGPT accounts linked to an Iranian influence operation that used to generate content to influence U.S. presidential election. | Photo Credit: Reuters

The story so far: OpenAI on Thursday (August 16, 2024) said it banned ChatGPT accounts linked to an Iranian influence operation that used the chatbot to generate content to influence the U.S. presidential election. The Microsoft-backed company said it identified and took down a "cluster of ChatGPT accounts" and that it was monitoring the situation.

What is Storm-2035?

OpenAI assigned the group the Storm-2035 moniker, and said the operation was made up of four websites that acted as news organisations. These news sites exploited issues like LGBTQ rights and Israel-Hamas conflict, to target U.S. voters. The sites also used AI tools to plagiarise stories and capture web traffic, per a Microsoft Threat Analysis Center (MTAC) report issued on August 9.

Some named sites included EvenPolitics, Nio Thinker, Westland Sun, Teorator, and Savannah Time. The operation allegedly targeted both liberal and conservative voters in the U.S.

How did the group use ChatGPT?

According to OpenAI, the operatives used ChatGPT to create long-form articles and social media comments that were then posted by several X and Instagram accounts.

(For top technology news of the day, subscribe to our tech newsletter Today's Cache)

AI chatbots such as ChatGPT can potentially assist foreign operatives fool gullible internet users by mimicking American users' language patterns, rehashing already existing comments or propaganda, and cutting down the time it takes to create and circulate plagiarised content meant to sway voters.

Apart from the upcoming U.S. presidential election, Storm-2035 operation covered world issues such Venezuelan politics, Latin rights in the U.S., the destruction in Palestine, Scottish independence, and Israel taking part in the Olympic Games. The network also exploited popular topics like fashion and beauty.

OpenAI shared screenshots of some of the news stories and social media posts it attributed to the operation; one article claimed that X was censoring former president Donald Trump's tweets, while separate social media posts asked users to "dump" Trump or Vice President Kamala Harris.

How severe is the impact of Storm-2035?

OpenAI has downplayed the severity of the incident, claiming that audiences did not engage much with the uploaded content on social media.

Using Brookings' BreakoutScale, which measures the impact of covert operations on a scale from 1 (lowest) to 6 (highest), the report shared this operation was at the low end of Category 2, meaning it was posted on multiple platforms, but there was no evidence that real people picked up or widely shared their content.

However, OpenAI stressed it had shared the threat information with "government, campaign, and industry stakeholders."

While OpenAI presented the discovery and disruption of the Iran-linked influence operation as a positive development, the use of generative AI tools by foreign operatives against U.S. voters is a gravely urgent issue that highlights multiple points of failure across OpenAI, X, Instagram, and the search engines ranking the sites.

Were there other similar issues OpenAI faced in the past?

In May, the AI firm posted a report revealing it had been working for over three months to dismantle covert influence operations that used its tools for generating comments on social media, articles in multiple languages, fake names and bios for social media accounts, and translating or proofreading text.

A Russian outfit that OpenAI called 'Bad Grammar,' used the Telegram to target Ukraine, Moldova, the Baltic States and the U.S.

Separately, another Russia-based operation titled 'Doppelganger,' an Israeli operation that OpenAI nicknamed 'Zeno Zeno,' a Chinese network called 'Spamouflage,' and an Iranian group called 'International Union of Virtual Media' or IUVM, used ChatGPT to write comments on social media platforms like X and 9GAG, and to post articles and news stories.

The investigation found that the content covered issues like Russia's invasion of Ukraine, the Gaza conflict, Indian and European elections, and the criticism of the Chinese government by Chinese dissidents or foreign governments.

Besides hunting down influence networks, OpenAI also found incidents of state-backed threat actors abusing AI to attack enemies.

Other serious cases exposing OpenAI's vulnerabilities followed. In July, the Microsoft-backed firm revealed that early last year, a hacker gained access to OpenAI's internal messaging systems and stole information related to the company's AI technologies. While the hacker was found to be an individual, the incident raised alarms that Chinese adversaries could easily do the same.

What is OpenAI doing to safeguard its tech?

While studying these cases, OpenAI found that its AI tools thankfully refused to generate text or images for some prompts due to the safeguards already built into them. The firm also developed AI-powered security tools to detect threat actors within days instead of weeks.

While not explicitly discussed by OpenAI, the AI company has become enmeshed with prominent figures from U.S. federal agencies or government bodies.

In June, OpenAI picked cybersecurity expert and retired U.S. Army General Paul M. Nakasone to be a part of its Board of Directors. Nakasone led the U.S. National Security Agency and has served in assignments with cyber units in the U.S., Korea, Iraq, and Afghanistan.

A couple of weeks ago, the firm also announced it will be teaming up with the U.S. AI Safety Institute, so that its next big foundational model GPT-5 can be previewed and tested by it.