ARTIFICIAL INTELLIGENCE

NSA official warns of hackers using AI to perfect their English in phishing schemes

NSA Cybersecurity Director Rob Joyce said the language used in hacking and phishing schemes was becoming more sophisticated and convincing.

Jan. 9, 2024, 8:08 PM GMT+1

By Kevin Collier

Hackers and propagandists are turning to generative artificial intelligence chatbots like ChatGPT to make their operations seem more convincing to native English speakers, a senior official at the National Security Agency said Tuesday.

Speaking at the International Conference on Cyber Security at Fordham University in New York, NSA Cybersecurity Director Rob Joyce said the spy agency has seen both cybercriminals and hackers who work for foreign intelligence agencies using such chatbots to seem more like native English speakers.

ADVERTISING

Vivez dans un lieu qui vous ressemble !

Espaces Atypiques - Sponsored

Visiter le Site

“We already see criminal and nation-state elements utilizing AI,” he said. “We’re seeing intelligence operators, we’re seeing criminals on those platforms.”

“They’re much better at English-language content today,” Joyce said.

While generative AI chatbots frequently produce inaccurate information, the technology has rapidly become extremely adept at mimicking convincing and grammatically correct writing.

Hacking operations often rely, in part, on phishing schemes used to trick people into providing their personal information. The U.S. government, researchers and tech companies have repeatedly accused countries like Russia of orchestrating online propaganda campaigns by creating accounts that falsely purport to be those of American users. Generative AI helps hackers perfect and generate their communications, which makes malicious online activity more convincing, Joyce said.

“One of the first things they’re doing is they’re just generating better English-language outreach to their victims, whether it’s phishing emails or something much more elaborative in the case of malign influence,” he said.

Joyce didn’t name any specific AI company, but he said the issue is widespread.

“They’re all subscribed to the big-name companies that we would expect, all of the generative AI models out there,” he said.

Most generative AI services say they forbid their products from being used for criminal activity. But doing so is often trivial. NBC News was able to easily persuade both OpenAI’s ChatGPT and Google’s Bard to produce convincing phishing emails Tuesday.

In a statement, a Google representative said: "We have policies and protections in place against the use of generating content for deceptive or fraudulent activities like phishing. While the use of generative AI to produce negative results is an issue across all LLMs, we’ve built important guardrails into Bard that we’ll continue to improve over time.”

An OpenAI spokesperson said in an emailed statement that "We have studied cyber applications of LLMs, and are funding research and development toward an evaluation suite for LLM cybersecurity capabilities."

LLMs, or large language models, are models trained on massive amounts of data.

On the other hand, AI is becoming a more helpful defensive cybersecurity tool, Joyce said.

In September, the NSA chief, Gen. Paul M. Nakasone, announced the creation of the AI Security Center, which the Defense Department called in a news release “the focal point for developing best practices, evaluation methodology and risk frameworks with the aim of promoting the secure adoption of new AI capabilities across the national security enterprise and the defense industrial base.”

Joyce said: “AI machine learning, deep learning, is absolutely making us better at finding malicious activity. It is helping us see things illuminated so it stands out from the norm.”

Kevin Collier

Kevin Collier is a reporter covering cybersecurity, privacy and technology policy for NBC News.