Cyber-defense systems seek to outduel criminals in AI race

October 25 2023, by Gopal Ratnam, CQ-Roll Call

Credit: Pixabay/CC0 Public Domain

Not long after generative artificial intelligence models like ChatGPT were introduced with a promise to boost economic productivity, scammers launched the likes of FraudGPT, which lurks on the dark web

promising to assist criminals by crafting a finely tailored cyberattack.

The cybersecurity firm Netenrich in July identified FraudGPT as a "villain avatar of ChatGPT" that helps craft spear-phishing emails, provides tools to break passwords, and writes undetectable malware or other malicious code.

And so the AI arms race was on.

Companies are embracing cyber-defenses based on generative AI hoping to outpace attackers' use of similar tools. But more effort is needed, experts warn, including to safeguard the data and algorithms behind the generative AI models, lest the models themselves fall victim to cyberattacks.

This month, IBM released survey results of corporate executives, in which 84% of respondents said they would "prioritize generative AI security solutions over conventional ones" for cybersecurity purposes. By 2025, AI-based security spending is expected to be 116% greater than in 2021, according to the survey that was based on responses from 200 CEOs, chief security officers and other executives at U.S.-based companies.

Top lawmakers already are concerned about the dangers that AI can pose to cybersecurity.

At a hearing of the Senate Intelligence Committee in September, Chairman Mark Warner, D-Virginia, said "generative models can improve cybersecurity, helping programmers identify coding errors and contributing toward safer coding practices … but with that potential upside, there's also a downside since these same models can just as readily assist malicious actors."

Separately, the Pentagon's Defense Advanced Research Projects Agency in August announced a competition to design AI-based tools that can fix bugs in commonly used software. The two-year contest is intended to create systems that can automatically defend any kind of software from attack.

IBM said it is developing cybersecurity solutions based on generative AI models to "improve the speed, accuracy and efficacy of threat detection and response capabilities and drastically increase productivity of security teams."

Detecting deviations

Darktrace, a cybersecurity firm with offices in the United States and around the world, is deploying custom-built generative AI models for cybersecurity purposes, said Marcus Fowler, the company's senior vice president for strategic engagements and threats.

The company is using AI to predict potential attacks and designing proprietary self-learning AI models that observe and understand "the behavior of the environment that they're deployed within," meaning a computer network's normal patterns of use in a corporate or government setting. It maps activities of individuals, peer groups, and outliers, said Fowler, who previously served at the CIA developing the agency's global cyber-operations.

The system then is able to detect "deviations from normal and provide a context for such deviations," allowing security experts to take action, he said.

The company also developed AI systems to study how security experts investigate a breach and create "an autonomous triaging capability" that automates the first 30 minutes or so of an investigation, allowing security

officials to take swift action when an attack or a breach is detected, Fowler said.

In addition to detecting anomalies and aiding in investigations of a cyberattack, AI tools ought to be useful in analyzing malware to determine the origins of attackers, said Jose-Marie Griffiths, president of Dakota State University, who previously served on the congressional National Security Commission on Artificial Intelligence.

"Reverse engineering a malware to identify who sent it, what was the intent, is one area where we haven't seen a lot" of use of AI tools, "but we could potentially see quite a bit of work, and that's an area we are interested in," Griffiths said, referring to the university's ongoing work.

While malware is mostly software code, hackers often include notes in their own language, either to themselves or others, about a particular line of code's function. Using AI to glean such messages, especially those written in languages other than English, could help sharpen attribution, Griffiths said.

Use of generative AI models to improve cybersecurity is gaining momentum, but security experts also must pay attention to safeguarding the generative AI models themselves because attackers could attempt to break into the models and their underlying data, Griffiths said.

Broader use of generative AI in cybersecurity could help ease chronic problems facing security experts, said John Dwyer, head of research at IBM's X-Force, the company's cybersecurity unit.

"Alert fatigue, talent shortage and mental health issues have sort of been associated with cybersecurity for a long time," Dwyer said. "And it turns out that we can apply [AI] technologies to really move the needle to help address some of these core problems that everyone's been dealing with."

Cybersecurity experts are burned out by being constantly on alert, doing repetitive tasks, "sifting through a bunch of hay looking for a needle," and either leaving the industry or confronting mental health challenges, Dwyer said.

Using AI models to offload some of those repetitive tasks could ease the workload, and allow security analysts to focus on high-value tasks, Dwyer said.

As with all advances in technology online, progress in legitimate uses on the publicly accessible parts of the web often is accompanied by a "much faster rate of growth" in the underwater or dark web, where criminals and hackers operate, Griffiths said. In the case of generative AI, as defenders rush to incorporate the tools in defense, the attackers are racing to use the same tools.

"That's unfortunately the battle we are in," she said. "It's going to be constant."