Feature

Alamy

SECURITY > VULNERABILITIES AND THREATS

# ChatGPT and Cybersecurity: The Good, the Bad, and the Careful

Despite its benefits, ChatGPT has opened a Pandora's box of security risks. Here's what to know about protecting your company from criminal exploitation of AI-based chatbots.

Karen D. Schwartz | Mar 15, 2023

However, these same chatbots pose threats when used to develop malware, create phishing attacks, and steal sensitive information.

Organizations can combat threats with authentication, user education, and AI risk management frameworks.

In the acclaimed 2014 film *Ex Machina*, a scientist builds a humanoid robot in female form with AI, and the plot centers around determining whether the robot is capable of genuine thought and consciousness. By the end of the film, viewers see the robotic woman escape her confines and board a plane to parts unknown. The rest is left up to the imagination.

Related: How Artificial Intelligence Will Evolve in 2023

Less than a decade ago, a plot like that was pure science fiction. Today, not so much. One of the most talked-about advances in AI is ChatGPT, technology that trains itself to learn what humans are asking and answers those questions in conversational form. It's based on GPT-3, the third generation of OpenAI's neural network machine learning model, and is more advanced than previous attempts at conversational AI. According to research from Stanford University, this version was trained on 570 gigabytes of text and can perform tasks it wasn't even specifically trained on, like translating sentences from one language to another, with few or even no training examples.

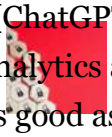
"It has an 'advanced Google search engine' kind of feel to it," explained Ketake Borade, a senior cybersecurity analyst at Omdia. "You can ask any question on earth to it, and it will answer you with whatever information it has."

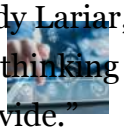
Generative AI technology has rapidly become a hot commodity. Since January alone, Microsoft invested billions in OpenAI and announced that it will incorporate the developer's language AI into its products, while Google introduced a conversational AI-based chat service called Bard. There is talk that Apple is working hard to develop AI-based applications.

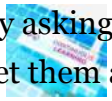
As ChatGPT continues to generate buzz, it's due in part to its potential for good and its potential for wreaking real havoc.

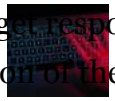
## The Good: How Companies Benefit From AI, Done Right

"[ChatGPT] takes things to another level," said Randy Lariar, practice director of big data, AI, and analytics at Optiv. "[You can] ask it anything you're thinking about and [get] an answer that's at least as good as an intern or entry-level worker could provide."


Key Strategies for Tackling Third-party Software Vulnerabilities
MAY 22, 2023


Look Out – Get-Rich-Quick Schemes Are Coming to AI
MAY 12, 2023


Cisco UCS X-Series Explains How Upskilling and Reskilling Drive Tech Careers
MAY 10, 2023


Generative AI: Cybercriminal's New Best Friend
MAY 08, 2023

By asking the right questions, Lariar said users can get responses that will either solve a problem or get them as much as 90% of the way to the completion of their tasks. It also allows employees to focus more on higher-level strategic ideas and customer experience.

Most importantly, perhaps, companies can use AI to strengthen cybersecurity defenses. For example, AI can infer patterns from incomplete or changed data, helping to reduce false positives, identify and respond to attacks, and create better security detections.







Security and IT professionals often work with code on one screen and Google on another to seek solutions to coding problems. "Now you've got ChatGPT, which takes that same kind of workflow, but you can now talk to it," Lariar said.

## The Bad: How AI-powered Chatbots Could Aid Cybercriminals

There is no way around it: Despite all its potential benefits to cybersecurity, ChatGPT gives threat actors more ways to do harm.

“If I was a competent threat actor, I’d know what to type [into ChatGPT],” said Guy Rosefelt, chief product officer for Sangfor Technologies' international division. “I could tell it that I want ransomware that does these things, based on this, using this algorithm and encryption protocol, and it will come out with a reasonable set of code I can then tweak and turn into something effective.”

Many believe that ChatGPT will help make phishing a bigger threat. Phishing is most successful when the emails or social media messages appear realistic, and ChatGPT can easily emulate the style and tone of a public figure, corporate executive, or company representative. For example, a bad actor could use an AI-based chatbot to craft an email from a public figure by researching that person’s writings and speeches. With that information, the chatbot could write an email in that person’s style, which could convince somebody to click on a link, buy something, or send money. Criminals can even send those emails from a legitimate account by hacking into that account first.

ChatGPT also might make it easier for bad actors to steal passwords, software codes, and sensitive corporate data, then use that information in a ransomware attack. Rosefelt described a case where someone posted on the dark web that they had used ChatGPT to write a script called infostealer, which could parse an organization’s information and steal credentials.

ChatGPT’s ability to perpetrate other types of malware is another concern. For example, a hacker could ask ChatGPT for tips on how to probe and penetrate networks or how to get around certain types of security controls. ChatGPT can even help hackers write malware if they ask the right questions, Lariar noted.

While these threats are real, the person requesting the actions still must have some knowledge of what to request. Without that background knowledge, it’s doubtful that the malware code would be effective.

## The Careful: Protecting Organizations From AI-based Threats

There is no doubt that ChatGPT is already being used by the bad guys. The key is to find ways to protect organizations from AI-equipped criminal efforts. Even companies using advanced chatbot technology for good must remain alert and well-informed.

### Authentication measures

In addition to stepping up awareness training that teaches users to be suspicious, Rosefelt said companies should strongly consider using some type of authentication mechanism to verify the

companies should strongly consider using some type of authentication mechanism to verify the source of emails and other communications, even if they appear to have come from the company's CEO. That might mean using a public/private key combination that requires matching keys. If it was a phishing email, the public key wouldn't work.

### User education

It's equally important to teach internal users how to use AI responsibly.

"We're still in the early days, but companies that are relying on AI more and more are likely to introduce new attack surfaces," Rosefelt explained.

For example, if a company develops a public-facing system that includes AI, hackers might be able to find out information the company didn't intend to make public. "As they bring AI into production, they need to think about the ways those tools could be exploited and promote best practices for how to use AI in production without opening holes."

### Frameworks for AI risk management

It's also wise to study AI frameworks like MITRE ATLAS and NIST's AI Risk Management Framework. Released in June of 2021, MITRE ATLAS is a knowledge base of tactics, techniques, and case studies that can help IT professionals protect their organizations. NIST's framework, released in January of 2023, aims to help organizations better manage risks associated with AI. The framework's companion, NIST AI RMF Playbook, provides tips on how to get the most out of the framework.

## ChatGPT's Potential for Harm and Protection in Cybersecurity

Because ChatGPT is still relatively new, the full scope of its risks remains unknown. At the same time, cybersecurity vendors have ChatGPT on their minds, and they are likely either to come up with standalone products or incorporate protection into their existing technologies. In addition, there is plenty of venture capital money pouring into developing new tools.

The most effective way to combat AI-based threats is with AI itself, Rosefelt noted. "You would need to have AI tools that could evaluate something to determine whether it is fake or not," he said.

That's the route that Rosefelt expects vendors, including Sangfor Technologies, to take. He said these tools will likely become available by the second half of 2023.

### About the author

*Karen D. Schwartz is a technology and business writer with more than 20 years of experience. She has written on a broad range of technology topics for publications including CIO, InformationWeek, GCN, FCW, FedTech, BizTech, eWeek and Government Executive.*