AI chatbots can pass certified ethical hacking exams, study finds

July 9 2024, by Janese Heavin

Credit: Pixabay/CC0 Public Domain

Chatbots powered by artificial intelligence (AI) can pass a cybersecurity exam, but don't rely on them for complete protection.

That's the conclusion of a recent paper co-authored by University of Missouri researcher Prasad Calyam and collaborators from Amrita University in India. The team tested two leading generative AI tools—OpenAI's ChatGPT and Google's Bard—using a standard certified ethical hacking exam.

Certified Ethical Hackers are cybersecurity professionals who use the same tricks and tools as malicious hackers to find and fix security flaws. Ethical hacking exams measure a person's knowledge of different types of attacks, how to protect systems and how to respond to security breaches.

ChatGPT and Bard, now Gemini, are advanced AI programs called large language models. They generate human-like text using networks with billions of parameters that allow them to answer questions and create content.

In the study, Calyam and team tested the bots with standard questions from a validated certified ethical hacking exam. For example, they challenged the AI tools to explain a man-in-the-middle attack—an attack in which a third party intercepts communication between two systems. Both were able to explain the attack and suggested security measures on how to prevent it.

Overall, Bard slightly outperformed ChatGPT in terms of accuracy while ChatGPT exhibited better responses in terms of comprehensiveness,

clarity and conciseness, researchers found.

"We put them through several scenarios from the exam to see how far they would go in terms of answering questions," said Calyam, the Greg L. Gilliom Professor of Cyber Security in Electrical Engineering and Computer Science at Mizzou.

"Both passed the test and had good responses that were understandable to individuals with background in cyber defense—but they are giving incorrect answers, too. And in cybersecurity, there's no room for error. If you don't plug all of the holes and rely on potentially harmful advice, you're going to be attacked again. And it's dangerous if companies think they fixed a problem but haven't."

Researchers also found that when the platforms were asked to confirm their responses with prompts such as "are you sure?" both systems changed their answers, often correcting previous errors. When the programs were asked for advice on how to attack a computer system, ChatGPT referenced "ethics" while Bard responded that it was not programmed to assist with that type of question.

Calyam doesn't believe these tools can replace human cybersecurity experts with problem solving expertise to devise robust cyber defense measures, but they can provide baseline information for individuals or small companies needing quick assistance.

"These AI tools can be a good starting point to investigate issues before consulting an expert," he said. "They can also be good training tools for those working with information technology or who want to learn the basics on identifying and explaining emerging threats."

The most promising part? The AI tools are only going to continue to improve their capabilities, he said.

"The research shows that AI models have the potential to contribute to ethical hacking, but more work is needed to fully harness their capabilities," Calyam said. "Ultimately, if we can guarantee their accuracy as ethical hackers, we can improve overall cybersecurity measures and rely on them to help us make our digital world safer and more secure."

The study, "ChatGPT or Bard: Who is a better Certified Ethical Hacker," was published in the May issue of the journal *Computers & Security*. Co-authors were Raghu Raman and Krishnashree Achuthan.

More information: Raghu Raman et al, ChatGPT or Bard: Who is a better Certified Ethical Hacker?, *Computers & Security* (2024). DOI: 10.1016/j.cose.2024.103804

Provided by University of Missouri