CyberTalk.org

ChatGPT Security Risks: A Guide for Cyber Security Professionals

How previously unknown chatbot risks could affect your business

Table of Contents

Introduction

The advancement of language models, like ChatGPT, heralds the beginning of a new era in human-machine collaboration. ChatGPT can offer human-like responses based on a vast knowledge base, and due to the machine learning model on which it was built, the technology will continually improve over time.

In general, business executives see the value of artificial intelligence, and some are exceptionally enthusiastic about what AI-based tools, like ChatGPT, can accomplish.

Presently, employees can use ChatGPT on a per-instance basis, or businesses can integrate ChatGPT into their own applications or platforms, such as a website or mobile app, to provide automated support or AI-powered features. API integrations are available through OpenAI (ChatGPT's parent company) and through third-parties.

Because third-party entities also offer pre-built chatbot-powered solutions, businesses can customize the tool in ways that integrate into specific workflows. Some third-party solutions may be sold as software-as-a-service products.

How influential figures perceive ChatGPT

Elon Musk spoke of ChatGPT as having "great promise" even if it comes with "great danger."

"Until now, artificial intelligence could read and write, but could not understand the content. The new programs like ChatGPT will make office jobs more efficient..."

- Bill Gates

"We're on the verge of a very interesting revolution—the AI revolution,"

- Gil Shwed, CEO of Check Point

"...there are a lot of benefits to be gained from this kind of technology and we're only beginning to scratch the surface, but we can't ignore the ramifications."

– Shishir Singh, CTO, Cybersecurity, BlackBerry

Regardless of how ChatGPT is integrated and applied in the business setting, the technology can provide powerful AI-based processing capabilities that boost efficiency, reduce costs, provide competitive advantages and offer new insights.

The challenge is how to overcome security obstacles.

As the use of ChatGPT and similar technologies expand, so do cyber security risks.

The four substantial areas of risk include:

People

Data Privacy

Malware

Data Breaches

Continue reading to understand how these risks could affect your business.

People

Every CISO and CIO knows that humans top the list when it comes to the most significant threat vectors and attack surfaces. In other words, people are error-prone and no ready-made solution exists to solve that issue.

By and large, employees have the best of intentions, and want to engage with ChatGPT in a way that's beneficial to their work, responsible, and respectful of workplace expectations. But non-technical employees may not inherently know what is or isn't acceptable in the way of chatbot use. Chances are that they haven't given it much thought—they just know that a chatbot can help provide results.

The "newness" of the technology may not 'mesh' with employees' pattern-recognition modalities. Employees may not realize that this 'shiny new thing' is still, in essence, a third-party website for which regular cyber security rules apply.

For example, an employee may unwittingly provide ChatGPT or a similar platform with specific financial information for clients, asking the technology to build out a report for them. In turn, corporate data may end up on non-enterprise servers. On said servers, the data may be under-secured or secured in a way that's not compliant with an organization's legal mandates. The ChatGPT end-user license agreement discusses this risk.

> Earlier this year, Amazon issued a company-wide warning pertaining to sharing information with OpenAI's chatbot. 'We wouldn't want its output to include or resemble our confidential information (and I've already seen instances where its output closely matches existing material)' read the warning.[1]

As a leader, advocate for responsible use of ChatGPT in the workplace. In a personable way, tell employees about what information can be and should not be shared with chatbots. Explain the reasoning and ensure that you can point to evidence that supports your decisions. Be sure to thank everyone for their cooperation with these new guidelines that are evolving in parallel with the technology itself.

[1] JPMorgan restricts ChatGPT usage..., Paul Farrel, Daily Mail, 22 February 2023

Data Privacy

According to ChatGPT itself, “Chatbots that use ChatGPT may collect and store sensitive information such as personal data, financial information, or health data. This information could potentially be accessed or stolen by unauthorized individuals, putting the privacy and security of individuals at risk.”

Collection and storage of sensitive information: Chatbots that use ChatGPT may collect and store sensitive information such as personal data, financial information, or health data. This information could potentially be accessed or stolen by unauthorized individuals, putting the privacy and security of individuals at risk.

To ensure maximal data privacy, leverage the following insights:

- Implement access controls in order to limit internal access to sensitive data. Apply user controls and authentication mechanisms. Encrypt sensitive data.

- AI-based applications that use ChatGPT and ChatGPT itself should be hosted on secure servers and storage systems. Make sure that your hosting and storage providers apply appropriate cyber security measures; from access controls, to encryption, to intrusion detection systems.

- Routinely update cyber security measures to ensure that they remain capable of fending off waves of advanced cyber threats. In so doing, organizations may wish to consider vulnerability assessments, penetration testing, and regular updates to security policies and procedures.

- When leveraging ChatGPT, businesses should be sure to pursue appropriate measures to protect sensitive data that may be collected by chatbots and other AI-based applications.

- Users of the system should not input any personal confidential information.

Malware

At this point, ChatGPT's capacity to produce malicious software code is limited, although extant. Cyber criminals can use chatbots to help them execute on malicious activities and in so doing, have been observed bypassing the chatbot's safeguards.

Check Point researchers have monitored dark web forums and found instances where cyber criminals were exchanging information about the use of the chatbot to “improve” malware code.

> **“...[ChatGPT] can improve the cost efficiency of sophisticated threats.”[2]**
>
> - Sergy Shykevich, Threat Intelligence Group Manager, Check Point

Reddit users are having discussions about “jailbreaks” (certain language prompts that successfully override the chatbot’s defenses), which cyber criminals have made use of.

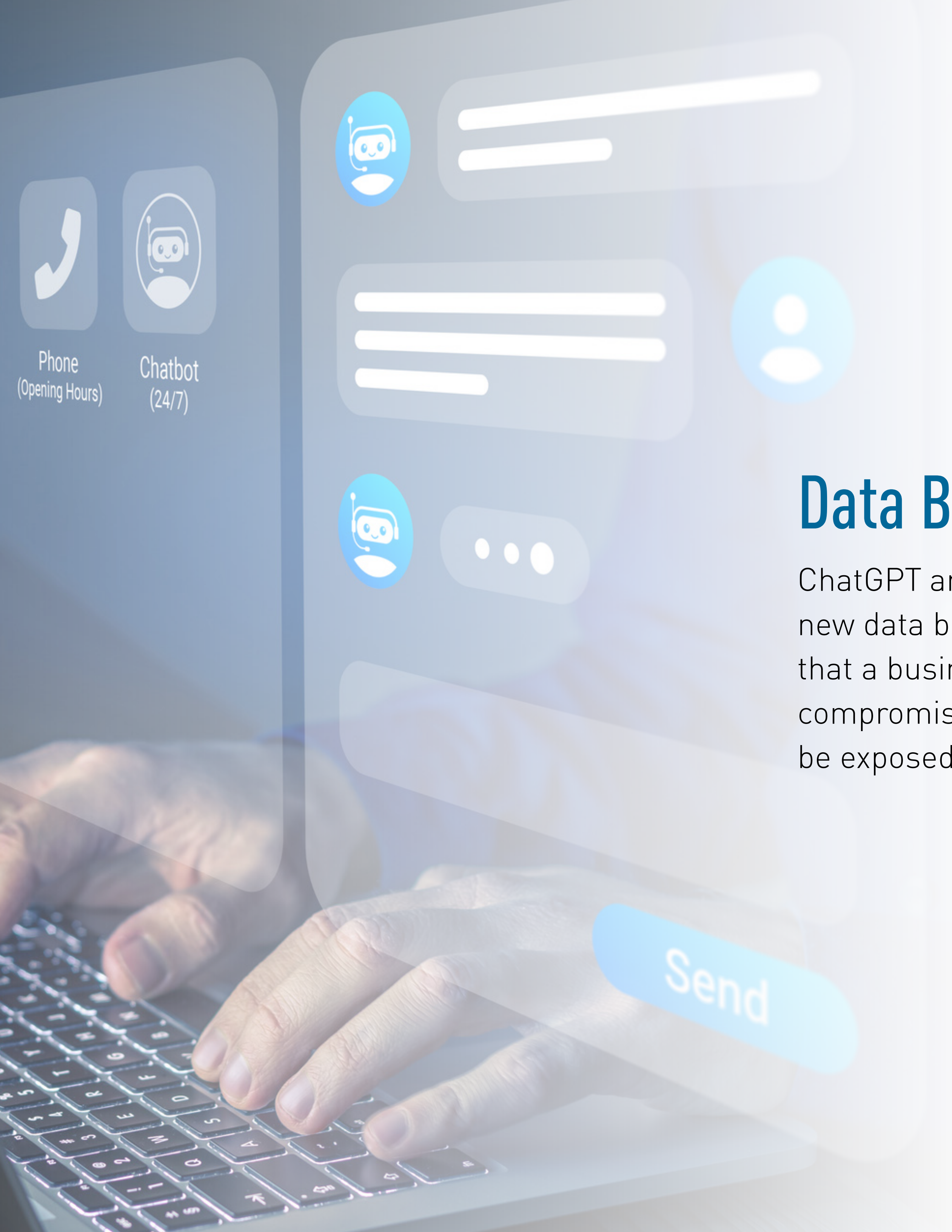
For instance, a person who asks the chatbot for a sample of code that encrypts files could potentially leverage the code in order to fast-track ransomware projects. While the chatbot will not write a complete ransomware script, the short-form sample material produced could still prove dangerous.

Cyber security researchers have also observed that asking ChatGPT for new pieces of code continuously can enable users to create highly evasive polymorphic malware. While this is not a new capability for cyber attackers, ChatGPT’s ability to manufacture code may enable low-skilled wanna-be cyber criminals to execute sophisticated attacks.

Since the discovery of serious cyber security concerns, ChatGPT’s parent company, OpenAI has worked to refine and redefine the chatbot’s capabilities. Nonetheless, cyber criminals can still use the program, and with that, they can scale-up malware deployment.

Because “there is no way that the abuse will be reduced to zero,” says Sergy Shykevich, threat intelligence group manager at Check Point, organizations should take corresponding precautions. To quickly stem the volume of malware-based threats in your environment, leverage artificial intelligence-based cyber security tools, which can help your business respond to and stop potential breaches. Fight AI threats with AI-powered tools.

[2] How hackers can abuse ChatGPT to create malware, Alexis Zacharakos, TechTarget, 22 February 2023



Data Breaches

ChatGPT and similar technologies introduce new data breach-related risks. In the event that a business's ChatGPT instance were compromised, sensitive information could be exposed to hackers.

According to OpenAI's privacy policy, the company collects individual IP addresses, browser types, settings and data on user interactions with the chatbot site. Other collected data may include the type of content that users engage with, features that users utilize, and actions that users take.

It also aggregates data about users' browsing activities over time and across divergent websites. OpenAI's policies also state that the company may share users' personal information with unspecified third-parties in order for the company to accomplish business objectives.

To mitigate the risk of data compromise that comes with chatbot use, businesses need to pursue appropriate security measures. These may include implementing encryption to protect sensitive data, limiting access to systems, and regularly monitoring systems for suspicious activity.

Address system weaknesses quickly and broadly take a proactive approach to security. When implementing and using artificially intelligent language-processing technologies, be sure to mitigate data compromise-related risks.

AI **Artificial** Intelligence

Conclusion

ChatGPT and similarly powerful tools are now an inescapable element to consider in developing a strong, resilient cyber security framework.

"It's a great technology—but, as always with new technology, there are risks and it's important to discuss them to be aware of them," says threat intelligence group manager at Check Point, Sergy Shykevich.[3]

As technology continues to advance, cyber security and IT leaders need to stay informed and to stay ahead of potential cyber security threats.

By learning about the risks associated with ChatGPT and similar technologies, and by taking precautions, we can ensure that powerful AI-based tools are used in the most secure ways possible. Protect your world.

Did you know?

New chatbots also offer a certain advantage to cyber security professionals.

These chatbots are particularly good at understanding code, and as a result, defenders may be able to use them to better understand malware.

For more executive-level insights into ChatGPT, please visit CyberTalk.org.

[3] ChatGPT and more: What AI chatbots mean for the future of cybersecurity, Danny Palmer, ZDNet, Feb 14 2023

CyberTalk.org