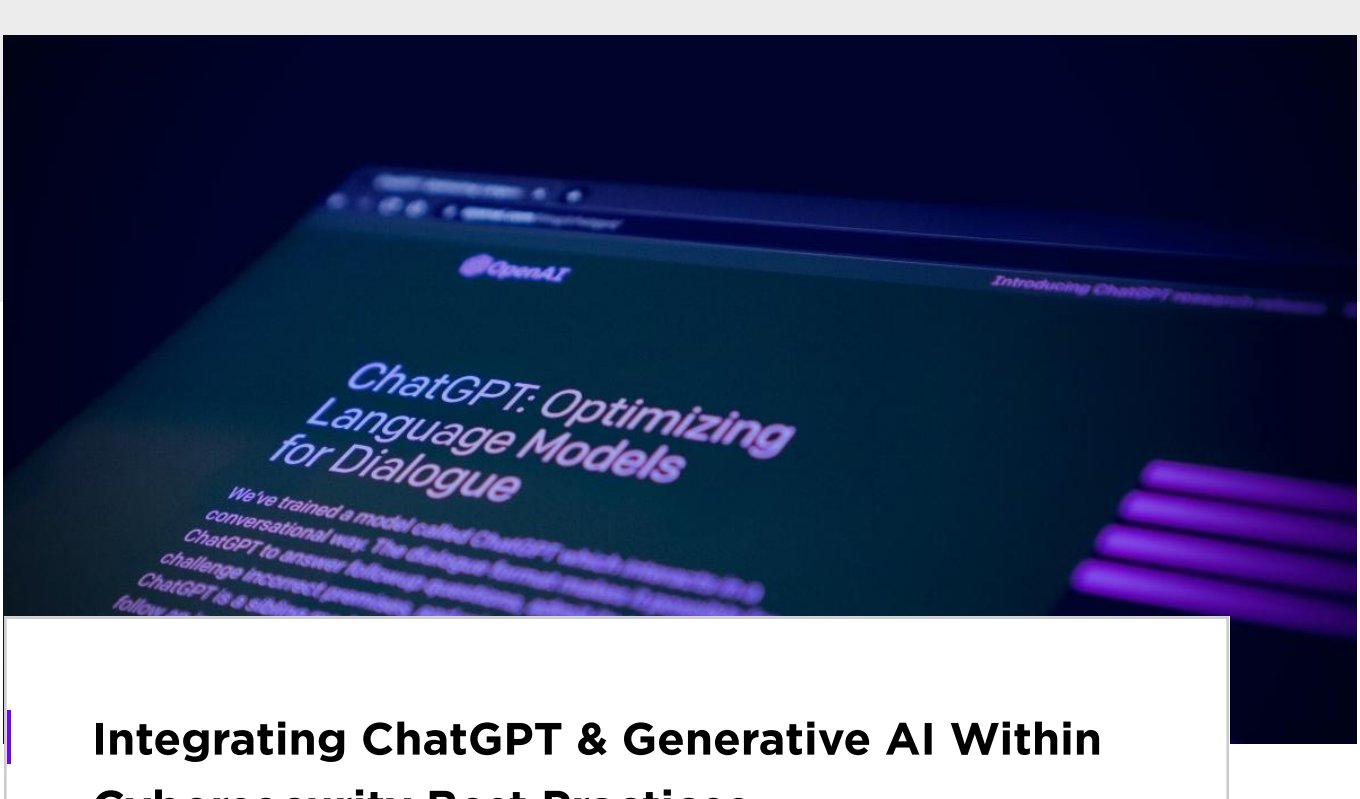


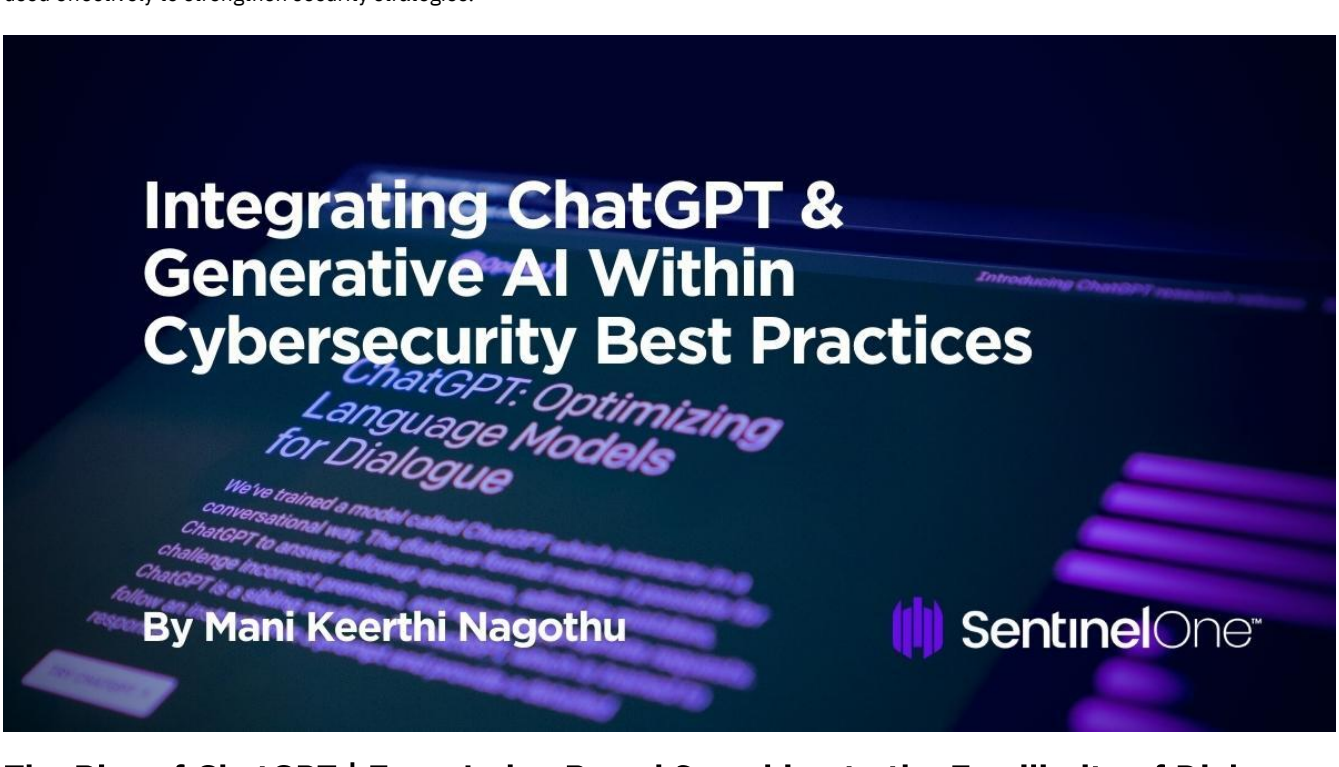
Integrating ChatGPT & Generative AI Within Cybersecurity Best Practices

April 5, 2023
by Mani Keerthi Nagothu

ChatGPT has generated a tremendous buzz since it was launched in November 2022. Over the last five months, the generative artificial intelligence (AI) chatbot has become the subject of many complex and ongoing discussions on its potential to impact the infosec community and cybersecurity landscape on the whole.

For security leaders, the use of AI can be a powerful tool, helping to ensure that their business's architecture can withstand all of the developing challenges in the threat landscape. Staying ahead of incoming threats means that leaders are looking past reactive, single-layer solutions that can no longer keep up with novel and increasingly sophisticated techniques of threat actors.

This blog post examines the use of ChatGPT in the context of cybersecurity and discusses its implications on information security. It describes how generative AI tools like ChatGPT can be safely implemented within the boundaries of cybersecurity best practices and be used effectively to strengthen security strategies.



The Rise of ChatGPT | From Index-Based Searching to the Familiarity of Dialogue

Over the span of two decades, Google Search has arguably shaped the way modern internet users discovered, consumed, and maneuvered through the vast store of information on the web.

Google's search engine provides information much like the index of a book rather than a table of contents. The search is largely manual and it is up to the user to sift through all the hits, find the accurate answer, make the connections between found information, and finally, to process how they should think about it.

Alternatively, ChatGPT's wildfire-like popularity has been credited to its ability to 'speak' in a natural-sounding language. When asked a question, ChatGPT draws from a massive amount of text data, including books, articles, and web pages to generate its response. Trained to be conversational, it then uses machine learning algorithms to reply in the same context as the rest of the dialogue. As a result, the responses are human-like, giving the user a familiar sense of two-way conversation as opposed to a one-way, index-based search that internet users have adopted in the era of Google.

Multimodality of Social Media & ChatGPT in Information Search

As the newest generation of internet users emerge, research is showing their preference for multimodal forms of communication and searching for information on social apps such as TikTok and Instagram. On these apps, information seems to come from 'first hand' sources, building an information exchange based on conversation and casual dialogue.

Though the technology behind ChatGPT is not new, it seems to ride on the coattails of this new preference for community-centric conversations between a 'source' and its consumer. Perhaps ChatGPT heralds a turning point in how users think about and approach data organization. Within the first two months of its launch, the AI bot had more than 100 million users and sees more than 13 million daily visitors as of 2023.

Can ChatGPT Be Safely Integrated in Cybersecurity Strategies & Processes?

Given the popularity of ChatGPT and how it is impacting information consumers globally, security leaders are giving more attention to such tools as a way to enhance their business. Though a powerful tool, it is important to assess how it can fit into organizational workflows in a safe and effective manner.

In most cases, end users have reported a positive experience with regard to the data output by the AI chatbot. However, the bot's parent company, OpenAI, has published various terms and safety policies, pointing users to the reality that ChatGPT currently does not offer any fact checks and responses provided should not be blindly trusted.

According to NIST's AI Risk Management Framework, an AI system can only be deemed trustworthy if it adheres to multiple criteria. These include being valid, reliable, accountable, transparent, fair with harmful biases managed, secure and resilient, explainable, interpretable, and safe.

Securing ChatGPT From a People Perspective

ChatGPT can be used safely when an organization's leaders and security teams work together to manage risks. From a people perspective, it is important to understand the ways that generative AI chatbots can be misused and how to defend against them.

Reduced Barriers for Entry for Adversaries

Researchers have found that ChatGPT can be leveraged by threat actors to generate malware commands or even create on-the-fly malware. While this goes strictly against OpenAI's content policy, there is indication that actors are actively working to bypass the chatbot company's restrictions by sharing their techniques in dark forums. Should these restrictions fall, the chatbot could be used by lower-level cyber criminals and script kiddies to generate or improve existing malicious code.

AI Phishing Emails

Given the dialogue-based nature of the chatbot, security experts hypothesize that threat actors could use ChatGPT to craft well-written phishing emails. Before, some of the common tell-tale signs of a phishing email included poor grammar, spelling mistakes, and odd or urgent tone of voice. If threat actors begin to leverage ChatGPT to create their social engineering content, they could increase their production of phishing emails and sound more convincing.

As it stands, ChatGPT is targeted just like many other platforms available on the market. Organizations can combat the potential risks that generative AI tools bring from a people-first approach. This means engaging employees in training and awareness programs on how bots like ChatGPT work, how to detect AI-generated content, and buckling down on identity-based cybersecurity measures.

Securing ChatGPT From a Process Perspective

An increasing number of business leaders are beginning to recognize that their employees need direction on how to use ChatGPT safely and how to reduce risk while doing so. To protect data privacy, many organizations are investing time in crafting ChatGPT-specific policies and processes. This may include recommendations and requirements surrounding code checking, brainstorming, content drafting, editing, and research.

These policies and processes may also cover how companies will implement quality control over content where ChatGPT has been involved in its lifecycle. Additionally, they can include any related contractual risks for using third-party generative AI tools to produce or process sensitive data and deliverables, managing inherent bias, and risks regarding privacy, consumer protection, intellectual property, and vendors.

Privacy Risks Associated with ChatGPT

Due to the capability of AI systems to aggregate information, there is a significant possibility that personally identifiable information (PII) could be used by the bot to provide outputs to the end user. End users are not restricted from inputting PII information into the AI bot, which is then used to aggregate information for future purposes.

Confidentiality Risks Associated with ChatGPT

End users might interact with the AI system by inputting confidential information of the organization and trying to gather a better understanding or output. For example, an end user can input the organization's security policy and ask an AI to word it in simpler terms. The output might be excellent with a greater structure than before, but the AI can collect this information for future responses.

Data Integrity Risks Associated with ChatGPT

There might be cases where the AI system-generated content may give output that differs from the original view or context and can lead to an inaccurate, incomplete, and biased output. Furthermore, relying on the output as the truth can place end users to use incorrect information.

Legal and Regulatory Risks Associated with ChatGPT

Data fed to the AI might contain copyright material, trade secrets, or confidential information, and responses output by the AI do not have the data owners' consent. End users need to consider whether they have the right to use or publish such material. There are also geographical laws and regulatory requirements that may need to be met when using data from the AI bot.

Reputational Risks Associated with ChatGPT

In the case where staff are using ChatGPT for producing content, it is important to be aware that tools are already available to recognize whether the content was produced using AI. Tools to recognize content generated by ChatGPT are not perfect yet, but utilities like OpenAI AI Text Classifier are improving rapidly and likely to become more widely adopted going forward.

Securing ChatGPT From a Technology Perspective

What makes ChatGPT appealing is the speed and the apparent sophistication of the output when posed with questions and commands. This can lead to an implicit trust of the technology underlying it. When organizations move towards using this technology in their organizations, they need to understand the dependencies in delivering the output as per their expectations, including the following areas.

Homogenization of Data Output

There is a potential that all the output produced by ChatGPT can be similar in terms of structure and writing style and the absence of human emotion in creating content. Mass production and propagation of ChatGPT output can lead to limited perspectives, dissuade users from exploring more angles and research, and discourage creative writing or more innovative problem solving.

Potential Costs

While the costs associated with tools like ChatGPT may seem attractive now as developers seek adoption, this may not remain the case in the future. Building dependencies on nascent 3rd party solutions needs to take into account the possibility that costs may rise unexpectedly in the future. Just recently, OpenAI's CEO announced that a professional version of the tool will offer higher limits and faster performance for those subscribed.

From a technological point of view, it is critical for organizations to be clear about what AI tools like ChatGPT can and cannot do. ChatGPT holds the ability to analyze vast amounts of data and find patterns to generate responses, but it cannot reason, think critically, or understand what is best for the business. It can, however, be a very powerful addition to human intelligence. The human element in the safe use of ChatGPT remains key to organizations that are starting to leverage AI more in their day-to-day work.

Conclusion

The benefits of ChatGPT are many, and there is no doubt that generative AI tools like it have proven to augment human tasks and make workflows and content production more efficient. As companies try to maximize the advantages of ChatGPT and use it for competitive advantage, it is important to note that the use of generative AI is still in its nascent stage for large-scale adoption.

When integrating ChatGPT into a businesses' cybersecurity strategy and processes, security leaders should consider the various risks across their people, processes, and technology. By putting the right safeguards in place, generative AI tools can be used to support existing security infrastructures.

SentinelOne continues to protect and support enterprises worldwide as they explore the benefits of innovative and new technologies. Learn more about how Singularity™ helps organizations autonomously prevent, detect, and recover from threats in real time by contacting us or requesting a demo.

Like this article? Follow us on LinkedIn, Twitter, YouTube or Facebook to see the content we post.

Read more about Cyber Security

- SentinelOne's Cybersecurity Predictions 2023 | What's Next?
- Apple's macOS Ventura | 7 New Security Changes to Be Aware Of
- SentinelOne Debuts at the Top of MITRE Engenuity ATT&CK® Deception Evaluation. See Why.
- Cyber Risks in the Education Sector | Why Cybersecurity Needs to Be Top of the Class
- Best-of-Breed Identity Threat Detection and Response Meets Best-of-Breed XDR
- Cybersecurity Sharing | An Infosec User's Guide to Getting Started on Mastodon