Can ChatGPT Be Safely Integrated in Cybersecurity Strategies & Processes?

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 conversations between a 'source' and its consumer. Perhaps ChatGPT heralds a turning point in how users think about and approach data.

Though the technology behind ChatGPT is not new, it seems to ride on the coattails of this new preference for community-centric sources, building an information exchange based on conversation and casual dialogue.

As the newest generation of internet users emerge, an inaccurate, incomplete, and biased output. Furthermore, relying on the output as the truth can place end users to use incorrect information.

The AI Text Classifier

AI holds the ability to analyze vast amounts of data and find patterns to generate responses, but it cannot reason, think critically, or be used safely. Tools to recognize content generated by ChatGPT are not perfect yet, but utilities like AI Text Classifier can be used to identify whether the content was produced using AI.

In the case where staff are using ChatGPT for producing content, it is important to be aware that tools are already available to recognize the data owners' consent. End users need to consider whether they have the right to use or publish such material. There are also recommendations and requirements surrounding code checking, brainstorming, content drafting, editing, and brainstorming.

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, but they should be aware of the potential privacy risks associated with ChatGPT.

Privacy Risks Associated with ChatGPT

Cyber criminals and script kiddies can use ChatGPT to create socially engineered content, such as phishing emails. The urgent tone of voice could be used by threat actors to increase their impact. Additionally, ChatGPT can be used to generate or improve existing malicious code, making it easier for cyber criminals to exploit vulnerabilities.

AI phishing emails can be difficult to detect, as they can sound more convincing than traditional phishing attempts. To protect data privacy, many organizations are investing time in crafting ChatGPT-specific policies and processes. This may include being valid, reliable, accountable, transparent, fair with harmful biases managed, secure and resilient, explainable, interpretable, process, and efficient.

In summary, ChatGPT is a powerful tool that can be used to enhance business processes. However, it is important to assess how it can fit into organizational workflows and how to reduce risk while doing so. To protect data privacy, many organizations are investing time in crafting ChatGPT-specific policies and processes.