ExtraHop

Detecting Data Leaks from Employee Use of ChatGPT with Reveal(x)

Ever since OpenAI released ChatGPT on November 30, 2022, use of AI as a service (AIaaS), including generative AI tools, has skyrocketed. Within five days of launching, ChatGPT exceeded 1 million users. By March, it reached 1 billion users. By comparison, it took TikTok eight years to reach 1 billion users.

As individuals look to tap the enormous productivity benefits that AIaaS promises, organizations have learned some hard lessons about the ease and speed with which employees can share sensitive or proprietary information with these public tools.

What employees using AI to accelerate code reviews and new drug discovery may not realize is that they're effectively putting confidential data in the public domain: once they share proprietary information with an AIaaS, the service continues to use it to process other people's requests–in some cases, indefinitely, depending on the AIaaS provider's terms of service.

The immediate data leakage and intellectual property risks center on users logging into the websites and APIs of these services and sharing confidential data. However, these risks increase as local deployments of AI systems flourish and people start connecting them to each other. Once an AI service begins sharing data directly with other AI services, the human oversight element that currently assesses whether that data should be shared is lost. The AI service is unlikely to understand the impact and potential consequences of sharing data it has access to and may not notify its human handlers that data has been shared outside of an organization.

Thus, the risk of IP loss and customer data leakage has made it imperative for organizations to understand the scope of AIaaS use across their businesses. Until now, organizations haven't had an easy way to audit employee use and potential misuse of these tools.

Data Protection From Rogue AI Use and Accidental Misuse

ExtraHop Reveal(x) has a new Threat Briefing for AI as a Service that helps organizations understand their risk exposure from employee use of ChatGPT.

Reveal(x) provides customers with visibility into the devices and users on their networks that are connecting to OpenAI domains. This capability is essential as organizations move quickly to adopt policies governing the use of large language models and generative AI tools, since it will give organizations a mechanism to audit compliance with those policies. ExtraHop is taking this step as part of a larger security platform approach, incorporating the AIaaS monitoring with our existing, industry-leading network detection and response (NDR) capabilities.

By tracking which devices are connecting to OpenAI domains, identifying the users associated with those devices and the amount of data those devices are sending to those domains, Reveal(x) enables organizations to assess the risk associated with their employees' ongoing use of ChatGPT.

In addition, because Reveal(x) shows the amount of data being sent to and received from these domains, security leaders can evaluate what falls within an acceptable range and what indicates potential IP loss. For example, simple user queries to a chatbot should fall within a range of bytes to kilobytes. If security teams see MBs of data flowing to these domains, that volume may signify employees are sending proprietary data with their query.

Organizations will be able to identify the type of data and individual files that employees are sending to OpenAI domains if the traffic in question is not encrypted and Reveal(x) is able to identify related data exfiltration and data staging detections.

Network Telemetry: Key to Exposing Data Leaks

Reveal(x) is able to provide this deep visibility and real-time detection because we use network packets as the primary data source for monitoring and analysis. Using a real-time stream processor, Reveal(x) transforms unstructured packets into structured wire data and analyzes payloads and content from OSI Layer 2–7 for complete network visibility. From device discovery to behavioral analysis, network telemetry is the immutable source of truth for understanding an organization's hybrid environment. Logs can tell you that two devices talked to each other, but Reveal(x) provides rich context about the communication.

At ExtraHop, we can't underscore the importance of this capability enough as organizations grapple with the popularity and proliferation of AIaaS and the data leakage risk associated with it. ExtraHop believes the productivity benefits of these tools outweigh the data leakage risks, provided organizations understand how these services will use their data (and how long they'll retain it), and provided organizations not only implement policies governing use of these services but also have a control like Reveal(x) in place that allows them to assess policy compliance and spot risks in real time.

See it in Action

Learn how Reveal(x) can help your organization assess the scope of data leaks and generative AI use in your organization.

SCHEDULE A PERSONALIZED DEMO

ABOUT EXTRAHOP NETWORKS

ExtraHop is the cybersecurity partner enterprises trust to reveal the unknown and unmask the attack. The ExtraHop Reveal(x) 360 platform is the only network detection and response solution that delivers the 360-degree visibility needed to uncover the cybertruth. When organizations have full network transparency with ExtraHop, they see more, know more, and stop more cyberattacks.

Learn more at **www.extrahop.com**

ExtraHop

info@extrahop.com
www.extrahop.com