Kiteworks Proactively Protects Confidential IP and Private Data from Exposure in Data-hungry Generation AI LLMs

Kiteworks' content-defined zero trust and digital rights management addresses rising concerns over sensitive content ingestion in large language model tools.

SAN MATEO, CA, US, August 17, 2023 /EINPresswire.com/ -- Kiteworks, which delivers data privacy and compliance for sensitive content communications through its Private Content Network, announced that the Kiteworks platform uses next-generation digital rights management (DRM) protection to protect critical corporate intellectual property (IP) and personally identifiable information (PII) from ingestion into a burgeoning number of large language model (LLM) tools built on generative artificial intelligence (AI).

> Kiteworks next-generation DRM capabilities enable businesses to track and control the sensitive content employees, contractors, and third parties ingest into generative AI LLMs."
>
> *Tim Freestone, CMO, Kiteworks*

A recent survey by Gartner found that enterprises list generative AI as their second-highest risk, pinpointing three primary risk aspects: 1) IP used as part of the training set and leveraged in outputs for other users, 2) PII and other sensitive personal data being used in AI tools that violate data privacy laws, and 3) bad actors using generative AI tools to accelerate the development of attacks and exploitations.

Sensitive content at risk includes 1) training data—that used to train the AI language model, 2) knowledgebase data—confidential, proprietary information used to generate responses from the generative AI LLM tool, and 3) confidential chatbot interactions in customer support, sales, and marketing scenarios where PII and other personal data information are entered, both intentionally and inadvertently, into the chat interface. Many organizations rank the risk as serious, enough that three-quarters of organizations are currently implementing or considering bans on ChatGPT and other generative AI applications within the workplace; 61% of those indicate the bans are intended for the long term or even permanently. (1)

Recent studies show there is a significant risk of sensitive content leakage into generative AI LLM tools:

• 15% (a number that is growing fast) of employees regularly post company data into generative AI LLMs, and one-quarter of that data is considered sensitive. (2)
• For workers using generative AI LLM tools, they use them an average of 36 times a day with 25% of the uses including a data paste. (3)
• The top categories of confidential information being inputted into generative AI LLMs include internal business data (43%), source code (31%), PII (12%), and customer data (9%). (4)
• There are an astounding 30,000 GPT-related projects on GitHub. (5)
• Only 28% of organizations have instituted processes to mitigate regulatory compliance risks. (6)
• Only 20% of organizations have instituted processes to mitigate PII being used in generative AI LLMs. (7)

Without content-based risk policies and controls in place, sensitive content leakage into the generative AI LLMs can be a serious threat for organizations. In addition to cybercriminals manipulating generative LLMs for malicious activity, security researchers believe training data extraction attacks could successfully recover verbatim texts, PII, and IP from generative AI models that cybercriminals hold for ransom. Loss of PII and protected health information (PHI), even if inadvertent, may violate data privacy regulations like GDPR, HIPAA, PIPEDA, PCI DSS, and numerous others that require public disclosure and notification. This can result in regulatory fines and penalties, brand denigration, diminished productivity, and decreased revenue.

“Generative AI LLMs present an urgent data protection challenge to organizations,” said Tim Freestone, CMO at Kiteworks. “The number of employees and contractors using generative AI LLMs is skyrocketing—and will continue to do so because of the immense competitive advantages it offers. And one must remember those within your network are just the cusp of the problem; most organizations send, share, receive, and store sensitive content with thousands of third parties.”

The good news is organizations using Kiteworks can track and control confidential information, such as trade secrets, customer data, PII, PHI, and financials, preventing it from being exposed to generative AI LLMs. Kiteworks delivers capabilities based on content policy risk:

• Low Risk: Content-defined Zero Trust. Use least-privilege access policies and controls for employees and third parties defined based on the sensitivity of content assets. Watermarking can be applied to alert users that specific content should not be used in generative AI LLMs.
• Moderate Risk: View-only DRM. Employ Kiteworks SafeView™ to disallow local copies of content so that employees and third parties cannot extract and upload the copy into a generative AI LLM. Additional policy can be set to expire access at a specified time period or passage of a specified number of views.
• High Risk: Next-generation DRM for Collaboration. Leverage next-generation DRM using Kiteworks SafeEdit™ to prevent data from leaving an organization’s network data center and repository while still empowering efficient collaboration with internal users and third parties. Kiteworks SafeEdit streams and transmits an editable video image to users.

“Kiteworks next-generation DRM capabilities enable businesses to track and control the sensitive content employees, contractors, and third parties ingest into generative AI LLMs,” said Freestone. “Our SafeView and SafeEdit capabilities let businesses track and control how end-users access and use confidential data, including what can and cannot be ingested in generative AI LLMs. With Kiteworks, businesses that have embraced generative AI LLMs can do so with confidence, while those that have banned their use can reevaluate their decision, knowing their sensitive data is protected.”

About Kiteworks
Kiteworks empowers organizations to effectively manage risk in every send, share, receive, and save of sensitive content. The Kiteworks platform provides customers with a Private Content Network that delivers content governance, compliance, and protection. The platform unifies, tracks, controls, and secures sensitive content moving within, into, and out of their organization, significantly improving risk management and ensuring regulatory compliance on all sensitive content communications. Kiteworks protects over 35 million end users for over 3,800 global enterprises and government agencies.

(1) Blackberry, “75% of Organizations Worldwide Set to Ban ChatGPT and Generative AI Apps on Work Devices,” Dark Reading, August 8, 2023, https://www.darkreading.com/endpoint/75-of-organizations-worldwide-set-to-ban-chatgpt-and-generative-ai-apps-on-work-devices.
(2) Stefanie Schappert, “Workers regularly post sensitive data into ChatGPT,” Cybernews, June 16, 2023, https://cybernews.com/security/workers-regularly-post-sensitive-data-into-chatgpt/.
(3) Ibid.
(4) Ibid.
(5) Elizabeth Montalbano, “Generative AI Projects Pose Major Cybersecurity Risk to Enterprises,” Dark Reading, June 28, 2023, https://www.darkreading.com/vulnerabilities-threats/generative-ai-projects-cybersecurity-risks-enterprises.
(6) Alex Singla, “The state of AI in 2023: Generative AI’s breakout year,” McKinsey, August 1, 2023, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.
(7) Ibid.

Patrick Spencer
Kiteworks
email us here
Visit us on social media:
Facebook
Twitter
LinkedIn
YouTube