NROC

Safely Boost Employee GenAI Usage

CISO Guide: Best Practices for Securing Public GenAI apps and LLMs in the Enterprise

Introduction

Generative AI applications (GenAI apps) and large language models (LLMs) are being rapidly embraced by employees across enterprises. A recent survey found that nearly 70% of enterprise employees are using publicly available apps like ChatGPT, Microsoft Copilot and Google Gemini to work faster, automate tasks and spark creativity.

However, this new technology challenges traditional security and compliance architectures. Unlike conventional software, Gen AI apps involve dynamic, unpredictable user interactions and data flows. The models can hallucinate, aggregate data in unexpected ways, and/or leak data. The providers have created a maze of data protection promises that leave many gaps in which the end user data may be used for training their AI models.

Every 100 prompts contain **32 instances of PII**.

Source: NROC Security user base

The below sections explain why traditional security models fall short for GenAI apps, the unique risks of enterprise AI use, and best practices for GenAI security.

Why Traditional Security Models Fall Short for GenAI apps

GenAI apps represent a **new breed of software where use cases are invented on the fly, user inputs are unpredictable, and any data can be used**. These are the challenges faced by legacy security tools:

- **Lack of visibility:** Secure web gateways or CASBs tend to block or allow all AI sites indiscriminately, without providing insights on how the organization uses GenAI. Most network monitoring cannot see the content of the prompts or responses. Most tools have no context of the unique attributes of GenAI apps, i.e. if the EULA gives the vendor rights to your data.

- **User-driven data exposure:** Unlike traditional apps, an LLM will accept virtually any prompt from the user, including proprietary data, personally identified information (PII), or source code. Conventional DLP systems looking for pre-labeled documents are unaware of end user copy/pasting or their regexes are noisy due to the volume of content. They often can only alert the SOC instead of guiding the end user.

- **Potentially inaccurate outputs:** LLMs often aggregate information inaccurately or otherwise hallucinate incorrect advice. This opens whole new questions: What AI created content can introduce more risk than benefit productivity? Where could an erroneous output slip into the organization with insufficient expert validation?

34% of AI-generated content included financial, legal or SW code topics that are most susceptible to hallucinations.

Source: NROC Security user base

- **LLM manipulation attacks vs. signature-based detections:** LLM's "threat" is in the content of its interactions, which isn't detectable by malware signatures or analysis of network traffic. For example, a crafty prompt could trick an AI chatbot into revealing data or performing an unsafe action. Input validations of traditional applications are largely absent with LLM-based apps.

- **Identity and access gaps:** Many enterprises have strict identity management (SSO, MFA) for corporate apps, but employees may use public GenAI apps and/or LLM services with personal credentials and accounts. At the same time, the data protection commitments of these public apps may only apply to users on corporate accounts – a loophole not accounted for in classic security architectures.

- **End-user engagement and friction:** The 'threat actor' here is more often a careless business end user than a malicious insider. AI is new and the service landscape is quite complex. Unsuitable tools by the security team tend to increase friction, which discourages the use of GenAI, or, at worst, causes the end users to use private laptops, where there is no visibility or protections.

Best Practices for Securing GenAI apps and LLMs in the Enterprise

To securely enable GenAI apps and LLMs used in the enterprise, CISO teams must go beyond the traditional playbook. Below are key best practices, aligned with the unique challenges discussed above:

- **Establish clear policies, training, and governance:** The foundation of GenAI and LLM security is a strong policy framework. Corporations need to define:

    - What employees can and cannot do with AI tools?
    - What data is off limits?
    - Which GenAI app and LLM services are approved?
    - What oversight is required for AI-generated output?

  Mostly importantly, involve your compliance, privacy, and legal teams early on to ensure these policies align with regulations. Existing laws, like GDPR, *do* apply to AI.

Once policies are in place, conduct training sessions to educate employees about the risks of careless AI use, such as data leaks or inaccuracies, and of the importance of adhering to guidelines. Finally, set up a governance team, like an AI committee or working group, to continuously oversee AI use cases, review incidents, and update policies as needed.

Make also use of the enterprise features offered by GenAI providers. Many of them offer a "no training" opt-out knob for business customers. Make sure this is stipulated in your policies but be realistic about the end users' ability to navigate the services, subscription types, or apply correct settings.

- **Gain visibility into GenAI app usage and data flows:** You can't secure what you can't see. Deeper insight than a list of accessed URLs is needed to answer the below questions:
    - What is the real extent of usage, by app, by org group? Any PII and classified content in prompts? What is created and how much of that is in 'specialized' content?
    - Is usage compliant with acceptable use policies? Where it is not?
    - What are the most common use cases – some that might warrant a specialist tool or showcasing internally?

  Solutions that process all GenAI web traffic can monitor AI interactions at the content level. This not only helps in risk detection but also in understanding the business value, essentially safely enabling most value adding use.

- **Protect sensitive data and prevent leakage:** Given the risks of sensitive prompts and dangerous outputs, it's essential to actively inspect all interactions. "Keep company sensitive content and PII away from any AI", should be enforced as an on-the-fly content redaction or stopped prompts. This pattern matching might benefit from integrating with your existing DLP dictionaries. Similarly, a chunk of source code in a prompt should be immediately flagged as a potential intellectual property leakage.

  A practical implementation is an "Are you sure?" pop-up alert, reminding the user of the corporate policy and requiring confirmation before continuing. Prompt guardrails should also look for signs of misuse. Any "ignore all your guidelines" or other manipulative prompts should be filtered out. The policy controls require sufficient granularity to differentiate between a paid subscription to a corporate-grade 'copilot' and a free consumer-oriented service such as Gemini.

- **Implement response guardrails:** Inspecting responses is equally important. The universally problematic output (hate speech, malware code) is often adequately filtered by protections in each GenAI app/LLM. You should limit what a particular tool is meant to be used for in your company. If an GenAI tool that is not approved for SW code creation outputs something that looks like software code, the result should be withheld from the user.

  AI-generated is susceptible for unspotted hallucinations. The risk is elevated for any content that requires expert validation, such as contractual/legal text or SW code. These guardrails can be soft. For

example, “Are you sure you want to use this?” may be enough that an end user ‘validates the output. Prompt/response rules should be tailored to different user groups and applications. That way one can direct the GenAI usage into use cases that maximize the value with minimal risk.

- **Enabling AI to work on your proprietary data (while keep some away from any AI):** Never expose more data than needed, but getting proprietary results from AI requires it to work on your proprietary data. Thus, the best practice mission reads: “enable the flow of the right data to the right AI”.

  Ideally, visibility and policy enforcement with AI interactions is based on the content of the document (e.g., resume, contract, invoice, transcript). Existing sensitivity labels often do not map to the AI use cases. An ideal solution can ‘learn’ what the important documents look like, spot them in real time and block/allow their flow. The best ones guide the users: “Let’s keep office documents in Copilot, shall we?”

  Start by understanding your use cases and identify what data is the right data. For real-time scanning, ask for modern similarity search-based algorithms to replace dictionaries or regexes.

- **Enforce authentication and access controls:** To maintain oversight and accountability, you need your employees to use GenAI apps with their corporate identity, not as anonymous internet users. The best way is to gate access to public GenAI apps and LLM services using your Single Sign-On (SSO). It is not possible to force end users to SSO into the GenAI app service, therefore authenticating them at a gateway is essential. Best practices block any users trying to access GenAI apps using private identities, when a corporate plan of the same is available.

  Tying usage to corporate identities has several benefits: it deters risky behavior, enables per-group policies and brings accurate incident attribution. With identities in place, you can also bring GenAI usage into your normal security operations (e.g., correlate AI events with user activity logs, apply user-based anomaly detection across GenAI services).

- **Monitor and log all interactions:** For policy enforcement and incident investigations, logging prompt and response content (in a privacy-conscious way) allows you to detect policy violations and/or unusual behavior. From a compliance perspective, the detailed records are becoming extremely useful and necessary. You may need to prove to customers that their data is safe. Regulators may ask how do you ensure AI decisions can be audited or reproduced.

  Monitoring is crucial for building trust in AI. Logs enable you to trace an AI generate content back to the prompt, service and user behind it. A best practice is to tag AI-generated content with some ‘serial number’ that ties back to logs. That way you can review AI artifacts for accuracy and close the learning loop back to prompting.

Conclusion

GenAI apps and LLMs are transformative technologies but they also come with new risks that are not addressed by traditional security architectures. Organizations need to blend new technical guardrails, policy governance, and user-centric measures to create a secure AI enablement framework.

The value of GenAI security is not only for reducing risk in the enterprise. When executed properly, **GenAI security produces evidence of compliance, builds trust in AI usage and provides strategy-defining insights into the best AI use cases in the organization. Best practice security for AI allows an organization empower employees to innovate with new tools and use cases, and thus, accelerate organizational learning and transformation.**

* * *

NROC Security was founded to help organizations safely boost employee GenAI usage by providing visibility, user authentication & governance, prompt and response guardrails, and data flow guards. Ready to get started? Schedule a demo for a customized consultation.