

Securing Your AI Pipeline – from Development to Production



How can you protect increasingly mission-critical workloads from new and evolving generative AI threats?

Gaining complete, end-to-end visibility into generative AI and traditional workloads has become almost impossible, but effective modern security requires a complete awareness of your environment.

Top generative AI security challenges

The Open Worldwide Application Security Project has identified six attacks unique to generative AI that go beyond infrastructure.

- 1 Prompt injection** manipulates LLMs with inputs. Attackers craft specific prompts to trick your models into ignoring their normal restrictions and generate content that your developers didn't intend.
- 2 Insecure output handling** occurs when LLM outputs expose your backend systems. These outputs could contain sensitive information or malicious content, such as malware code snippets or links that expose internal systems.
- 3 Training data poisoning** introduces vulnerabilities or biases. Attackers inject malicious or misleading data into your training dataset to introduce vulnerabilities or biases and influence outputs.
- 4 Sensitive information disclosure** occurs when responses unintentionally reveal your confidential data. Without proper safeguards, users could enter a prompt that returns data they aren't authorized to see.
- 5 Insecure LLM plugin design** comes from insufficient application controls. Most generative AI applications are lengthy and complex and can include foundation models, tools, libraries, frameworks, databases, and multiple APIs. When linked, any of these tech stacks could contain vulnerabilities that enter your application.
- 6 Excessive agency** results from excessive functionality, permissions, or autonomy granted to an LLM-based system. Controlling generative AI output while ensuring creativity and relevance is a complex balance.

AI security starts with control— but it doesn't end there

Beyond your data center confines, your security teams can no longer control what teams might deploy or where. Your workloads may be vulnerable due to overlooked security settings, misconfigurations, or unsecured applications, resulting in unexpected attack paths.

AI workloads require the same security controls as traditional workloads at the platform and application levels. These controls must span clouds, development platforms, storage, and virtual machines. You need visibility into all areas of your environment to avoid threats, including:

- Credential theft
- Data leakage
- Intellectual property theft
- Resource abuse
- Ransomware

An effective security solution provides complete visibility into cloud environments and workloads, including the specialized development platforms and data that generative AI requires. The solution should be able to build an AI capability inventory with a real-time view of services to monitor security health. This view must include misconfigurations, external exposure, sensitive data, and identity risks.

With this complete visibility, you can proactively uncover potential security issues by pinpointing vulnerabilities in the underlying resources of your AI application supply chain.

How to fully secure your generative AI workloads

You need a security solution with two key capabilities to prevent attacks.

First, the solution must provide visibility into every environment along your entire AI pipeline, including models, training data, and infrastructure. Second, it must include intelligence to detect vulnerabilities and prioritize high-value and sensitive assets. This intelligence comes from fully understanding your environment to expose vulnerabilities and recognize sensitive data, such as personally identifiable information (PII) or intellectual property (IP).



Key security capabilities must prioritize:

- **Complete visibility along the AI pipeline:** Protect against data leakage and poisoning. Reveal attack paths that expose your AI models to external organizations.
- **Data filtering:** Ensure you've correctly configured your web application firewalls. Filter and monitor HTTP traffic between applications and the internet.
- **Data loss prevention:** Don't allow unauthorized access to sensitive data.
- **Input guardrails:** Vet prompts before adding them to your LLMs so the output meets your standards, security policies, and privacy policies.
- **Output validation:** Validate your LLM output format, content, and structure. Verify they don't contain malicious code, links, or sensitive information.

Wiz protects AI workloads across multiple cloud environments and platforms, including Vertex AI or any of Google Cloud's environments — VMs, Kubernetes, or serverless — against attacks.


Innovate securely with Wiz on Google Cloud Vertex AI

Help your data scientists and engineers to focus on deploying more AI applications.

For example, Vertex AI is Google Cloud's fully managed, unified AI development platform that enables you to securely build, deploy, and scale generative AI models and applications. With Wiz and Google Cloud, you can gain complete visibility into all Vertex AI dependencies.

Wiz agentlessly provides an inventory of all Vertex AI dependencies, and the Wiz Security Graph provides a real-time visual summary of the services you're using. Identify misconfigurations, external exposure, sensitive data, identity risks on Vertex AI services, and vulnerabilities and secrets in the Google Compute Engine behind Vertex AI.

Wiz and Google Cloud deliver capabilities that help your team build and deploy AI applications quickly and securely. From Vertex AI to other clouds and platforms, Google Cloud and Wiz are here to support your AI journey.



Vertex AI is Google Cloud's fully managed, unified AI development platform that enables you to securely build, deploy, and scale generative AI models. With Wiz and Google Cloud, you can gain complete visibility into all Vertex AI dependencies.



[Learn what Wiz on Google Cloud can do for you.](#)

