8

oligo

Effective Ways to Secure GenAI Applications and Agentic Workflows

INTRODUCTION

Generative AI has the potential to fundamentally transform just about every industry. GenAI applications can help enterprises improve productivity, boost economics, and accelerate business agility. But they also introduce challenges for AppSec and SecOps teams.

Traditional software vulnerability scanning tools and runtime application security solutions, designed to protect conventional applications and services, don't adequately defend GenAI applications against modern threats. Adversaries can exploit inherent large language model (LLM) and vision-language model (VLM) vulnerabilities to disrupt business-critical systems, steal confidential data, or disseminate misinformation. Security leaders must implement new systems and practices to prevent attacks on GenAI applications.

This eBook reviews common LLM and VLM security concerns and offers actionable advice for protecting GenAI implementations. You will learn eight best practices for securing GenAI apps, safeguarding agentic workflows, and reducing risk, including:

BEST PRACTICE#1

Focus on Securing Live Production Workloads

The Challenge

Conventional "left-side" AppSec solutions like software composition analysis (SCA) tools are designed to identify vulnerable components within an application's codebase. They don't provide real-time, contextual visibility into dynamic GenAI system behavior and cannot detect anomalous activity symptomatic of a contemporary attack on a live GenAI application. In one notable example, Oligo Research discovered that thousands of organizations, including some of the world's largest companies, were unknowingly running vulnerable PyTorch TorchServe instances in production, putting them at serious risk. PyTorch is one of the most popular AI frameworks in the world.

Recommendation

The most significant risks associated with GenAI emerge when models are deployed in production environments and interact with real-world data and business-critical systems. **Take a "right-side," real-time approach to securing agentic AI workflows at runtime.** Use an Application Detection and Response (ADR) solution to monitor and control production GenAI agents, LLMs, and VLMs in real-time. ADR solutions are specifically designed to detect and respond to threats targeting applications. Best-of-breed ADR solutions provide deep visibility into application-layer activity, helping security teams identify anomalies, prevent exploitation, and mitigate attacks in real time.

Clever threat actors can manipulate GenAI prompts to maliciously influence a model's behavior. By monitoring a model's inputs and outputs at runtime, you can automatically identify irregular activity, block suspicious function calls, and stop attackers in their tracks.

BEST PRACTICE#2

Apply the Principle of Least Privilege to GenAI Agents

The Challenge

Adversaries can exploit over-permissioned LLM agents to gain access to SQL databases, enterprise knowledge graphs, data lakes, and other resources to exfiltrate data or carry out attacks. The Open Worldwide Application Security Project (OWASP) includes excessive agency on its 2025 Top Ten Most Critical LLM Application Vulnerabilities list (LLM06:2025).

Recommendation

Apply the principle of least privilege to GenAI just like you would to any other application or user. Use an Application Detection and Response solution to govern agent actions in real-time at the function-call level for ultimate security.
With an ADR solution, you can:

- → **Automatically** detect and block inappropriate function calls to stop attacks
- → **Ensure** GenAI agents execute only the essential functions required to perform their intended tasks
- → **Tightly** control access to sensitive data and critical systems
- → **Prevent** non-privileged agents from writing or deleting database records or issuing SQL commands

BEST PRACTICE#3

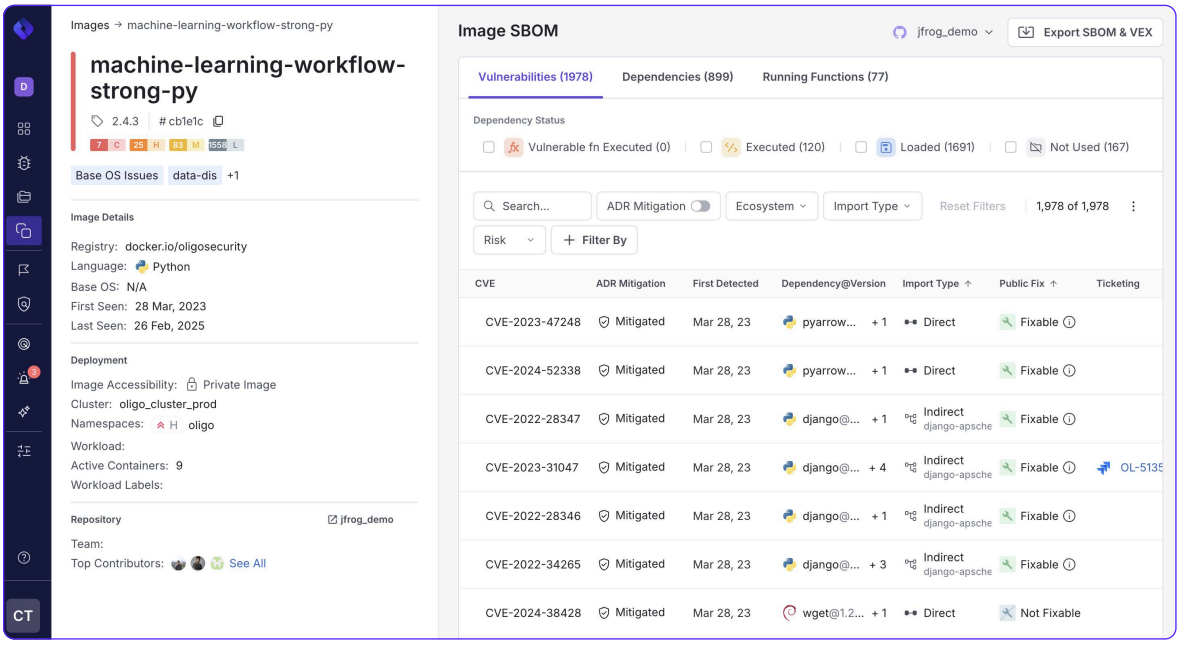
Identify Your Use of GenAI in First-Party Apps

The Challenge

Developers are adopting GenAI at a rapid pace, which makes it difficult for organizations to track its deployment and to identify and resolve GenAI system vulnerabilities. For example, a large enterprise may have several development teams working on different projects, each utilizing different GenAI models. These teams might be unaware of each other's implementations, leading to fragmented oversight and increased security risks. Vulnerabilities can easily go unnoticed as GenAI models are updated and added to the mix.

Recommendation

Use an ADR solution to definitively identify which GenAI models are being used, in which first-party apps, and how they're being used. Leading ADR solutions provide live AI bills of material (AI-BOMs) that let you easily determine which AI application libraries are running and executed in production, with contextual information.



ADR solutions include real-time AI and machine learning (ML) BOMs.

BEST PRACTICE#4

Validate and Control Model Output

The Challenge

Threat actors can use LLM-generated content (e.g., code, markdown, commands, SQL queries) for malicious purposes. Examples include:

- → Cross-site scripting (XSS) attacks or cross-site request forgery (CSRF) attacks to hijack browser sessions or steal data
- → SQL injection attacks to access, delete, or modify database records
- → Issuing shell commands to execute remote code on backend systems and inference servers to disrupt services or exfiltrate data
- → Server-side request forgery (SSRF) attacks to steal data or breach backend systems

In addition, GenAI hallucinations (unintended or erroneous outputs) can potentially disclose confidential data or introduce security vulnerabilities. OWASP includes insecure output handling—inadequate validation, management, and sanitization of LLM-generated content—on its 2025 Top Ten for LLM Applications list (LLM05:2025).

Recommendation

Use an ADR solution to inspect and control GenAI agent, LLM, and VLM outputs in real-time. Block questionable function calls to prevent attacks and mitigate adverse hallucination effects.

Carefully review code that connects the model with application business logic. If an application uses the model's output to make decisions, the behavior of the LLM or VLM will influence how the application operates. **Application security engineers should thoroughly review the code that handles the model output.**

BEST PRACTICE#5

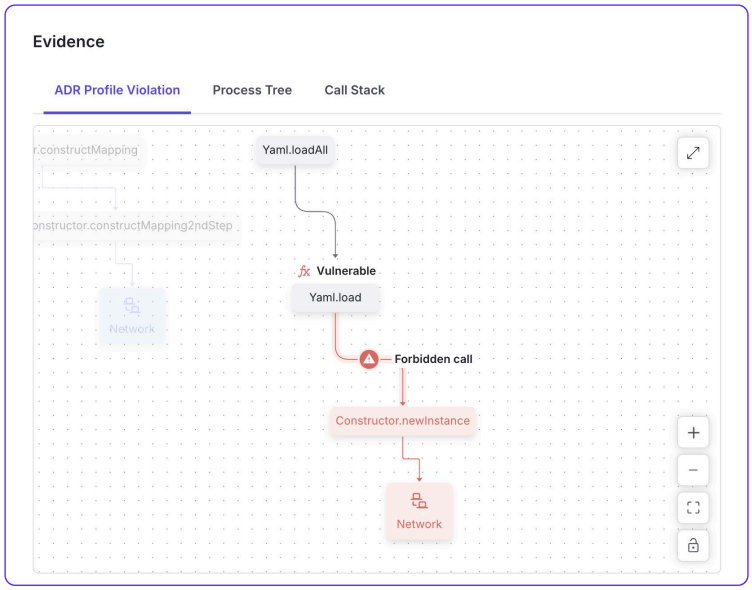
Reduce Prompt Injection Risks

The Challenge

Adversaries can manipulate user prompts, tricking GenAI systems into leaking sensitive data, spreading misinformation, or performing other unintended actions. Prompt injection attacks are relatively easy to wage because they do not require illicit access to GenAI system components. They are difficult to detect because they masquerade as legitimate user activity. OWASP includes prompt injection on its 2025 Top Ten for LLM Applications list (LLM01:2025).

Recommendation

Use an ADR solution to effectively identify unusual behavior and automatically block suspect function calls. Filter model inputs and outputs for additional protection. To be truly foolproof, take a "man-in-the-middle approach, using an LLM proxy like Meta Purple Llama to inspect and filter requests and responses in real time. The open-source Purple Llama project includes a free tool called Prompt Guard that you can use to filter out high-risk prompts. Prompt Guard provides out-of-the-box protection against a wide range of prompt injection attacks and is easily extensible.



Function-level call tracing that gives you detailed visibility into GenAI application behavior and agentic workflows.

BEST PRACTICE#6

Mitigate Shadow Vulnerabilities

The Challenge

Shadow vulnerabilities are security weaknesses that exist in software but are not documented in CVE databases. They are very common in open-source AI projects. Conventional scanning tools and runtime protection solutions (e.g., CWPP or DAST tools), which are designed to spot documented CVEs, can't detect shadow vulnerabilities lurking in GenAI applications. (And even once a CVE is discovered it can take weeks for the security community to qualify and document its behavior.)

Recommendation

Use an ADR solution to uncover and mitigate shadow vulnerabilities in production GenAI applications. Unlike conventional AppSec tools that require prior knowledge of a CVE or written rules established by the security community, contextual-aware ADR solutions work by automatically identifying and blocking anomalous application behavior indicative of an attack.

Learn more

Read how Oligo Research discovered ShadowRay, the first known attack campaign targeting AI workloads observed in the wild.

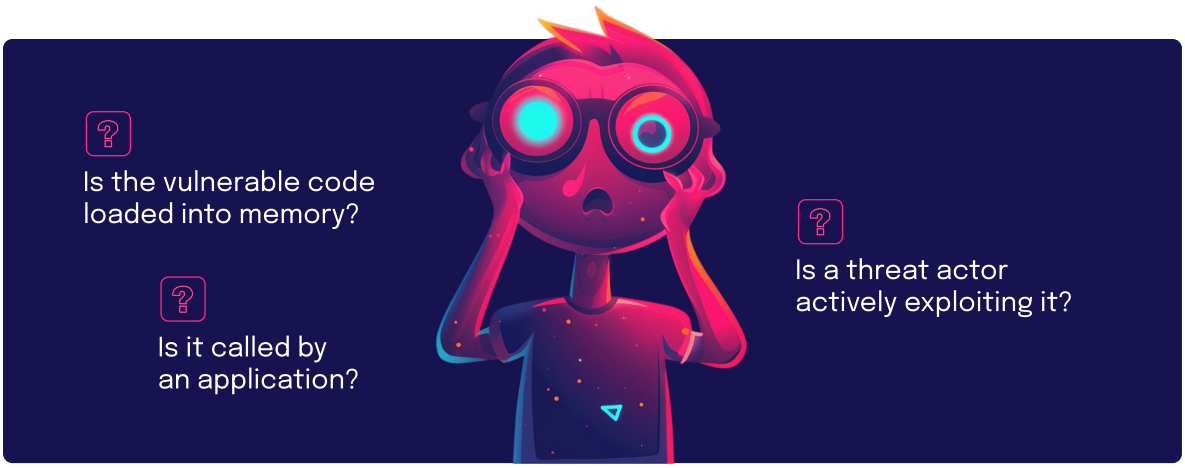
ShadowRay exploits a shadow vulnerability in Ray, a widely used open-source AI framework.

BEST PRACTICE#7

Protect Inference Servers at Runtime

The Challenge

Most popular inference servers rely on open-source technologies. Open-source software is laden with CVEs and shadow vulnerabilities and susceptible to supply-chain attacks and other application-level exploits. OWASP includes supply chain vulnerabilities on its 2025 Top Ten for LLM Applications list (LLM03:2025). Traditional "left-side" security solutions like SCA tools can identify vulnerabilities in open-source libraries, but they can't tell you if those vulnerabilities present a genuine risk to the business. Is the vulnerable code loaded into memory? Is it called by an application? Is a threat actor actively exploiting it?



Recommendation

Treat AI just like any other application in your supply chain. **Use an ADR solution to determine conclusively if vulnerable open-source code is running in memory or called by a GenAI application.** Defend inference servers against supply chain attacks and other threats by proactively blocking dubious actions at the library and function level.

BEST PRACTICE#8

Avoid Data Leakage

The Challenge

Organizations may unintentionally supply confidential data to LLMs and VLMs. Data leakage can damage your company's reputation, put users at risk, and lead to costly regulatory fines and legal settlements. OWASP includes Sensitive Information Disclosure on its list of the top ten most critical LLM application vulnerabilities (LLM02:2025).

Recommendation

Implement robust data input policies and controls. Ensure only authorized personnel can input data. Validate incoming data to be sure it doesn't contain confidential information. Remove, anonymize, or redact sensitive information.

Institute robust data output controls. Implement filtering to review and sanitize model output before it is shared externally. Continuously monitor and audit model outputs to ensure compliance with data privacy regulations like HIPAA and GDPR.

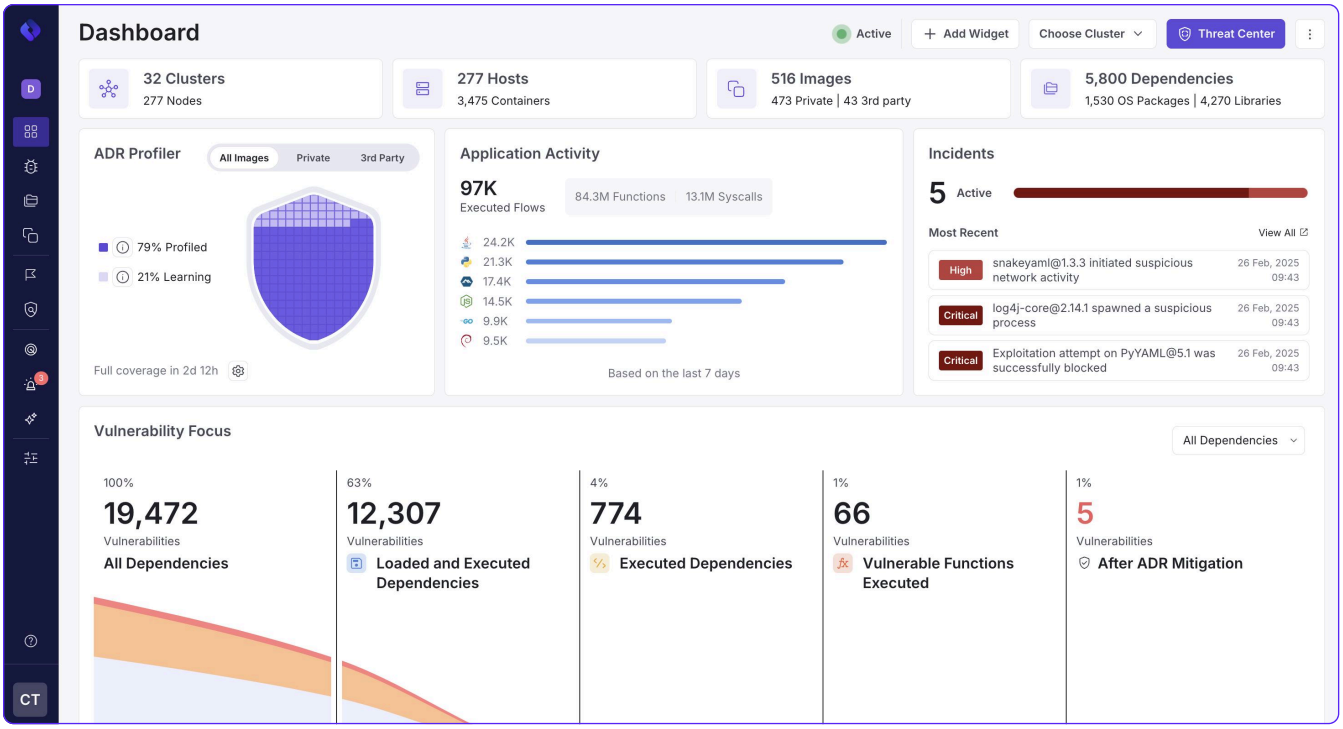
CONCLUSION

GenAI can help your company increase automation and accelerate the pace of business, but it also creates new opportunities for threat actors. Traditional application security solutions can't protect GenAI systems against today's sophisticated attackers.

GenAI requires a fresh approach to security. An approach that closely monitors and governs agent and model behavior in real-time to identify and thwart advanced attacks on GenAI systems. By following our eight best practices and using an ADR solution to inspect and control agentic AI workflows at runtime, you can strengthen your security posture, reduce risk, and make the most of your AI investments.

Protect GenAI Applications and Secure Agentic Workflows with Oligo

Oligo ADR makes it easy to detect and respond to in-progress attacks against any application including GenAI applications. Only Oligo uses innovative deep application inspection technology to observe and profile the runtime behavior of every dependency in every GenAI application and system component you build, buy, or use. Oligo automatically identifies and blocks suspicious activity symptomatic of an attack in real-time. It enables you to apply the principle of least privilege to GenAI agents and to strictly govern LLM and VLM output.



Oligo's intuitive browser-based GUI and comprehensive API let you instantly assess your GenAI application security posture, identify active exploits, and view key threat intelligence data.

Next Steps

To learn how Oligo can help your company secure GenAI applications and safeguard agentic workflows, get a personalized demo today: https://www.oligo.security/demo.

About Oligo

Oligo protects applications against attackers with the industry's leading Application Detection and Response platform. With deep application inspection through real-time monitoring and context-aware analysis, Oligo enables customers to instantly see all of the vulnerabilities in their environments, identify those that matter most, and stop application-based attacks in their tracks.

For more information, visit **www.oligo.security**