OWASP® | GenAI SECURITY PROJECT
TOP 10 FOR LLM AND GENERATIVE AI

Generative AI and Agentic Security Solutions Landscape

Q2/Q3 2025

Version 1.1

License and Usage

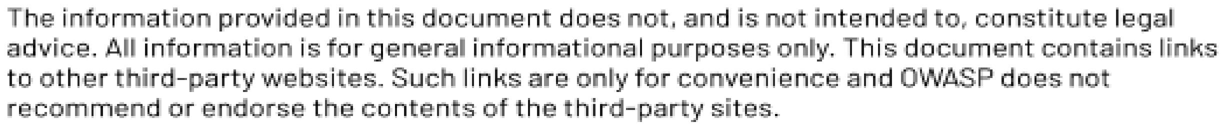
The information provided in this document does not, and is not intended to, constitute legal advice. All information is for general informational purposes only. This document contains links to other third-party websites. Such links are only for convenience, and OWASP does not recommend or endorse the contents of the third-party sites.

This document is licensed under Creative Commons, CC BY-SA 4.0

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.
- Under the following terms:
    - Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner but not in any way that suggests the licensor endorses you or your use.
    - Attribution Guidelines - must include the project name as well as the name of the asset Referenced
        - OWASP Top 10 for LLMs - LLMSecOps Solutions Landscape
        - OWASP Top 10 for LLMs - CyberSecurity Solution and LLMSecOps Landscape Guide
- ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Link to full license text: https://creativecommons.org/licenses/by-sa/4.0/legalcode



The information provided in this document does not, and is not intended to, constitute legal advice. All information is for general informational purposes only. This document contains links to other third-party websites. Such links are only for convenience and OWASP does not recommend or endorse the contents of the third-party sites.

Table of Content

Letter from the lead author

Why we created this companion resource

The creation of this document was initiated after we discussed as a core team that while the OWASP Top 10 List for LLMs and Generative AI List provided a great list of risks and potential mitigations, it fell short on providing the next level of guidance. This is in part due to the structure of what makes OWASP top 10 list so popular. This is being concise and focused to highlight the top risks and mitigation for a certain application space. There were more than enough candidates to go beyond 10, but the focus of 10 we felt essential to be able to ensure practical focused guidance. Deviating from the traditional OWASP Top 10 format would bloat the document and impact its concise listing.

Adopting a solutions approach for the project

While the Top 10 list for LLM and Gen AI provides the list Top 10 Risk and Mitigations, we felt it beneficial go further than traditional Top 10 Lists and to take a solutions approach and help connect the Top 10 Risks to the opens source and commercial security solutions organizations could look to to help address the Top 10 Risks for LLMs and Generative AI in a practical way.

In addition, since the Gen AI security landscape is moving so quickly, covering a range of new application types from static prompt augmentation, through RAG, plugins and Agentic Ai architectures, we saw a range of new security solutions emerging and wanted to be able to provide a regularly updated resource to identify the solution that could be used to address these new architectures and application risks highlighted in the Top 10 for LLM and Gen AI List.

Structuring the document

To organize the solutions, we chose to leverage and document the application types and the LLM/GenAI Ops and SecOps lifecycle and categories to provide an actionable way to both organize the solutions and map them to the Top 10 for LLM and Gen AI, which we would update quarterly. To accompany this document we also decided to publish an online directory. We hope this solution guide is helpful in implementing your own strategy for secure LLM and Gen AI adoption within your organization.

- **Scott Clinton**
  Co-Chair OWASP GenAI Security Project
  & AI, Security Solutions Initiative Lead

Who Is This Document For?

This document is tailored for a diverse audience comprising developers, AppSec professionals, DevSecOps and MLSecOps teams, data engineers, data scientists, CISOs, and security leaders who are focused on developing strategies to secure Large Language Models (LLMs) and Generative AI applications. It provides a reference guide of the solutions available to aid in securing LLM applications, equipping them with the knowledge and tools necessary to build robust, secure AI applications.

Objectives

This document is intended to be a companion to the OWASP Top 10 for Large Language Model (LLM) Applications List and the CISO Cybersecurity & Governance Checklist. Its primary objective is to provide a reference resource for organizations seeking to address the identified risks and enhance their security programs. While not designed to be an all-inclusive resource, this document offers a researched point of view based on the top security categories and emerging threat areas. It captures the most impactful existing and emerging categories. By categorizing, defining, and aligning applicable technology solution areas with the emerging LLM and generative AI threat landscape, this document aims to simplify research efforts and serve as a solutions reference guide.

Scope

The scope of this document is to create a shared definition of solution category areas that address the security of the LLM and generative AI life cycle, from development to deployment and usage. This alignment supports the OWASP Top 10 List For LLMs outcomes and the CISO Cybersecurity and Governance Checklist. To achieve this, the document will create an initial framework and category descriptors, utilizing both open-source solutions and providing mechanisms for solution providers to align their offerings with specific coverage areas as examples to support each category.

The document adheres to several key rules to maintain its integrity and usefulness:

- **Vendor-Agnostic and Open Approach:** It maintains a neutral stance, avoiding recommendations of one technology over another, instead providing category guidance with choices and options.
- **Straightforward, Actionable Guidance:** The document offers clear, actionable advice that organizations can readily implement.
- **Coordinated Knowledge Graph:** It includes coordinated terms, definitions, and descriptions for key concepts.
- **Point to Existing Standards**: Where existing standards or sources of truth are available, the document references these instead of creating new sources, ensuring consistency and reliability.

Introduction

With the growth of Generative AI adoption, usage, and application development comes new risks that affect how organizations strategize and invest. As these risks evolve, so do risk mitigation solutions, technologies, frameworks, and taxonomies. To aid security leaders in prioritization, conversations about emerging technology and solution areas must be aligned appropriately to clearly understood business outcomes for AI security solutions. The business outcomes of AI security solutions must be properly defined to aid security leaders in budgeting

Many organizations have already invested heavily in various security tools, such as vulnerability management systems, identity and access management (IAM) solutions, endpoint security, Dynamic Application Security Testing (DAST), observability platforms, and secure CI/CD (Continuous Integration/Continuous Deployment) tools, to name a few. However, these traditional security tools may not be sufficient to fully address the complexities of AI applications, leading to gaps in protection that malicious actors can exploit. For example, traditional security tools may not sufficiently address the unique data security and sensitive information disclosure protection in the context of LLM and Gen AI applications. This includes but is not limited to the challenges of securing sensitive data within prompts, outputs, and model training data, and the specific mitigation strategies such as encryption, redaction, and access control mechanisms.

Emergent solutions like LLM Firewalls, AI-specific threat detection systems, secure model deployment platforms, and AI governance frameworks attempt to address the unique security needs of AI/ML applications. However, the rapid evolution of AI/ML technology and its applications has driven an explosion of solution approaches, which has only added to the confusion faced by organizations in determining where to allocate their security budgets.

Defining the Security Solutions Landscape

There have been many approaches to characterizing the solutions landscape for Large Language Model tools and infrastructure. In order to develop a solutions landscape that focuses on the security of LLM applications across the lifecycle from planning, development, deployment, and operation, there are four key areas of input we have focused on to develop both a definition for Large Language Model DevSecOPs and related solutions landscape categories.

Landscape Considerations

Application Types and Scope - which impacts the people, processes, and tools needed based on the complexity of the application and the LLM environment, as-a-service, self-hosted, or custom-built.

Emerging LLMSecOps Process - while this is a work in progress, many are looking to adapt and adopt existing DevOps and MLOps and associated security practices. We expect our definition to evolve as the development processes for LLM applications begin to mature.

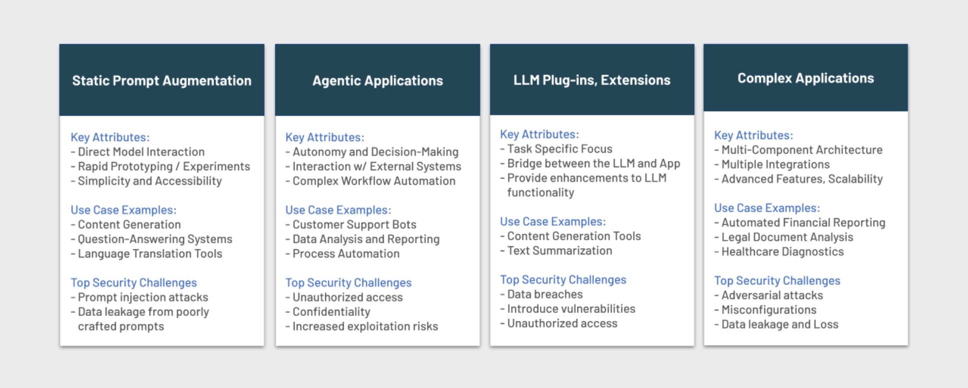
Threat and Risk Modeling - understanding the risks posed by LLM systems, application usage, or misuse like those outlined in the OWASP Top 10 for LLMs and Generative AI Applications, are key to understanding which solutions are best suited to improve the security posture and combat a range of attacks.

Tracking Emerging Solutions - many existing security solutions are adapting to support LLM development workflows and use cases however given the nature of new threats and evolving technology and architectures new types of LLM-specific security solutions will be necessary.

LLM Application Categories, Security Challenges

Organizations have been leveraging Machine Learning in applications for decades. This often required detailed expertise in Data Science and extensive model training. Generative AI has changed this. Specifically, Large Language Models (LLMs) have made machine learning technology widely accessible. The ability to dynamically interact in plain language has opened the door for the creation of a new class of data-driven applications and application integrations. Furthermore, usage is no longer limited to the highly skilled efforts of traditional developers and data scientists. Pre-trained models enable nearly anyone to perform complex computational tasks, regardless of prior exposure to programming or security. Organizations have been leveraging Machine Learning in applications for decades including Natural Language Processing (NLP) models that often require detailed expertise in Data Science and extensive model training.

With the advent of transformers technology enabling generative capabilities combined with the ease of access for pre-trained as-a-service models like ChatGPT and other as-a-service, Four major categories of LLM Application Architecture emerged; Prompt-centric, AI Agents, Plug-ins/extensions, and complex generative AI application where the LLM plays a key role in a larger application use case.



Static Prompt Augmentation	Agentic Applications	LLM Plug-ins, Extensions	Complex Applications
Key Attributes: - Direct Model Interaction - Rapid Prototyping / Experiments - Simplicity and Accessibility	Key Attributes: - Autonomy and Decision-Making - Interaction w/ External Systems - Complex Workflow Automation	Key Attributes: - Task Specific Focus - Bridge between the LLM and App - Provide enhancements to LLM functionality	Key Attributes: - Multi-Component Architecture - Multiple Integrations - Advanced Features, Scalability
Use Case Examples: - Content Generation - Question-Answering Systems - Language Translation Tools	Use Case Examples: - Customer Support Bots - Data Analysis and Reporting - Process Automation	Use Case Examples: - Content Generation Tools - Text Summarization	Use Case Examples: - Automated Financial Reporting - Legal Document Analysis - Healthcare Diagnostics
Top Security Challenges - Prompt injection attacks - Data leakage from poorly crafted prompts	Top Security Challenges - Unauthorized access - Confidentiality - Increased exploitation risks	Top Security Challenges - Data breaches - Introduce vulnerabilities - Unauthorized access	Top Security Challenges - Adversarial attacks - Misconfigurations - Data leakage and Loss

(figure: Application Categories & Summary Attributes)

Having a common view of typical LLM application architectures, including agents, models, LLMs, and the ML application stack, is crucial for defining and aligning the application stack, security model, and application offerings. Below, we have provided a short description of key characteristics, use cases, and security challenges for each application category.

Static Prompt Augmentation Applications

These applications involve specific static natural language inputs to guide the behavior of a large language model (LLM) toward generating the desired output. This technique optimizes the interaction between the user and the model by fine-tuning the phrasing, context, and instructions given to the LLM. These applications allow users to accomplish a wide range of tasks by simply refining how they ask questions or provide instructions.

Key Characteristics

- Human to model / model to human interaction and response
- Static prompt augmentation
- Flexibility and Creativity
- Simplicity and Accessibility
- Rapid Prototyping and Experimentation

Use Case Examples

- Experimentation/Rapid Prototyping
- Content Generation Tools
- Text Summarization Applications
- Question-Answering Systems
- Language Translation Tools
- Chatbots and Virtual Assistants

Security Challenges

- Prompt-based applications face security risks like prompt injection attacks and data leakage from poorly crafted prompts. Lack of context or state management can lead to unintended outputs, increasing misuse vulnerability. User-generated prompts may cause inconsistent or biased responses, risking compliance or ethical violations. Ensuring prompt integrity, robust input validation, and securing the LLM environment are crucial to mitigate these risks.

Agentic Applications

These applications leverage Large Language Models (LLMs) to autonomously or semi-autonomously perform tasks, make decisions, and interact with users or other systems. These agents are designed to act on behalf of users, handling complex processes that often involve multiple steps, integrations, and real-time decision-making. They operate with a level of autonomy, allowing them to complete tasks without constant human intervention.

Key Characteristics

- Autonomy and Decision-Making
- Interaction with External Systems
- State Management and Memory
- Complex Workflow Automation
- Agent to Agent, Agent to Environment communications
- Human-Agent Collaboration

Use Case Examples

- Virtual Assistants
- Customer Support Bots
- Process Automation Agents
- Data Analysis and Reporting Agents
- Intelligent Personalization Agents
- Coding and deep research agents
- Security and Compliance Agents

Security Challenges

- Agent applications, with their autonomy and access to various systems, must be carefully secured to prevent misuse. They face security challenges like unauthorized access, increased exploitation risks due to interaction with multiple systems, and vulnerabilities in decision-making processes. If someone gains control of an autonomous agent, the consequences could be severe, especially in critical systems. Ensuring robust access controls and encryption methods to protect against this is essential. Ensuring data integrity and confidentiality is critical, as agents often handle sensitive information it is important to secure data at all stages, including at -rest, in motion, and access through secured APIs. Their autonomy also poses risks of unintended or harmful decisions without oversight. Robust authentication, authorization, encryption, monitoring, and fail-safe mechanisms are essential to mitigate these security risks. Observability and Traceability solutions that monitor the entire lifecycle of the Agents (Design, Development, Deployment, and Visibility on decision-making) must be considered to ensure real-time corrections using a humans-in-the-loop process can be enforced.

Model Plug-ins, Extensions

Plug-ins are extensions or add-ons that integrate LLMs into existing applications or platforms, enabling them to provide enhanced or new functionalities. Plug-ins typically serve as a bridge between the LLM and the application, facilitating seamless integration, such as adding a language model to a word processor for grammar correction or integrating with customer relationship management (CRM) systems for automated email responses.

While it can be sometimes difficult to draw the line between Agents and plug-ins or extensions which are often components of larger applications, one measure is the way it is deployed and used. For example, a plug-in would be a pre-built agen designed for reuse that you call explicitly, through an API, or as part of an LLMs plugin or extension framework vs. custom code running in the background on a periodic basis.

Key Characteristics

- Modularity and Flexibility
- Seamless Integration
- Task Specific Focus
- Ease of Deployment and Use
- Rapid Updates and Maintenance

Use Case Examples

- Content Generation Tools
- Text Summarization Applications

Security Challenges

- Plugins interacting with sensitive data or critical systems must be carefully vetted for security vulnerabilities. Poorly designed or malicious plugins can cause data breaches or unauthorized access. LLM plugins face challenges like compatibility issues, where updates can introduce vulnerabilities, and integration with sensitive systems increases the risk of data leaks. Ensuring secure API interactions, regular updates, and robust access controls is crucial. Resource-intensive plugins may degrade performance, risking exploitation.

Complex Applications

Complex applications are sophisticated software systems that deeply integrate Large Language Models (LLMs) as a central component to provide advanced functionalities and solutions. These applications are characterized by their comprehensive scope, scalability, and the integration of multiple technologies and components. They are typically designed to solve intricate problems, often in enterprise environments, and require extensive development, engineering, and ongoing maintenance efforts.

Key Characteristics

- Multi-component architectures are designed to process prompts from other non-human systems.
- Often use multiple integrations, including other models.
- Multi-Component Architecture
- Scalability and Performance
- Advanced Features and Customization
- End-to-End Workflow Automation

Use Case Examples

- Legal Document Analysis Platforms
- Automated Financial Reporting Systems
- Customer Service Platforms
- Healthcare Diagnostics

Security Challenges

- Complex LLM applications face major security challenges due to their integration with multiple systems and extensive data handling. These include API vulnerabilities, data breaches, and adversarial attacks. The complexity increases the risk of misconfigurations, leading to unauthorized access or data leaks. Managing compliance across components is also difficult. Robust encryption, access controls, regular security audits, and comprehensive monitoring are essential to protect these applications from sophisticated threats and ensure data security.

Model (LLM, etc) Development and Consumption Models

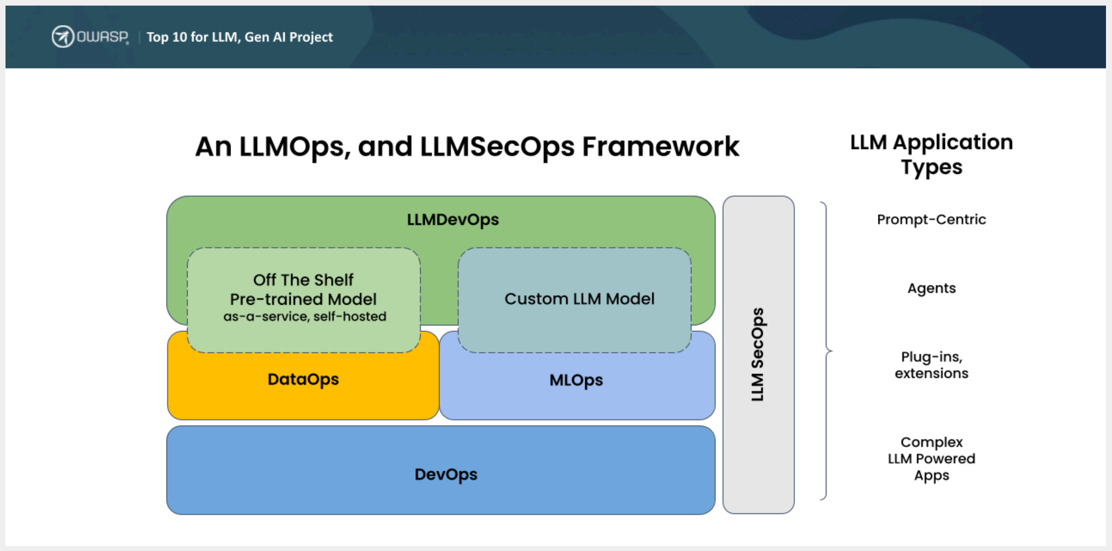
One of the first considerations for an organization is deciding upon the approach to leveraging LLM capabilities based on the type of application and goals for the project. Today, developers have a choice of two primary deployment models when implementing LLM and generative model-based applications and systems.

Create a New Model: The training process for custom LLMs is intensive, often involving domain-specific datasets and extensive fine-tuning to achieve desired performance levels. This approach is more akin to MLOps building ML models from the ground up, with detailed data analysis, collection formatting, cleaning, and labeling. One of the benefits of this approach is that you know the lineage and source of the data the model is built on and can attest directly to its validity and fit. However, a major downside is the resources, cost, and expertise necessary to build, train, and verify a model that meets the project objectives. Custom LLMs provide tailored solutions optimized for specific tasks and domains, offering higher accuracy and alignment with an organization's specific needs.

Consume and Customize Existing Models: Pre-trained (foundation) models, whether self-hosted or offered as a service, such as with ChatGPT, Bert and others on the other hand provide a more accessible entry point for organizations. These models can be quickly deployed via APIs, allowing for rapid solution validation and integration into existing systems. The LLMOps process in this scenario emphasizes customization through fine-tuning with specific datasets, ensuring the model meets the application's unique requirements, followed by robust deployment and monitoring to maintain performance and security.

GenAIOps and GenAISecOps Defined

Having a common view of typical LLM application architectures, including agents, models, LLMs, and the ML application stack, is crucial for defining and aligning the application stack and security model.



OWASP | Top 10 for LLM, Gen AI Project

An LLMOps, and LLMSecOps Framework

(figure: LLMOps related Operations Process for Data, Machine Learning and DevOps)

A Quick Ops Primer - Foundation for LLMOps

DevOps, which emphasizes collaboration, automation, and continuous integration and deployment (CI/CD), has laid the groundwork for efficient software development and operations. By streamlining the software development lifecycle, DevOps enables rapid and reliable delivery of applications, fostering a culture of collaboration between development and operations teams.

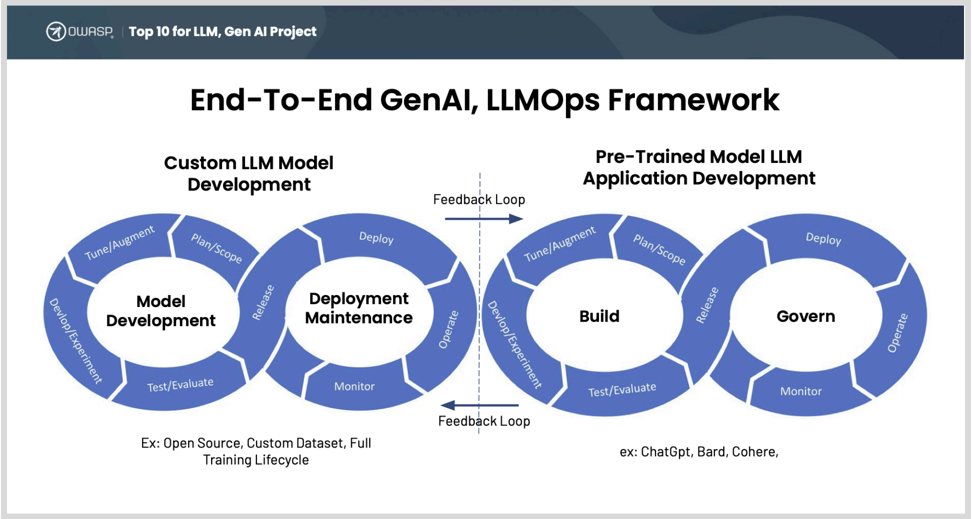
DataOps builds on DevOps, where data pipelines are managed with similar automation, version control, and continuous monitoring, ensuring data quality and compliance across the data lifecycle. MLOps also extends the DevOps principles to machine learning, focusing on the unique challenges of model development, training, deployment, and monitoring. Utilizing DevOps as a foundation ensures that both DataOps and MLOps inherit a robust infrastructure that prioritizes efficiency, scalability, security, and faster innovation in data-driven and machine learning applications.

MLOps and DataOps are foundational to LLMOps because they establish the critical processes and infrastructure needed for managing the lifecycle of large language models (LLMs). DataOps ensures that data pipelines are efficiently managed, from data collection and preparation to storage and retrieval, providing high-quality, consistent, and secure data that LLMs rely on for training and inference. MLOps extends these principles by automating and orchestrating the machine learning lifecycle, including model development, training, deployment, and monitoring.

LLMOps and MLOps, while rooted in the same foundational principles of lifecycle management, diverge significantly in their focus and requirements due to the specific demands of large language models (LLMs). LLMOps encompasses the complexities of training, deploying, and managing LLMs, which require substantial computational resources and sophisticated handling. LLMOps ensure that LLMs are efficiently integrated into production environments, monitored for performance and biases, and updated as needed to maintain their effectiveness. This holistic approach ensures that the deployment and operation of LLMs are streamlined, scalable, and secure, including considerations for data validation and provenance to ensure that the data used for training and fine-tuning LLMs is trustworthy and free from tampering. This can include techniques for data auditing and verification.

LLMOps Life Cycle Stages - Foundation for LLMDevSecOps

As mentioned earlier in this document, to align security solutions for LLM applications for our solution guide we are using the LLMOps process to define the solution categories so that they align with the challenges developers are facing in developing and deploying LLM-based applications.



OWASP | Top 10 for LLM, Gen AI Project

End-To-End GenAI, LLMOps Framework

Custom LLM Model Development

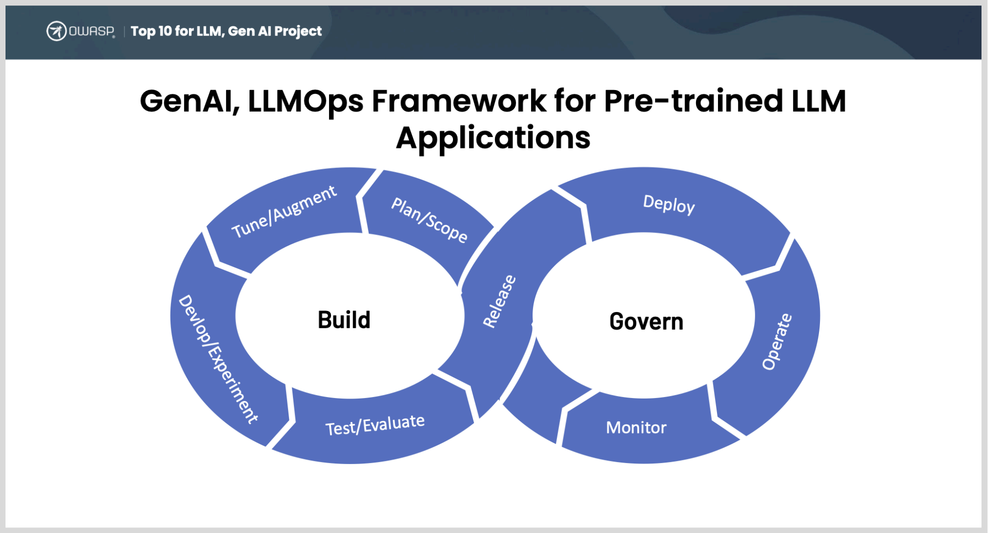
Pre-Trained Model LLM Application Development

Ex: Open Source, Custom Dataset, Full Training Lifecycle

ex: ChatGpt, Bard, Cohere,

(figure: Combined LLM Custom and LLM Pre-Trained Image)

The LLMOps processes differ significantly between using pre-trained LLM models for application development and creating custom LLM models from scratch using open-source and custom datasets, which inherit more from MLOps practices with some additions. We first need to define the stages, the typical developer tasks, and the security steps at each stage of the life cycle.



OWASP | Top 10 for LLM, Gen AI Project

GenAI, LLMOps Framework for Pre-trained LLM Applications

Build

Tune/Augment

Plan/Scope

Develop/Experiment

Test/Evaluate

Release

Govern

Deploy

Operate

Monitor

(figure: LLMops Pre-Trained Process and Steps)

These phases we have defined include: Scope/Plan, Model Fine-Tuning/Data Augmentation, Test/Evaluate, Release, Deploy, Operate, Monitor, and Govern. Of course, this is an iterative approach, whether you are practicing waterfall, agile, or a hybrid approach each of these steps can be leveraged.

Scoping/Planning

The focus is on defining the application's goals, understanding the specific needs the LLM will address, and determining how the pre-trained model will be integrated into the larger system. This stage involves gathering requirements, assessing potential ethical and compliance considerations, and setting clear objectives for performance, scalability, and user interaction. The outcome is a detailed project plan that outlines the scope, resources, and timelines needed to implement the LLM-powered application successfully.

Typical Activities:

LLMOps	LLMSecOps
● Data Suitability ● Model Selection ● Requirements Gathering (business, technical, and data) ● Task Identification ● Task Suitability	● Access Control and Authentication Planning ● Compliance and Regulatory Assessment ● Data Privacy and Protection Strategy ● Early Identification of Sensitive Data ● Third-Party Risk Assessment (Model, Provider, etc.) ● Threat Modeling

Data Augmentation and Fine-Tuning

The focus is on customizing the pre-trained model to better suit the specific application needs. This involves augmenting the original dataset with additional domain-specific data, enhancing the model's ability to generate accurate and contextually relevant responses. Fine-tuning is then conducted by retraining the LLM on this enriched dataset, optimizing its performance for the intended use case. This stage is critical for ensuring that the LLM adapts effectively to the unique challenges of the target domain, improving both accuracy and user experience with fewer instances of hallucination.

Typical Activities:

LLMOps	LLMSecOps
● Data Integration ● Retrieval Augmented Generation (RAG) ● Fine Tuning ● In-context Learning and Embeddings ● Reinforcement Learning with Human Feedback	● Data Source Validation ● Secure Data Handling ● Secure Data Pipeline ● Secure vector database ● Secure Output Handling ● Adversarial Robustness Testing ● Model Integrity Validation (ex: serialization scanning for malware) ● Vulnerability Assessment

Application Development and Experimentation

The focus shifts to integrating the fine-tuned model into the application's architecture. This stage involves building the necessary interfaces, user interactions, and workflows that leverage the LLM's capabilities. Developers experiment with different configurations, testing the model's performance within the application and refining the integration based on user feedback and real-world scenarios. This iterative process is crucial for optimizing the user experience and ensuring the LLM functions effectively within the broader application context.

Typical Activities:

LLMOps	LLMSecOps
● Agent Development ● Experimentation, Iteration ● Prompt Engineering	● Access, Authentication, and Authorization (MFA) ● Experiment Tracking ● LLM & App Vulnerability Scanning ● Model and Application Interaction Security ● SAST/DAST/ IAST ● Secure Coding Practices ● Secure Library/Code Repository ● Software Composition Analysis

Test and Evaluation

At this stage in the LLM SDLC and Ops process, the focus is on rigorously assessing the application's performance, security, and reliability. This stage involves conducting comprehensive testing, including functional, security, and usability tests, to ensure the LLM integrates seamlessly with the application and meets all defined requirements. Evaluation metrics are used to measure the model's accuracy, response times, and user interactions, allowing for fine-tuning and adjustments. This phase is crucial for identifying and resolving any issues before the application is deployed to production, ensuring it operates effectively and securely in real-world environments.

Typical Activities:

LLMOps	LLMSecOps
● Evaluate the model on validation and test datasets. ● Integration Testing ● Perform bias and fairness checks. ● Stress / Performance Testing ● Use cross-validation and other techniques to ensure robustness. ● Validate the model's interpretability and explainability.	● Adversarial Testing ● Application Security Orchestration and Correlation ● Bias and Fairness Testing ● Final Security Audit ● Incident Simulation, Response Testing ● LLM Benchmarking ● Penetration Testing ● SAST/DAST/IAST ● Vulnerability Scanning ● Available Agent Scanning

Release

The focus shifts to deploying the finalized application to the production environment. This stage involves finalizing the deployment strategy, configuring the infrastructure for scalability and security, and ensuring that all components, including the LLM, are integrated and functioning as intended. Critical tasks include setting up monitoring and alerting systems, conducting a final security review, and preparing for user onboarding. The goal is to ensure a smooth and secure transition from development to production, making the application available to users with minimal risk and downtime.

Typical Activities:

LLMOps	LLMSecOps
● Enable continuous delivery of model updates ● Integrate security checks and automated testing in the pipeline. ● Package the model for deployment (e.g., using Docker, Kubernetes). ● Set up CI/CD pipelines to automate application and model training, testing, and deployment.	● AI/ML Bill of Materials (BOM) ● Digital Model\Dataset Signing ● Model Security Posture Evaluation ● Secure CI/CD pipeline ● Secure Supply Chain Verification ● Static and Dynamic Code Analysis ● User Access Control Validation ● Model Serialization Defenses

Deploy

The focus is on securely launching the LLM and its associated components into the production environment. This stage involves configuring the deployment infrastructure for scalability and reliability, ensuring that all security measures are in place, and validating the integration of the LLM with other application components. Key activities include setting up real-time monitoring, conducting final checks to prevent any vulnerabilities, and implementing fallback mechanisms to ensure continuous operation. The goal is to smoothly transition from development to live operation, ensuring that the application is ready to handle real-world usage.

Typical Activities:

LLMOps	LLMSecOps
● Infrastructure Setup ● Integrate with existing systems or applications. ● Model and App Deployment ● Set up APIs or services for access ● User access and role management ● Agent Permission and Ownership Control ● Agentic Registry	● Compliance Verification ● Deployment Validation ● Digital Model\Dataset Signing Verification ● Encryption, Secrets management ● LLM Enabled Web Application Firewall ● Multi-factor Authentication ● Network Security Validation ● Secrets Management ● Secure API Access ● Secure Configuration ● User and Data Privacy Protections

Operate

The focus at this stage in the LLM SDLC and Ops process is on managing and maintaining the application in a live production environment. This stage involves continuous monitoring of the application's performance, security, and user interactions to ensure it operates smoothly and securely. Key activities include responding to incidents, applying updates or patches, and refining the model based on real-world data and feedback. The goal is to maintain high availability, optimize performance, and ensure the application remains secure and effective over time.

Typical Activities:

LLMOps	LLMSecOps
● Feedback Collection ● Iterative Enhancements ● Model Maintenance ● Performance Management ● Scalability and Infrastructure Management ● User Support and Issue Resolution	● Adversarial Attack Protection ● Automated Vulnerability Scanning ● Data Integrity and Encryption ● LLM Guardrails ● LLM Incident Detection and Response ● Patch Management ● Privacy, Data Leakage Protection ● Prompt Security ● Runtime Application Self-Protection ● Secure Output Handling ● Anomaly Detection in Agent Chains ● Runtime Agent Policy Validation

Monitor

The focus at this stage is on continuously observing the application's performance, security, and user interactions in real-time. This stage involves tracking key metrics, detecting anomalies, and ensuring the LLM model and application components are functioning as expected. Monitoring also includes gathering data for ongoing improvement, identifying potential issues before they impact users, and maintaining compliance with security and operational standards. The goal is to ensure the application remains stable, secure, and efficient throughout its lifecycle.

Typical Activities:

LLMOps	LLMSecOps
● Automate retraining processes based on new data. ● Detect and respond to model drift or degradation. ● Manage model versioning and rollback if necessary ● Monitor model performance (e.g., latency, accuracy, user interactions).	● Adversarial Input Detection ● Model Behavior Analysis ● AI/LLM Secure Posture Management ● Patch and Update Alerts ● Regulatory Compliance Tracking ● Security Alerting ● Security Metrics Collection ● User Activity Monitoring ● Agents Activity Monitoring ● Observability ● Data Privacy and Protection ● Ethical Compliance

Govern

At this stage in the LLMOps process, the focus is on establishing and enforcing policies, standards, and best practices to ensure the application operates securely and ethically throughout its lifecycle. This stage involves setting governance frameworks that oversee data usage, model management, compliance, and security controls. Key activities include auditing, risk management, and ensuring the application adheres to regulatory requirements and organizational policies.

Typical Activities:

LLMOps	LLMSecOps
● Conduct regular audits for compliance (e.g., GDPR, CCPA). ● Data Governance ● Document model decisions, datasets used, and model versions. ● Implement model governance frameworks.	● Bias and Fairness Oversight ● Compliance Management ● Data Security Posture Management ● Incident Governance ● Risk Assessment and Management ● User/Machine Access audits ● Agent Action Audit

Agentic AI Application Context for GenAI SecOps

Why an Agentic AI Application Context for SecOps?

As GenAI systems evolve from single-turn LLM calls to fully agentic architectures where multiple autonomous agents negotiate tasks via protocols such as A2A and invoke external services through MCP plug-in layers the classic DevOps and SecOps playbooks must likewise mature.

Agentic AI introduces a new application layer of autonomous agents, but it's built on the same foundational stack—data pipelines, model training, evaluation, and serving. These layers are interdependent: you can't secure agents without securing the models and infrastructure beneath them. As with web apps evolving from static pages, security must now expand to cover agent behaviors, inter-agent trust, and tool invocation while maintaining traditional AI/ML safeguards.

Adjusting our DevOps and SecOps frameworks to recognise agent patterns, reasoning graphs, and protocol governance lets teams preserve the reliability, compliance, and auditability hard-won in traditional AI stacks while confidently layering on the complex interdependencies that power next-generation autonomous applications. Below we've leveraged the same GenAIOps/ SecOps Framework and employed the Agentic AI Context to help organizations build an integrated approach to AI and Gen AI security

Scope/Plan - Agentic Apps

Agentic DevOps	Agentic SecOps
● Define the business goal and translate into agent goals & roles ● Choose model families (chat-LLM vs. multimodal) & hosting mode. ● Define agent architecture patterns (single, hierarchical, swarm) ● Identify external services and tooling ● Design inter-agent communication and tool workflows ● Select memory pattern (short-term context vs long-term e.g. vector DB). ● Create initial threat model and Service Level Objectives.	● Conduct agentic threat modeling (referencing the threat modeling approach from the GenAI Security Project - Agentic Security Initiative) ● Identify system-wide non-human identities (NHIs) and determine authentication protocols (e.g., SPIFFE, mTLS). ● Draft policies for agent privilege boundaries, tool scopes (e.g., MCP), and delegation logic. ● Define controls for memory scoping, isolation, and long-term persistence rules.

Data Augmentation & Fine-Tuning - Agentic Apps

Agentic DevOps	Agentic SecOps
● Collect domain-specific corpora that agents will reference during planning & reflection. ● Generate tool-schema embeddings so planners can choose the right action. ● Fine-tune/refine LLM on task-specific dialogues that include multi-step reasoning traces (ReAct, Tree-of-Thought). ● Populate seed "agent memory" (company knowledge, rules).	● Scan datasets for prompt-poisoning, biased instructions, or encoded policy bypasses. ● Validate RLHF traces for ethical alignment, adversarial manipulation, or leakage of secrets. ● Register data lineage and provenance in immutable logs. ● Apply differential privacy or obfuscation on sensitive knowledge injected into agent memory. ● Agent Action Audit

Development & Experimentation - Agentic Apps

Agentic DevOps	Agentic SecOps
● Implement agent loops (Observe-Plan-Act-Reflect) with frameworks such as LangGraph / AutoGen. ● Build manager-worker graphs; encode delegation policies. ● Wire plugins for each external API (e.g., MCP) and enforce input/output schemas. ● Prototype interagent protocol (e.g. A2A) handshake and capability negotiation. ● Iterate on prompts, system instructions, and guard-functions; run sandbox tests.	● Perform SAST/DAST on agent planning code, tool wrappers, and plugin interfaces. ● Harden agent loop logic against infinite loops, unsafe function routing, and unauthorized self-modification. ● Validate connector (e.g., MCP) contracts (input/output schemas and permissions). ● Implement policy enforcement hooks in Frameworks ○ e.g. LangGraph, CrewAI, or Semantic Kernel flows.

Test & Evaluation - Agentic Apps

Agentic DevOps	Agentic SecOps
● Spin up synthetic multi-agent arenas to stress-test negotiation, bidding and consensus flows. ● Run goal-drift, prompt-injection, and resource-exhaustion scenarios against the planner. ● Benchmark reflection latency and memory-poisoning resilience. ● Validate generated tool calls in a sandbox for RCE / over-scope.	● Available Agent Scanning ● Conduct adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Run multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls—particularly code execution or cloud API triggers.

Release - Agentic Apps

Agentic DevOps	Agentic SecOps
● Package agent graphs, plugins, policies, and memory snapshots ● Generate Model & Tool SBOMs; sign artefacts (Sigstore). - shared responsibility ● Publish agent capability-cards to an internal A2A registry.	● Generate and verify model + agent + tool SBOMs - shared responsibility ● Sign model weights, plugin manifests, and memory snapshots. ● Ensure policy bundles (e.g., OPA/Rego) are cryptographically validated at deploy time. ● Register all agents in an internal trust registry with capability descriptors.

Deploy - Agentic Apps

Agentic DevOps	Agentic SecOps
● Provision vector DB, memory store, tool side-cars, and service-mesh with mTLS for A2A traffic. ● Apply least-privilege IAM roles to every agent (non-human identities). ● Load initial long-term memory and register agents with discovery service. ● Enable runtime guardrails / LLM firewall	● Enforce zero-trust policies between agents, tools, and external APIs via mTLS and fine-grained RBAC. ● Rotate all shared secrets, keys, and tokens with ephemeral, scoped credentials. ● Apply runtime guardrails (e.g., LLM firewalls, tool allowlists) before production traffic is enabled. ● Configure inter-agent authorization policies based on capabilities and roles

Operate - Agentic Apps

Agentic DevOps	Agentic SecOps
● Run SRE playbooks: auto-scale inference pods, rotate keys/tokens, prune memory. ● Collect feedback / RLHF traces; schedule periodic self-evaluation tasks. ● Trigger automated reflection or human-in-the-loop when agent confidence drops. ● - Orchestrate inter-agent workflows.	● Monitor agent memory mutation patterns for drift, poisoning, or unauthorized overwrites. ● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions. ● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● Runtime guardrails & moderation; anomalous tool use.

Monitor - Agentic Apps

Agentic DevOps	Agentic SecOps
● Stream agent-step telemetry via OpenTelemetry; correlate tool errors with planning nodes. ● Track KPIs: goal-completion rate, average reasoning depth, vector-store growth, inter-agent latency. ● Alert on anomaly patterns (looping, hallucination cascades, excessive privilege use)..	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies like goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.

Govern - Agentic Apps

Agentic DevOps	LLMSecOps
● Maintain registry of agent versions, roles, and approved tools; enforce retirement policy. ● Run quarterly attestation of A2A trust graph and MCP connector scopes. ● Archive immutable logs for audit; map evidence to EU AI Act / NIST RMF controls. ● Periodically review alignment metrics and update constitutional rules.	● Enforce role- and task-based access policies across agent populations and their tool access. ● Automate agent versioning, expiration, and rotation policies. ● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. ● Automate goal alignment audits, including adversarial review of long-term agent memory.

Red Teaming Solutions for GenAI Systems and Applications

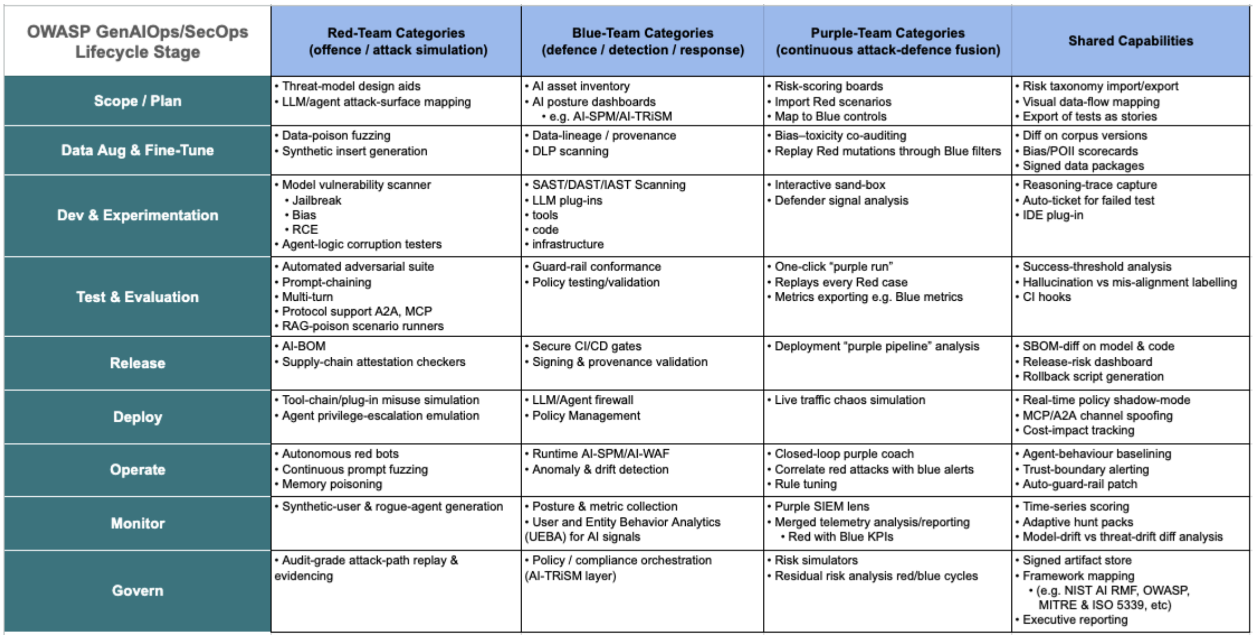
Red Teaming Solution Framework for Generative AI

As GenAI systems adopt layered architectures—with foundation models at the core and higher-order patterns like RAG and Agentic AI layered above—red teaming must evolve accordingly. The OWASP LLM & GenAI Security Solutions Guide emphasizes a full lifecycle approach, where security testing spans planning, development, deployment, and ongoing operation.

Traditional red teaming has focused on static LLM behavior (e.g., jailbreaks or offensive outputs), but Agentic AI introduces a new application layer with dynamic, autonomous behavior: agents that plan, delegate, access tools (via MCP), and collaborate (via A2A). These agent-driven workflows bring new threat surfaces—goal hijacking, tool misuse, memory poisoning, inter-agent deception—that cannot be fully tested with prompt injection alone.

To address this, red teaming must be embedded across the GenAI lifecycle. During design, teams should simulate misuse of agent goals and reasoning paths. During development, they must test for insecure plugin integration and unsafe memory flows. This evolution also means red teaming tools need to evolve as well including capabilities that include reasoning-step tracing, agent orchestration simulation, plugin misuse emulation, and memory poisoning injection

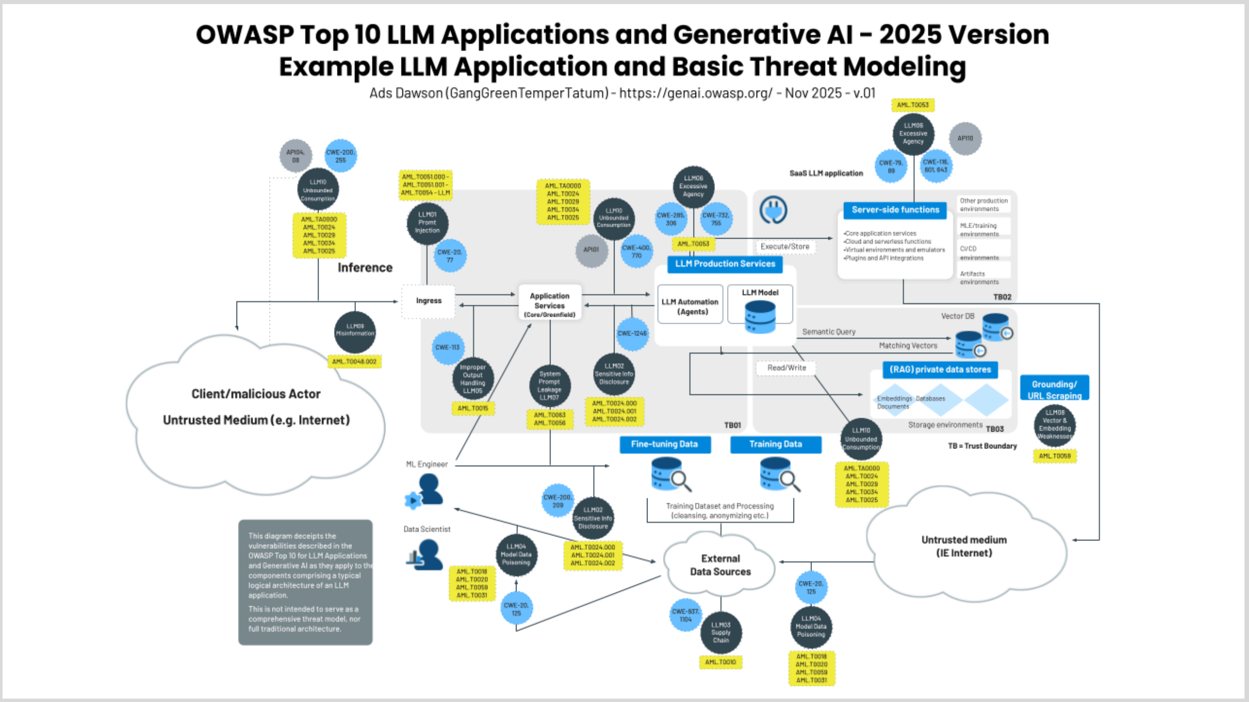
Below we have aligned the new capabilities required for Red Teaming solutions—aligned with OWASP's



OWASP GenAIOps/SecOps Lifecycle Stage	Red-Team Categories (offence / attack simulation)	Blue-Team Categories (defence / detection / response)	Purple-Team Categories (continuous attack-defence fusion)	Shared Capabilities
Scope / Plan	• Threat-model design aids • LLM/agent attack-surface mapping	• AI asset inventory • AI posture dashboards • e.g. AI-SPM/AI-TRiSM	• Risk-scoring boards • Import Red scenarios • Map to Blue controls	• Risk taxonomy import/export • Visual data-flow mapping • Export of tests as stories
Data Aug & Fine-Tune	• Data-poison fuzzing • Synthetic insert generation	• Data-lineage / provenance • DLP scanning	• Bias–toxicity co-auditing • Replay Red mutations through Blue filters	• Diff on corpus versions • Bias/POII scorecards • Signed data packages
Dev & Experimentation	• Model vulnerability scanner • Jailbreak • Bias • RCE • Agent-logic corruption testers	• SAST/DAST/IAST Scanning • LLM plug-ins • tools • code • infrastructure	• Interactive sand-box • Defender signal analysis	• Reasoning-trace capture • Auto-ticket for failed test • IDE plug-in
Test & Evaluation	• Automated adversarial suite • Prompt-chaining • Multi-turn • Protocol support A2A, MCP • RAG-poison scenario runners	• Guard-rail conformance • Policy testing/validation	• One-click "purple run" • Replays every Red case • Metrics exporting e.g. Blue metrics	• Success-threshold analysis • Hallucination vs mis-alignment labelling • CI hooks
Release	• AI-BOM • Supply-chain attestation checkers	• Secure CI/CD gates • Signing & provenance validation	• Deployment "purple pipeline" analysis	• SBOM-diff on model & code • Release-risk dashboard • Rollback script generation
Deploy	• Tool-chain/plug-in misuse simulation • Agent privilege-escalation emulation	• LLM/Agent firewall • Policy Management	• Live traffic chaos simulation	• Real-time policy shadow-mode • MCP/A2A channel spoofing • Cost-impact tracking
Operate	• Autonomous red bots • Continuous prompt fuzzing • Memory poisoning	• Runtime AI-SPM/AI-WAF • Anomaly & drift detection	• Closed-loop purple coach • Correlate red attacks with blue alerts • Rule tuning	• Agent-behaviour baselining • Trust-boundary alerting • Auto-guard-rail patch
Monitor	• Synthetic-user & rogue-agent generation	• Posture & metric collection • User and Entity Behavior Analytics (UEBA) for AI signals	• Purple SIEM lens • Merged telemetry analysis/reporting • Red with Blue KPIs	• Time-series scoring • Adaptive hunt packs • Model-drift vs threat-drift diff analysis
Govern	• Audit-grade attack-path replay & evidencing	• Policy / compliance orchestration (AI-TRiSM layer)	• Risk simulators • Residual risk analysis red/blue cycles	• Signed artifact store • Framework mapping • (e.g. NIST AI RMF, OWASP, MITRE & ISO 5339, etc) • Executive reporting

Reference Spreadsheet: OWASP Landscape Context for Red Teaming

Mapping to the OWASP Top 10 for LLM Threat Model



OWASP Top 10 LLM Applications and Generative AI - 2025 Version
Example LLM Application and Basic Threat Modeling
Ads Dawson (GangGreenTemperTatum) - https://genai.owasp.org/ - Nov 2025 - v.01

This diagram deceipts the vulnerabilities described in the OWASP Top 10 for LLM Applications and Generative AI as they apply to the components comprising a typical logical architecture of an LLM application.

This is not intended to serve as a comprehensive threat model, nor full traditional architecture.

(figure: OWASP Top 10 for LLM Application architecture and Threat Model)

Having a common view of typical LLM application architectures, including agents, models, LLMs, and the ML application stack, is crucial for defining and aligning the application stack and security model. By leveraging the application architecture from the OWASP Top 10 for LLMs, we can align appropriate security solutions with the specific risks and mitigation areas identified in the OWASP Top 10. This alignment ensures a comprehensive and cohesive approach to addressing the unique security challenges posed by LLM applications.

Application Services

An LLM application service uses large language models to process and generate human-like text for tasks like chatbots, translation, and content creation. It integrates with data agents, APIs, and security measures to ensure seamless, secure, and efficient AI-driven services, managing the model lifecycle from training to deployment.

Production Services

Production services deploy and manage large language models for real-time applications, ensuring high performance, scalability, and security. These services handle model training, versioning, and monitoring, integrating with APIs and security frameworks to deliver reliable apps like chatbots and translation services in a production environment.

Training Datasets & Processing

Training datasets consist of vast, diverse text sources, including books, articles, and web content. To ensure quality and consistency, these datasets undergo preprocessing steps like tokenization, cleaning, and normalization.

Downstream Services

Downstream services utilize the output of language models for applications such as chatbots, content generation, sentiment analysis, and automated translations. These services integrate LLM capabilities to enhance user interactions and data processing

External data sources

External data sources include web crawling through search engine APIs, remote datastores, and third-party APIs. They provide additional context and up-to-date information, enhancing the model's accuracy and relevance by supplementing the pre-trained data with real-time, domain-specific insights.

OWASP Gen AI Security Solutions Landscape

The LLM security solutions landscape leverages the LLMSecOps framework and integrates seamlessly with the LLMOps processes, encompassing Scope/Plan, Model Fine-Tuning/Data Augmentation, Test/Evaluate, Release, Deploy, Operate, Monitor, and Govern stages. This framework ensures that security is embedded at every phase of the LLM lifecycle, addressing unique challenges posed by LLM applications, including prompt-based interfaces, automation agents, LLM extensions, and complex LLM-driven applications.

The landscape includes both traditional security controls extended to support LLM Models, applications, and workloads, as well as specialized security solutions designed for LLM environments. While not intended to be a comprehensive list it provides a guiding framework for security professionals looking to integrate security controls and address the LLM Application Top 10 security risks as part of the LLM application and operations lifecycle.

Emerging GenAI/LLM-Specific Security Solutions

The architecture and approaches for LLMs and Generative AI applications are still in their infancy, introducing new challenges that extend beyond the scope of traditional security and DevSecOps practices, often operating in unpredictable and dynamic environments where traditional security controls may fall short in addressing specific risks such as prompt injection, adversarial manipulation, and ethical biases.

We have begun to see new solutions emerging that address these security gaps and have attempted to capture them in the table below. We will continue to update our list as new solutions appear. These categories are typically early in development, but can have immediate benefits.

Security Solutions	Description
LLM Firewall	An LLM firewall is a security layer specifically designed to protect large language models (LLMs) from unauthorized access, malicious inputs, and potentially harmful outputs. This firewall monitors and filters interactions with the LLM, blocking suspicious or adversarial inputs that could manipulate the model's behavior. It also enforces predefined rules and policies, ensuring that the LLM only responds to legitimate requests within the defined ethical and functional boundaries. Additionally, the LLM firewall can prevent data exfiltration and safeguard sensitive information by controlling the flow of data in and out of the model.
LLM Automated Benchmarking (includes vulnerability scanning)	LLM-specific benchmarking tools are specialized tools designed to identify and assess security weaknesses unique to large language models (LLMs). These capabilities include detecting potential issues such as prompt injection attacks, data leakage, adversarial inputs, and model biases that malicious actors could exploit. The scanner evaluates the model's responses and behaviors in various scenarios, flagging vulnerabilities that traditional security tools might overlook.
LLM Guardrails	LLM guardrails are protective mechanisms designed to ensure that large language models (LLMs) operate within defined ethical, legal, and functional boundaries. These guardrails help prevent the model from generating harmful, biased, or inappropriate content by enforcing rules, constraints, and contextual guidelines during interaction. LLM guardrails can include content filtering, ethical guidelines, adversarial input detection, and user intent validation, ensuring that the LLM's outputs align with the intended use case and organizational policies.
AI Security Posture Management	AI-SPM has emerged as a new industry term promoted by vendors and analysts to capture the concept of a platform approach to security posture management for AI, including LLM and GenAI systems. AI-SPM focuses on the specific security needs of these advanced AI systems. Focused on the models themselves traditionally. The stated goal of this category is to cover the entire AI lifecycle—from training to deployment—helping to ensure models are resilient, trustworthy, and compliant with industry standards. AI-SPM typically provides monitoring and address vulnerabilities like data poisoning, model drift, adversarial attacks, and sensitive data leakage.
Agentic AI App Security	Agentic AI architectures and application patterns are still emerging, new Agentic security solutions have already started to appear. It's unclear given this immaturity what the unique priorities for securing Agentic apps are. Our project has ongoing research in this area and will be tracking this emerging solution area

LLM & Generative AI Security Solutions

The security solutions matrix below is based on the LLMSecOps lifecycle, and mapping it to the OWASP Top 10 for LLMs and Generative AI offers a targeted approach to assessing security controls. This matrix helps identify gaps by aligning security tools with OWASP's key risks at each stage, such as adversarial attacks and data leakage.

By cross-referencing existing security measures with the specific needs of LLM and Generative AI applications, organizations can ensure comprehensive coverage and strengthen their security posture across the entire development process.

GEN AI SECURITY SOLUTIONS LANDSCAPE - ONLINE DIRECTORY

https://genai.owasp.org/ai-security-solutions-landscape/

Visit the online directory to see the latest solutions listing

The solution landscape of open source projects and proprietary offerings will be updated quarterly in this document to ensure the community maintains a reasonably updated reference list. We are also maintaining an on-line director on the project website to provide the most up to date listings. These listings are community and research sourced.

Solution listings may be submitted online by companies, projects or individuals. Submissions will be reviewed for accuracy before publishing. Below is an outline of the solution matrix maintained in the document with definitions for each area.

Solution Landscape Matrix Definitions

EXAMPLE				
Solution **(Project, Product, Service)**	**Type** **(Open Source, Proprietary)**	**Project, Company**	**Gen AI/LLMSecOps** **Category Coverage**	**Top 10 for LLM** **Risk Coverage**
Project/Product Name Create hyperlink to the project/product	Open Source	Open Source Project Name, Company Name	List of covered security control categori es provided within each stage	List of the LLM Top 10 Risks Covered by the solution. Use "LLM_All" for all categories.

Gen AI Landscape Solution Matrix

SCOPING/PLANNING				
Solution	Type	Project/Company	Gen AI/LLMSecOps	Top 10 for LLM Risk Coverage
StrideGPT	Open Source	StrideGPT	● Threat Modeling	LLM_All
MitreAtlas	Proprietary	Mitre	● Threat Modeling	LLM_All
Data Command Center	Proprietary	Securiti AI	● Access Control and Authentication Planning ● Compliance and Regulatory Assessment ● Data Privacy and Protection Strategy ● Early Identification of Sensitive Data ● Third-Party Risk Assessment (Model, Provider, etc)	LLM_All
Blueteam AI Gateway	Proprietary	Blueteam AI	● Access Control and Authentication Planning ● Compliance and Regulatory Assessment ● Data Privacy and Protection Strategy ● Early Identification of Sensitive Data ● Third-Party Risk Assessment (Model, Provider, etc)	LLM01, LLM04, LLM05, LLM06, LLM09
Palo Alto Networks AI Runtime Security	Proprietary	Palo Alto Networks	● Early Identification of Sensitive Data	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● Compliance and Regulatory Assessment	LLM01, LLM02, LLM03, LLM04,

			• Data Privacy and Protection Strategy • Early Identification of Sensitive Data, • Third-Party Risk Assessment (Model, Provider, etc) • Threat Modeling	LLM05, LLM07, LLM08, LLM09
Seezo Security Design Review	Proprietary	Seezo	• Threat Modeling	LLM01, LLM02, LLM07
PILLAR : An AI-powered Privacy Threat Modeling tool	Open Source	P.I.L.L.A.R	• Threat Modeling	LLM04, LLM05, LLM06
Pillar Security	Proprietary	Pillar Security	• Early Identification of Sensitive Data	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
Microsoft Defender for Cloud AI-SPM	Proprietary	Microsoft	• Access Control and Authentication Planning, • Compliance and Regulatory Assessment, • Data Privacy and Protection Strategy, • Early Identification of Sensitive Data, • Third-Party Risk Assessment (Model, Provider, etc), • Threat Modeling	LLM04, LLM08
SpiceDB	Open Source	AuthZed	• Access Control and Authentication Planning, • Data Privacy and Protection Strategy	LLM01, LLM02, LLM06, LLM07, LLM08, LLM10
Noma Security	Proprietary	Noma Security	• Compliance and Regulatory Assessment, • Third-Party Risk Assessment (Model, Provider, etc)	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Prediction Guard	Proprietary	Prediction Guard	• Data Privacy and Protection Strategy, • Early Identification of Sensitive Data	LLM01, LLM02, LLM04, LLM05, LLM06

DATA AUGMENTATION AND FINE-TUNING

Solution	Type	Project/Company	Gen AI/LLMSecOps	Top 10 for LLM Risk Coverage
Cloaked AI	Proprietary	IronCore Labs	• Secure Data Handling	LLM06
Unstructred.io	Proprietary	Unstructured.io	• Secure Data Handling	LLM_All
Data Command Center	Proprietary	Securiti AI	• Secure Data Handling • Secure Output Handling	LLM_All
Decisionbox	Open Source	Blueteam AI	• Data Source Validation • Secure Data Handling • Secure Output Handling	LLM02, LLM03, LLM05
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	• Secure Data Handling • Secure Output Handling • Vulnerability Assessment	LLM01, LLM02, LLM03, LLM04, LLM05, LLM07, LLM08, LLM09
Pillar Security	Proprietary	Pillar Security	• Adversarial Robustness Testing	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
Prediction Guard	Proprietary	Prediction Guard	• Secure Data Handling, • Secure Output Handling, • Model Integrity Validation (ex: serialization scanning for malware), • Vulnerability Assessment	LLM02, LLM03, LLM05, LLM06

DEVELOPMENT AND EXPERIMENTATION

Solution	Type	Project/Company	Gen AI/LLMSecOps	Top 10 for LLM Risk Coverage
Aqua Security	Proprietary	Aqua Security	● SAST, DAST & IAST ● Secure Library/Code Repository ● Software Composition Analysis ● Secure Library/Code Repository	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
Cloaked AI	Proprietary	IronCore Labs	● Secure Data Handling	LLM02, LLM08
Fickling	Open Source	Trail of Bits	● Pickle Library ● Malicious Run-time File Detection	LLM03
PrivacyRaven	Open Source	Trail of Bits	● Privacy testing library for AI models ● Malicious Run-time File Detection	LLM02,LLM03, LLM04,
Pangea Sanitize	Proprietary	Pangea	● Model And Application Interaction Security ● Secure Coding Practices	LLM02, LLM03, LLM05, LLM06
Pangea Authorization	Proprietary	Pangea	● Access, Authentication And Authorization (MFA) ● Model And Application Interaction Security ● Secure Coding Practices	LLM04, LLM06, LLM07, LLM08, LLM10
Pangea Authentication	Proprietary	Pangea	● Access, Authentication And Authorization (MFA), ● Model And Application Interaction Security, ● Secure Coding Practices	LLM04, LLM07, LLM10
Pangea Redact	Proprietary	Pangea	● Model And Application Interaction Security, ● Secure Coding Practices	LLM01, LLM02, LLM03, LLM06

PurpleLlama CodeShield	Open Source	Meta-PurpleLlama	● Insecure Code Generation	LLM05
Pangea Data Guard	Proprietary	Pangea	● Model And Application Interaction Security, ● Secure Coding Practices	LLM02, LLM03, LLM07, LLM10
Pangea Prompt Guard	Proprietary	Pangea	● Model And Application Interaction Security, ● Secure Coding Practices	LLM01, LLM03
Cisco AI Validation	Proprietary	Cisco Systems	● LLM & App Vulnerability Scanning, ● Model and Application Interaction Security	LLM01, LLM03, LLM04, LLM05, LLM06, LLM09
Mend AI	Proprietary	Mend.io	● LLM & App Vulnerability Scanning ● Model And Application Interaction Security ● SAST/DAST/IAST ● Secure Coding Practices ● Secure Library/Code Repository ● Software Composition Analysis	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Data Command Center	Proprietary	Securiti AI	● Access ● Authentication and Authorization (MFA), ● Model and Application Interaction Security	LLM_All
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● LLM & App Vulnerability Scanning	LLM01, LLM02, LLM03, LLM04, LLM05, LLM07, LLM08, LLM09
Operant 3D Runtime Defense	Proprietary	Operant AI	● LLM & App Vulnerability Scanning ● Model and Application Interaction Security ● Secure Coding Practices	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
TrojAI Detect	Proprietary	TrojAI	● LLM & App Vulnerability Scanning ● Model and Application Interaction Security ● SAST/DAST/ IAST	LLM01, LLM02, LLM03, LLM04, LLM06, LLM09, LLM10

Pillar Security	Proprietary	Pillar Security	• LLM & App Vulnerability Scanning, • Model and Application Interaction Security, • Software Composition Analysis	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
SpiceDB	Open Source	AuthZed	• Access, • Authentication and Authorization (MFA)	LLM01, LLM02, LLM04, LLM06, LLM07, LLM08, LLM10
Infosys Responsible AI Toolkit	Open Source	InfoSys	• LLM & App Vulnerability Scanning, • Model and Application Interaction Security	LLM_All
Noma Security	Proprietary	Noma Security	• LLM & App Vulnerability Scanning, • Model and Application Interaction Security, • SAST/DAST/ IAST, • Secure Coding Practices, • Secure Library/Code Repository, • Software Composition Analysis	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
AIandMe	Proprietary	AIandMe	• LLM & App Vulnerability Scanning, • Model and Application Interaction Security, • Secure Coding Practices	LLM01, LLM02, LLM04, LLM07, LLM10
Privacy-focused Code Scanner for AI Applications	Proprietary	HoundDog.ai, Inc.	• LLM & App Vulnerability Scanning, • Model and Application Interaction Security, • SAST/DAST/ IAST, • Secure Coding Practices	LLM01, LLM02, LLM05, LLM08

TEST AND EVALUATION				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Top 10 for LLM Risk Coverage**
LLM Vulnerability Scanner	Open Source	Garak.AI	● LLM Vulnerability Scanning	LLM01
Prompt Foo	Open Source	Prompt Foo	● Adversarial Testing, ● Application Security Orchestration and Correlation, ● Bias and Fairness Testing, ● Final Security Audit, ● LLM Benchmarking, ● Penetration Testing, ● SAST/DAST/IAST, ● Vulnerability Scanning	LLM_All
Modelscan	Open Source	Protect AI	● Penetration Testing ● Vulnerability Scanning	LLM01
CyberSecEval	Open Source	Meta	● Adversarial Testing ● LLM Benchmarking ● Vulnerability Scanning	LLM01, LLM02, LLM07, LLM08, LLM09, LLM10
Cisco AI Validation	Proprietary	Cisco Systems	● Final Security Audit, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM03, LLM04, LLM05, LLM06, LLM09
Enkrypt AI	Proprietary	Enkrypt AI	● Adversarial Testing, ● Bias and Fairness Testing, ● Final Security Audit, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing, ● SAST/DAST/IAST, ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10

Harmbench	Open Source	Harmbench	● Adversarial Testing ● Bias And Fairness Testing ● Incident Simulation ● Response Testing ● LLM Benchmarking ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM06, LLM08, LLM09
Aqua Security	Proprietary	Aqua Security	● Adversarial Attack Protection ● SAST/DAST/IAST ● Secure CI/CD Pipeline ● Secure Library/Code Repository ● Software Composition Analysis ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
Prompt Fuzzer	Open Source	Prompt Security	● Adversarial Testing, ● Bias And Fairness Testing, ● Incident Simulation, ● Response Testing	LLM01, LLM02, LLM03, LLM06
Pillar Security	Proprietary	Pillar Security	● Adversarial Testing, ● Incident Simulation, ● Response Testing, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
ZenGuard AI	Proprietary	ZenGuard AI	● Adversarial Testing, ● Penetration Testing	LLM_All
Giskard	Open Source	Giskard	● Adversarial Testing, ● Bias and Fairness Testing ● LLM Benchmarking, ● Vulnerability Scanning	LLM01, LLM02, LLM06, LLM08, LLM09
Data Command Center	Proprietary	Securiti AI	● Bias and Fairness Testing ● Final Security Audit ● LLM Benchmarking	LLM_All
TrojAI Detect	Proprietary	TrojAI	● Adversarial Testing ● Bias and Fairness Testing ● Final Security Audit ● Incident Simulation ● Response Testing ● LLM Benchmarking ● Penetration Testing ● SAST/DAST/IAST	LLM01, LLM02, LLM03, LLM04, LLM06, LLM09, LLM10
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● Final Security Audit,	LLM01, LLM02, LLM03, LLM04,

			● Vulnerability Scanning	LLM05, LLM07, LLM08, LLM09
Recon	Proprietary	Protect AI	● Adversarial Testing ● Bias and Fairness Testing ● LLM Benchmarking ● Penetration Testing ● SAST/DAST/IAST ● Vulnerability Scanning	LLM01, LLM02, LLM04, LLM06, LLM07, LLM08, LLM09
Citadel Lens	Proprietary	Citadel AI	● Adversarial Testing ● Bias and Fairness Testing ● LLM Benchmarking	LLM01, LLM02, LLM06
LangCheck	Open Source	Citadel AI	● Adversarial Testing ● Bias and Fairness Testing ● LLM Benchmarking	LLM01, LLM02, LLM06
Vulcan	Proprietary	AIFT	● Adversarial Testing, ● Bias and Fairness Testing ● Final Security Audit ● Incident Simulation ● Response Testing ● LLM Benchmarking ● Vulnerability Scanning	LLM01, LLM02, LLM04, LLM06, LLM08, LLM09
Watchtower	Open Source	BoschAIShield	● Adversarial Testing ● Penetration Testing ● SAST/DAST/IAST ● Vulnerability Scanning	LLM03, LLM05, LLM06
AIShield AISpectra	Proprietary	AIShield,Powered by Bosch	● Adversarial Testing ● LLM Benchmarking ● Penetration Testing ● SAST/DAST/IAST ● Vulnerability Scanning	LLM01, LLM03, LLM05, LLM06, LLM10
Mindgard	Proprietary	Mindgard	● Adversarial Testing ● Final Security Audit ● LLM Benchmarking ● Penetration Testing ● SAST/DAST/IAST ● Vulnerability Scanning	LLM01, LLM02, LLM04, LLM06, LLM08, LLM09, LLM10
Adversa AI Red Teaming Platform	Proprietary	Adversa AI	● Adversarial Testing, ● Final Security Audit, ● LLM Benchmarking, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10

AIandMe	Proprietary	AIandMe	● Adversarial Testing, ● Incident Simulation, ● Response Testing, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM02, LLM04, LLM07, LLM10
AiFort	Proprietary	KELA	● Adversarial Testing, ● Bias and Fairness Testing, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing	LLM01, LLM02, LLM04, LLM05, LLM06, LLM08, LLM09
AIM Supervisor	Proprietary	AIM Intelligence	● Adversarial Testing, ● Bias and Fairness Testing, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
CalypsoAI	Proprietary	CalypsoAI	● Adversarial Testing, ● Application Security Orchestration and Correlation, ● Bias and Fairness Testing, ● Final Security Audit, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing, ● SAST/DAST/IAST, ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM10
DeepTeam	Open Source		● Adversarial Testing, ● Application Security Orchestration and Correlation, ● Bias and Fairness Testing, ● LLM Benchmarking, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
DryRun Security	Proprietary	DryRun Security	● Final Security Audit, ● SAST/DAST/IAST, ● Vulnerability Scanning	LLM01, LLM02, LLM08
Dynamo AI	Proprietary	Dynamo AI	● Adversarial Testing, ● LLM Benchmarking, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM06, LLM09

Fujitsu GenAI Security Framework (LLM Vulnerability Scanner and Guardrails)	Proprietary	Fujitsu Limited	● Adversarial Testing, ● Bias and Fairness Testing, ● Vulnerability Scanning	LLM_All
Infosys Responsible AI Toolkit	Open Source	InfoSys	● Adversarial Testing, ● Application Security Orchestration and Correlation, ● Bias and Fairness Testing, ● Final Security Audit, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing, ● SAST/DAST/IAST, ● Vulnerability Scanning	LLM_All
Noma Security	Proprietary	Noma Security	● Adversarial Testing, ● Incident Simulation, ● Response Testing, ● Penetration Testing, ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
OWASP Top 10 for LLM	Proprietary	SplxAI	● Adversarial Testing, ● Application Security Orchestration and Correlation, ● Bias and Fairness Testing, ● Final Security Audit, ● Incident Simulation, ● Response Testing, ● LLM Benchmarking, ● Penetration Testing, ● SAST/DAST/IAST, ● Vulnerability Scanning	LLM_All
Prediction Guard	Proprietary	Prediction Guard	● LLM Benchmarking, ● Vulnerability Scanning	LLM01, LLM02, LLM03, LLM05, LLM06
SAIFE X RedTeam	Proprietary	Eroun&Company	● Adversarial Testing, ● LLM Benchmarking	LLM01, LLM02, LLM03, LLM07, LLM10
Straiker AI	Proprietary	Straiker Inc	● Adversarial Testing, ● Application Security Orchestration and Correlation, ● Bias and Fairness Testing, ● LLM Benchmarking, ● Penetration Testing	LLM01, LLM02, LLM05, LLM06, LLM07, LLM09, LLM10

Trend Vision One™	Proprietary	Trend Micro	● Adversarial Testing, ● LLM Benchmarking, ● Vulnerability Scanning	LLM01, LLM02, LLM05, LLM06, LLM10
VeriGenAI	Proprietary	VeriGenAI	● Adversarial Testing	LLM_All

RELEASE				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Top 10 for LLM Risk Coverage**
Cisco AI Validation	Proprietary	Cisco Systems	● Model Security Posture Evaluation ● Secure Supply Chain Verification	LLM01, LLM03, LLM04, LLM05, LLM06, LLM09
Data Command Center	Proprietary	Securiti AI	● Model Security Posture Evaluation ● User Access Control Validation	LLM_All
Palo Alto Networks AI Runtime Security	Proprietary	Palo Alto Networks	● AI/ML Bill of Materials (BOM)	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● Model Security Posture Evaluation ● Secure Supply Chain Verification	LLM01, LLM02, LLM03, LLM04, LLM05, LLM07, LLM08, LLM09
CycloneDX	Open Source	CycloneDX	● LLM/ML BOM Generation	LLM05
Aqua Security	Proprietary	Aqua Security	● SAST, DAST & IAST ● Secure Library/Code Repository ● Software Composition Analysis ● Secure Library/Code Repository	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
Legit Security - AI-SPM	Proprietary	Legit Security	● AI Generated Code Detection	LLM05
Pillar Security	Proprietary	Pillar Security	● AI/ML Bill of Materials (BOM)	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
Noma Security	Proprietary	Noma Security	● AI/ML Bill of Materials (BOM), ● Model Security Posture Evaluation, ● Secure CI/CD pipeline,	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06,

			● Secure Supply Chain Verification, ● Static and Dynamic Code Analysis	LLM07, LLM09, LLM10
The CalypsoAI Inference Platform	Proprietary	CalypsoAI	● AI/ML Bill of Materials (BOM), ● Digital Model Signing, ● Model Security Posture Evaluation, ● Secure CI/CD pipeline, ● Secure Supply Chain Verification, ● Static and Dynamic Code Analysis, ● User Access Control Validation	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM10

DEPLOY				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Top 10 for LLM Risk Coverage**
Cisco AI Runtime	Proprietary	Cisco Systems	● LLM Enabled Web Application Firewall ● User and Data Privacy Protections	LLM01, LLM02, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
PurpleLlama CodeShield	Open Source	Meta	●	LLM02
Data Command Center	Proprietary	Securiti AI	● Compliance Verification ● Multi-factor Authentication ● Secure Configuration ● User and Data Privacy Protections	LLM_All
Blueteam AI Gateway	Proprietary	Blueteam AI	● Deployment Validation, ● Encryption, ● Secrets management, ● LLM Enabled Web Application Firewall, ● Secrets Management, ● Secure API Access, ● Secure Configuration, ● User and Data Privacy Protections	LLM01, LLM04, LLM06, LLM09
Palo Alto Networks AI Runtime Security	Proprietary	Palo Alto Networks	● Compliance Verification ● Network Security Validation ● User and Data Privacy Protections	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Operant 3D Runtime Defense	Proprietary	Operant AI	● Secure API Access ● Secure Configuration ● User and Data Privacy Protections	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
TrojAI Detect	Proprietary	TrojAI	● Compliance Verification ● LLM Enabled Web Application Firewall	LLM01, LLM02, LLM04, LLM06, LLM10

			● User and Data Privacy Protections	
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● Compliance Verification, ● Encryption ● Secrets management ● User and Data Privacy Protections	LLM01, LLM02, LLM03, LLM04, LLM05, LLM07, LLM08, LLM09
AI Trust Platform	Proprietary	Preamble	● Secure Configuration, ● User and Data Privacy Protections	LLM01, LLM02, LLM03, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
IronCore Labs Cloaked AI	Open Source	IronCore Labs	● Encryption, ● Secrets management	LLM06
Infosys Responsible AI Toolkit	Open Source	InfoSys	● LLM Enabled Web Application Firewall, ● User and Data Privacy Protections	LLM_All
Noma Security	Proprietary	Noma Security	● LLM Enabled Web Application Firewall	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Lasso for Applications	Proprietary	Lasso	● Compliance Verification, ● LLM Enabled Web Application Firewall, ● User and Data Privacy Protections	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Lasso for Employees	Proprietary	Lasso	● Compliance Verification, ● LLM Enabled Web Application Firewall, ● User and Data Privacy Protections	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Aim Security	Proprietary	Aim Security	● Compliance Verification, ● Deployment Validation, ● LLM Enabled Web Application Firewall, ● Secure Configuration, ● User and Data Privacy Protections	LLM_All

Prediction Guard	Proprietary	Prediction Guard	● Secrets Management, ● Secure API Access, ● Secure Configuration, ● User and Data Privacy Protections	LLM04, LLM10
Teleport	Open Source		● Compliance Verification, ● Encryption, Secrets management, ● LLM Enabled Web Application Firewall, ● Multi-factor Authentication, ● Network Security Validation, ● Secrets Management, ● Secure API Access, ● Secure Configuration, ● User and Data Privacy Protections	LLM01, LLM02, LLM06, LLM07, LLM08, LLM10

OPERATE				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Top 10 for LLM Risk Coverage**
AI Blue Team	Proprietary	NRI SecureTechnologies	● Adversarial Attack Protection, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Secure Output Handling	LLM01, LLM02, LLM04, LLM06, LLM08, LLM09
Aim AI Security Platform	Proprietary	Aim Security	● Adversarial Attack Protection, ● Automated Vulnerability Scanning, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy , ● Data Leakage Protection, ● Prompt Security, ● Runtime Application Self-Protection, ● Secure Output Handling	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08
Cisco AI Runtime	Proprietary	Cisco Systems	● Adversarial Attack Protection, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Runtime Application Self-Protection, ● Secure Output Handling	LLM01, LLM02, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Data Command Center	Proprietary	Securiti AI	● Adversarial Attack Protection, ● Data Integrity and Encryption, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Secure Output Handling	LLM_All

Blueteam AI Gateway	Proprietary	Blueteam AI	● Adversarial Attack Protection, ● Data Integrity and Encryption, ● LLM Guardrails, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Runtime Application Self-Protection, ● Secure Output Handling	LLM01, LLM04, LLM06, LLM09
LLM Guard	Open Source	Protect AI	● Privacy, Data Leakage Protection ● Prompt Security, ● Adversarial Attack Protection	
Llama Guard	Open Source	Meta	● LLM Guardrails	LLM01, LLM02, LLM06, LLM07
Palo Alto Networks AI Runtime Security	Proprietary	Palo Alto Networks	● Adversarial Attack Protection, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Secure Output Handling	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
TrojAI Detect	Proprietary	TrojAI	● Adversarial Attack Protection, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Runtime Application Self-Protection, ● Secure Output Handling	LLM01, LLM02, LLM04, LLM06, LLM10
ZenGuard AI	Proprietary	Zenguard.ai	● Adversarial Attack Protection, ● Automated Vulnerability Scanning, ● LLM Guardrails, ● Privacy ● Data Leakage Protection ● Prompt Security ● Secure Output Handling	LLM_All
ZenGuard AI	Proprietary	Zenguard.ai	● Adversarial Attack Protection, ● LLM Guardrails, ● Prompt Security	LLM01, LLM02, LLM06
Aqua Security	Proprietary	Aqua Security	● Adversarial Attack Protection, ● Adversarial Testing, ● Automated Vulnerability Scanning,	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06,

			• Data Leakage Protection, • LLM Guardrails, • Penetration Testing, • Privacy, • Prompt Security, • Secure Output Handling	LLM07, LLM08, LLM09, LLM10
AI Trust Platform	Proprietary	Preamble	• Adversarial Attack Protection, • LLM Guardrails, • LLM Incident Detection and Response, • Privacy, • Data Leakage Protection, • Prompt Security, • Runtime Application Self-Protection, • Secure Output Handling	LLM01, LLM02, LLM03, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
dyana	Open Source	Dreadnode	• Automated Vulnerability Scanning, • LLM Incident Detection and Response, • Runtime Application Self-Protection	LLM03, LLM04
DynamoGuard	Proprietary	Dynamo AI	• Adversarial Attack Protection, • LLM Guardrails, • Privacy, • Data Leakage Protection, • Prompt Security	LLM01, LLM06, LLM09
F5 AI Gateway	Proprietary	F5	• Adversarial Attack Protection, • LLM Guardrails, • Prompt Security, • Secure Output Handling	LLM01, LLM02, LLM05, LLM10
Infosys Responsible AI Toolkit	Open Source	InfoSys	• Automated Vulnerability Scanning, • LLM Guardrails, • Privacy, • Data Leakage Protection, • Prompt Security, • Secure Output Handling	LLM01, LLM02, LLM07, LLM08, LLM09
Insight For Webserver (IWS)	Proprietary	Infotect Security	• LLM Incident Detection and Response, • Privacy, • Data Leakage Protection, • Secure Output Handling	LLM01, LLM02, LLM05, LLM06, LLM07
IronCore Labs Cloaked AI	Open Source	IronCore Labs	• Data Integrity and Encryption	LLM06

Knostic	Proprietary	Knostic	● LLM Guardrails, ● Privacy, ● Data Leakage Protection, ● Secure Output Handling	LLM02, LLM05, LLM06
Microsoft Security - Secure and Govern AI	Proprietary	Microsoft	● Adversarial Attack Protection, ● Automated Vulnerability Scanning, ● Data Integrity and Encryption, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Secure Output Handling	LLM01, LLM02, LLM04, LLM05, LLM06, LLM08, LLM09
Noma Security	Proprietary	Noma Security	● Adversarial Attack Protection, ● Automated Vulnerability Scanning, ● LLM Guardrails, ● Prompt Security, ● Runtime Application Self-Protection, ● Secure Output Handling	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Pillar Security	Proprietary	Pillar Security	● Adversarial Attack Protection, ● Automated Vulnerability Scanning, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Runtime Application Self-Protection	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
Prediction Guard	Proprietary	Prediction Guard	● Adversarial Attack Protection, ● LLM Guardrails, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Secure Output Handling	LLM01, LLM02, LLM05, LLM06
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● Automated Vulnerability Scanning, ● Data Integrity and Encryption, ● LLM Guardrails	LLM01, LLM02, LLM03, LLM04, LLM05, LLM07, LLM08, LLM09
Skyrelis	Proprietary	Skyrelis	● LLM Guardrails, ● LLM Incident Detection and Response, ● Prompt Security, ● Runtime Application Self-Protection	LLM01, LLM03, LLM04, LLM06, LLM09, LLM10

Straiker AI	Proprietary	Straiker AI	● Adversarial Attack Protection, ● Data Integrity and Encryption, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Secure Output Handling	LLM01, LLM02, LLM05, LLM06, LLM07, LLM09
Trend Vision One™	Proprietary	Trend Micro	● Adversarial Attack Protection, ● Automated Vulnerability Scanning, ● Data Integrity and Encryption, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Patch Management, ● Privacy, ● Data Leakage Protection, ● Prompt Security, ● Runtime Application Self-Protection, ● Secure Output Handling	LLM01, LLM02, LLM03, LLM05, LLM06, LLM08, LLM10
WebOrion® Protector Plus	Proprietary	Cloudsine Pte Ltd	● Adversarial Attack Protection, ● LLM Guardrails, ● LLM Incident Detection and Response, ● Privacy, ● Data Leakage Protection, ● Prompt Security	LLM01, LLM02, LLM05, LLM07, LLM08, LLM09, LLM10

MONITOR				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Top 10 for LLM Risk Coverage**
Aim AI Security Platform	Proprietary	Aim Security	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08
AIShield Guardian	Proprietary	AISheild,Powered by Bosch	● Adversarial Input Detection, ● AI/LLM Secure Posture Management, ● Security Alerting, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM01, LLM02, LLM04, LLM06, LLM07, LLM08, LLM10
Blueteam AI Gateway	Proprietary	Blueteam AI	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Patch and Update Alerts, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM01, LLM04, LLM06, LLM09
Cisco AI Validation	Proprietary	Cisco Systems	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management,	LLM01, LLM03, LLM04, LLM05, LLM06, LLM09

			• Regulatory Compliance Tracking	
Data Command Center	Proprietary	Securiti AI	• Adversarial Input Detection, • Model Behavior Analysis, • AI/LLM Secure Posture Management, • Regulatory Compliance Tracking, • Security Alerting, • User Activity Monitoring, • Data Privacy and Protection, • Ethical Compliance	LLM_All
HiddenLayer AISec Platform	Proprietary	HiddenLayer, Inc	• Adversarial Input Detection, • Model Behavior Analysis, • AI/LLM Secure Posture Management, • Regulatory Compliance Tracking, • Security Alerting, • User Activity Monitoring, • Observability, • Data Privacy and Protection	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
Lakera	Proprietary	Lakera	• Adversarial Input Detection, • Regulatory Compliance Tracking, • Security Alerting, • Security Metrics Collection, • Data Privacy and Protection, • Ethical Compliance	LLM_All
Layer	Proprietary	Protect AI	• Adversarial Input Detection, • Model Behavior Analysis, • AI/LLM Secure Posture Management, • Security Alerting, • Security Metrics Collection, • User Activity Monitoring, • Observability, • Data Privacy and Protection	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09
Operant 3D Runtime Defense	Proprietary	Operant AI	• Adversarial Input Detection, • Model Behavior Analysis, • AI/LLM Secure Posture Management, • Regulatory Compliance Tracking, • Security Alerting,	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10

			● Security Metrics Collection, ● Observability, ● Data Privacy and Protection	
Palo Alto Networks AI Runtime Security	Proprietary	Palo Alto Networks	● Adversarial Input Detection, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● Observability, ● Data Privacy and Protection	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Data Privacy and Protection	LLM01, LLM02, LLM03, LLM04, LLM05, LLM07, LLM08, LLM10
PromptGuard	Open Source	Meta	● Adversarial Input Detection	LLM01
SPLX.AI	Proprietary	Brand Engagement Networks	● Adversarial Input Detection, ● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Security Metrics Collection, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM_All
TrojAI Detect	Proprietary	TrojAI	● Adversarial Input Detection, ● Model Behavior Analysis, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● Data Privacy and Protection	LLM01, LLM02, LLM04, LLM06, LLM10
AISec Platform	Proprietary	Hidden Layer	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Security Alerting, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection	LLM01, LLM02, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10

Aqua Security	Proprietary	Aqua Security	• AI/LLM Secure Posture Management	LLM04, LLM06, LLM10
AI Trust Platform	Proprietary	Preamble	• Adversarial Input Detection, • AI/LLM Secure Posture Management, • Security Alerting, • Security Metrics Collection, • User Activity Monitoring, • Observability, • Data Privacy and Protection, • Ethical Compliance	LLM01, LLM02, LLM03, LLM05, LLM06, LLM07, LLM08, LLM09, LLM10
AIandMe	Proprietary	AIandMe	• Adversarial Input Detection, • Patch and Update Alerts, • Regulatory Compliance Tracking, • Security Alerting, • User Activity Monitoring, • Observability, • Ethical Compliance	LLM01, LLM02, LLM04, LLM07, LLM10
AiFort	Proprietary	KELA	• AI/LLM Secure Posture Management, • Regulatory Compliance Tracking, • Security Alerting, • Observability, • Data Privacy and Protection, • Ethical Compliance	LLM01, LLM02, LLM04, LLM05, LLM06, LLM08, LLM09
Apex Security AI	Proprietary	Apex Security AI	• Adversarial Input Detection, • Model Behavior Analysis, • AI/LLM Secure Posture Management, • Regulatory Compliance Tracking, • Security Alerting, • User Activity Monitoring, • Observability, • Data Privacy and Protection, • Ethical Compliance	LLM_All
DynamoGuard	Proprietary	Dynamo AI	• Regulatory Compliance Tracking, • User Activity Monitoring, • Observability, • Ethical Compliance	LLM01, LLM06, LLM09

Fiddler AI	Proprietary	Fiddler AI	● Adversarial Input Detection, ● Model Behavior Analysis, ● Security Alerting, ● Observability, ● Data Privacy and Protection	LLM01, LLM02, LLM04, LLM07, LLM09
GuardionAI	Proprietary	GuardionAI	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07
Insight For Webserver (IWS)	Proprietary	Infotect Security	● Security Alerting, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM01, LLM02, LLM05, LLM06, LLM07
LLMInspect	Proprietary	EUNOMATIX	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM_All
Noma Security	Proprietary	Noma Security	● Adversarial Input Detection, ● AI/LLM Secure Posture Management, ● Data Privacy and Protection	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Pillar Security	Proprietary	Pillar Security	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Data Privacy and Protection,	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10

			● Ethical Compliance	
Prediction Guard	Proprietary	Prediction Guard	● Security Metrics Collection, ● User Activity Monitoring, ● Observability	LLM01, LLM02, LLM04, LLM06
Skyrelis	Proprietary	Skyrelis	● Adversarial Input Detection, ● Model Behavior Analysis, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection	LLM01, LLM02, LLM04, LLM05, LLM07
Straiker AI	Proprietary	Straiker AI	● Adversarial Input Detection, ● AI/LLM Secure Posture Management, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability	LLM01, LLM02, LLM05, LLM06, LLM07, LLM09
Teleport	Open Source	Teleport	● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking, ● Security Alerting, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection	LLM01, LLM02, LLM06, LLM07, LLM08, LLM10
The CalypsoAI Inference Platform	Proprietary	CalypsoAI	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Patch and Update Alerts, ● Regulatory Compliance Tracking, ● Security Alerting, ● Security Metrics Collection, ● User Activity Monitoring, ● Observability, ● Data Privacy and Protection, ● Ethical Compliance	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM10
Trend Vision One™	Proprietary	Trend Micro	● Adversarial Input Detection	LLM01

GOVERN				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Top 10 for LLM Risk Coverage**
AI Verify	Open Source	AI Verify Foundation	- Bias and Fairness Oversight - Risk Assessment and Management	LLM_All
Aim AI Security Platform	Proprietary	Aim Security	- Compliance Management, - Data Security Posture Management, - Risk Assessment and Management, - User/Machine Access audits	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08
Blueteam AI Gateway	Proprietary	Blueteam AI	- Bias and Fairness Oversight, - Compliance Management, - Data Security Posture Management, - User/Machine Access audits	LLM01, LLM04, LLM06, LLM09
Cisco AI Validation	Proprietary	Cisco Systems	- Compliance Management, - Risk Assessment and Management	LLM01, LLM03, LLM04, LLM05, LLM06, LLM09
Data Command Center	Proprietary	Securiti AI	- Bias and Fairness Oversight, - Compliance Management, - Data Security Posture Management, - Incident Governance, - Risk Assessment and Management, - User/Machine Access audits	LLM_All
Palo Alto Networks AI Runtime Security	Proprietary	Palo Alto Networks	- Compliance Management, - Incident Governance, - Risk Assessment and Management	LLM01, LLM02, LLM03, LLM04, LLM06, LLM07, LLM08, LLM09, LLM10
Prisma Cloud AI-SPM	Proprietary	Palo Alto Networks	- Compliance Management, - Data Security Posture Management,	LLM01, LLM02, LLM03, LLM04,

			● Risk Assessment and Management	LLM05, LLM07, LLM08, LLM09
Prompt Security	Proprietary	Prompt Security	● Bias and Fairness Oversight, ● Compliance Management, ● Data Security Posture Management, ● Incident Governance, ● Risk Assessment and Management, ● User/Machine Access audits	LLM_All
Tumeryk, AI Trust Score	Proprietary	Tumeryk, Inc.	● Bias and Fairness Oversight, ● Compliance Management, ● Data Security Posture Management, ● Incident Governance, ● Risk Assessment and Management	LLM01, LLM02, LLM05, LLM06, LLM09, LLM10
Unbound Security	Proprietary	Unbound Security	● Compliance Management, ● Data Security Posture Management, ● Incident Governance	LLM01, LLM02, LLM05, LLM08
Lasso Secure Gateway for LLMs	Proprietary	Lasso Security (Silver Sponsor)	● LLM Secure Gateway	LLM01, LLM02
AI Security & Governance	Proprietary	Securiti (Silver Sponsor)	● Model Discovery ● Model Risk Management	LLM03, LLM06, LLM09
Cranium Platform and AI Trust Hub	Proprietary	Cranium	● Compliance Management, ● Data Security Posture Management, ● Incident Governance, ● Risk Assessment and Management	LLM_All
DynamoGuard	Proprietary	Dynamo AI	● Compliance Management, ● Risk Assessment and Management	LLM01, LLM06, LLM09
Insight For Webserver (IWS)	Proprietary	Infotect Security	● Compliance Management, ● Data Security Posture Management, ● Incident Governance, ● Risk Assessment and Management	LLM01, LLM02, LLM05, LLM06, LLM07

Noma Security	Proprietary	Noma Security	● Compliance Management, ● Risk Assessment and Management	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM09, LLM10
Pillar Security	Proprietary	Pillar Security	● Bias and Fairness Oversight, ● Compliance Management, Incident Governance, ● Risk Assessment and Management	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM08, LLM10
Pomerium	Open Source	Pomerium	● Compliance Management, ● Data Security Posture Management, ● User/Machine Access audits	LLM01, LLM02, LLM06, LLM10
Skyrelis	Proprietary	Skyrelis	● Compliance Management, ● Data Security Posture Management, ● Risk Assessment and Management, ● User/Machine Access audits	LLM01, LLM02, LLM04, LLM05, LLM07
Teleport	Open Source	Teleport	● Compliance Management, ● Data Security Posture Management, ● Incident Governance, ● Risk Assessment and Management, ● User/Machine Access audits	LLM01, LLM02, LLM06, LLM07, LLM08, LLM10
The CalypsoAI Inference Platform	Proprietary	CalypsoAI	● Bias and Fairness Oversight, ● Compliance Management, ● Data Security Posture Management, ● Incident Governance, ● Risk Assessment and Management, ● User/Machine Access audits	LLM01, LLM02, LLM03, LLM04, LLM05, LLM06, LLM07, LLM10

Agentic AI Security Solutions and SecOps, Risks and Mitigations Coverage

Agentic AI Security Solutions

AGENTIC AI - SCOPING/PLANNING				
Solution	Type	Project/Company	Gen AI/LLMSecOps	Agentic Risk Coverage
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	● Conducting Agentic Threat Modeling, ● Support for Gen AI Security Project - Agentic Security Threat Modeling Approach, ● Identify system-wide non-human Identities and Auth Protocols, ● Draft policy for Agent privilege boundaries, ● Draft policy Agent for tool scopes, ● Draft policy for delegation logic	T02, T03, T08, T13
Noma Security	Proprietary	Noma Security	● Conducting Agentic Threat Modeling, ● Identify system-wide non-human Identities and Auth Protocols, ● Draft policy for Agent privilege boundaries, ● Draft policy Agent for tool scopes, ● Draft policy for delegation logic, ● Define controls for memory scoping, isolation and long-term persistence	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15
Pillar Security	Proprietary	Pillar Security	● Conducting Agentic Threat Modeling, ● Support for Gen AI Security Project - Agentic Security Threat Modeling Approach	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Straiker AI	Proprietary	Straiker	● Conducting Agentic Threat Modeling ● Draft policy for Agent privilege boundaries, ● Draft policy Agent for tool scopes ● Draft policy for delegation logic	T01, T02, T03, T04, T05, T06, T07, T09, T11, T12, T13, T14, T15
Zenity	Proprietary	Zenity	● Conducting Agentic Threat Modeling, ● Support for Gen AI Security Project - Agentic Security Threat Modeling Approach ● Identify system-wide non-human Identities and Auth Protocols,	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15

			● Draft policy for Agent privilege boundaries, ● Draft policy Agent for tool scopes	
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Conducting Agentic Threat Modeling, ● Support for Gen AI Security Project - Agentic Security Threat Modeling Approach, ● Identify system-wide non-human Identities and Auth Protocols, ● Draft policy for Agent privilege boundaries, ● Draft policy Agent for tool scopes, ● Draft policy for delegation logic, ● Define controls for memory scoping, isolation and long-term persistence	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15

AGENTIC AI - DATA AUGMENTATION AND FINE-TUNING				
Solution	Type	Project/Company	Gen AI/LLMSecOps	Agentic Risk Coverage
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	● Apply differential privacy or obfuscation on sensitive data injected into agent memory ● Agent Action Audit	T02, T03, T08, T13
Noma Security	Proprietary	Noma Security	● Apply differential privacy or obfuscation on sensitive data injected into agent memory ● Agent Action Audit	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Pillar Security	Proprietary	Pillar Security	● Apply differential privacy or obfuscation on sensitive data injected into agent memory ● Agent Action Audit ● Apply PII and Sensitive data masking injected into agent components	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Zenity	Proprietary	Zenity	● Apply differential privacy or obfuscation on sensitive data injected into agent memory ● Agent Action Audit	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Apply differential privacy or obfuscation on sensitive data injected into agent memory ● Agent Action Audit	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15

DEVELOPMENT AND EXPERIMENTATION				
Solution	Type	Project/Company	Gen AI/LLMSecOps	Agentic Risk Coverage
Noma Security	Proprietary	Noma Security	● Perform SAST/DAST on agent planning code, tool wrappers, and plugin interfaces. ● Harden agent loop logic against infinite loops, unsafe function routing, and unauthorized self-modification. ● Validate connector (e.g., MCP) contracts (input/output schemas and permissions). ● Implement policy enforcement hooks in Frameworks (e.g. LangGraph, CrewAI, Others)	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15
Pensar	Proprietary	Pensar	● Perform SAST/DAST on agent planning code, tool wrappers, and plugin interfaces. ● Harden agent loop logic against infinite loops, unsafe function routing, and unauthorized self-modification.	T01, T02, T03, T04, T06, T07, T11, T13, T14
Pillar Security	Proprietary	Pillar Security	● Perform SAST/DAST on agent planning code, tool wrappers, and plugin interfaces. ● Validate connector (e.g., MCP) contracts (input/output schemas and permissions). ● Implement policy enforcement hooks in Frameworks (e.g. LangGraph, CrewAI, Others)	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
TrojAI	Proprietary	TrojAI	● Harden agent loop logic against infinite loops, unsafe function routing, and unauthorized self-modification. ● Validate connector (e.g., MCP) contracts (input/output schemas and permissions).	T05, T06, T07, T13, T15

			● Implement policy enforcement hooks in Frameworks (e.g. LangGraph, CrewAI, Others)	
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Perform SAST/DAST on agent planning code, tool wrappers, and plugin interfaces. ● Harden agent loop logic against infinite loops, unsafe function routing, and unauthorized self-modification. ● Validate connector (e.g., MCP) contracts (input/output schemas and permissions). ● Implement policy enforcement hooks in Frameworks (e.g. LangGraph, CrewAI, Others)	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Mindgard	Proprietary	Mindgard	● Perform SAST/DAST on agent planning code, tool wrappers, and plugin interfaces.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T14, T15
Straiker AI	Proprietary	Straiker	● Implement policy enforcement hooks in Frameworks (e.g. LangGraph, CrewAI, Others)	T01, T02, T03, T04, T05, T06, T07, T09, T11, T12, T13, T14, T15

AGENTIC AI – TEST AND EVALUATION				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Agentic Risk Coverage**
Adversa AI Red Teaming platform	Proprietary	Adversa AI	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Validate agent decisions against expected goal plans.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T15
Agentic Radar	Open Source		● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
ai&me	Proprietary	ai&me	● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Validate agent decisions against expected goal plans.	T01, T02, T04, T06, T07, T11
Citadel Lens	Proprietary	Citadel AI	● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Validate agent decisions against expected goal plans.	T01, T02, T04, T05, T06, T07, T08, T10, T13, T15
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	● Available Agent Scanning	T02, T03, T08, T13

Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
HiveTrace	Proprietary	HiveTrace	● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T02, T03, T04, T06, T07, T12
InspectRAG	Proprietary	Eunomatix	● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage.	T03
Mend AI	Proprietary	mend.io	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T05, T06, T07, T08, T09, T11, T12, T13
Mindgard	Proprietary	Mindgard	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T05, T06, T07, T09, T10, T12, T15

Noma Security	Proprietary	Noma Security	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15
Pillar Security	Proprietary	Pillar Security	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
SplxAI Platform	Proprietary	SplxAI	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Straiker	Proprietary	Straiker	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans.	T01, T02, T03, T04, T05, T06, T07, T09, T11, T12, T13, T14, T15

			● Sandboxed testing of all tool calls, code execution, cloud API triggers	
Trend Vision One™	Proprietary	Trend Micro	● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage.	T01, T02, T05, T06, T07, T09, T11
TrojAI	Proprietary	TrojAI	● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Validate agent decisions against expected goal plans.	T05, T06, T07, T13, T15
Vulcan	Proprietary	Vulcan	● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage.	T02, T03, T04, T05, T06, T07, T09, T11, T12
Zenity	Proprietary	Zenity	● Available Agent Scanning, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Validate agent decisions against expected goal plans.	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
ARTEMIS	Proprietary	Repello AI	● Available Agent Scanning, ● Agent Penetration Testing, ● Adversarial red-teaming: goal drift, prompt injection, hallucination chaining, and over-permissioned tool usage. ● Multi-agent scenario simulations for collusion, misalignment, or deception detection. ● Validate agent decisions against expected goal plans. ● Sandboxed testing of all tool calls, code execution, cloud API triggers	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15

AGENTIC AI – RELEASE				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Agentic Risk Coverage**
SplxAI Platform	Proprietary	SplxAI	- Generate and verify model + agent + tool SBOMs - shared responsibility	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	- Generate and verify model + agent + tool SBOMs - shared responsibility	T02, T03, T07, T13
Noma Security	Proprietary	Noma Security	- Generate and verify model + agent + tool SBOMs - shared responsibility, - Register all agents in an internal trust registry	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15
Pillar Security	Proprietary	Pillar Security	- Generate and verify model + agent + tool SBOMs - shared responsibility, - Register all agents in an internal trust registry	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Zenity	Proprietary	Zenity	- Generate and verify model + agent + tool SBOMs - shared responsibility, - Register all agents in an internal trust registry	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15

AGENTIC AI – DEPLOY

Solution	Type	Project/Company	Gen AI/LLMSecOps	Agentic Risk Coverage
AI Security Platform	Proprietary	Pangea	● Enforce zero-trust policies between agents, tools, and external APIs, ● Rotate all shared secrets, keys, and tokens with ephemeral, scoped credentials. ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T06, T07, T08, T12
Cequence AI Gateway	Proprietary	Cequence Security	● Enforce zero-trust policies between agents, tools, and external APIs, ● Apply and manage runtime Guardrails	T02
Citadel Lens	Proprietary	Citadel AI	● Apply and manage runtime Guardrails	T01, T02, T04, T05, T06, T07, T08, T10, T13, T15
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Enforce zero-trust policies between agents, tools, and external APIs, ● Rotate all shared secrets, keys, and tokens with ephemeral, scoped credentials. ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
GuardionAI	Proprietary		● Enforce zero-trust policies between agents, tools, and external APIs, ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
HiveTrace	Proprietary	HiveTrace	● Enforce zero-trust policies between agents, tools, and external APIs, ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T02, T03, T04, T06, T07, T12

LlamaFirewall	Open Source	Meta	● Enforce zero-trust policies between agents, tools, and external APIs, ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T06, T07, T10, T12, T14, T15
MCP Secure Gateway	Open Source		● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T05, T08, T10, T12
Noma Security	Proprietary	Noma Security	● Enforce zero-trust policies between agents, tools, and external APIs, ● Rotate all shared secrets, keys, and tokens with ephemeral, scoped credentials. ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15
Pillar Security	Proprietary	Pillar Security	● Enforce zero-trust policies between agents, tools, and external APIs, ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Pomerium	Open Source		● Enforce zero-trust policies between agents, tools, and external APIs, ● Rotate all shared secrets, keys, and tokens with ephemeral, scoped credentials. ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T02, T03, T09, T12, T13, T14, T15
SplxAI Probe Platform	Proprietary	SplxAI	● Apply and manage runtime Guardrails	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Trend Vision One™	Proprietary	Trend Micro	● Apply and manage runtime Guardrails	T01, T05, T06, T07, T08
TrojAI	Proprietary	TrojAI	● Apply and manage runtime Guardrails	T01, T02, T06, T07, T11, T12, T13, T14, T15
Straiker	Proprietary	Straiker	● Enforce zero-trust policies between agents, tools, and external APIs, ● Apply and manage runtime Guardrails	T01, T02, T03, T04, T05, T06, T07, T09, T11,

				T12, T13, T14, T15
Zenity	Proprietary	Zenity	● Enforce zero-trust policies between agents, tools, and external APIs, ● Rotate all shared secrets, keys, and tokens with ephemeral, scoped credentials. ● Apply and manage runtime Guardrails, ● Configure Inter-agent authorization policies, capabilities, and roles	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15

AGENTIC AI - OPERATE

Solution	Type	Project/Company	Gen AI/LLMSecOps	Agentic Risk Coverage
AI Blue Team	Proprietary	NRI SecureTechnologies, Ltd.	● Detect task replay, infinite delegation, or hallucination loops. ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T03, T04, T06, T07, T11, T15
Aim AI Security Platform	Proprietary	Aim Security	● Monitor agent memory mutation patterns for drift, ● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T03, T04, T05, T06, T07, T08, T09, T12, T13, T14
Citadel Lens	Proprietary	Citadel AI	● Monitor agent memory mutation patterns for drift, ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T04, T05, T06, T07, T08, T10, T13, T15
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● LLM Incident Detection and Response	T02, T03, T07, T13
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Monitor agent memory mutation patterns for drift, ● Detect task replay, infinite delegation, or hallucination loops.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15

			● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	
GuardionAI	Proprietary		● Monitor agent memory mutation patterns for drift, ● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
HiveTrace	Proprietary	HiveTrace	● Runtime guardrails & moderation; anomalous tool use	T02, T03, T04, T06, T07, T12
Microsoft Defender for Cloud	Proprietary	Microsoft	● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T02, T04, T07, T15
Mindgard	Proprietary	Mindgard	● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions	T01, T02, T03, T04, T05, T06, T07, T08, T9, T10, T15
Noma Security	Proprietary	Noma Security	● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15

Pillar Security	Proprietary	Pillar Security	● Monitor agent memory mutation patterns for drift, ● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
SplxAI Probe Platform	Proprietary	SplxAI	● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● Runtime guardrails & moderation; anomalous tool use	T02, T05, T07, T10, T11, T13, T15
Straiker	Proprietary	Straiker	● Monitor agent memory mutation patterns for drift, ● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T01, T02, T03, T04, T05, T06, T07, T09, T11, T12, T13, T14, T15
TrojAI	Proprietary	TrojAI	● Runtime guardrails & moderation; anomalous tool use	T01, T02, T06, T07, T11, T12, T14, T15
Zenity	Proprietary	Zenity	● Monitor agent memory mutation patterns for drift, ● Detect task replay, infinite delegation, or hallucination loops. ● Enable human-in-the-loop (HITL) override thresholds on high-risk or ambiguous actions, ● Continuously scan loaded plugins for CVEs and privilege escalation vectors. ● LLM Incident Detection and Response, ● Runtime guardrails & moderation; anomalous tool use	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15

AGENTIC AI – MONITOR				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Agentic Risk Coverage**
Cisco AI Validation	Proprietary	Cisco Systems	● Adversarial Input Detection, ● Model Behavior Analysis, ● AI/LLM Secure Posture Management, ● Regulatory Compliance Tracking	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
AI Blue Team	Proprietary	NRI SecureTechnologies , Ltd.	● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter.	T01, T02, T03, T04, T06, T07, T11, T15
AI Trust Score™	Proprietary	Tumeryk INC	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes.	T02, T04, T05, T06, T07, T08, T10, T12, T13, T14, T15
ARGUS	Proprietary	Repello AI	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Citadel Lens	Proprietary	Citadel AI	● Correlate telemetry from agent step tracing, tool execution, and message logs.	T01, T02, T04, T05, T06, T07, T08, T10, T13, T15
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	● Correlate telemetry from agent step tracing, tool execution, and message logs.	T01, T02, T03, T06, T13

			● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter.	
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Fiddler AI	Proprietary	Fiddler AI	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes.	T01, T02, T05, T06, T07, T12, T13, T14, T15
GuardionAI	Proprietary		● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
HiveTrace	Proprietary	HiveTrace	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter.	T02, T03, T04, T06, T07, T12

			● Audit reflection accuracy by comparing stated and observed planning outcomes.	
Insights For WebServers (IWS)	Proprietary	Infotect Security Pte Ltd	● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter.	T01, T02, T03, T05, T06, T07, T09, T15
Metomic	Proprietary	Metomic	●	T02, T15
Microsoft Defender for Cloud	Proprietary	Microsoft	●	T02, T06, T15
Noma Security	Proprietary	Noma Security	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T15
Pangea AI Security Platform	Proprietary	Pangea	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T03, T06, T07, T08, T09, T12
Pillar Security	Proprietary	Pillar Security	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter.	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
SplxAI Probe Platform	Proprietary	SplxAI	● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter.	T02, T03, T06, T07, T09, T11, T13, T14, T15

Straiker	Proprietary	Straiker	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T03, T04, T05, T06, T07, T09, T11, T12, T13, T14, T15
Tenable AI (Apex acquired by Tenable)	Proprietary	Tenable AI	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes.	T01, T02, T03, T04, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
TrojAI	Proprietary	TrojAI	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T01, T02, T06, T07, T11, T12, T13, T14, T15
Zenity	Proprietary	Zenity	● Correlate telemetry from agent step tracing, tool execution, and message logs. ● Alert on anomalies; e.g., goal reversal, unexpected plan depth, adversarial-input, excessive tool usage, or rapid inter-agent chatter. ● Audit reflection accuracy by comparing stated and observed planning outcomes. ● Use immutable logs (e.g., Sigstore, Immudb) for forensic readiness.	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15

AGENTIC AI - GOVERN				
Solution	**Type**	**Project/Company**	**Gen AI/LLMSecOps**	**Agentic Risk Coverage**
AgenticTrust by HUMAN Security	Proprietary	HUMAN Security	● Enforce role- and task-based access policies across agent populations and their tool access.	T03, T05, T06, T07, T09, T13
Cortex Cloud AI-SPM	Proprietary	Palo Alto Networks	● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001.	T02, T03, T07, T13, T14
Enkrypt AI Security and Compliance Platform	Proprietary	Enkrypt AI	● Enforce role- and task-based access policies across agent populations and their tool access. ● Automate agent versioning, expiration, and rotation policies. ● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. ● Automate goal alignment audits, including adversarial review of long-term agent memory.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Fiddler AI	Proprietary	Fiddler AI	● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001.	T01, T02, T05, T06, T12, T13, T14, T15
GuardionAI	Proprietary	GuardionAI	● Enforce role- and task-based access policies across agent populations and their tool access. ● Automate agent versioning, expiration, and rotation policies. ● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. ● Automate goal alignment audits, including adversarial review of long-term agent memory.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, T15
Noma Security	Proprietary	Noma Security	● Enforce role- and task-based access policies across agent populations and their tool access.	T01, T02, T03, T04, T05, T06, T07, T08, T09,

			- Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. - Automate goal alignment audits, including adversarial review of long-term agent memory.	T10, T11, T12, T15
Pangea AI Security Platform	Proprietary	Pangea	- Enforce role- and task-based access policies across agent populations and their tool access. - Automate agent versioning, expiration, and rotation policies. - Automate goal alignment audits, including adversarial review of long-term agent memory.	T01, T02, T03, T06, T07, T08, T09, T12
Pillar Security	Proprietary	Pillar Security	- Enforce role- and task-based access policies across agent populations and their tool access. - Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. - Automate goal alignment audits, including adversarial review of long-term agent memory.	T01, T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15
Prompt Security (Prompt for Agentic AI)	Proprietary	Prompt Security	- Enforce role- and task-based access policies across agent populations and their tool access. - Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. - Automate goal alignment audits, including adversarial review of long-term agent memory.	T02, T03, T04, T06, T07, T10, T11, T12, T13, T14
SplxAI Probe Platform	Proprietary	SplxAI	- Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001.	T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13
Straiker	Proprietary	Straiker	- Enforce role- and task-based access policies across agent populations and their tool access. - Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001.	T01, T02, T03, T04, T05, T06, T07, T09, T11, T12, T13, T14, T15

			● Automate goal alignment audits, including adversarial review of long-term agent memory.	
TrojAI	Proprietary	TrojAI	● Enforce role- and task-based access policies across agent populations and their tool access. ● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001.	T01, T02, T06, T07, T11, T12, T13, T14, T15
Unbound	Proprietary	Unbound	● Enforce role- and task-based access policies across agent populations and their tool access. ● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001.	T02, T04, T06
Zenity	Proprietary	Zenity	● Enforce role- and task-based access policies across agent populations and their tool access. ● Automate agent versioning, expiration, and rotation policies. ● Align control evidence with frameworks like EU AI Act, NIST AI RMF, and ISO/IEC 42001. ● Automate goal alignment audits, including adversarial review of long-term agent memory.	T02, T03, T04, T06, T07, T08, T09, T11, T12, T13, T14, T15

Acknowledgements

Lead Authors

Scott Clinton
Ads Dawson
Jason Ross
Heather Linn

Contributors

Andy Smith
Arun John
Aurora Starita
Bryan Nakayama
Dennys Pereira
Emmanuel Guilherme
Fabrizio Cilli
Garvin LeClaire
Helen Oakley
Ishan Anand
Jason Ross
Marcel Winandy
Markus Hupfauer
Migel Fernandes
Mohit Yadav
Rachel James
Rico Komenda
Talesh Seeparsan
Teruhiro Tagomori
Todd Hathaway
Ron F. Del Rosario
Vaibhav Malik

Reviewers

Andy Smith
Arun John
Aurora Starita
Blanca Rivera Campos
Bryan Nakayama
Dan Guido
Dennys Pereira
Emmanuel Guilherme
Fabrizio Cilli
Garvin LeClaire
Heather Linn
Helen Oakley
Ishan Anand
Jason Ross
Joshua Berkoh
Krishna Sankar
Marcel Winandy
Markus Hupfauer
Migel Fernandes
Mohit Yadav
Rachel James
Rammohan Thirupasur
Rico Komenda
Rammohan Thirupasur
Talesh Seeparsan
Teruhiro Tagomori
Todd Hathaway
Ron F. Del Rosario
Vaibhav Malik

OWASP Top 10 for LLM Project Sponsors

We appreciate our Project Sponsors, funding contributions to help support the objectives of the project and help to cover operational and outreach costs augmenting the resources the OWASP.org foundation provides. The OWASP Top 10 for LLM and Generative AI Project continues to maintain a vendor neutral and unbiased approach. Sponsors do not receive special governance considerations as part of their support. Sponsors do receive recognition for their contributions in our materials and web properties.

All materials the project generates are community developed, driven and released under open source and creative commons licenses. For more information on becoming a sponsor Visit the Sponsorship Section on our Website to learn more about helping to sustain the project through sponsorship.

Project Sponsors:

GenAI SECURITY PROJECT — GOLD PROJECT SPONSOR — TOP 10 for LLM and Generative AI

ByteDance
HIDDENLAYER
Mend.io
snyk
paloalto
TREND

GenAI SECURITY PROJECT — SILVER PROJECT SPONSOR — TOP 10 for LLM and Generative AI

splx AI
Calypso AI
Acuvity
pangea
Cobalt
LAKERA
securiti
ActiveFence
Prompt:
TROJ.AI
Lasso SECURITY
PROTECTO
PROMPTARMOR
Synack
CT

Sponsors list, as of publication date. Find the full sponsor list here.

References

- Andreesen/Horowitz. (n.d.). Emerging architectures for LLMs. A16Z. https://a16z.com/emerging-architectures-for-llm-applications/
- Databricks. (n.d.). LLM architecture. Google Drive. https://drive.google.com/file/d/166D_Pyt3iDu18xGl3qAoMza0cq-Y52AX/view?usp=drive_link
- Protect AI. (n.d.). What is in AI Zeroday? Protect AI Blog. https://protectai.com/blog/what-is-in-ai-zeroday
- Insight Partners. (n.d.). LLMOps & MLOps: What you need to know. Insight Partners. https://www.insightpartners.com/ideas/llmops-mlops-what-you-need-to-know/
- Software Engineering Institute. (n.d.). Application of large language models (LLMs) in software engineering: Overblown hype or disruptive change? SEI Insights. https://insights.sei.cmu.edu/blog/application-of-large-language-models-llms-in-software-engineering-overblown-hype-or-disruptive-change/
- Salesforce. (2023, August 3). SDLC for prompts: The next evolution in enterprise AI development. Salesforce DevOps. https://salesforcedevops.net/index.php/2023/08/03/sdlc-for-prompts-the-next-evolution-in-enterpriseai-development/
- Valohai. (n.d.). LLMOps: Everything you need to know. Valohai Blog. https://valohai.com/blog/llmops/
- Smart Bridge. (n.d.). AI done right: Streamline development & boost value with LLMOps. Smart Bridge. https://smartbridge.com/ai-done-right-streamline-development-boost-value-llmops/
- Neptune AI. (n.d.). MLOps tools & platforms landscape. Neptune AI Blog. https://neptune.ai/blog/mlops-tools-platforms-landscape
- IBM. (n.d.). All the Ops: DevOps, DataOps, MLOps, and AIOps. IBM Developer. https://developer.ibm.com/articles/all-the-ops-devops-dataops-mlops-and-aiops/
- Arxiv. (2024). A comprehensive study on large language models and their security risks. Arxiv. https://arxiv.org/abs/2406.10300
- Cloud Security Alliance. (n.d.). CSA large language model (LLM) threats taxonomy. Cloud Security Alliance. https://cloudsecurityalliance.org/artifacts/csa-large-language-model-llm-threats-taxonomy
- Sapphire Ventures. (n.d.). GenAI infra startups. LinkedIn. https://www.linkedin.com/posts/sapphirevc_genai-infra-startups-activity-7186724761400442883-Xt3D
- AIMultiple. (n.d.). LLM security tools. AIMultiple. https://research.aimultiple.com/llm-security-tools/

- GitLab – "What is Agentic AI: Understanding AI agents for DevOps and security"
  URL: https://about.gitlab.com/topics/agentic-ai/
- DevOps.com – "Why You Shouldn't Forget Workflows With Agentic AI Systems"
  URL: https://devops.com/why-you-shouldnt-forget-workflows-with-agentic-ai-systems/
- Microsoft DevBlogs – "Agentic DevOps in action: Reimagining every phase…"
  URL: https://devblogs.microsoft.com/blog/reimagining-every-phase-of-the-developer-lifecycle
- Medium (INI8 Labs) – "Building Agentic AI Frameworks for DevOps Workflows"
  URL:https://medium.com/@INI8labs/building-agentic-ai-frameworks-for-devops-workflows-c4c1ae16239f1

Project Supporters

Project supporters lend their resources and expertise to support the goals of the project.

HADESS
KLAVAN
Precize
AWS
Snyk
Astra Security
AWARE7 GmbH
iFood
Kainos
Aigos
Cloud Security Podcast
Trellix
Coalfire
HackerOne
IBM
Bearer
Bit79
Stackarmor
Cohere
Quiq
Lakera
Credal.ai
Palosade
Prompt Security
NuBinary
Balbix
SAFE Security
BeDisruptive
Preamble
Nexus
PromptArmor
Exabeam
Modus Create
IronCore Labs
Cloudsec.ai
Layerup
Mend.io
Giskard
BBVA
RHITE
Praetorian
Cobalt
Nightfall AI