

Updated on 10/02/2026

Sign Up

OWASP LLM GenAI Training

2 days (14 hours)

Overview

The OWASP Top 10 for LLM Applications framework identifies 10 major risk categories for applications that incorporate language models. These include prompt injections, data disclosure, compromise of source data, and excessive permissions granted to AI agents —it helps identify threats and select the appropriate safeguards for each architecture.

Our OWASP Top 10 LLM 2026: Generative AI Security training will help you understand the 10 risks outlined in the 2026 edition and apply key security measures to applications that incorporate an LLM, a RAG (Retrieval-Augmented Generation) system, or an agent-based AI architecture.

You'll learn to identify attack surfaces, protect data and document access, validate model outputs, and manage tool calls according to the principle of least privilege.

The training emphasizes defense in depth: combining multiple layers of protection to reduce risks and limit the consequences of an attack, without relying solely on instructions given to the model for security.

Workshops in a sandbox environment allow you to observe vulnerabilities, apply safeguards, and evaluate their limitations. The training also covers security testing through attack simulations, resource consumption management, monitoring, and the integration of controls into a DevSecOps approach.

Objectives

- Identify the 10 risk categories from the OWASP LLM Top 10 2026 and determine their relevance based on the architecture under study
- Develop a threat model for an application incorporating an LLM, RAG, or AI agents
- Apply safeguards to limit prompt injections and the disclosure of sensitive data

- Implement controls on RAG pipelines, model outputs, and tool calls
- Evaluate the protections through security testing and prioritize fixes based on their impact
- Define consumption controls, monitoring controls, and security validation criteria to be integrated into a DevSecOps approach

Target Audience

- Application developers integrating LLMs
- Software and AI architects
- DevSecOps engineers
- Application security (AppSec) engineers
- Security analysts with a background in development

Prerequisites

- Experience in application development and the ability to read and modify Python code
- Knowledge of HTTP communications, APIs, and the JSON format
- Understanding of the principles of authentication, authorization, and secret management
- Basic understanding of LLMs, prompts, and RAG (document retrieval used to enrich the model's context)

Technical Prerequisites

- Computer with a 4-core processor, 16 GB of RAM, and 100 GB of available disk space for workshops run locally
- Windows 11 with WSL 2, macOS, or Linux, with the necessary permissions to install tools
- Docker environment with Docker Compose set up and running for workshops run locally
- Python, Git, and a code editor such as Visual Studio Code, in the versions specified prior to the session
- Stable internet connection allowing access to the lab services and the model used

Agenda for our OWASP LLM GenAI - Top 10 training, 2026 edition

[Day 1 - Morning]

Identifying risks and securing the model's environment

- Presentation of the OWASP LLM Top 10 2026: scope, risk categories, and application to LLM, RAG, and AI agent architectures
- Building a threat model: components, data flows, trust boundaries, sensitive assets, and impacts

- LLM01:2026 Prompt Injection: Direct and Indirect Prompt Injections: External Content, Documents, Tool Responses, Multimodal Inputs, and Persistent Memory
- Reducing Injection Risks: Handling Untrusted Content, Limitations of Filters and instructions, application controls, and mitigating consequences
- LLM02:2026 Sensitive Information Disclosure: minimizing data transmitted to the model, protecting secrets, and controlling logs
- LLM08:2026 Hidden Context Exposure: exposure of internal context: system prompts, internal instructions and information; managing secrets and permissions outside the model
- Hands-on workshop: analyze an architecture, test injection and disclosure scenarios, then compare the application's behavior with and without protections.

[Day 1 - Afternoon]

Securing components, data, and RAG

- LLM04:2026 Supply Chain: Supply chain security: models, dependencies, datasets, and third-party tools; provenance, integrity, versions, and component inventory
- LLM05:2026 Data and Model Poisoning: , training and fine-tuning data; source validation and change traceability
- LLM09:2026 Vector and Embedding Weaknesses: vulnerabilities related to vector bases and embeddings: unauthorized access, data exposure, and compartmentalization flaws
- Protection of the document ingestion chain: source verification, document integrity, access metadata, and handling of untrusted content
- Compliance with permissions during search and before documents are added to the model; isolation of users, organizations, caches, and memory
- Hands-on workshop: analyze the risks of an RAG pipeline, experiment with document ingestion and access controls, then verify their effects and limitations.

[Day 2 - Morning]

Controlling the outputs and actions of AI agents

- LLM03:2026 Excessive Agency: excessive capabilities for action: functionalities, permissions, and autonomy; applying the principle of least privilege to the tools and identities used
- LLM10:2026 Improper Output Handling: Inadequate handling of outputs: risks of injection, code execution, and data leakage when used by other components
- Output validation: schemas for structured data, business rules, context-appropriate encoding, and verification of generated code; distinction between format compliance and content security
- Supervision of tool, API, and connector calls, including MCP: authorized operations, argument validation, and application-side permission checks
- Limiting the consequences of unauthorized actions: execution isolation, network access restrictions, and human validation of sensitive actions with a clear display of the exact action
- Alignment with the OWASP Top 10 for Agent-Based Applications: distinction between scopes and identification of additional risks related to agents, their memory, and their interactions
- Hands-on workshop: Evaluate the outputs and actions of an application equipped with tools, apply validation and permission controls, and then verify that unauthorized operations are blocked.

[Day 2 - Afternoon]

Evaluating protections and preparing for exploitation

- LLM07:2026 Misinformation: incorrect or misleading information: verifying responses and sources, limitations of RAG, and human validation based on the consequences of an error
- LLM06:2026 Unbounded Consumption: uncontrolled consumption: depletion of resources, cost overruns, and misuse of model access; quotas, token limits, maximum durations, and caps on tool calls and agent loops
- Preparation of security tests through attack simulation: scenarios tailored to the identified risks, expected results, repeated testing, and adaptation of attacks to existing safeguards
- DevSecOps integration: non-regression testing, validation criteria, and review of changes to models, prompts, corpora, and tools
- Incident monitoring and response: logging without unnecessary exposure of sensitive data, alerts, halting processing, and rolling back to a validated version
- Prioritization of fixes: test results, impacts, mitigating measures, and residual risks
- Hands-on workshop: assessing an application's security, verifying usage limits, and document a risk assessment, tested safeguards, and priority fixes.

Target Audience

This training is intended for both individuals and companies—large or small—that wish to train their teams in new, advanced IT technology or to acquire specific industry knowledge or modern methods.

Assessment at the Start of Training

The pre-training assessment complies with Qualiopi quality standards. Upon final registration, the learner receives a self-assessment questionnaire that allows us to evaluate their estimated proficiency with various types of technologies, as well as their expectations and personal goals for the upcoming training, within the limits imposed by the selected format. This questionnaire also allows us to anticipate certain connection or internal security issues within the company (intra-company or virtual classroom) that could pose challenges for monitoring and ensuring the smooth running of the training session.

Teaching Methods

Hands-On Training: 60% Practical, 40% Theoretical. Training materials distributed in digital format to all participants.

Organization

The course alternates between theoretical input from the instructor—supported by examples and reflection sessions—and group work.

Assessment

At the end of the session, a multiple-choice quiz is administered to verify that participants have successfully acquired the skills.

Certification

A certificate will be issued to each trainee who has completed the entire training program.