CS-116 · AI Security

OWASP Top 10 for LLM Applications

A community reference that names the leading security risks in applications built on large language models.

Executive Summary

The OWASP Top 10 for LLM Applications is a community-driven reference that names the most important security risks facing software built on large language models. It gives builders, security teams, and governance professionals a shared vocabulary for the threats that generative AI introduces, from manipulated input to unsafe handling of output. This sheet describes the project at a conceptual level so teams can use it as a checklist without copying its text.

What It Is

The OWASP Top 10 for LLM Applications is a freely available, community-maintained list published by OWASP, the same nonprofit behind the widely used web application security list. It focuses specifically on the new risks that appear when an application is built around a large language model rather than traditional code. The project describes each risk, explains why it matters, and points to mitigations. It is periodically updated as the field changes. Because it is vendor-neutral and built by many contributors, it has become a common reference point for anyone securing or governing AI applications. This sheet summarizes the project's purpose and the categories of risk it addresses in our own words and does not reproduce its specific item numbers or titles.

Why It Matters

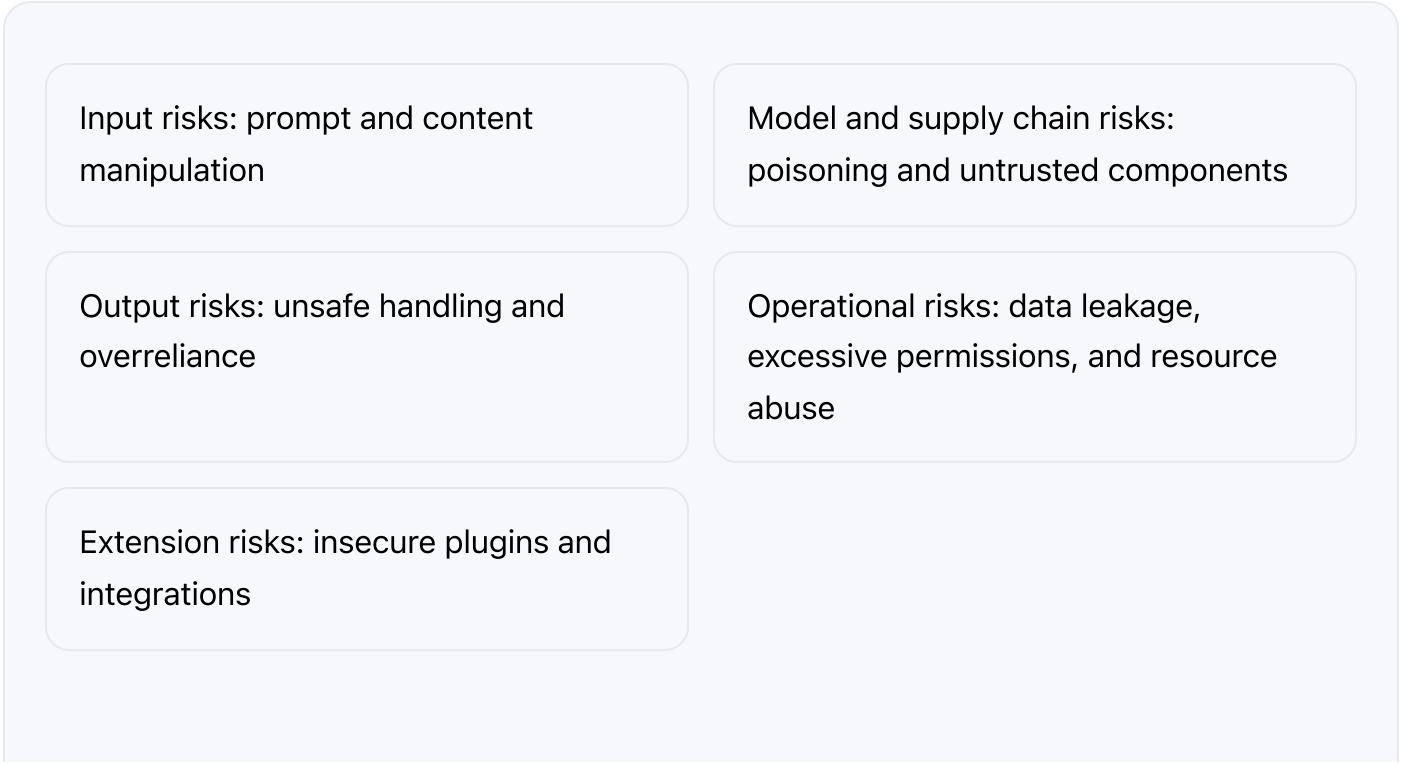
Generative AI applications fail in ways that classic security checklists do not fully capture, because the model blends instructions and data, connects to tools, and can produce output that is confidently wrong. Without a shared reference, teams tend to miss whole categories of risk.

The OWASP Top 10 for LLM Applications gives organizations a starting checklist to evaluate their AI systems, a common language for security and governance to talk to each other, and a way to show diligence to auditors and customers. For governance, risk, and compliance professionals, it is a practical bridge between abstract AI risk and concrete controls, and it pairs naturally with broader frameworks like the NIST AI Risk Management Framework.

How It Works

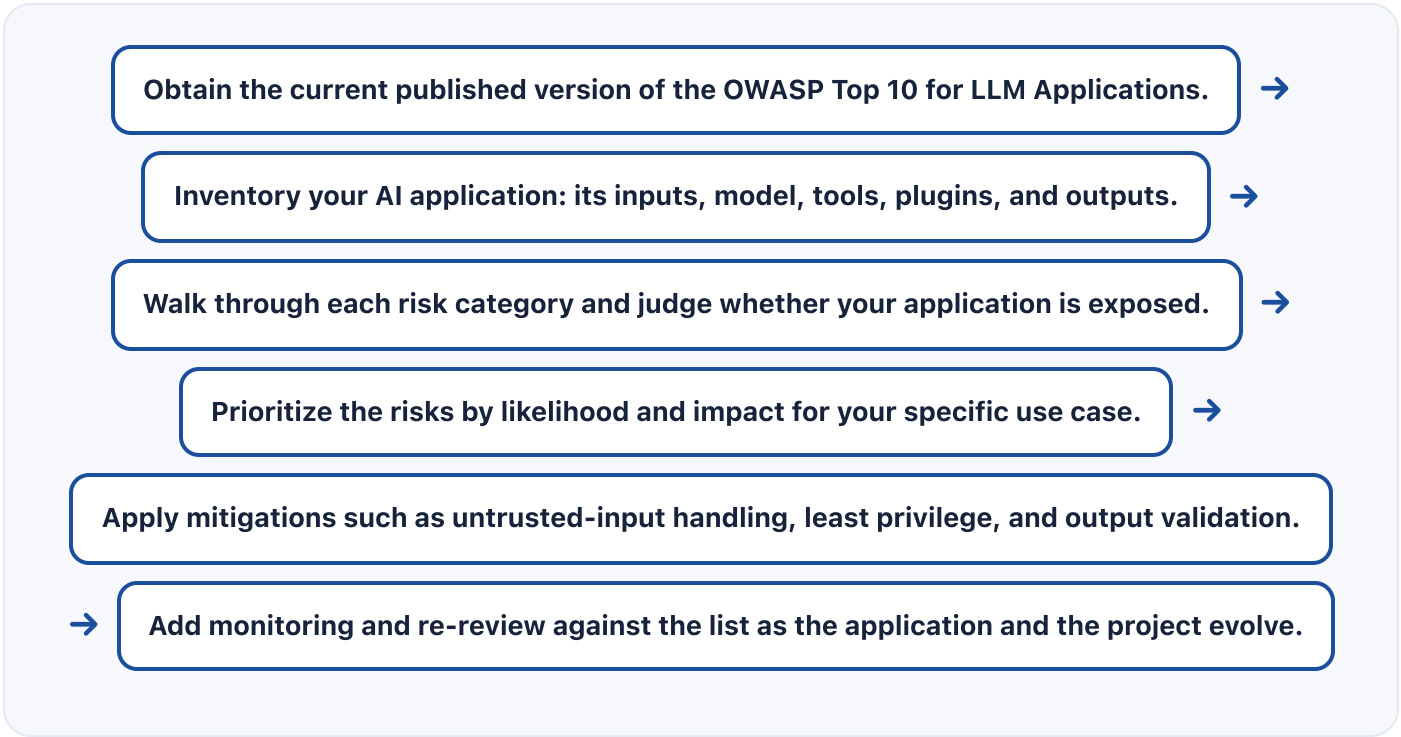
In practice, teams use the list as a lens on an AI application. The risks it covers can be grouped conceptually into a few areas. Input risks include manipulation of the prompt or of content the model reads, the concern at the heart of prompt injection. Model and supply chain risks include training data poisoning and the use of tampered or untrusted models and components. Output and downstream risks include unsafe handling of model output and overreliance on results that may be wrong. Operational risks include leaking sensitive information, excessive permissions granted to the model and its tools, resource exhaustion, and weaknesses in plugins or extensions. A team walks through each category, checks whether their application is exposed, and applies mitigations such as treating input as untrusted, enforcing least privilege, validating output, and adding monitoring. Because the project is updated over time, teams should work from the current published version and confirm the exact items there rather than relying on memory.

Architecture Diagram



The risks group conceptually into input, model and supply chain, output, and operations; work through each against your application.

Visual Workflow



Common Attacks

- Prompt and content manipulation that overrides the model's intended instructions
- Training data poisoning or use of tampered, untrusted models and components
- Unsafe handling of model output that flows into code, queries, or other systems
- Sensitive information disclosure through model responses or configuration leakage
- Abuse of excessive permissions or insecure plugins to make the model take real actions

Common Mistakes

- Treating the list as a one-time audit instead of a living checklist
- Working from memory or an outdated version rather than the current published items
- Focusing only on prompt injection and ignoring supply chain and output risks
- Checking the boxes without applying the mitigations behind them

- Assuming a traditional web application checklist already covers these AI-specific risks

Best Practices

- Use the current published version as a baseline checklist for every AI application
- Map each risk category to concrete controls in your own environment
- Combine it with a broader framework such as the NIST AI Risk Management Framework
- Treat all input and retrieved content as untrusted and enforce least privilege on tools
- Validate output before it is trusted and add monitoring for abuse
- Re-review as the application changes and as the project publishes updates

Quick Checklist

- [x] The current version of the list has been obtained and reviewed
- [x] The AI application's inputs, model, tools, plugins, and outputs are inventoried
- [x] Each risk category has been assessed for exposure
- [x] Mitigations are applied and mapped to the relevant risks
- [x] The list is combined with a broader AI risk framework
- [x] Monitoring is in place and the review is repeated as things change

Recommended Tools

LLM guardrails and filtering layer
Enforces input and output policy for the input and output risk categories

AI gateway or proxy
Centralizes access control, least privilege, and logging for model traffic

Model and data provenance tracking
Supports the model and supply chain risk categories

LLM security testing and evaluation tooling
Probes an application against the listed risk categories

Industry Standards

OWASP Top 10 for LLM Applications
The reference itself, describing the leading risks for language model applications

NIST AI Risk Management Framework (AI RMF, AI 100-1)
Broader framework that pairs with the list through Govern, Map, Measure, and Manage

MITRE ATLAS
Complements the list with adversarial techniques for AI threat modeling

Career Relevance

The OWASP Top 10 for LLM Applications is a near-universal reference for AI security engineers securing model-powered software, and it is quickly becoming expected knowledge for AI governance analysts and AI risk managers who translate its categories into policy and risk decisions. GRC analysts cite it when assessing AI products and vendors, because it offers a recognized, vendor-neutral checklist. Familiarity with it is a strong, practical signal in interviews for the roles AI-Governance-Jobs.com serves.

Interview Questions

- What is the OWASP Top 10 for LLM Applications, and who maintains it?
- How does it differ from the classic OWASP list for web applications?
- Can you describe the broad categories of risk it covers at a conceptual level?
- How would you use the list to assess a new AI application?
- How does it fit alongside the NIST AI Risk Management Framework?

Related Certifications

OWASP resources and training for LLM application security (as available)

ISC2 or ISACA AI security offerings

CompTIA Security+ (for the security foundations)

Further Reading

- OWASP Top 10 for LLM Applications
- NIST AI Risk Management Framework
- MITRE ATLAS

Key Takeaways

- The OWASP Top 10 for LLM Applications names the leading security risks in language model software.
- It is a free, vendor-neutral, community-maintained reference that is updated over time.
- Its risks group conceptually into input, model and supply chain, output, and operations.
- Use the current published version as a living checklist, not a one-time audit.
- It pairs with the NIST AI Risk Management Framework and is expected knowledge across AI security and GRC roles.

FAQ

▶ Is the OWASP Top 10 for LLM Applications the same as the classic OWASP Top 10?

▶ Is the list enough on its own to secure an AI application?

▶ Why does this sheet not print the exact item numbers and titles?

Related Careers

RELATED ROLES

Ai Security Engineer

Ai Governance Analyst

Ai Risk Manager

Grc Analyst

RELATED CERTIFICATIONS

OWASP resources and training for LLM application security

ISC2 or ISACA AI security offerings (as available)

CompTIA Security+ (for the security foundations)

CURRENT OPENINGS

Live roles are on the job board.

Browse all jobs

GUIDES & SALARY

Career guides | Role roadmaps | Salary data | Career tools

SUGGESTED LEARNING PATH

1. Ground the basics with CS-001 Cybersecurity
2. Study this sheet: **OWASP Top 10 for LLM Applications**
3. Go deeper: Prompt Injection
4. Go deeper: Model Poisoning
5. Validate it: work toward **OWASP resources and training for LLM application security**
6. Find the role: browse current openings

RELATED SHEETS

CS-111 Prompt Injection | CS-112 Model Poisoning | CS-114 AI Hallucinations | CS-115 Secure AI Adoption | CS-001 Cybersecurity