

AI security

Oláh Norbert



DECRIPT
— UNIVERSITY OF DEBRECEN —



Requirements

- A presentation on a topic of your choice in information security (self-selected topic, 10 minutes)
- Protecting a Cryptographic Algorithm
- Maximum of three absences



OWASP Top 10 for LLM Applications

- Open Worldwide Application Security Project (OWASP)
- The OWASP Foundation launched on December 1st, 2001, becoming incorporated as a United States non-profit charity on April 21, 2004.
- Lots of useful tools:
 - Zed Attack Proxy
 - Juice Shop, Proactive Controls
 - Software Assurance Maturity Model (SAMM)
 - Application Security Verification Standard (ASVS)



OWASP Top 10 for LLM Applications

- Zed Attack Proxy: The most widely used web application scanner in the world.
- Juice Shop: not a secure web application! It can be used in security trainings, awareness presentations, CTFs and as a guinea pig for safety tools! Juice Shop includes the full OWASP Top Ten vulnerabilities, as well as many other security flaws found in real applications!
- Proactive Controls: OWASP Top Ten Proactive Controls describes the most important controls and control categories that all architects and developers should absolutely, 100% include in every project.



OWASP Top 10 for LLM Applications

- **Software Assurance Maturity Model (SAMM):** software security maturity model
 - provides an effective and measurable way for organisations of all types to analyse and improve their software security posture.
 - It supports the entire software lifecycle, including development and procurement, and is technology and process independent.
 - It is deliberately designed to be flexible and risk-based.
- **Application Security Verification Standard (ASVS):** provides a basis for testing technical security audits of web applications, and a list of requirements for secure development for developers.
 - Authentication and authorization checks, data protection and encryption checks, input validation and protection against code injections, detection of business logic errors, API security checks
 - ASVS Level 1 – basic security, for all applications.
 - ASVS Level 2 – stronger controls for applications handling sensitive data.
 - ASVS Level 3 – critical systems (e.g. financial, healthcare).



OWASP AI Security and Privacy Guide

Source:

<https://owasp.org/www-project-ai-security-and-privacy-guide/>

https://owaspai.org/docs/ai_security_overview/

<https://genai.owasp.org/resource/owasp-genai-llm-top-10-2026/>

<https://www.lakera.ai/blog/owasp-top-10-for-large-language-model-applications-guide>

https://www.youtube.com/watch?v=ABmWHnFrMqI&ab_channel=OWASPFoundation

<https://www.hackerone.com/vulnerability-management/owasp-llm-vulnerabilities>



OWASP Top 10 for LLM Applications

- Structure:
 - A description of vulnerability
 - Common examples of vulnerability
 - How to prevent it
 - Examples of attack scenarios
 - References



OWASP Top 10 for LLM Applications

- Structure:
 - A description of vulnerability
 - Common examples of vulnerability
 - How to prevent it
 - Examples of attack scenarios
 - References

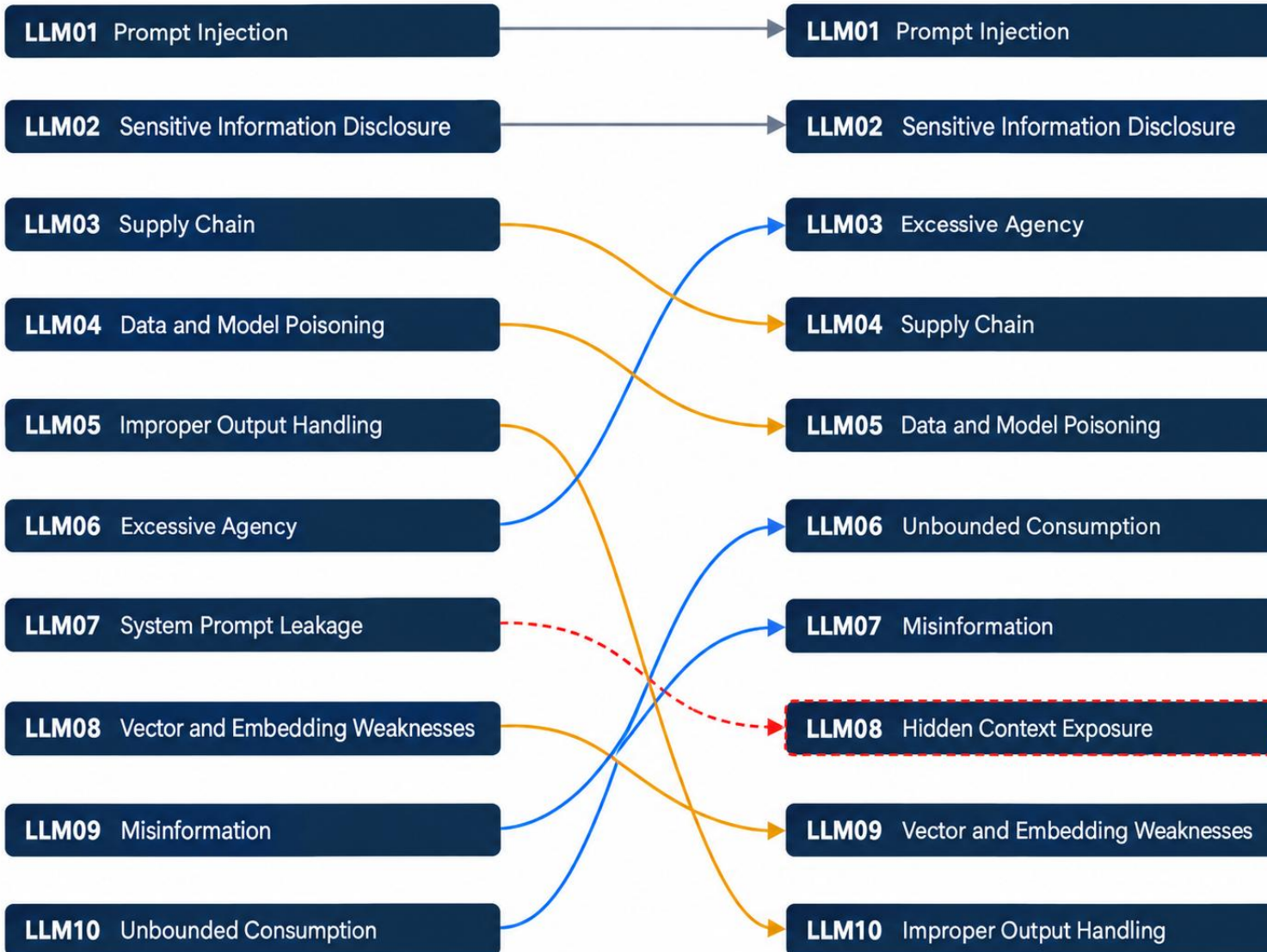


OWASP Top 10 for LLM Applications 2025

1. Prompt Injection
2. Sensitive Information Disclosure
3. Supply Chain
4. Data and Model Poisoning
5. Improper Output Handling
6. Excessive Agency
7. System Prompt Leakage
8. Vector and Embedding Weaknesses
9. Misinformation
10. Unbounded Consumption

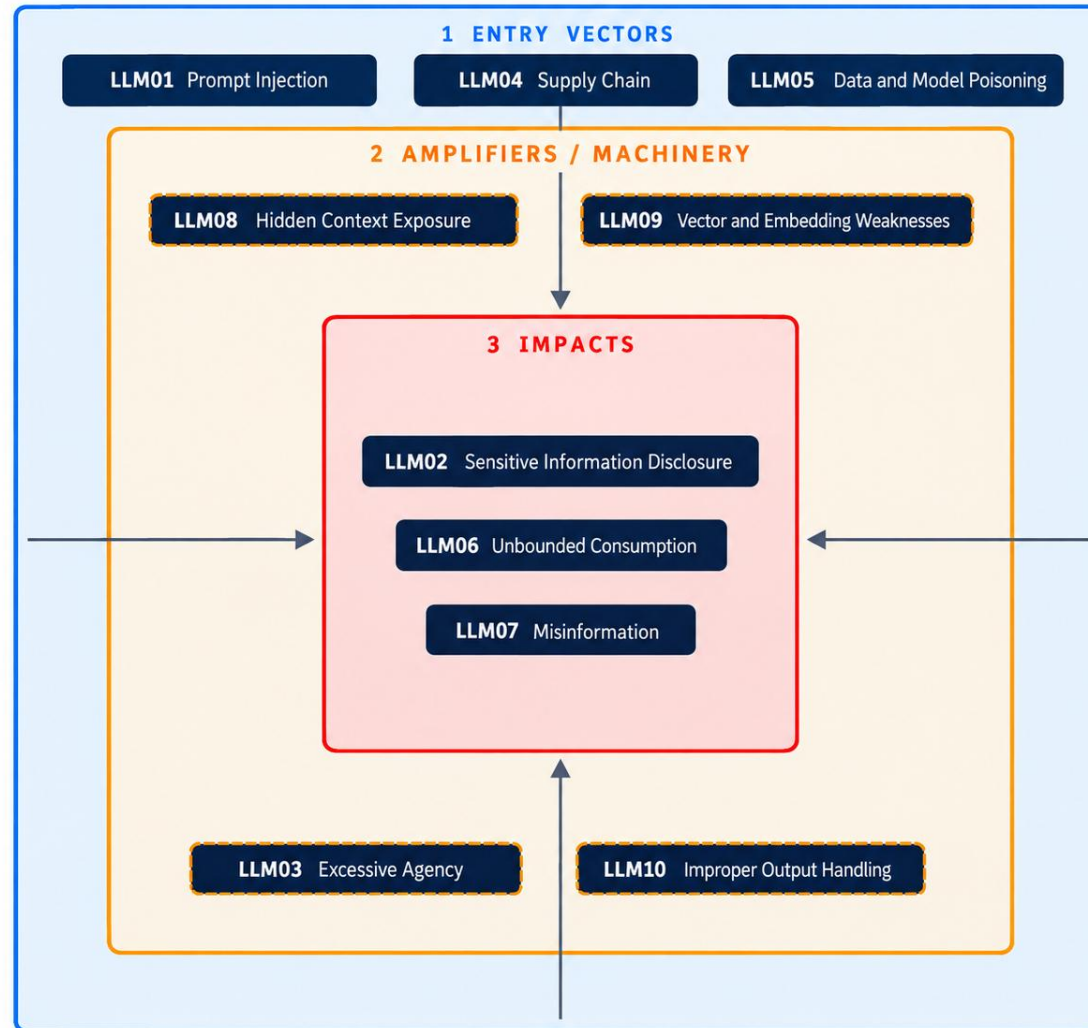


What's New in the 2026 Top 10



Concentric Threat Flow

OWASP GenAI/LLM Top 10 (2026): Concentric Threat Flow



Attack flows inward. Defense pushes outward.

Dashed chips sit in more than one layer.



OWASP Top 10 vs. LLM Top 10

Classic Web (2025)	LLM Top 10 (2026)	Connection
Broken Access Control	Excessive Agency	Unauthorised operations and access
Security Misconfiguration	Hidden Context Exposure	Configuration & instruction exposure
Software Supply Chain Failures	Supply Chain	Supply chain risks & component tampering
Injection (SQLi, XSS stb.)	Prompt Injection	Input manipulation / Instruction-data mixing
Insecure Design	Improper Output Handling	Design errors, unvalidated outputs
Software or Data Integrity Failures	Sensitive Information Disclosure	Data integrity & data protection breach



OWASP Top 10 vs. LLM Top 10

Classic Web (2025)	LLM Top 10 (2026)	Connection
Identification & Auth Failures	Excessive Agency	Identity & privilege abuse
Data Integrity Failures	Data & Model Poisoning	Poisoned data, integrity
Logging & Monitoring Failures	Misinformation	Lack of recognition of malfunction
Mishandling of Exceptional Conditions	Unbounded Consumption	Resource overuse & abnormal state handling
-	Vector & Embedding Weaknesses	New, LLM-specific risk



Thank you for your attention!

