About the MCP Top 10

As AI systems become increasingly integrated into software supply chains, enterprise applications, and security infrastructure, the need for structured, secure, and interpretable model interaction layers is paramount. The Model Context Protocol (MCP) is emerging as a framework to define the operational, contextual, and behavioral boundaries of AI models. However, with the power and flexibility of MCPs comes a new class of vulnerabilities and attack surfaces that remain underexplored.

This OWASP Top 10 for MCP outlines the most critical security concerns arising in the lifecycle of MCP-enabled systems—spanning from model misbinding, context spoofing, and prompt-state manipulation to insecure memory references and covert channel abuse. These risks are amplified in scenarios involving agentic AI, model chaining, multi-modal orchestration, and dynamic role assignment.

By mapping the top 10 MCP-related vulnerabilities and offering concrete recommendations for secure design, implementation, and auditing practices, this project aims to equip AI developers, ML engineers, and security practitioners with the insights necessary to build context-aware and attack-

MCP Top 10 Information

Incubator Project
Type of Project
Version 0.0.0
Server
Client

Downloads or Social Links

Download
Meetup

Code Repository

repo

Change Log

and industry feedback.

Road Map

Road Map Phase 1 – Drafting Create an initial draft of requirements that cover the industry aspects.

Phase 2 – Community Review and Feedback Publish the draft in a public repository for the community to review. Inputs from the community

Phase 3 – Beta Release and Pilot Testing - We are here right now Release a “beta” version of MCP Top 10. Gather feedback on usability and coverage.

Next Phase

Phase 4 – Final Release Incorporate feedback from pilot testing.

Phase 5 – Continuous Improvement Periodically release updated versions

Top 10

- MCP01:2025 - Token Mismanagement & Secret Exposure
- MCP02:2025 - Privilege Escalation via Scope Creep
- MCP03:2025 - Tool Poisoning
- MCP04:2025 - Software Supply Chain Attacks & Dependency Tampering
- MCP05:2025 - Command Injection & Execution
- MCP06:2025 - Prompt Injection via Contextual Payloads

Vandana Verma Sehgal
Liran Tal

Upcoming OWASP Global Events

OWASP Global AppSec EU 2026 - Vienna, Austria

- June 22-26, 2026

OWASP Global AppSec USA 2026 - San Francisco, CA

- November 2-6, 2026

OWASP Global AppSec EU 2027 - Vienna, Austria

- June 21-25, 2027

OWASP Global AppSec USA 2027 - Atlanta, GA

- September 20-24, 2027

OWASP Global AppSec EU 2028 - Vienna, Austria

- June 19-23, 2028

- MCP10:2025 - Context Injection & Over-Sharing

Overview

Title	Description
MCP01 - Token Mismanagement & Secret Exposure	Hard-coded credentials, long-lived tokens, and secrets stored in model memory or protocol logs can expose sensitive environments to unauthorized access. Attackers may retrieve these tokens through prompt injection, compromised context, or debug traces, leading to full compromise of connected systems.
MCP02 - Privilege Escalation via Scope Creep	Temporary or loosely defined permissions within MCP servers often expand over time, granting agents excessive capabilities. An attacker exploiting weak scope enforcement can perform unintended actions such as repository modification, system control, or data exfiltration.
MCP03 -	Tool poisoning occurs when an adversary compromises the tools, plugins, or their outputs that an AI

MCP04 - Software Supply Chain Attacks & Dependency Tampering	A compromised dependency can alter agent behavior or introduce execution-level backdoors.
MCP05 - Command Injection & Execution	Command injection occurs when an AI agent constructs and executes system commands, shell scripts, API calls, or code snippets using untrusted input whether from user prompts, retrieved context, or third-party data sources without proper validation or sanitization.
MCP06 - Prompt Injection via Contextual Payloa	This risk is analogous to classic injection attacks (e.g., XSS, SQLi), but in the MCP world the "interpreter" is the model and the "payload" is text (or any content that becomes text after OCR/ processing). Because models are designed to follow natural-language

MCP07 - Insufficient Authentication & Authorization	authorization occur when MCP servers, tools, or agents fail to properly verify identities or enforce access controls during interactions. Since MCP ecosystems often involve multiple agents, users, and services exchanging data and executing actions, weak or missing identity validation exposes critical attack paths.
MCP08 - Lack of Audit and Telemetry	Limited telemetry from MCP servers and agents impedes investigation and incident response. Maintain detailed logs of tool invocations, context changes, and user-agent interactions with immutable audit trails.
MCP09 - Shadow MCP Servers	“Shadow MCP Servers” refer to unapproved or unsupervised deployments of Model Context Protocol instances that operate outside the organization’s formal security governance.Much like Shadow IT, these rogue MCP nodes are often spun up by developers, research teams, or data scientists for experimentation, testing, or convenience frequently using

MCP10 - Context Injection & Over-Sharing	(MCP), “context” represents the working memory that stores prompts, retrieved data, and intermediate outputs across agents or sessions. When context windows are shared, persistent, or insufficiently scoped, sensitive information from one task, user, or agent may be exposed to another. This phenomenon known as context over-sharing turns convenience into a liability.

Data Sources

Licensing

Project Leaders

- Vandana Verma Sehgal (Twitter: @infosecvandana)

Edit on GitHub

Spotlight: Checkmarx