# OWASP LLM TOP 10 V2 + MITRE ATLAS V5.4

Monthly Attack Battery + Quarterly Deep-Scenario Red-Team · Continuous Coverage

## OWASP LLM TOP 10 V2 — MONTHLY ATTACK BATTERY

9 RISKS COVERED

| RISK | TECHNIQUE | FOUNDATION | STANDARD | ENTERPRISE |
|---|---|---|---|---|
| **LLM01** Prompt Injection | Direct + Indirect (XPIA) + Multi-turn manipulation | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + custom + continuous |
| **LLM02** Insecure Output | Tool-output sanitization + cross-context bleed | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + continuous |
| **LLM03** Training Data Poisoning | Customer-supplied training corpus poisoning | ○ Quarterly deep only | ✓ Quarterly deep | ✓ Quarterly deep + continuous |
| **LLM05** Supply Chain | MCP server provenance + tool-schema integrity + dependency confusion | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + continuous |
| **LLM06** Sensitive Info Disclosure | PHI / NPI / MNPI exfiltration via tool responses + retrieval leakage | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + continuous |
| **LLM07** Insecure Plugin Design | MCP plugin authorization scope + tool-argument sanitization | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + continuous |
| **LLM08** Excessive Agency | Tool over-invocation + approval-gated destructive + audit review | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + continuous |
| **LLM09** Overreliance | Hallucination detection + groundedness regression + clinician/decisioner reliance | ✓ Monthly battery | ✓ Monthly + custom | ✓ Monthly + continuous |
| **LLM10** Model Theft | Model extraction probing + prompt-extraction attacks | ○ Quarterly deep only | ✓ Quarterly deep | ✓ Quarterly deep + continuous |

✓ **Included monthly** ○ **Quarterly deep only** **Custom** = additional evaluator authoring STANDARD+ **Continuous** = scheduled PyRIT ENTERPRISE

**MONTHLY OWASP BATTERY · CONTINUOUSLY REFRESHED · CUSTOMER OWNS EVALUATORS**

OWASP LLM Top 10 v2 evolves quarterly · jailbreak corpora publish bypasses weekly · static test suites decay in 90 days

# MITRE ATLAS V5.4 · 11 AGENT-SPECIFIC TTPS

Added March 2026 · Monthly TTP simulation + Quarterly deep-scenario · Anthropic MCP Working Group intelligence

| TECHNIQUE | TTP / TACTIC | COVERAGE |
|---|---|---|
| T1<br>Publish Poisoned AI Agent Tool | TA0001 Initial Access · MCP server poisoning | **Monthly + Quarterly Deep**<br>ALL TIERS |
| T2<br>Escape to Host | TA0008 Lateral Movement · Container escape from MCP runtime | **Monthly + Quarterly Deep**<br>ALL TIERS |
| T3<br>Memory Manipulation | TA0040 Impact · Agent memory corruption + context-window injection | **Monthly + Quarterly Deep**<br>ALL TIERS |
| T4<br>Tool Poisoning Chain | TA0006 Credential Access via tool-leaked secrets | **Monthly + Quarterly Deep**<br>ALL TIERS |
| T5<br>Cross-Agent Adversarial Handoff | Foundry A2A protocol abuse · Cross-tenant pivots | **Quarterly Deep**<br>ALL TIERS · CONTINUOUS (ENTERPRISE) |
| T6<br>RAG Corpus Adversarial Injection | TA0042 ML Model Access · Foundry IQ corpus poisoning | **Monthly + Quarterly**<br>FOUNDATION+ |
| T7-11<br>5 Additional ATLAS v5.4 Techniques | Adversarial probing + Information Integrity HIGH | **Monthly battery rotation**<br>ALL TIERS |

## ANTHROPIC MCP SECURITY WORKING GROUP INTELLIGENCE

Kriv is **CPN-approved participant** · Direct early access to draft hardening guidance

**WORKING GROUP DRAFT**
Feb 2026 hardening guidance
Q2 2026 quarterly updates

→

**KRIV INTELLIGENCE FEED**
Standard tier monthly
Enterprise tier priority

→

**CUSTOMER PROTECTION**
Applied to monthly battery
+ quarterly deep · Pre-emptive

## QUARTERLY DEEP-SCENARIO RED-TEAM SCOPE · ALL TIERS

**Q1**
**Business-Logic Abuse**
Chained tool exploitation + workflow manipulation

**Q2**
**Poisoned-Tool Supply Chain**
MCP server compromise simulation + tool-registry abuse

**Q3**
**Cross-Agent Adversarial Handoffs**
A2A protocol abuse + cross-tenant pivots + multi-cloud federation

**Q4**
**Custom Scenario** Enterprise
Customer-specific threat model · 3-day immersive