Appendix: OWASP Agentic Risks Mapped to Quality Gates

OWASP's 2026 Top 10 for Agentic Applications catalogs common failure and attack patterns in autonomous AI systems. This appendix maps the current OWASP risk names to the quality gate tiers in this book. The mapping is interpretive: OWASP defines the risks; the gate-tier alignment below is the author's.

#	OWASP Risk	What It Means	Primary Gate Tier(s)	Primary Triangle Vertex
1	Agent Goal Hijack	The agent's objective is redirected by prompt injection or malicious instructions	Tier 1 (ground against a source of truth) + Tier 3 (permission boundaries)	Verification quality
2	Tool Misuse	The agent uses a tool in a harmful or unintended way	Tier 2 (hard invariants) + Tier 3 (approval and sandbox rules)	Verification quality
3	Identity and Privilege Abuse	The agent or its tools have more access than the task requires	Tier 3 (least privilege, scoped credentials)	Verification quality
4	Agentic Supply Chain Vulnerabilities	A model, package, extension, or MCP/tool dependency is compromised	Tier 0 (scan early) + Tier 3 (approved registries and trust boundaries)	Verification quality
5	Unexpected Code Execution	The agent executes code paths or commands the operator did not intend	Tier 2 (forbidden actions as invariants) + Tier 3 (sandbox and approval)	Verification quality
6	Memory and Context Poisoning	The agent's memory or context is manipulated, causing safety or task drift	Tier 4 (behavioral baselines and drift monitoring)	Verification quality
7	Insecure Inter-Agent Communication	Agent-to-agent messages are unauthenticated, unverified, or unsafe to trust	Tier 3 (identity, provenance, encryption)	Verification quality
8	Cascading Failures	Multi-step or multi-agent systems amplify local failures into broader outages	Tier 2 (containment invariants) + Tier 4 (behavioral monitoring)	Verification quality + Cost

#	OWASP Risk	What It Means	Primary Gate Tier(s)	Primary Triangle Vertex
9	Human-Agent Trust Exploitation	The system manipulates or over-leverages human trust	Tier 3 (approval boundaries) + Tier 4 (behavior and outcome monitoring)	Intent clarity + Verification quality
10	Rogue Agents	An agent continues operating outside intended goals or controls	Tier 2 (kill-switch invariants) + Tier 3 (containment) + Tier 4 (behavioral detection)	Verification quality + Cost

COVERAGE BY TIER

No single gate tier covers the OWASP list. The tiers are cumulative because the risks are layered:

- **Tier 0** is strongest on supply-chain and pre-execution scanning
- **Tier 2** adds hard safety boundaries and failure containment
- **Tier 3** is load-bearing for identity, privilege, tool use, and communication trust
- **Tier 4** is necessary for drift, memory poisoning, cascading behavior, and human-trust failures that only become visible over time

In practice, Tier 3 appears in most rows because the OWASP 2026 list is heavily about trust boundaries: who the agent is, what it can touch, and what it is allowed to believe.

USING THIS TABLE

Run this mapping against your current gate infrastructure quarterly. For each OWASP risk, ask: do we have the corresponding gate tier in place, and is the affected Triangle vertex being measured? Any row where both answers are "no" is a risk finding.

The OWASP Top 10 for Agentic Applications is a living document. Review it as agent capabilities and attack surfaces evolve. The gate tiers are stable; the specific risks they need to catch will change.

Source: OWASP, "Top 10 for Agentic Applications for 2026." https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/