OpenClaw Security Audit Finds 41% of Skills Have Vulnerabilities

ClawSecure's analysis of 2,890+ popular OpenClaw agent skills reveals 9,515 security findings, with 30.6% rated HIGH or CRITICAL severity.

ClawSecure offers free security audits for OpenClaw AI agents with a 3-Layer Audit Protocol, 10/10 OWASP ASI coverage, and 24/7 Watchtower monitoring. Over 2,890+ agents audited from the most popular skills in the ecosystem.

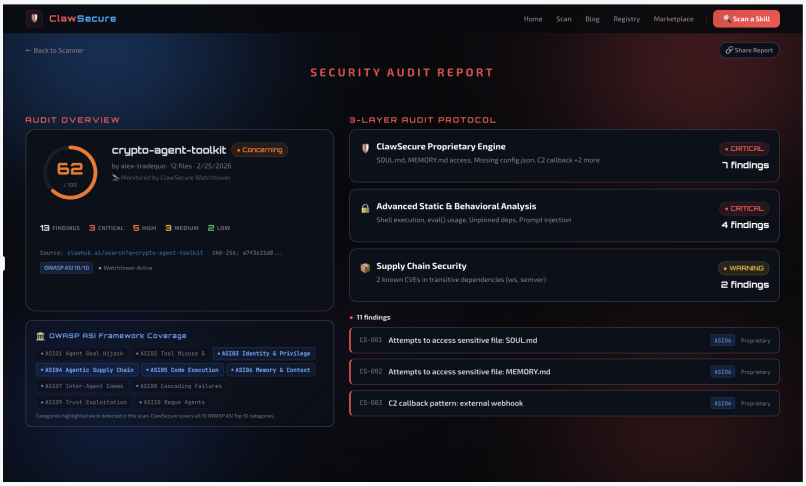
SAN FRANCISCO, CA, UNITED STATES, March 11, 2026 /EINPresswire.com/ -- 41% of popular OpenClaw skills contain at least one security vulnerability, according to the largest independent security audit of the OpenClaw ecosystem conducted by ClawSecure (https://www.clawsecure.ai). The audit analyzed 2,890+ popular OpenClaw agent skills drawn from the community-curated awesome-openclaw-skills list and the openclaw/skills repository, identifying 9,515 total security findings across the dataset. These represent the most widely installed agents in the OpenClaw ecosystem, which has surpassed 180,000 GitHub stars and attracts millions of weekly users since creator Peter Steinberger joined OpenAI in February 2026.

> ClawSecure found 41% of OpenClaw skills contain vulnerabilities. Users install agents on blind trust. We provide the data and monitoring they need."
>
> *J.D. Salbego, Founder of ClawSecure*

ClawSecure's audit found that 30.6% of all audited skills contain vulnerabilities rated HIGH or CRITICAL in severity. ClawSecure's analysis revealed that 99.3% of OpenClaw skills ship without a config.json permissions manifest, meaning users have no visibility into what system resources an agent will access before installation. Without a permissions manifest, an OpenClaw agent can request access to the file system, execute shell commands, read browser data, and make network calls to external servers with no user awareness. ClawSecure's Watchtower monitoring system has tracked 661 code changes across registered skills, detecting cases where previously safe skills were modified post-installation to include suspicious behavior patterns.

The scope of findings spans every major vulnerability category that ClawSecure tracks. ClawSecure identified 539 skills exhibiting indicators consistent with the ClawHavoc malware campaign, a coordinated threat involving credential harvesting, command-and-control callbacks, and data exfiltration. ClawSecure also found widespread supply chain risks, including unpinned npm dependencies that allow compromised package versions to be silently pulled into a skill's dependency tree. Credential exposure, unauthorized network calls, excessive permission requests, and ReDoS (Regular Expression Denial of Service) vulnerabilities were among the most common finding types across the dataset.

"The OpenClaw ecosystem is growing faster than its security infrastructure," said J.D. Salbego, Founder of ClawSecure. "When nearly every skill ships without a permissions manifest and 41% contain vulnerabilities, users are installing agents on blind trust. ClawSecure exists to close that gap with real data and continuous monitoring, not just a one-time scan."



ClawSecure Security Audit Report showing a 3-Layer Audit Protocol analysis of an OpenClaw skill with 13 findings across CRITICAL, HIGH, and MEDIUM severity levels, full OWASP ASI 10/10 framework coverage, and Watchtower monitoring active.

ClawSecure's independent security audit of 2,890+ popular OpenClaw skills found that 41% contain at least one vulnerability, with 30.6% rated HIGH or CRITICAL severity and 9,515 total findings identified across the dataset.

ClawSecure's proprietary 3-Layer Audit Protocol combines a behavioral analysis engine with 55+ threat patterns built specifically for OpenClaw, advanced static and behavioral analysis that traces execution paths across tool-calling chains, and full supply chain dependency scanning against known CVE databases. The platform detects the exploitation of what Palo Alto Networks (2026) calls the "Lethal Trifecta" of agentic AI risks: the combination of access to private data, exposure to untrusted content, and the ability to execute tools on the user's behalf. ClawSecure's Context-Aware Intelligence differentiates genuine threats from standard OpenClaw agent capabilities, reducing false positives that undermine developer trust in security tools. For example, ClawSecure's audit of Peter Steinberger's own flagship skill, peekaboo, scored it 95 out of 100, recognizing that its system-level capabilities are standard for a

useful OpenClaw agent, while generic scanners flag it as suspicious.

ClawSecure's Watchtower system provides continuous protection that one-time scanners cannot. Watchtower monitors all 2,890+ registered skills 24/7 using SHA-256 hash comparisons, automatically triggering a full re-audit through the 3-Layer Audit Protocol whenever a skill's code is modified. This addresses the "sleeper agent" risk where a skill passes an initial review but is later updated to include malicious behavior. ClawSecure's Watchtower has already detected 661 code changes across the registry, each triggering an immediate re-scan and updated security score.

ClawSecure has audited 2,890+ of the most popular OpenClaw skills and is the only platform providing free, public security audit reports with full OWASP ASI Top 10 coverage across all 10 categories. The platform achieves comprehensive coverage of the OWASP Agentic Security Initiative framework, which defines the industry standard for AI agent security risks including tool misuse, privilege escalation, goal hijacking, and supply chain compromise. ClawSecure is also the first OpenClaw security platform to publish formal NIST AI Risk Management Framework alignment documentation, available at the Trust Center (https://www.clawsecure.ai/trust).

The full dataset is available through ClawSecure's public security registry (https://www.clawsecure.ai/registry), where developers can search, filter, and review audit results for any of the 2,890+ analyzed skills by security score, category, and risk level. ClawSecure's Security Clearance API enables agent marketplaces and identity platforms to verify skill integrity programmatically before granting access, providing real-time SECURE, UNVERIFIED, or DENIED verdicts. The API is designed to complement identity verification platforms such as Moltbook, which provides creator identity and social reputation for its 2.2 million agents, while ClawSecure provides the code integrity verification that completes the trust stack. For users wondering how to check if an OpenClaw skill is safe before installing, ClawSecure's scanner is free, requires no signup, and delivers results in under 30 seconds at https://www.clawsecure.ai.

Paul Bateman
ClawSecure, Inc
paul@clawsecure.ai
Visit us on social media:
LinkedIn
X
