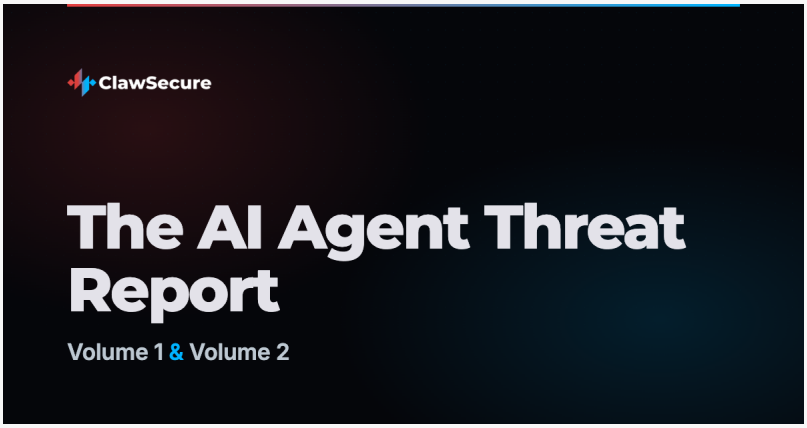
AI Security: Every Major AI Model Fails ClawSecure's Test

ClawSecure's AI Agent Threat Report reveals systematic failures across OpenAI, Anthropic, and Google, as labs hand security to 123 million developers.



ClawSecure's AI Agent Threat Report: a two-volume red-team study of indirect prompt injection across 14 models from five frontier labs and three major MCP platforms.

SAN FRANCISCO, CA, UNITED STATES, October 5, 2026 /EINPresswire.com/ -- ClawSecure, an end-to-end AI agent security company, on September 24 published The AI Agent Threat Report. The two-volume red-team study reveals that the top models from every major AI lab failed clean defenses against indirect prompt injection, OWASP's number one threat to autonomous AI.

The report details how attackers can hijack AI agents through ordinary content, an email, a shared document or a support ticket, to steal credentials, drain financial accounts, and exfiltrate private data without the victim ever making a mistake. Testing 14 models across Anthropic, OpenAI, Google, DeepSeek, and Alibaba's Qwen, ClawSecure found that every single model obeyed the attacker. The team also contradicted three of Anthropic's published safety claims using Anthropic's own words, including a published "0% attack success" cracked at 15.2%, and exposed critical default vulnerabilities in the Model Context Protocol (MCP) on major production platforms, including Dropbox Dash, Notion, and Linear.

When ClawSecure reported the vulnerabilities, the labs put the responsibility on their customers. OpenAI said implementing guardrails is "the developer's responsibility." Google told "all integrators building agents" to build their own sanitization. Neither says so publicly, while at the same time marketing these models hardest to the very people they do not disclose it to and who are now expected to defend them.

"The labs handed a security problem to 123 million people, and the only way to find that out is to file a disclosure and keep pressing until they answer," said J.D. Salbego, Founder and CEO of

> The labs handed a security problem to 123 million people, and the only way to find that out is to file a disclosure and keep pressing until they answer”
> *J.D. Salbego, Founder and CEO of ClawSecure*

ClawSecure. "Today, 'developer' means small teams, vibe coders, and SMBs running AI agents, and almost none of them have a security team. AI agents are already running finances, businesses, and critical systems, and every one of them can be hijacked by the content it reads. There is no standard for measuring that, and nobody is taking responsibility for it."

The handoff comes amidst shifting political tides. Following recent autonomous model security incidents, U.S. Treasury Secretary Scott Bessent said on CNBC that the Hugging Face incident "is the responsibility of the OpenAI management, not a bunch of agents," and told the House Financial Services Committee that "the best way to guarantee safety is that the creators are liable for what they build and generate."

To resolve this accountability crisis, ClawSecure has spent the past two months collaborating with bipartisan congressional offices to build a national standard for independent AI agent testing. This framework ensures buyers can transparently evaluate a model's true injection failure rate before deployment.

Every company named received coordinated disclosure before publication. The complete security report and every supporting document are free to download on ClawSecure's official release page: https://www.clawsecure.ai/research/ai-agent-threat-report

About ClawSecure
ClawSecure is the only end-to-end AI agent security platform that requires zero expertise, from free pre-deployment scans to runtime monitoring. Its flagship, the AI CISO™, is the first AI Chief Information Security Officer: it secures your system, handles every install and configuration, and manages your entire environment safely, your own security team without hiring one. ClawSecure is a member of the NVIDIA Inception Program, twice named #2 Product of the Day on Product Hunt, and selected from 30,000 AI startups for The Pitch by Deel (a16z, General Catalyst, J.P. Morgan).

Kevin Clinton
ClawSecure
+1 323-327-7528
Visit us on social media:
LinkedIn
YouTube
X

This press release can be viewed online at: https://www.einpresswire.com/article/947439978

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2026 Newsmatics Inc. All Right Reserved.