OpenClaw: We Are Not Ready

On the Uncontrolled Rise of Autonomous AI Agents,
the Asymmetry of Creation and Governance,
and the Geopolitics of Inaction

Opinion Article

February 2026

> *"Once men turned their thinking over to machines in the hope that this would set them free. But that only permitted other men with machines to enslave them."*
>
> — Frank Herbert, *Dune* (1965)

Abstract

In January 2026, a single open-source developer released OpenClaw, an autonomous AI agent capable of executing system commands, managing files, reading emails, and acting on behalf of users across multiple platforms. Within weeks, the project accumulated over 150,000 GitHub stars, was installed on tens of thousands of machines worldwide, and triggered emergency security advisories from CrowdStrike, Cisco, Bitdefender, Trend Micro, and other major cybersecurity firms. This article examines what OpenClaw reveals about society's preparedness for autonomous AI agents. Beginning with a firsthand account of one developer's attempt to evaluate the tool's safety, the article extrapolates from OpenClaw's architecture to scenarios of weaponized agentic AI, and concludes with an analysis of why the current geopolitical landscape — characterized by deregulation, strategic competition, and the erosion of multilateral cooperation — is structurally incompatible with the governance frameworks that autonomous AI demands. The central argument is not that OpenClaw itself is uniquely dangerous, but that it serves as a proof of concept demonstrating that a single individual can now deploy autonomous software agents into millions of lives with no institutional checkpoint along the way — and that we lack the technical, legal, and political infrastructure to manage the consequences.

Keywords: *OpenClaw, agentic AI, autonomous agents, AI safety, prompt injection, open-source security, AI governance, hybrid warfare, geopolitics of AI*

Contents

1. Part I: The Personal — A Developer's Reckoning

> *"It's a free, open source hobby project that requires careful configuration to be secure. It's not meant for non-technical users."*[1]

1.1. The Decision Not to Install

Last week, I decided not to install the most popular AI tool on the internet. It took me several hours of research, consulting security reports from five major cybersecurity firms, and planning sandboxed execution environments just to reach that conclusion. I am a software developer. I work daily with WSL[2], Docker[3], and containerized environments. I understand networking, file permissions, and system isolation. Most people who installed OpenClaw in those same hours did none of this evaluation.

The tool in question was OpenClaw[4] — an open-source personal AI assistant that had, by the time I encountered it, already accumulated over 150,000 stars on GitHub and was being celebrated across social media as "Jarvis for real" and "AI with hands." The premise was compelling: a locally-hosted agent that connects to large language models[5] and can autonomously execute tasks — send emails, manage calendars, run terminal commands, organize files — all through messaging apps like WhatsApp, Telegram, and Discord.

What first drew my attention was the claim of privacy. OpenClaw runs on your own hardware, its creators emphasized. Your data stays local. For a developer tired of sending proprietary code to cloud services, this was attractive. But as I investigated further, the picture darkened considerably.

1.2. What the Security Community Found

The warnings came from virtually every major cybersecurity firm, arriving within days of OpenClaw's viral adoption.

---

[1] Peter Steinberger, creator of OpenClaw, in a statement to CNBC. See: CNBC, February 2, 2026.

[2] Windows Subsystem for Linux (WSL): A compatibility layer that allows running Linux environments natively on Windows. See: Wikipedia: Windows Subsystem for Linux.

[3] Docker: A platform for developing, shipping, and running applications inside lightweight, isolated containers. See: Wikipedia: Docker.

[4] OpenClaw (formerly Clawdbot and Moltbot) is a free and open-source autonomous AI agent developed by Peter Steinberger. See: Wikipedia: OpenClaw and GitHub repository.

[5] Large Language Model (LLM): A type of artificial intelligence model trained on vast amounts of text data, capable of understanding and generating human language. See: Wikipedia: Large language model.

CrowdStrike[6] warned that if employees deploy OpenClaw on corporate machines and leave it misconfigured, "it could be commandeered as a powerful AI backdoor agent capable of taking orders from adversaries." They released an enterprise-wide search-and-removal content pack — the kind of tool typically reserved for actual malware.

Cisco[7] was blunt: "From a capability perspective, OpenClaw is groundbreaking. From a security perspective, it's an absolute nightmare." Their researchers found that a popular third-party skill called "What Would Elon Do?" was functionally malware, silently exfiltrating user data to an external server.

Bitdefender[8] scanned the ClawHub skill registry[9] and found nearly 900 malicious entries — approximately 20% of all packages — with some accounts uploading poisoned skills every few minutes using automated scripts.

Trend Micro[10] identified that misconfigurations in OpenClaw instances had exposed millions of records, including API tokens, email addresses, private messages, and credentials for third-party services.

A critical vulnerability (CVE-2026-25253)[11] was discovered that allowed attackers to fully hijack a user's OpenClaw instance by tricking them into visiting a malicious website. The attack exploited a cross-site WebSocket hijacking[12] flaw because OpenClaw's server did not validate connection origins.

1.3. The Privacy Illusion

Perhaps most revealing was the gap between OpenClaw's privacy claims and the architectural reality. While the agent framework runs locally, it requires a cloud-based LLM to function effectively. When OpenClaw processes your documents — summarizes your tax files, reviews your code, drafts an email from your notes — the content of those files is transmitted to whichever cloud provider you have configured: Anthropic, OpenAI, or others. The privacy is partial at best. The agent logic stays local; your data does not.

Running a fully local model is technically possible using tools like Ollama[13], but the trade-offs are severe: local models require significant hardware resources (particularly

[6]CrowdStrike: A leading American cybersecurity technology company. See: CrowdStrike Advisory on OpenClaw.

[7]See: Cisco Blog: "Personal AI Agents like OpenClaw Are a Security Nightmare".

[8]Bitdefender: A Romanian cybersecurity company. See: Bitdefender Technical Advisory on OpenClaw.

[9]ClawHub: OpenClaw's community-driven registry of plugins (called "skills") that extend the agent's capabilities. Analogous to npm for Node.js or PyPI for Python.

[10]See: Trend Micro: "Viral AI, Invisible Risks: What OpenClaw Reveals About Agentic Assistants".

[11]CVE (Common Vulnerabilities and Exposures): A standardized system for identifying publicly known cybersecurity vulnerabilities. See: Wikipedia: CVE. The OpenClaw vulnerability is documented at SecurityWeek.

[12]WebSocket: A communication protocol providing full-duplex channels over a single TCP connection. Cross-site WebSocket hijacking occurs when a server fails to validate the origin of WebSocket connections, allowing malicious websites to establish unauthorized connections. See: Wikipedia: WebSocket.

[13]Ollama: An open-source tool for running large language models locally on personal hardware. See: Wikipedia: Ollama.

GPU memory), deliver noticeably lower quality than frontier cloud models, and are impractical for most users. The “privacy-first” marketing, while not technically false, creates expectations that the architecture cannot fully deliver.

1.4. The Sandbox That Almost Was

Before deciding against installation, I spent considerable time designing a safe testing environment. The plan involved running OpenClaw inside a Docker container on an isolated internal network (`--network=none` or Docker’s `--internal` flag), with read-only bind mounts[14] to specific source code folders, a local LLM running via Ollama on the same network to avoid any cloud data transmission, and no connections to real email, calendar, or messaging accounts.

This configuration — which I believe represents the minimum viable security posture for experimenting with OpenClaw — requires familiarity with Docker networking, Linux file permissions, GPU passthrough, and container orchestration. It is not reasonable to expect this of the average user who encounters OpenClaw through an enthusiastic social media post.

The gap between what OpenClaw requires for safe operation and what most users are equipped to provide is, in microcosm, the central problem this article addresses.

[14] Bind mount: A Docker mechanism that maps a specific host directory into a container, optionally as read-only, allowing controlled file sharing between the host and the containerized application.

2. Part II: The Extrapolation — From Personal Assistant to Digital Weapon

> *"Agents are not only going to change how everyone interacts with computers. They're also going to upend the software industry, bringing about the biggest revolution in computing since we went from typing commands to tapping on icons."*[15]

2.1. The Architecture of Autonomy

OpenClaw's significance extends far beyond its immediate use as a personal assistant. It serves as a proof of concept — perhaps inadvertently — for a category of software that society has not yet reckoned with: autonomous agents capable of acting in the physical and digital world with minimal human oversight.

The architecture is straightforward. An OpenClaw agent receives a high-level goal, decomposes it into subtasks, identifies or acquires the necessary tools (skills), executes them sequentially or in parallel, maintains persistent memory across sessions, and communicates results through existing messaging infrastructure. Every one of these capabilities has legitimate, productive applications. But the same architecture, directed by different intent, enables something far more troubling.

To understand why, consider the components individually:

Autonomous task execution. OpenClaw does not wait for step-by-step instructions. It receives an objective, plans an approach, and executes across multiple systems. This is precisely the operational profile of an autonomous military or intelligence agent[16].

Dynamic skill acquisition. With ClawHub, the agent can search for and install new capabilities on demand. A military analogue would be a centralized command registry from which agents pull tactics, exploits, or operational modules in real time, adapting to conditions as they evolve.

Persistent memory. Configuration and interaction history enable adaptive behavior across sessions. The agent learns, remembers context, and refines its approach. It does not forget its objectives. It does not require rest.

Distribution through civilian infrastructure. OpenClaw propagates through WhatsApp, Telegram, Discord, and Signal — applications used by billions. A weaponized

[15] Bill Gates, "AI is about to completely change how you use computers," November 2023. See: GatesNotes.

[16] Autonomous agent: In the context of AI, a software system that can perceive its environment, make decisions, and take actions independently to achieve specified goals. See: Wikipedia: Intelligent agent.

agent could spread through the same channels, indistinguishable from legitimate traffic until activated.

Near-zero marginal cost. One developer built OpenClaw. Deploying thousands of instances requires only compute resources and API keys. The asymmetry between the cost of deploying such agents and the cost of defending against them is extreme.

2.2. The "Digital Soldier" Scenario

The scenario that follows is not science fiction. It requires no technological breakthrough beyond what OpenClaw has already demonstrated. It requires only a change in intent.

Imagine a state or non-state actor that forks[17] OpenClaw's open-source codebase and modifies it for offensive operations. The reasoning engine is replaced with a fine-tuned[18] open-source model — Llama, Qwen, DeepSeek — trained on military doctrine, social engineering techniques, and cyber exploitation methods. This is well within current capabilities; fine-tuning an open-source LLM requires only data and modest compute.

The skill registry is replaced with a military equivalent: a command hub distributing reconnaissance modules, phishing payloads, propaganda generation tools, and infrastructure exploitation kits. Each agent pulls capabilities as needed, adapting to its operational environment.

These agents are deployed at scale across compromised or co-opted machines. They communicate through encrypted messaging platforms, blending with civilian traffic. They maintain persistent memory of their targets, learning patterns and vulnerabilities over time. They are, for all practical purposes, immortal — operational as long as the underlying hardware and network connectivity persist.

This is not a hypothetical leap. Every component exists. OpenClaw demonstrated the integration. What separates a personal assistant from a digital soldier is not architecture but intent[19].

2.3. The Lethal Trifecta

Security researchers at HiddenLayer have identified what they call the "lethal trifecta"[20] of agentic AI: exposure to untrusted content, access to private data, and the ability to

[17] Fork: In software development, creating an independent copy of an existing project's source code to develop it in a different direction. See: Wikipedia: Fork.

[18] Fine-tuning: The process of further training a pre-trained AI model on a specialized dataset to adapt it for specific tasks or domains. See: Wikipedia: Fine-tuning.

[19] The dual-use problem — where the same technology can serve both beneficial and harmful purposes — is well-established in fields from nuclear physics to biotechnology. See: Wikipedia: Dual-use technology.

[20] See: Dark Reading: "OpenClaw's Gregarious Insecurities Make Safe Usage Difficult".

communicate externally. Any system possessing all three, without adequate security controls, is inherently dangerous.

OpenClaw possesses all three by design — because these are precisely the capabilities that make it useful. This is the fundamental tension at the heart of agentic AI: the features that create value are the same features that create risk. There is no configuration that preserves full utility while eliminating the attack surface. Every mitigation involves a trade-off.

In a weaponized context, this trifecta becomes even more dangerous. An offensive agent that can ingest intelligence from external sources, access compromised systems, and exfiltrate data through legitimate communication channels represents a new category of threat — one that operates within authorized permissions, leaves minimal forensic evidence, and is difficult to distinguish from normal system behavior.

2.4. The Unsolvable Problem at the Core

The most sobering aspect of OpenClaw's security challenges is that the deepest vulnerability — prompt injection[21] — is not an OpenClaw-specific bug. It is an unsolved problem across the entire AI industry.

Prompt injection occurs when malicious instructions are embedded in data that the agent is designed to process — emails, documents, webpages, chat messages. Because LLMs cannot reliably distinguish between legitimate instructions from the user and malicious instructions embedded in content, the agent may treat both identically. HiddenLayer demonstrated this concretely: simply asking OpenClaw to summarize a malicious webpage caused it to download and execute a shell script with full system permissions.

No major AI company — not Anthropic, not OpenAI, not Google — has a general solution to prompt injection. Until this foundational problem is resolved, every autonomous AI agent that processes external data and takes actions based on that processing carries an inherent, irreducible risk.

[21] Prompt injection: An attack technique where malicious instructions are embedded in data that an AI model processes, causing it to perform unintended actions. This is an unsolved problem in AI safety as of 2026. See: Wikipedia: Prompt injection.

3. Part III: The Geopolitics — Confrontation Where Cooperation Is Needed

> *"By mid-2026, at least one major global enterprise will fall to a breach caused or significantly advanced by a fully autonomous agentic AI system."*[22]

3.1. The Asymmetry of Creation and Governance

Peter Steinberger is an Austrian software engineer. Working essentially alone, he created a tool that within weeks required emergency responses from some of the world's largest cybersecurity companies, was deployed on tens of thousands of machines across multiple continents, spawned a social network with 1.6 million AI bots[23], and generated cryptocurrency scams, fraudulent distribution sites, and corporate security incidents.

This would have been impossible a decade ago. The convergence of three forces — frontier AI models available as API calls, messaging platforms providing instant global distribution, and viral culture replacing traditional marketing — means that a single person can now deploy autonomous software agents into millions of lives with no institutional checkpoint along the way.

The gap between the power to create and the capacity to govern the consequences of that creation has never been wider. And it is widening.

3.2. The Cooperation Deficit

History offers precedents for governing transformative and dangerous technologies. Nuclear weapons produced arms control treaties[24]. Biological weapons led to the Biological Weapons Convention[25]. Even at the depths of the Cold War, adversaries recognized that certain technologies demanded cooperative frameworks because the alternative — unconstrained proliferation — threatened all parties.

---

[22] Michael Freeman, Head of Threat Intelligence at Armis, in predictions for 2026. See: SecurityWeek: "Cyber Insights 2026".

[23] Moltbook: A social media platform designed exclusively for AI agents, created by an OpenClaw user. It attracted over 1.6 million registered bots and 7.5 million AI-generated posts. See: Nature: "OpenClaw AI chatbots are running amok".

[24] See the Treaty on the Non-Proliferation of Nuclear Weapons (1968): Wikipedia: NPT.

[25] See: Wikipedia: Biological Weapons Convention.

Autonomous AI agents, and the agentic frameworks that enable them, present a comparable challenge. Yet the global response is moving in precisely the opposite direction.

3.3. AI as Trade Commodity vs. AI as Strategic Risk

The framing of artificial intelligence in current policy discourse is predominantly economic. AI is treated as a competitive asset — a driver of productivity, a source of industrial advantage, a front in the trade wars between the United States, China, and the European Union. This framing has specific consequences for safety governance.

The current US administration has pursued a policy of aggressive deregulation of artificial intelligence, rolling back safety requirements and executive orders from the prior administration[26]. The explicit rationale is competitiveness: regulations are framed as obstacles to American dominance in AI, and their removal is presented as necessary to win what is characterized as an AI race with China.

This framing creates a structural problem. If the world's leading AI power treats safety governance as an impediment to competitiveness, it signals to every other nation that restraint is a disadvantage. Why would China, the European Union, or any other actor voluntarily constrain its AI development if the United States has declared that constraints are liabilities?

The result is a race to the bottom[27] in which the absence of norms becomes the norm. This is precisely the opposite of what autonomous AI requires.

3.4. The Open-Source Paradox

Open-source software has been, for decades, one of the most productive forces in technology. The ability to inspect, modify, and share code has driven innovation, improved security through transparency, and democratized access to powerful tools. OpenClaw itself is open source, and its creator has been transparent about its limitations.

But the agentic AI paradigm breaks assumptions that held true for traditional open-source software. Being able to read the source code of an autonomous agent does not protect you from prompt injection, because the vulnerability exists in the LLM's behavior, not in the agent's code. Community code review does not scale when a project goes from obscurity to 150,000 stars in 72 hours. And the open availability of the codebase means that anyone — including state actors, criminal organizations, and

[26] In January 2025, President Trump signed an executive order revoking the Biden administration's AI safety executive order (EO 14110), which had established reporting requirements, safety evaluations, and risk management guidelines for AI systems.

[27] Race to the bottom: A socio-economic phenomenon where competition between entities (nations, companies) leads to progressively lower standards of regulation, safety, or labor protections. See: Wikipedia: Race to the bottom.

individuals with malicious intent — can fork, modify, and deploy weaponized variants without attribution.

The open-source AI community has built extraordinary things. But it has not yet developed governance mechanisms appropriate to software that acts autonomously, accesses sensitive data, and takes real-world actions on behalf of users.

3.5. The Temporal Incompatibility

There is a structural mismatch between the speed at which autonomous AI agents can be developed and deployed and the speed at which human institutions — regulatory bodies, legislative processes, international negotiations, democratic debate — can respond.

OpenClaw went from zero to 150,000 GitHub stars in days. Critical vulnerabilities were exploited within weeks. Scam campaigns launched within hours of each name change. Meanwhile, the AI regulatory frameworks being discussed in the European Union, the United States, and other jurisdictions operate on timescales of months to years.

This is not a problem that can be solved by making regulation faster. Even dramatically accelerated regulatory processes would still operate orders of magnitude more slowly than the deployment cycles of viral open-source software. The question is whether we need entirely new institutional designs[28] rather than faster versions of existing ones.

3.6. What Global Cooperation Would Require

Preventing the weaponization of agentic AI frameworks would require, at minimum:

International norms on autonomous AI agents. Analogous to the norms governing autonomous weapons systems currently under discussion at the United Nations[29], but extended to cover software agents operating in cyberspace. The challenge: verification is nearly impossible. A weaponized AI agent runs on a laptop and looks identical to a legitimate assistant.

Shared vulnerability disclosure frameworks. When critical flaws are discovered in widely deployed agentic systems, coordinated disclosure across national boundaries could limit exploitation windows. The challenge: this requires trust between nations that are simultaneously competing for AI supremacy.

[28] Some scholars have proposed "regulatory sandboxes," real-time monitoring systems, and adaptive governance frameworks as alternatives to traditional legislative processes for fast-moving technologies. See: Wikipedia: Regulatory sandbox.

[29] See the UN Convention on Certain Conventional Weapons (CCW) discussions on Lethal Autonomous Weapons Systems (LAWS): Wikipedia: Lethal autonomous weapon.

Compute and infrastructure monitoring. Large-scale deployment of agent swarms requires compute resources that could, in principle, be monitored. The challenge: OpenClaw runs on a consumer laptop or a $5/month virtual private server. The compute floor for deployment is effectively zero.

Supply chain integrity for AI skill registries. Centralized or federated registries with mandatory security scanning, code signing, and provenance tracking. The challenge: as Bitdefender documented, malicious actors were already overwhelming ClawHub's community oversight with automated uploads within weeks of OpenClaw's release.

Each of these measures faces the same fundamental obstacle: they require international cooperation at precisely the moment when the major AI powers are choosing confrontation.

4. Conclusion: The Door That Cannot Be Closed

When I decided not to install OpenClaw, I was not making a statement about Peter Steinberger's work, which is technically impressive and transparently documented. I was making a statement about the state of the world into which his work was released.

We live in a moment where a single developer's hobby project can become a global security event within days. Where the same architecture that powers a helpful personal assistant can, without modification, serve as the skeleton of an autonomous digital weapon. Where the plugin ecosystem of a productivity tool can be poisoned at a rate of hundreds of malicious entries per day. And where the geopolitical order is moving away from, rather than toward, the cooperative frameworks that might govern these capabilities.

OpenClaw did not create these conditions. It revealed them. The tool is a mirror, reflecting the gap between what technology now makes possible and what society is prepared to manage.

The uncomfortable truth is that there is no single solution. Technical measures — sandboxing, skill verification, prompt injection defenses — are necessary but insufficient. Regulatory frameworks are essential but temporally mismatched. International cooperation is indispensable but politically improbable in the current climate.

What remains is the obligation to see clearly. To recognize that the age of autonomous AI agents is not approaching — it has arrived, propelled by viral adoption and open-source accessibility. To understand that the same tools that promise to liberate us from digital drudgery can, with trivial modification, be turned to purposes their creators never intended. And to insist, despite the pressures of competition and the seductions of convenience, that the question "are we ready?" deserves an honest answer.

We are not ready. The first step toward readiness is admitting it.

> *"Security is a process, not a product. Also, don't trust lobsters with shell access."*
>
> — OpenClaw Security Documentation

5. References

1. Astrix Security. “OpenClaw & MoltBot: The First AI Agent Security Nightmare.” February 2026. https://astrix.security/learn/blog/openclaw-moltbot-the-rise-chaos-and-security-nightmare-of-the-first-real-ai-agent/
2. Bitdefender. “Technical Advisory: OpenClaw Exploitation in Enterprise Networks.” February 2026. https://businessinsights.bitdefender.com/technical-advisory-openclaw-exploitation-enterprise-networks
3. Cisco. “Personal AI Agents like OpenClaw Are a Security Nightmare.” February 2026. https://blogs.cisco.com/ai/personal-ai-agents-like-openclaw-are-a-security-nightmare
4. CrowdStrike. “What Security Teams Need to Know About OpenClaw, the AI Super Agent.” February 2026. https://www.crowdstrike.com/en-us/blog/what-security-teams-need-to-know-about-openclaw-ai-super-agent/
5. Dark Reading. “OpenClaw’s Gregarious Insecurities Make Safe Usage Difficult.” February 2026. https://www.darkreading.com/application-security/openclaw-insecurities-safe-usage-difficult
6. eSecurity Planet. “OpenClaw and the Growing Security Risks of Agentic AI.” February 2026. https://www.esecurityplanet.com/threats/openclaw-and-the-growing-security-risks-of-agentic-ai/
7. Herbert, Frank. *Dune*. Chilton Books, 1965.
8. Hostinger. “OpenClaw Security: Risks, Best Practices, and a Checklist.” February 2026. https://www.hostinger.com/tutorials/openclaw-security
9. IBM. “OpenClaw, Moltbook and the Future of AI Agents.” February 2026. https://www.ibm.com/think/news/clawdbot-ai-agent-testing-limits-vertical-integration
10. Nature. “OpenClaw AI chatbots are running amok — these scientists are listening in.” February 2026. https://www.nature.com/articles/d41586-026-00370-w
11. Scientific American. “OpenClaw is an open-source AI agent that runs your computer.” February 2026. https://www.scientificamerican.com/article/moltbot-is-an-open-source-ai-agent-that-runs-your-computer/
12. SecurityWeek. “Cyber Insights 2026: Malware and Cyberattacks in the Age of AI.” February 2026. https://www.securityweek.com/cyber-insights-2026-malware-and-cyberattacks-in-the-age-of-ai/
13. SecurityWeek. “Vulnerability Allows Hackers to Hijack OpenClaw AI Assistant.” February 2026. https://www.securityweek.com/vulnerability-allows-hackers-to-hijack-openclaw-ai-assistant/
14. Steinberger, Peter. OpenClaw GitHub Repository. https://github.com/openclaw/openclaw

15. The Register. “OpenClaw ecosystem still suffering severe security issues.” February 2026. https://www.theregister.com/2026/02/02/openclaw_security_issues/
16. Trend Micro. “Viral AI, Invisible Risks: What OpenClaw Reveals About Agentic Assistants.” February 2026. https://www.trendmicro.com/en_us/research/26/b/what-openclaw-reveals-about-agentic-assistants.html
17. CNBC. “From Clawdbot to Moltbot to OpenClaw: Meet the AI agent generating buzz and fear globally.” February 2026. https://www.cnbc.com/2026/02/02/openclaw-open-source-ai-agent-rise-controversy-clawdbot-moltbot-moltbook.html
18. USCS Institute. “What is AI Agent Security Plan 2026?” https://www.uscsinstitute.org/cybersecurity-insights/blog/what-is-ai-agent-security-plan-2026-threats-and-strategies-explained
19. CyberArk. “AI Agents and Identity Risks: How Security Will Shift in 2026.” December 2025. https://www.cyberark.com/resources/blog/ai-agents-and-identity-risks-how-security-will-shift-in-2026
20. Darktrace. “The State of AI Cybersecurity 2026.” February 2026. https://www.darktrace.com/blog/the-state-of-ai-cybersecurity-2026
21. Stellar Cyber. “Top Agentic AI Security Threats in 2026.” December 2025. https://stellarcyber.ai/learn/agentic-ai-securiry-threats/
22. DigitalOcean. “What is OpenClaw? Your Open-Source AI Assistant for 2026.” https://www.digitalocean.com/resources/articles/what-is-openclaw
23. Wikipedia contributors. “OpenClaw.” Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/OpenClaw