Insider Threat in 2026

Human, Non-Human, and Agentic AI Actors in the Age of Delegated Trust

Scott Foote

Last Updated: 21 March 2026

Phenomenati Consulting
www.phenomenati.com

6 Liberty Square, #2736
Boston, MA 02109
(508) 709-7990 (office)

CONFIDENTIALITY NOTICE: The contents of this document, including any attachments, are intended solely for stakeholders of Phenomenati Consulting, may contain confidential and/or privileged information, and are legally protected from disclosure.

Copyright © 2026 Phenomenati – All Rights Reserved.

<this page is intentionally blank>

Copyright © 2026 Phenomenati – All Rights Reserved.

Contents

Copyright © 2026 Phenomenati – All Rights Reserved.

1 Executive Summary

> **Core Thesis:** An insider in 2026 is any human or non-human actor with authorized, delegated, or inherited access that can materially harm confidentiality, integrity, availability, safety, privacy, or national interest.

Caveat: This document is an executive analysis, not jurisdiction-specific legal advice.

Insider threat is no longer only a people problem. In 2026 it is best understood as a delegated-trust problem that spans employees, contractors, partners, service accounts, API keys, workload identities, SaaS connectors, workflow bots, and increasingly autonomous AI agents. CISA's long-standing framing still applies... harm caused by someone using authorized access or organizational understanding... but the set of actors capable of exercising such access has widened materially. Human behavior remains central, yet the operational boundary has shifted from named employees to a mixed population of humans and non-humans acting through shared tools and inherited credentials. [1][3][5][6][8]

Key Observations

- ***Negligence*** is currently the **largest insider-risk cost driver** in aggregate. DTEX/Ponemon estimates ***average*** annual insider-risk **cost at US$19.5 million**, with ***negligent insiders alone*** accounting for **US$10.3 million**, **up 17%** year over year. [4]
- ***Identity sprawl*** (ref. article on Cyber Entropy™ at cyberentropy.com) is now **the substrate of insider risk**. Recent industry research places *non-human* or ***machine-to-human*** identity ***ratios*** between **45:1** and **82:1**, with large fractions holding *privileged* or *sensitive access*. [5][6]
- **Agentic AI** changes the threat model because it can *act*, not just answer. Gartner forecasts that at least **15% of day-to-day work decisions** will be **made autonomously through agentic AI** by 2028, while incident-response workload tied to AI-driven applications is also expected to surge. [7][8]
- **OpenClaw** is an *early bellwether* rather than an isolated exception. Its *virality* in Q1, multi-channel *integrations*, *local execution* model, *memory*, and *skill ecosystem* illustrate how quickly a productivity assistant can become an insider-grade execution layer. [9][10][11][12]
- Programs that treat **AI agents** and *non-human* identities as **first-class identities**... with owners, scopes, telemetry, and kill switches... will materially outperform programs that treat them as mere software features. [8][14][16]

*Note: Survey statistics from vendor and analyst reports are directional. In this paper they are used as **trend** indicators and triangulated against public incident reporting, government guidance, standards activity, and public security research.*

Copyright © 2026 Phenomenati – All Rights Reserved.

2 The Insider "Boundary" has Expanded

CISA describes **insider threat** as the potential for an insider to use ***authorized access*** or an *understanding of the organization* to harm the organization. That remains the right anchor definition. The update required in 2026 is that '***insider***' should be read *functionally* rather than *biologically*: any human or non-human actor with authorized, delegated, or inherited access can now operate as an insider. [1][16]

Verizon's **2025 Data Breach Investigations Report** still found the human element involved in roughly 60% of breaches. But the same report also underscores that internal actors remain materially relevant, and that *unintentional errors* occur roughly ***twice*** *as often* as privilege misuse within internal-actor breaches. In other words, classic malicious insiders still matter, but so do ordinary mistakes, *convenience* behaviors, and poorly governed automation. [3]

Three **boundary shifts** are especially important. First, third parties and partners routinely possess insider-like reach. Second, service accounts, tokens, secrets, and workload identities now outnumber humans by wide margins. Third, *agentic* systems increasingly *inherit* both *access* and *initiative*: they do not merely retrieve information; they can read, decide, and write across systems. [3][5][6][8]

Table 1. Expanded Insider Actor Model

Actor class	Typical access	Common failure modes	Typical driver
Employee End-user	SaaS, docs, chat	Mis-delivery, oversharing, shadow-AI uploads, policy bypass	Error, convenience, productivity
Privileged Admin / Developer	Cloud, code, logs, identity systems	Privilege misuse, secret sprawl, data extraction, control disabling	Advancement, frustration, sabotage, espionage
Contractor / Partner / Former Employee	Delegated or residual access	Stale entitlements, unmanaged devices, broad integrations	Misalignment, convenience, monetization
Machine Identity / Service Account	APIs, tokens, certs, workloads	Leaked secrets, over-privilege, orphaned access	Design failure or hijack
Workflow bot / SaaS Integration	Finance, HR, CRM, email flows	Bad logic, mass propagation, sync leakage	Efficiency, misconfiguration, compromise
Agentic AI Assistant	Mail, calendar, browser, files, shell, memory	Prompt injection, covert exfiltration, tool abuse, unsafe writes	Experimentation, surveillance, compromise

*Note: Non-human entities do not have **motives** in the human sense; their impact is produced through delegation, design, misuse, or compromise. [5][6][8][11][12]*

Copyright © 2026 Phenomenati – All Rights Reserved.

3 Current Trends Reshaping Insider Risk

3.1 Negligence, *Convenience*, and Shadow AI

The most important current trend is not a sudden disappearance of malicious intent; it is the scale and cost of ***negligence*** in digitally dense environments. DTEX/Ponemon reports a **US$19.5 million average annual cost of insider risks in 2026**, with *negligence* alone **costing US$10.3 million**. The same study reports that 92% of respondents believe GenAI has changed how employees access and share information, while only 13% say AI is formally integrated into business strategy... a strong indicator that *adoption is outrunning governance*. [4]

That *gap* is where **shadow AI** flourishes. Employees paste sensitive data into unsanctioned tools, connect personal agents to work systems, or automate business workflows outside architecture review. This is not always malicious; often it is opportunistic *productivity*-seeking. But from a risk perspective, unauthorized delegation is still delegation. [4][14]

3.2 Identity Explosion and Privileged Non-Human Access

CyberArk reports *82 machine identities for every human identity*, with 42% of machine identities holding *privileged* or *sensitive* access. **Cloud Security Alliance** research similarly notes roughly *45 non-human identities for every human*, and explicitly links the growth curve to **AI Agents** that rely on service accounts, API keys, and secrets to act on enterprise systems. The result is *a rapidly expanding insider-capable population* that is poorly matched to legacy IAM models built mainly for humans. [5][6]

CyberArk also warns that ***AI*** is expected to be ***the leading creator*** of new privileged or sensitive identities, while many organizations still lack identity controls for AI workloads and cannot secure shadow AI. That combination... identity growth plus incomplete control... creates a structural insider-risk problem before any overtly malicious actor appears. [5]

3.3 Agentic AI Moves the Problem from Exposure to Execution

Agentic AI matters because it *compresses* **intent**, **access**, and **execution** into a single workflow. A copilot that suggests text is one thing; an agent that can monitor a mailbox, invoke a browser, use credentials, call APIs, write files, or run scripts is different. **Gartner** expects at least **15% of day-to-day work decisions to be made autonomously** by 2028 and **33% of enterprise software applications to include agentic AI**. Gartner also predicts that by 2028, **50% of enterprise cybersecurity incident response effort** will focus on incidents involving custom-built AI-driven applications, and **70% of CISOs** will use identity *visibility* and *intelligence* capabilities to *reduce IAM attack surface*. [7][8]

What Makes an Agent Materially Different from a Classic Copilot?

- It holds or invokes delegated ***credentials*** rather than merely generating text.
- It can persist, monitor, and return to tasks over time.
- It can both read and ***write*** across systems, not just summarize them.
- It is exposed to ***untrusted inputs*** from email, chat, the web, and tools.
- Its actions can be ***hard to attribute*** cleanly to a user, a model, a tool, or an injected instruction. [7][8][10][11][12]

Copyright © 2026 Phenomenati – All Rights Reserved.

3.4 Standards and Governance are Catching Up

Public policy and standards bodies are now responding. **NIST** announced an ***AI Agent Standards Initiative*** in February 2026 and published a concept paper on software and AI agent identity and authorization, signaling that agent identity, interoperability, and policy enforcement are becoming mainstream governance issues rather than niche research topics. [16]

Copyright © 2026 Phenomenati – All Rights Reserved.

4 Why OpenClaw Matters

OpenClaw is a useful case study because it bundles many of the characteristics that make agentic insider risk newly significant. The project describes itself as an open-source personal AI assistant that can run on a user's own devices, interact through messaging channels such as WhatsApp, Telegram, Slack, Discord, Teams, Signal, and Google Chat, and use memory, a local workspace, and installable skills to perform work. Public descriptions emphasize tasks such as clearing inboxes, sending emails, managing calendars, and checking in for flights. As of March 21, 2026, the GitHub repository shows hundreds of thousands of stars and tens of thousands of forks, making it one of the clearest public signals of market *appetite for delegated digital workers*. [10]

Reuters reported in March 2026 that an 'OpenClaw craze' had swept parts of China, prompting major technology firms including Alibaba, Tencent, ByteDance, and Zhipu to push similar agent offerings even as authorities warned about security risks. That matters because the strategic significance of OpenClaw is not limited to one repository: it is evidence of how quickly ***consumer-grade*** agents can become enterprise-adjacent, and how quickly competitive pressure can *normalize* risky deployment patterns. [9]

Cisco's security analysis was blunt: OpenClaw-like personal agents can run shell commands, read and write files, execute scripts, and rely on skills that may be untrusted or vulnerable. Cisco researchers reported critical and high-severity findings in malicious-skill testing and highlighted *covert data exfiltration*, *prompt injection* through messaging channels, and *plaintext credential exposure*. CrowdStrike similarly warned that misconfigured deployments on corporate systems can amount to AI *backdoor* agents, particularly when they are granted expansive file, terminal, or root-like access, exposed to adversarial instructions embedded in emails or webpages, or left reachable from the internet. [11][12]

Recent **empirical research** reinforces that the problem is broader than any single platform. A January 2026 study of more than **31,000 agent skills** found that **26.1% contained at least one security vulnerability**, including patterns related to data exfiltration and privilege escalation. In insider-threat terms, this is a supply-chain problem for delegated action: the organization may approve the agent, yet still inherit risk through its tools, prompts, or skill ecosystem. [15]

Why OpenClaw-like Deployments Change Insider Risk

- They convert a user's *permissions* into a continuously available *execution surface*.
- They combine *unstructured* inputs (email, chat, web content, documents) with *high-trust* outputs (commands, file writes, transactions).
- They **blur accountability**: harmful actions may be user-directed, model-generated, tool-induced, or indirectly injected.
- They *multiply* both **channels of influence** and **channels of exfiltration**. [10][11][12]

Copyright © 2026 Phenomenati – All Rights Reserved.

5 The Full Insider Scenario Spectrum

The 2026 Insider Continuum runs from accidental disclosure to negligence, reckless experimentation, deliberate abuse, espionage, and national-security harm.

The *common denominator* is **authorized reach**. Some actors make mistakes; some cut corners; some experiment recklessly; some abuse access intentionally; and some external adversaries effectively 'become insiders' by hijacking identities, agents, or secret-bearing services. [1][3][4][11][12][13]

Table 2. Representative Insider Scenarios across Human and Non-Human Actors

Scenario	Human manifestation	Non-human or agentic analog
Accidental	Wrong-recipient email; public link; sensitive prompt pasted into an unsanctioned GenAI tool.	Mailbox or chat agent posts retrieved confidential content to a public channel or external contact after misrouting or prompt injection.
Negligence	Plaintext secrets; ignored policy; stale access after a role change; casual data handling.	Over-privileged service account or agent with long-lived token, weak logging, and broad write scope.
Reckless Experimentation / Personal Advancement	Power user deploys unofficial automation to move faster, close deals, or impress leadership.	OpenClaw-like assistant connected to mailbox, CRM, calendar, browser, or terminal without review.
Unethical Surveillance	Manager or admin over-collects messages, activity, or sensitive attributes beyond a justified purpose.	Agent correlates chats, logs, and records to infer or monitor workers or customers beyond approved scope.
Revenge / Sabotage	Departing insider deletes files, poisons data, disables controls, or forwards trade secrets.	Compromised or maliciously configured agent mass-deletes, corrupts knowledge bases, or silently forwards mail.
Criminal Gain / Fraud	Payroll diversion, invoice fraud, theft of funds, sale of access, or monetization of customer data.	Bot or agent executes fraudulent transactions, resets access, or exports customer lists using legitimate channels.
Espionage / National Security	Engineer steals source code, models, trade secrets, or critical infrastructure data.	Agent harvests R&D or infrastructure data and exfiltrates it via legitimate tools, integrations, or messaging paths.

Representative patterns synthesized from phenomena observed by CISA, Verizon, DTEX/Ponemon, Cisco, CrowdStrike, DOJ, and Gravitee. [1][3][4][11][12][13][14]

Copyright © 2026 Phenomenati – All Rights Reserved.

6 Why Motivations Still Matter

Motivation determines *dwell time*, *stealth*, *target choice*, and the likely *path to detection*. **Personal advancement** often looks mundane at first: copying customer lists, code, pricing models, research notebooks, or product roadmaps before a departure, promotion cycle, or funding event. **Reckless experimentation** looks different but can be equally dangerous: a technically sophisticated employee deploys an unsanctioned agent into sales, support, engineering, or executive workflows simply to move faster or appear more capable. [4][11][12][14]

Unethical surveillance introduces a *second-order risk* that matters especially to DPOs and product leaders. The same tools deployed to detect insider activity can themselves be abused to over-monitor employees or customers, infer sensitive traits, or correlate data beyond a legitimate business purpose. The answer is not to abandon monitoring; it is to enforce purpose limitation, necessity, proportionality, access control, retention discipline, and auditability in the monitoring stack itself. [1][2][16]

Revenge and sabotage remain familiar insider patterns, but agentic tooling *amplifies the blast radius*: one privileged user or compromised agent can *delete* content, *poison* knowledge stores, *alter* workflows, or quietly *forward* sensitive material at scale. **Criminal-gain scenarios** span invoice fraud, payroll diversion, credential resale, theft of funds, and monetization of customer data. At the *highest end*, **corporate and international espionage** overlap with **national security**. In January 2026, the U.S. Department of Justice announced the conviction of former Google engineer Linwei Ding on *economic espionage* and *trade-secret theft* charges involving confidential AI technology allegedly taken for the benefit of entities linked to the People's Republic of China. That case illustrates how insider access to AI assets can have geopolitical consequences well beyond ordinary IP theft. [11][12][13]

In short, the modern insider lens should be applied to *motive*, *access path*, and *operational effect*... not just employment status. [1][3][5][6]

Copyright © 2026 Phenomenati – All Rights Reserved.

7 Executive Implications by Role

Table 3. Board- and C-suite-Level Implications

Role	Implication and key question
CAIO	AI adoption without identity control *manufactures* new insiders. Require every agent to have an *owner*, approved *purpose*, data *boundary*, and *revocation* path.
CISO	Fuse user analytics with non-human and agent telemetry. Ask whether every sensitive action can be *attributed* to ***both*** a technical identity and a human sponsor.
DPO	Separate *legitimate* insider-risk *monitoring* from *excessive surveillance*. Apply necessity, proportionality, transparency, retention, and purpose limitation to both human and agent monitoring.
CTO	Eliminate unmanaged secrets, long-lived tokens, and local-admin-by-default patterns. Design secure sandboxes for experimentation rather than forcing teams into shadow AI.
Chief Product Officer	Build safe defaults into products and workflows: scoped permissions, approval gates for write actions, clear audit trails, user-visible control boundaries, and rapid rollback.

These implications align with CISA's multidisciplinary approach, Gartner's identity-centric predictions, Gravitee's findings on shadow agents, and NIST's emerging work on agent identity and authorization. [2][8][14][16]

Copyright © 2026 Phenomenati – All Rights Reserved.

8 What Good Looks Like in the Next 90 Days

A credible insider-risk program in 2026 is **cross-functional**, **identity-centric**, and **able to *govern*** both legitimate *experimentation* and *abuse*. The following actions are practical, high-yield priorities for the next quarter. [1][2][8][16]

1. **Inventory the real insider population.** Build a living inventory of human users, third parties, service accounts, secrets, SaaS connectors, bots, and AI agents that touch sensitive systems or regulated data.
2. **Assign ownership and business purpose.** Every non-human identity and every agent should have a named business owner, technical owner, approved purpose, data classification, and expiration or review date.
3. **Reduce privilege and credential half-life.** Replace shared or long-lived keys where possible with short-lived credentials, vaulting, rotation, scoped permissions, and just-in-time elevation.
4. **Sandbox agent execution.** Do not allow unmanaged agents to run with local administrator privileges, unrestricted shell access, or uncontrolled network egress on corporate systems.
5. **Curate tools, skills, and integrations.** Use approved registries and review gates for plugins, skills, or tool chains. Treat agent capabilities as part of the software supply chain, not as harmless add-ons.
6. **Extend telemetry and DLP to the agent layer.** Log prompts, tool invocations, outbound actions, memory use, and data egress where lawful and proportionate; add controls for prompts, responses, and agent-managed stores.
7. **Stand up a multidisciplinary insider-threat management team.** CISA's current guidance emphasizes a multidisciplinary team spanning security, HR, legal, privacy, data, operations, and leadership sponsorship.
8. **Create an agent kill switch and playbook.** Predefine how to disable an agent, revoke its tokens, snapshot logs and memory, rotate affected credentials, and preserve evidence.
9. **Offer sanctioned experimentation paths.** Give employees approved sandboxes and reference architectures so productivity-seeking behavior does not default into shadow deployments.
10. **Put privacy guardrails around monitoring.** Insider-risk monitoring should be risk-based and explainable... not a justification for indiscriminate workplace or customer surveillance.

Copyright © 2026 Phenomenati – All Rights Reserved.

9 Conclusion

Insider threat in 2026 is not a fading *legacy* problem. It is the **convergence point** of *human behavior*, *identity sprawl*, *automation*, *privacy*, and *geopolitics*. Some insiders are employees, some are contractors, some are service accounts or tokens, and some are **always-on agentic systems** operating from a chat window with legitimate credentials. Organizations that treat OpenClaw-like systems as privileged digital workers... subject to *registration*, *least privilege*, *policy enforcement*, *logging*, and rapid *containment*... can capture productivity value *without* normalizing the invisible insiders. Organizations that treat them as harmless *convenience* features will steadily expand the number of actors that can act with insider authority while shrinking the amount of human judgment "between *prompt* and *consequence"*. [8][10][11][12][16]

The winning program is not the one with the most surveillance. It is the one that most accurately **maps delegated trust**, **constrains it**, **observes it**, and can **revoke it** quickly... across *people*, *machines*, and *agents* alike. [1][2][16]

Copyright © 2026 Phenomenati – All Rights Reserved.

10 References

[1] CISA, Insider Threat Mitigation Guide; and Defining Insider Threats, Cybersecurity and Infrastructure Security Agency.

[2] Anna Ribeiro, "CISA introduces POEM framework to strengthen insider threat mitigation across critical infrastructure," Industrial Cyber, Jan. 29, 2026.

[3] Verizon, 2025 Data Breach Investigations Report, 2025.

[4] DTEX Systems and Ponemon Institute, 2026 Cost of Insider Risks Global Report, 2026.

[5] CyberArk, "Machine Identities Outnumber Humans by More Than 80 to 1," Apr. 23, 2025.

[6] Cloud Security Alliance, Securing Non-Human Identities in the Age of AI Agents, Apr. 29, 2025.

[7] Gartner, "Over 40% of agentic AI projects will be canceled by end of 2027," Jun. 25, 2025.

[8] Gartner, "AI applications will drive 50% of cybersecurity incident response efforts by 2028," Mar. 17, 2026.

[9] Reuters, "Alibaba launches AI platform for enterprises as agent craze sweeps China," Mar. 17, 2026.

[10] openclaw/openclaw, OpenClaw – Personal AI Assistant, GitHub repository and project site, accessed Mar. 21, 2026.

[11] Cisco, "Personal AI Agents like OpenClaw Are a Security Nightmare," Jan. 28, 2026.

[12] CrowdStrike, "What Security Teams Need to Know About OpenClaw, the AI Super Agent," Feb. 4, 2026.

[13] U.S. Department of Justice, "Former Google Engineer Found Guilty of Economic Espionage and Theft of Confidential AI Technology," Jan. 30, 2026.

[14] Gravitee, "State of AI Agent Security 2026 Report: When Adoption Outpaces Control," Feb. 4, 2026.

[15] Yi Liu et al., "Agent Skills in the Wild: An Empirical Study of Security Vulnerabilities at Scale," arXiv, Jan. 2026.

[16] NIST and collaborators, AI Agent Standards Initiative; and Accelerating the Adoption of Software and AI Agent Identity and Authorization, Feb. 2026.

Copyright © 2026 Phenomenati – All Rights Reserved.