

The New Face of Insider Risk

Why Trusted Access, Artificial Intelligence, and Point-in-Time Controls Call for a Governed, Human-Centered Approach

Introduction

When the Department of Defense stood up its insider threat management and analysis center in 2016, the problem was easier to picture: a disgruntled employee, a bad actor inside the walls. A decade later, insider risk is no longer a single person. It is an ecosystem of employees, remote workers, contractors, borrowed and synthetic identities, and increasingly AI tools and agents, all operating under trusted access. The perimeter that security programs were built to defend has effectively disappeared.

Most security stacks are not broken. They do exactly what they were designed to do, and most organizations have invested heavily and well. The difficulty is that the stack answers a technical question about a user inside the environment, while insider risk is a human question that largely plays out outside it. The gap between those two things is not a detection problem an organization can buy its way out of. It is a governance and enablement problem, and this paper sets out how leading organizations are closing it.

Insider Risk Starts With Ordinary Circumstances

Most insider incidents do not begin with a sophisticated adversary or with malicious intent. In the most recent Ponemon Institute study, negligent and mistaken insiders accounted for 13.8 of the 25.4 incidents the average organization experienced in a year. The problem is less a villain to be profiled than a set of conditions a workforce moves in and out of every day. Four stand out.

- **Financial stress:** Carnegie Mellon's CERT program lists financial difficulty, or an unexplained extreme change in finances, as a behavioral indicator in its Common Sense Guide to Mitigating Insider Threats. It is also one of the few indicators almost entirely invisible to a technical stack.
- **Burnout and fatigue:** Judgment degrades under sustained load, and the highest-performing, highest-access people are frequently the most loaded.
- **Personal disruption:** A divorce, a serious illness, or a family emergency can change behavior overnight while nothing in the environment changes at all.
- **Broad access on small teams:** In a lean organization, one trusted employee often holds access across engineering, cloud infrastructure, customer data, and intellectual property that a large enterprise would split across several people and approval chains.

Samsung illustrates the modern shape of the problem. In 2023, shortly after generative AI tools were permitted internally, Bloomberg reported that staff had uploaded sensitive source code to a public model, and the company restricted generative AI use on company devices. None of it was malicious. Well-intentioned employees trying to work faster created exposure that could not be pulled back.

The Growing Cost to Industry

The Ponemon Institute's 2026 Cost of Insider Risks Global Report, which surveyed 8,750 IT and security practitioners across 354 organizations, puts the average annual cost of insider risk at \$19.5 million per

organization in 2025, up 20 percent over two years. It is the annual cost per organization across all incidents, not the cost of a single incident, and it is large because of volume: 25.4 incidents a year, roughly one every two weeks. Every participating organization had already experienced a material insider event, so the figures describe the shape of the problem rather than a base rate.

Negligence is the largest share by volume, at 13.8 incidents and \$10.3 million a year. Malicious insiders account for 6.3 incidents and \$4.7 million, and credential theft for 5.3 incidents and \$4.5 million. Per incident, however, credential theft is now the most expensive category at \$842,462, ahead of negligence at \$747,107 and malicious insiders at \$742,125, and its incident count rose from 4.8 the prior year. The intuition that the malicious insider is the costliest is no longer what the data shows. The stolen identity is. In all three categories, the activity happens under valid access.

Insider category	Incidents per year	Annual cost	Cost per incident
Negligent or mistaken	13.8	\$10.3M	\$747,107
Malicious	6.3	\$4.7M	\$742,125
Credential theft	5.3	\$4.5M	\$842,462
All categories	25.4	\$19.5M	

Exhibit 1. Average per organization, 2025. Source: Ponemon Institute, 2026 Cost of Insider Risks Global Report.

Containment is improving, from 86 days in 2023 to 81 in 2024 and 67 in 2025, yet only 13 percent of incidents are contained within 30 days, and the clock drives the cost. Organizations that contain incidents in under 30 days average \$14.2 million a year; those that take more than 90 days average \$21.9 million.

Artificial Intelligence Changes the Equation

In ClearForce's view, artificial intelligence is accelerating insider risk on several fronts at once. It sharpens targeting, makes social engineering far more convincing, and introduces a new class of actor inside the network. In the FBI's words, attackers are leveraging AI to craft highly convincing voice or video messages and emails to enable fraud schemes against individuals and businesses alike.

The progression is visible in two cases. In 2019, The Wall Street Journal reported that the chief executive of a UK energy company wired roughly \$243,000 after a phone call from what he believed was the chief executive of his German parent company. The voice was synthetic, yet according to the company's insurer he recognized his boss's slight German accent and the melody of his speech. Five years later the tools had matured. In early 2024, a finance employee in the Hong Kong office of the British engineering firm Arup received a request for a confidential transaction. He was suspicious, so he did what security training teaches and verified it on a video call, where he saw and heard colleagues he recognized. They were deepfakes. Arup confirmed to CNN that fake voices and images were used, and he made 15 transfers totaling about \$25 million. The verification step itself defeated him. He followed the procedure, and the technology outran the control.

AI also changes how adversaries select the people they approach. The pattern ClearForce sees is systematic: identify an organization of interest, identify the roles that hold the access required, identify who fills those roles, and then match those individuals against breached personal data to find financial, legal, or personal pressure that makes them approachable through social engineering, coercion, or bribery. AI tools make each of those steps faster and cheaper for an adversary.

Inside the organization, AI creates risk from well-meaning employees and from the tools themselves. In the Ponemon research, 92 percent of organizations said generative AI has changed how employees share

information, 73 percent worry that unauthorized AI use is creating data exfiltration paths, and only 18 percent have fully integrated AI governance into their insider risk programs. AI agents compound the issue as they take on tasks and accumulate access much as employees do: 44 percent of respondents believe malicious use of AI agents will increase the risk of data theft, while only 19 percent classify AI agents as equivalent to human insiders.

Finally, AI is reshaping the workforce itself. Pew Research Center found that 52 percent of U.S. workers feel worried about the future impact of AI in the workplace, and 32 percent believe it will lead to fewer job opportunities for them. ClearForce expects that anxiety to add to the resentment and disengagement that have long been associated with insider risk, compounding the conditions described above.

The Governance Mismatch

Risk can change every day. Most controls do not. In nearly every organization, trust is assessed at a handful of discrete moments: a pre-hire check, a hire, a role change, a periodic review, and a departure. Between them sit visibility gaps. Inside a gap an employee can experience financial distress, a criminal or legal event, or undisclosed outside employment; access can change materially; a concern can surface with a manager and go no further; and a contractor in the supply chain can change hands entirely. The question a board should ask is simple: what does management actually know between the checks? Point-in-time controls did not fail. They answered the question they were asked, at the moment it was asked. Two cases show the limit of that answer.

KnowBe4, a security awareness training company, disclosed in July 2024 that it had hired a remote software engineer who was in fact a North Korean operative. He used a valid but stolen U.S. identity and a photograph that began as stock imagery and was altered with AI. He passed four video interviews, and the background check came back clear because the identity was real. Shortly after his company workstation arrived, he began loading malware. KnowBe4's security team detected and contained the device, no data was lost, and the company published a full account. Every control worked as designed, and the answer was still wrong.

The scale behind that case became clear in July 2025, when Christina Chapman of Arizona was sentenced to 102 months in federal prison for running a laptop farm from her home on behalf of North Korean IT workers. Using 68 stolen identities, those workers were placed at 309 U.S. companies, including Fortune 500 firms, an aerospace manufacturer, and an American carmaker, and the scheme generated more than \$17.1 million in illicit revenue between October 2020 and October 2023. Each of those companies ran a pre-hire process. When a nation-state can get itself hired onto a payroll, the debate about where the perimeter sits is over.

What the Security Stack Sees, and What It Does Not

Inside the environment, visibility is strong: user activity monitoring, SIEM, EDR and endpoint telemetry, data loss prevention, identity and access management, and physical access control. Outside it sit signals that can precede insider incidents: an arrest or court activity, significant financial stress, undisclosed or conflicting employment, public risk indicators, and employee concerns never connected to any cyber activity at all. A system can show a trusted user behaving normally while risk develops somewhere else entirely.

The GE Power case makes the point. Engineer Xiaoqing Zheng used steganography to hide stolen turbine technology files inside an innocuous digital photograph and emailed it from his GE account to his personal account. He was convicted of conspiracy to commit economic espionage, sentenced in January 2023 to two years in federal prison, and the conviction was affirmed on appeal in 2024. The network saw an employee email himself a picture: no policy violation, no volume anomaly, no suspicious destination. There was no rule to write.

The Harder Problem Is Enablement, Not Detection

Most organizations already hold more relevant human risk signals than their operating model can responsibly use. HR holds performance concerns, workplace complaints, and disclosures. Cyber holds technical telemetry. Corporate security holds physical access and incident history. Legal and compliance hold ethics reports and investigations. Self and peer reporting exists, and so does appropriate external and public information. That is six repositories, six owners, and six legal postures, usually with no single documented authority to bring them together. What is missing is policy, authority, process, and cross-functional structure.

The case of Jack Teixeira, the Massachusetts Air National Guardsman sentenced to 15 years for leaking classified documents on Discord, shows the cost. The Department of the Air Force Inspector General found that members of his unit had information about as many as four instances of questionable intelligence-seeking activity, that three individuals intentionally failed to report what they knew, fearing security officials might overreact, and that what was eventually reported was substantially minimized. Fifteen personnel were disciplined and the wing commander was relieved of command. Every signal existed. The structure to validate, contextualize, and escalate them did not. That is a governance failure well before it is a technology failure.

Connecting Human and Technical Signals

An effective model connects both sides before an incident develops. On the human side are criminal, financial, and employment changes, employee reporting, public indicators, and the context that explains changing behavior, since much of what surfaces will be a person having a hard year rather than a person becoming a risk. On the technical side is the existing security investment, which nothing here replaces. Between them is assessment: bring the signals together, establish context, investigate when warranted, and intervene early.

The operational value is concrete. When network monitoring flags questionable activity on an account, the credentials have either been compromised externally or an insider is involved. An external compromise calls for shutting the account down immediately; an insider issue may call for a discreet investigation. Network signals alone often cannot tell the two apart. Knowing whether the account holder shows no human risk indicators, or several, lets a team make that call faster and with far more confidence.

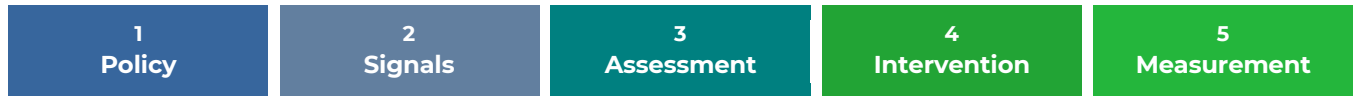
The goal is not surveillance. It is earlier awareness, fair review, and documented action.

A Five-Part Governance Framework

Moving from something that sounds like monitoring to a governed, compliant process of continuous evaluation rests on five components a board can name and hold management accountable for.

-  **Policy:** Which risks matter, to whom, why, and on what authority. Without an answer to the authority question, nothing downstream is defensible.
-  **Signals:** What information is appropriate, relevant, and permissible. Permissible is where FCRA, EEOC considerations, state privacy law, and consent posture live.
-  **Assessment:** How information is validated and put into context. Skipping this step is how organizations act on a name match rather than a person.
-  **Intervention:** Who acts, at what threshold, and how it is documented. A threshold not written down in advance is set by whoever is in the room.

 **Measurement:** Whether leadership can demonstrate control effectiveness, not merely activity. What the program learns then changes the policy, and the loop closes.



Policy > Signals > Assessment > Intervention > Measurement > Policy

A board does not need to understand the underlying technology. It needs to know that all five components exist, that each has a named owner, and that the loop closes.

Practices Leading Organizations Are Adopting

Organizations that are ahead on insider risk are pairing the framework with practical changes to policy, contracts, and reporting.

- **Updated conduct policies:** Codes of conduct now require self-reporting of criminal events and, for access-sensitive roles, of major derogatory financial events such as liens and judgments.
- **Evergreen consent:** Onboarding documentation authorizes appropriate background review over the course of employment rather than only at hire, consistent with FCRA disclosure and authorization requirements and applicable state law.
- **Supply chain flow-down:** Contract language requires that contractors with digital or physical access are screened properly and rechecked after initial screening, with security teams taking a clearer role in confirming it happens.
- **Remote and dual employment:** Policies address undisclosed concurrent employment. Part of that risk is productivity, but the insider risk is intellectual property loss when a full-time employee is also working for a competitor.
- **Modern reporting channels:** Traditional whistleblower hotlines are often underused, so organizations are adopting tools that make self, peer, and anonymous reporting easier and more visible.

What It Looks Like in Practice

At one critical infrastructure organization, an anonymized ClearForce deployment, leaders concluded that their pre-hire controls were sound but left a post-hire visibility gap. They implemented a policy-driven, consent-based continuous evaluation model jointly owned by Security, HR, and Legal. The monitored population was defined by role, access, and impact before any signal was collected, which helps establish the nexus between a role and the risk it presents. Signals were filtered to the events that violate defined policy. In ClearForce's data from this deployment, roughly 3 percent of the monitored population generated a reportable alert, and about 10 percent of those alerts met the reportable-offense threshold after filtering. Each was confirmed as a genuine match, put in context, and reviewed under a documented workflow. Within 30 days of implementation the organization had a validated high-risk case that would otherwise have sat in a visibility gap until the next periodic review, or until it became an incident. Management also gained trend analytics by charge type, policy, role, and location, turning individual personnel matters into a program it could manage.

Continuous visibility is not about more alerts; it is about isolating the few that require action and defensibly setting aside the rest.

Culture Comes First

None of this works in an organization that cuts corners under schedule pressure, rationalizes early signals away, waits for damage before acting, or treats security as another department's problem. Governance without culture produces a binder. Culture without governance produces good intentions and no record. Culture comes first because it determines whether people report at all. In the Teixeira case, colleagues who saw the behavior held back what they knew. When employees trust that reporting leads to a fair and proportionate response, and ideally to help for a struggling colleague rather than automatic punishment, they are more willing to speak up.

That same principle turns insider risk management into employee support. Many organizations offer employee assistance programs that go underused because people do not ask for help. Seeing financial distress early creates the chance to course-correct, retain a good employee, and reduce risk in one motion. Employees are also more supportive than leadership often expects: across ClearForce deployments in regulated industries, first-time consent capture routinely lands in the mid-90 percent range when the program is transparent and the consent is real.

Questions Leadership Should Be Able to Answer

- **Authority:** Which function owns the decision to combine HR, security, cyber, and legal signals, and where is that authority documented?
- **AI exposure:** Is AI governance integrated into the insider risk program, and are AI agents with network access treated as identities to be governed?
- **Thresholds:** Are intervention thresholds written down in advance, with a named owner and a documented response?
- **Hiring integrity:** Can the hiring process detect a valid but stolen identity, and what happens if one is already on the payroll?

Build Trust Before You Need It

Samsung, Arup, KnowBe4, Chapman, GE, and Teixeira span different industries, motives, and levels of sophistication, yet each involved trusted access used in ways no perimeter tool was built to catch. The lesson is consistent: the earlier the human signal is seen, and the more defensibly an organization can act on it, the smaller the incident. Insider risk is not a question of if. The difference between organizations is whether they detect it in days or in months, and whether what they do next is documented and fair.

Security is built in layers of continually improving capability. The question for leadership is not whether the current stack is good. It is which layer is missing, and whether the governance exists to use it responsibly once it is added.

Build trust before you need it.

About ClearForce. ClearForce is a human risk signals technology company founded in 2016 and based in Tysons Corner, Virginia. Its patented Resolve™ platform helps federal agencies, critical infrastructure operators, financial institutions, and the defense industrial base continuously and lawfully identify behavioral risk and misconduct, with privacy and governance built in. Learn more at www.clearforce.com.

Sources

1. Ponemon Institute, *2026 Cost of Insider Risks Global Report*, sponsored by DTEX, February 2026. Measurement year 2025; 8,750 practitioners in 354 organizations. Annual, per-incident, incident-count, and containment figures. ponemonsullivanreport.com/2026/05/2026-cost-of-insider-risks-global
2. DTEX (report sponsor), *Insider Risk Costs Hit \$19.5M USD Per Year as AI Creates New Blind Spots*, press release via GlobeNewswire, February 24, 2026. Generative AI, unauthorized AI, AI governance, and AI agent findings. globe.newswire.com
3. Carnegie Mellon University Software Engineering Institute, CERT Division, *Common Sense Guide to Mitigating Insider Threats*, Fifth Edition, distributed by the National Insider Threat Task Force (archive copy dated February 9, 2018). archive.dni.gov/files/NCSC/documents/nittf/20180209-CERT-Common-Sense-Guide-Fifth-Edition.pdf
4. Bloomberg, *Samsung Bans ChatGPT, Google Bard, Other Generative AI Use by Staff After Leak*, May 2, 2023; TechCrunch, *Samsung Bans Use of Generative AI Tools Like ChatGPT After April Internal Data Leak*, May 2, 2023. techcrunch.com/2023/05/02/samsung-bans-use-of-generative-ai-tools-like-chatgpt-after-april-internal-data-leak
5. CNN Business, *Arup Revealed as Victim of \$25 Million Deepfake Scam Involving Hong Kong Employee*, May 16, 2024. edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk; Fortune, May 17, 2024.
6. Forbes, *A Voice Deepfake Was Used to Scam a CEO Out of \$243,000*, September 3, 2019, and Dark Reading, September 4, 2019, both reporting The Wall Street Journal's account, which drew on the company's insurer, Euler Hermes. forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000
7. Federal Bureau of Investigation, San Francisco Field Office, *FBI Warns of Increasing Threat of Cyber Criminals Utilizing Artificial Intelligence*, May 8, 2024. fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence
8. Pew Research Center, *U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace*, February 25, 2025. Survey of 5,273 employed U.S. adults, October 2024. pewresearch.org/social-trends/2025/02/25/u-s-workers-are-more-worried-than-hopeful-about-future-ai-use-in-the-workplace
9. KnowBe4, *How a North Korean Fake IT Worker Tried to Infiltrate Us*, July 23, 2024, and *North Korean Fake IT Worker FAQ*. blog.knowbe4.com/how-a-north-korean-fake-it-worker-tried-to-infiltrate-us
10. U.S. Department of Justice, U.S. Attorney's Office for the District of Columbia, sentencing of Christina Marie Chapman in the \$17 million IT worker fraud scheme, July 24, 2025. justice.gov/usao-dc/pr/arizona-woman-sentenced-17m-it-worker-fraud-scheme-illegally-generated-revenue-north
11. U.S. Department of Justice, *New York Man Charged With Theft of Trade Secrets*; U.S. Attorney's Office for the Northern District of New York, *Former GE Power Engineer Sentenced for Conspiracy to Commit Economic Espionage*, January 3, 2023; *United States v. Zheng*, No. 23-6070 (2d Cir. August 27, 2024). law.justia.com/cases/federal/appellate-courts/ca2/23-6070/23-6070-2024-08-27.html
12. U.S. Department of Justice, *Former Air National Guardsman Sentenced to 15 Years in Prison for Unlawfully Disclosing Classified National Defense Information*, November 12, 2024. justice.gov/usao-ma/pr/former-air-national-guardsman-sentenced-15-years-prison-unlawfully-disclosing-classified
13. Department of the Air Force Inspector General, *Report of Investigation (S9691)*, December 11, 2023, af.mil/Portals/1/documents/2023SAF/UD_ROI_-_11_Dec_23.pdf; U.S. Air Force, *Results of Investigation into A1C Teixeira's Unit*, December 11, 2023.
14. ClearForce, anonymized customer deployment metrics and consent capture rates observed across regulated industry deployments. First-party data.