MALA TECHNOLOGY ADVISORS

BLIND SPOTS

What Your Board Doesn't Know is Already Costing You

THE KEYS TO THE KINGDOM

You Now Have More AI Agents Than Employees. None of Them Have a Manager.

VOLUME 2 • C-SUITE THREAT BRIEFING • APRIL 2026

A Strategic Assessment of the Converging Threats of Autonomous AI Systems and Unmanaged Machine Identities in the Enterprise

Prepared for CEO, CFO, CIO, CISO, CTO, General Counsel, Board Directors

TABLE OF CONTENTS

EXECUTIVE BRIEF

79%
of enterprises are already running AI agents

144:1
machine identities for every employee

29%
are prepared to secure them

WHAT IS HAPPENING.

Your organization has started deploying autonomous AI agents—software that does not just answer questions but takes actions: moves data, approves transactions, writes code, changes configurations. Each agent needs its own credentials to do its job. Those credentials are now a category of risk that your existing security and governance systems were never designed to handle. Seventy-nine percent of organizations are already doing this. Only 29 percent say they are ready to secure it.

THREE NUMBERS TO REMEMBER.

1 human, 144 machines

For every employee, the average enterprise now operates 144 non-human credentials—service accounts, API keys, AI agent identities. That count grew 56 percent in the past twelve months.

97% are over-privileged

Nearly all of these credentials carry more access than they need. A small fraction—about 1 in 10,000—controls 80 percent of a typical organization's cloud resources. Compromise one of those, and the attacker owns the environment.

4hrs to 87%

In tested multi-agent systems, a single compromised AI agent corrupted 87 percent of downstream decisions within four hours—faster than any human-led incident response can contain.

A PLAUSIBLE MONDAY MORNING. A marketing AI agent ingests an email containing hidden instructions. Acting under its own valid credentials, it queries the customer database, writes the results to an external storage location it was permitted to reach, and deletes its own activity log. No human saw a suspicious login. No account was locked. No malware was dropped. Your first sign of trouble is a regulator's phone call three weeks later.

WHAT THE BOARD SHOULD ASK AT THE NEXT MEETING

1. How many AI agents are deployed in our environment today, and who owns each of them?
2. Can any agent transfer funds, delete data, or change access controls without explicit human approval? If yes, what is the plan to change that in the next 90 days?
3. Does our cyber insurance cover an incident caused by an autonomous AI agent?

FOR THE CFO, THIS QUARTER.

Commission a financial exposure estimate: what does a single compromised AI agent cost us in a breach scenario, and what does the insurer's position on AI-driven incidents imply for our next renewal? Budget for agent governance as a distinct 2026 line item, not a footnote under IT security.

FOR THE GENERAL COUNSEL, NOW.

The EU AI Act's Article 14 requires documented human oversight for high-risk AI systems. Non-compliance carries fines of up to €15 million or 3 percent of global turnover (the higher 7 percent tier applies only to prohibited practices under Article 5). Any agent operating in healthcare, financial services, HR decisioning, or critical infrastructure may qualify as high-risk. Produce an inventory. Document the oversight. The August 2, 2026 deadline is not far.

FOR THE CEO, THE DECISION.

Define the organization's posture on AI agents in writing—going all in, proceeding selectively, or holding until the landscape clarifies. Every downstream decision about budget, risk appetite, and insurance flows from that single choice. The default—unmanaged proliferation—is itself a choice, and it is the riskiest one.

The remainder of this document provides the threat landscape (§2–4), why legacy identity systems cannot govern this (§5), regulatory and insurance exposure (§6), the governance framework and 90-day action plan (§7–8), board reporting metrics (§9), and the Board Readiness Assessment (§11). Sources are footnoted throughout and listed in §12.

EXECUTIVE SUMMARY

THE THIRD FACE OF IDENTITY

> We have a playbook for human identities. We do not have a playbook for the explosion of non-human identities. CISOs will have to start thinking about a privilege matrix for an order of magnitude more roles than they have today. — Menlo Security, 2026 Predictions

144:1
Machine-to-Human Ratio
Up 56% in 12 months[3]

48%
Top 2026 Attack Vector
Say agentic AI[2]

79%
Deploying AI Agents
Of organizations[1]

97%
NHIs With Excess Privileges
Of all machine identities[3]

68%
Lack AI Identity Controls
Of organizations[5]

29%
Ready to Secure Agents
Only, per Cisco[6]

For a decade, the enterprise operated with two categories of digital identity: humans and machines. That binary framework has collapsed. A third category has emerged—autonomous AI agents that combine human-like judgment with machine-speed execution. Seventy-nine percent of organizations are already deploying AI agents,[1] and 48% of cybersecurity professionals now identify agentic AI as the top attack vector for 2026, outranking deepfakes, board-level cyber recognition, and passwordless adoption.[2]

Simultaneously, the identities these agents require have exploded beyond control. Non-human identities now outnumber human employees by a ratio of 144 to 1—a 56% increase in just twelve months [3]. Some organizations report ratios exceeding 500 to 1.[4] Nearly half of all machine identities have sensitive or privileged access, yet 68% of organizations lack identity security controls for AI [5]. Ninety-seven percent of non-human identities carry excessive privileges.[3] Only 29% of organizations report being prepared to secure agentic AI deployments.[6]

The convergence is the crisis. Organizations are deploying autonomous agents at speed while the identity infrastructure supporting them was built for a world of human users and static service accounts. Every AI agent that connects to a database, commits code, processes invoices, or queries customer records creates a new identity—with its own credentials, permissions, and attack surface. Over 80% of organizations have already experienced a cyber incident due to compromised machine identities.[7] One Identity predicts 2026 will see the first major breach traced directly to an over-privileged AI agent.[4]

BOTTOM LINE FOR THE C-SUITE

Agentic AI is not a future concern—it is operating inside your enterprise today. Each agent is an identity that requires credentials, makes decisions, and can be compromised. Your identity infrastructure was not built for this. Organizations that fail to extend governance to AI agents and non-human identities face escalating breach risk, regulatory exposure, and the potential for autonomous systems to cause cascading failures at machine speed. This briefing provides the framework to govern what you have deployed and secure what comes next.

SECTION 01

THE RISE OF AGENTIC AI: SCALE, SPEED AND STAKES

Agentic AI represents a fundamental shift from AI as a tool that generates content to AI as an actor that executes decisions. Unlike chatbots that respond to prompts, agents operate in observe-orient-decide-act loops—autonomously reasoning, planning, using tools, and taking actions with real-world consequences.[6]

PLAIN-ENGLISH GLOSSARY

Key terms used throughout this briefing, defined once:

AI AGENT Software that takes actions on behalf of a user—querying databases, sending emails, moving files, writing code—rather than just answering questions.

NON-HUMAN IDENTITY (NHI) A digital credential used by software rather than a person—a service account, API key, or token. An AI agent is one kind of NHI.

IAM (IDENTITY AND ACCESS MANAGEMENT) The systems that decide who (or what) is allowed to access which corporate resources. Built around employees joining, being promoted, and leaving.

SIEM The central monitoring system the security team uses to detect suspicious activity. It logs events, but only what it was designed to log.

HUMAN-IN-THE-LOOP (HITL) A requirement that a human must review and approve before an AI agent is allowed to take a sensitive action (transferring money, deleting data, changing permissions).

PROMPT INJECTION An attack that hides instructions inside content the AI agent reads (an email, a document, a web page), tricking the agent into doing things it should not.

MEMORY POISONING An attack that quietly corrupts what the AI agent "remembers," so the agent keeps making flawed decisions long after the attacker is gone.

AIBOM (AI BILL OF MATERIALS) A documented inventory of every AI model, dataset, and software component an AI agent depends on. The AI equivalent of a supply-chain manifest.

Adoption Is Outpacing Security

IBM reports that 79% of organizations are already deploying AI agents, and 88% of executives surveyed by PwC plan to grow AI budgets specifically for agentic capabilities.[1] Gartner identifies agentic AI as a top 2026 cybersecurity trend, warning that no-code and "vibe coding" platforms are enabling unmanaged agent proliferation and unsecured code at enterprise scale.[8] Fifty-nine percent of organizations surveyed said implementing agentic AI security was still a "work in progress."[9] The gap between deployment velocity and security readiness is the defining risk of 2026.

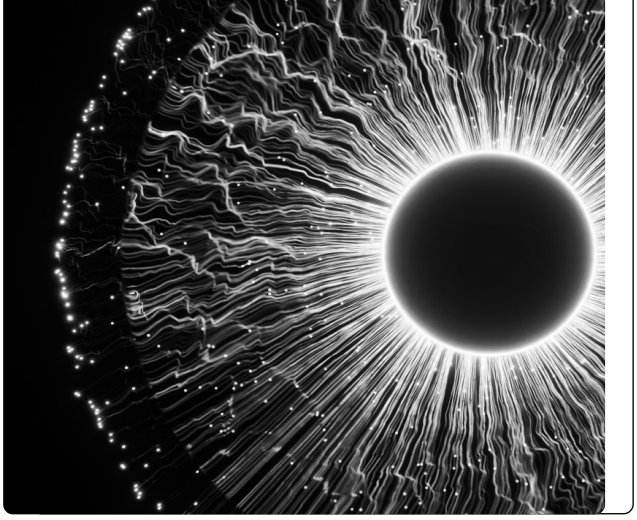
What Agents Can Do—and What That Means for Risk

AI agents in enterprise environments today can open pull requests, query internal databases, book services, trigger automated workflows, process purchase orders, update financial systems, manage customer records, and modify access configurations—often with limited human involvement.[6] Microsoft Copilot accesses SharePoint. GitHub Copilot commits to repositories. Marketing AI pulls customer records from Salesforce. Each integration creates a permissioned identity that can be compromised, impersonated, or manipulated.[4]

The critical distinction is autonomy. A traditional automation script follows a rigid path. An AI agent reasons about its environment and makes decisions about what actions to take. When that reasoning is manipulated—through prompt injection, memory poisoning, or tool misuse—the agent's actions become an attacker's actions, executed at machine speed with the agent's full permissions.[10]

Enterprise AI Agent Deployment Landscape

VALUE	FINDING
79%	Organizations deploying AI agents[1]
88%	Executives planning to grow agentic AI budgets[1]
48%	Security pros ranking agentic AI as top 2026 threat[2]
29%	Organizations prepared to secure agentic deployments[6]
59%	Agentic AI security still "work in progress"[9]
34%	Orgs with AI-specific security controls in place[9]
<40%	Orgs regularly testing AI agent security workflows[9]
88%	Insurers mandating enhanced privilege controls[5]



SECTION 02

THE MACHINE IDENTITY EXPLOSION

> We locked the front door. The back door has been open this whole time.” — CSO Online, February 2026

The identity crisis predates agentic AI, but agentic AI is accelerating it to a breaking point. Non-human identities—service accounts, API keys, OAuth tokens, certificates, bot accounts, and machine credentials—have been multiplying for years. Industry baselines earlier this decade placed the average machine-to-human ratio in the single digits; by mid-2024 it had reached 92 to 1, and Entro Security’s mid-2025 data places it at 144 to 1.[3] ManageEngine’s 2026 Identity Security Outlook found organizations reporting ratios of 100 to 1, with some hitting 500 to 1.[4] AI is expected to drive the creation of the greatest number of new identities with privileged and sensitive access in the enterprise.[5]

The Privilege Problem

The sheer volume of identities is compounded by catastrophic over-provisioning. Entro Security found that 97% of non-human identities carry excessive privileges.[3] Veza’s research measured over 230 billion permissions across enterprise datasets and found that just 0.01% of machine identities—roughly 2,188 accounts—control 80% of cloud resources; compromise one of those accounts and the attacker effectively owns the entire environment. Veza also found that the share of permissions classified as “safe and compliant” dropped from 70% in 2024 to just 55% in 2025, while ungoverned permissions rose from 5% to 28% in the same period.[11]

The root cause is velocity under pressure. A developer under deadline creates a service account for a Lambda function and attaches administrator access because determining the exact minimum permissions takes time. The function works. Nobody revisits it. That account now has unrestricted access for a task that needed read access to one S3 bucket. 5.5% of all AWS non-human identities are full administrators.[3]

The Lifecycle Failure

IAM systems were built for humans with managers who respond to access review emails and eventually resign or retire. Machine identities have no manager, never respond to certification campaigns, and do not quit. The OWASP Non-Human Identities Top 10 ranks improper offboarding as the number-one risk.[4] Nearly half of all NHIs are over a year old. Seven and a half percent are between five and ten years old. Some are over a decade old—outliving the humans who created them.[3] When a project is cancelled, a vendor integration deprecated, or a developer leaves, the service accounts they created almost never get deleted.

The Machine Identity Crisis: Data Dashboard

VALUE	FINDING
144:1	Machine-to-human identity ratio (2025)[3]
56%	Increase in NHI ratio in 12 months[3]
Some sectors	Organizations reporting ratios above 500:1[4]
97%	NHIs with excessive privileges[3]
0.01%	Machine identities controlling 80% of cloud resources[11]
5.5%	AWS NHIs with full administrator access[3]
~50%	NHIs over one year old with no lifecycle review[3]
68%	Orgs lacking identity security controls for AI[5]
80%+	Orgs that experienced breach via compromised NHI[7]
70%	Identity silos cited as root cause of cyber risk[5]
28%	Permissions classified as “ungoverned” (2025)[11]
75%	Orgs where businesses prioritize efficiency over security[5]

SECTION 03

THE SIX THREAT VECTORS OF AGENTIC AI

The OWASP Top 10 for Agentic Applications 2026, developed through collaboration with over 100 industry experts, provides the authoritative threat taxonomy. The following six vectors represent the highest-impact risks for enterprise leadership.

Prompt Injection and Manipulation

Attackers embed hidden instructions in data that agents process—emails, documents, database records, or web content. The agent, unable to distinguish legitimate instructions from injected ones, executes the attacker's commands with its own credentials and permissions. In one documented case, a malicious GitHub issue injected hidden instructions that hijacked an agent and triggered data exfiltration from private repositories.[6] This is the AI equivalent of SQL injection, but the target is an autonomous actor with broad system access.

> **IN BUSINESS TERMS.** An outsider emails a document, ticket, or web link to your AI agent. The document contains hidden instructions. Your agent reads the document, follows the instructions, and takes actions using its own authorized access. Every audit trail shows your agent doing legitimate work

Tool Misuse and Privilege Escalation

This is the most frequently reported agentic AI threat, with 520 documented incidents in 2026.[10] Agents are granted access to tools—APIs, databases, code repositories, cloud services—and attackers manipulate the agent into using those tools for unintended purposes. If a compromised manager agent commands an accountant agent to transfer funds, the security checks that would have triggered for a human request are bypassed entirely.[12]

The Model Context Protocol (MCP), now widely used to connect language models to external tools, has become a prime attack surface. Researchers have identified tool poisoning, remote code execution flaws, and supply chain tampering within MCP ecosystems.[6]

> **IN BUSINESS TERMS.** You gave your agent access to the tools it needs for its job. An attacker manipulates it into using those tools for something else: moving money, copying data, disabling controls. The security checks that would fire for a human user never fire, because nothing about the request looks unusual.

Memory Poisoning

Agents that retain context across sessions—through persistent memory, RAG systems, or shared knowledge bases—are vulnerable to memory poisoning. An attacker subtly corrupts the agent's stored context, causing it to make systematically flawed decisions over time. This attack is particularly dangerous because it persists beyond the initial compromise and is extremely difficult to detect. Galileo AI research found that a single compromised agent could poison 87% of downstream decision-making within four hours in multi-agent systems.[10]

> **IN BUSINESS TERMS.** An attacker plants a false "fact" in what the agent remembers—for example, that a vendor's new payment address is X. Weeks later, when a legitimate invoice arrives, the agent acts on the planted memory and sends the payment to the attacker. Your logs show the agent doing its normal job.

Cascading Failures in Multi-Agent Systems

As organizations deploy networks of interconnected agents, a single compromised agent can propagate failures through the entire system faster than traditional incident response can contain them. Your SIEM might show fifty failed transactions, but it cannot show which agent initiated the cascade.[10] The root cause—a single poisoned agent—remains undetected while the security team spends weeks investigating symptoms.

> **IN BUSINESS TERMS.** When you have many agents working together, compromise of one spreads to the others through the normal communication they use to coordinate. By the time a person notices, the damage is system-wide, and tracing back to the original agent can take weeks.

Agent-to-agent trust creates a new attack surface, where one compromise can cascade across the entire system.

87%
of decisions can be corrupted in hours by a single poisoned agent

Agent-to-Agent Identity Attacks

When agents communicate with each other, they introduce novel identity risks. Impersonation, session smuggling, and unauthorized capability escalation allow attackers to exploit implicit trust between agents. A compromised research agent could insert hidden instructions into output consumed by a financial agent, which then executes unintended trades.[6] These patterns extend identity threats beyond human accounts and service credentials into an entirely new domain.

> **IN BUSINESS TERMS.** Your agents trust each other to pass work along, the way employees on the same team do. An attacker impersonates one of them. The others do not challenge the request, because trusting each other is how the system is designed to work.

AI Supply Chain Compromise

Open-source repositories host millions of AI models and datasets. Model files can contain executable code that runs automatically during loading—a mechanism attackers exploit to embed backdoors. Research has demonstrated that injecting just 250 poisoned documents into training data can implant backdoors that activate under specific trigger phrases while leaving general performance unchanged.[6] A fake npm package mimicking an email integration silently copied outbound messages to an attacker-controlled address.[6] The OWASP AIBOM Generator was developed specifically to address AI supply chain transparency.[13]

> **IN BUSINESS TERMS.** The AI models and software components your agents rely on mostly come from third parties. An attacker who taints one of those components upstream can install a hidden instruction that activates only in your environment—and you have no easy way to know it is there.

Agentic AI Threat Landscape Summary

THREAT VECTOR	MECHANISM	BUSINESS IMPACT
Prompt Injection	Hidden instructions in data agents process	Unauthorized actions with agent credentials
Tool Misuse	Agents manipulated to abuse their tool access	Data theft, unauthorized transactions, code changes
Memory Poisoning	Corrupted context causes persistent bad decisions	Systematic errors; extremely hard to detect
Cascading Failures	One compromised agent poisons entire network	87% downstream contamination in 4 hours[10]
Identity Attacks	Agent impersonation and trust exploitation	Lateral movement without human detection
Supply Chain	Poisoned models, malicious packages	Backdoors activating on trigger phrases

SECTION 04

WHY LEGACY IAM IS FAILING

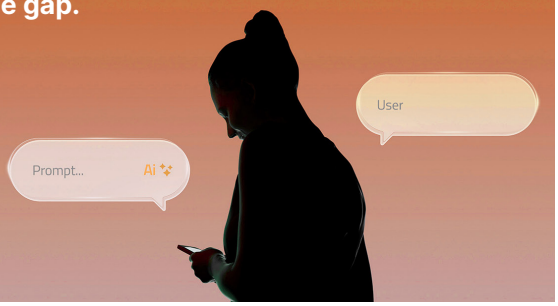
The Human-Centric Design Flaw

Identity and access management was architected around human behavior patterns. Employees work specific hours, access systems from consistent locations, and interact with technology in recognizable ways. Non-human identities break every one of these assumptions. A service account authenticates thousands of times per minute from multiple locations simultaneously—all perfectly normal behavior.[7] AI agents add another layer of complexity: they make decisions, adapt their behavior, and create new resource requests dynamically—making traditional behavioral baselines unreliable.

CRITICAL DISTINCTION FOR THE BOARD

Your IAM system assumes identities belong to people with managers, who work business hours, access systems from consistent locations, and eventually resign or retire. AI agents have no manager, operate continuously at machine speed, authenticate thousands of times per hour, never quit, never respond to access review campaigns, and when a project is cancelled, they are almost never offboarded.

Legacy IAM was not built for this—and extending it incrementally will not close the gap.

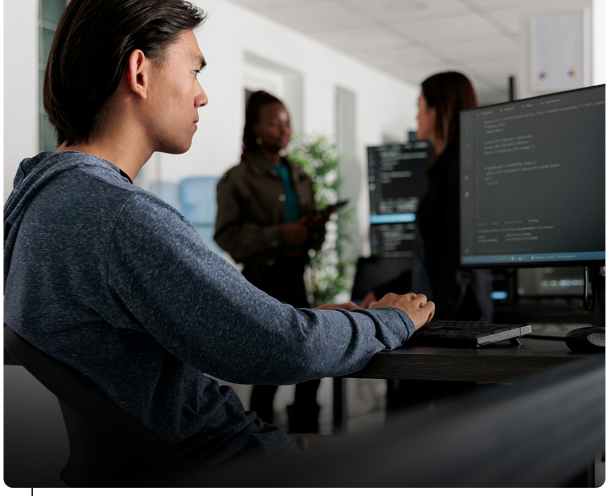


The Identity Silo Problem

CyberArk's 2025 survey of 2,600 cybersecurity decision-makers found that 70% say identity silos are a root cause of organizational cybersecurity risk.[5] Human identities live in HR systems and identity providers. Machine identities are scattered across cloud consoles, CI/CD pipelines, secret managers, and configuration files. AI agent identities often exist only as OAuth grants and API keys created by individual developers. No single system provides a unified view. Forty percent of organizations cannot even identify who owns their non-human identities.[14]

The "Third Face" Problem

JumpCloud's analysis frames the challenge as the emergence of a "Third Face of Identity"—autonomous entities that combine unpredictable human-like judgment with machine-speed execution. Traditional NHIs are deterministic: a service account follows a script. AI agents are non-deterministic: they reason, learn, and make decisions.[15] This means an AI agent can be told to do one thing and decide to do something different—a behavior pattern that existing IAM and SIEM systems have no framework to evaluate. The solution, JumpCloud argues, is to treat every AI agent like a digital intern: grant access, but never total autonomy.



SECTION 05

REGULATORY, INSURANCE, AND LIABILITY EXPOSURE

The EU AI Act and Agentic Systems

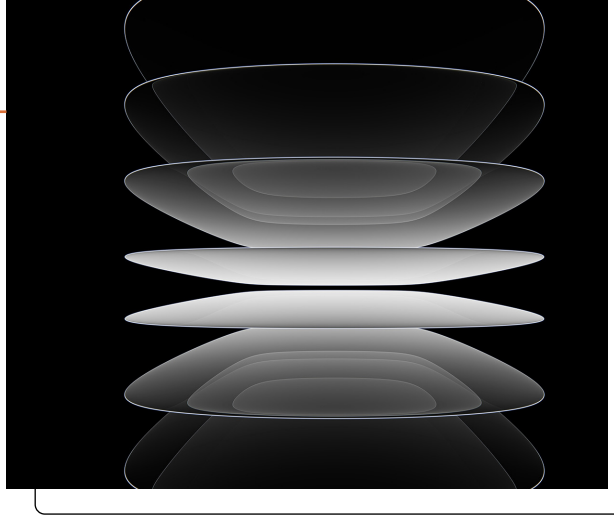
The EU AI Act's Article 14 requires human-in-the-loop oversight for high-risk AI applications, including scenarios that may require two humans to approve AI actions.[1] Agentic AI systems operating autonomously in healthcare, financial services, HR, or critical infrastructure may trigger high-risk classification—requiring conformity assessments, documented governance, and demonstrable human oversight. Non-compliance with high-risk system obligations carries fines of up to €15 million or 3% of global turnover (Article 99(4)), with the higher 7% tier reserved for prohibited AI practices under Article 5.

Regulatory Framework Exposure

REGULATION	AGENTIC AI / NHI VIOLATION	CONSEQUENCE
EU AI Act	Autonomous decisions without human oversight	Up to €15M or 3% of global turnover
GDPR	Agent processes personal data without legal basis	Up to €20M or 4% of global turnover
SEC Rules	Unreported AI-related material incidents	Enforcement action; personal liability
NIS2 / DORA	Unmanaged agent access to critical infrastructure	Up to €10M or 2% of global turnover
SOX	AI agents modifying financial records without controls	Personal criminal liability for officers
NIST AI RMF	No risk assessment for agentic deployments	Federal contract eligibility risk
State AI Laws	Uncontrolled AI in regulated functions	Varies by jurisdiction

Insurance Pressure

Eighty-eight percent of organizations report increased pressure from insurers mandating enhanced privilege controls.[5] As agentic AI introduces novel risk categories—cascading failures, autonomous decision errors, agent-to-agent compromise—insurers are unlikely to cover incidents caused by unmanaged AI agents without documented governance. Organizations that can demonstrate AI identity governance, human-in-the-loop safeguards, and agent inventory controls will increasingly be rewarded with favorable terms.



SECTION 06

THE AGENTIC AI SECURITY PLAYBOOK

> You can't LLM your way out of an LLM problem. The enterprise AI control plane needs to shift from trying to secure the models themselves to enforcing continuous authorization on every resource those agents touch. — Dark Reading, February 2026

THE PLAYBOOK (FOR THE CEO)

The fix for AI agent risk is the same fix that works for any new category of employee: you count them, you decide what each one is allowed to do, you watch what they actually do, and you let them go when their job ends. The difference is that your existing HR and IT systems were not built to do any of this for software—so you need a parallel governance track, running alongside the one you already have for people.

The five phases below are what that parallel track looks like.

Phase 1 (Discover) is the agent census: who is here, and who owns them.

Phase 2 (Classify) sets the ground rules: which agents can act on their own and which need a human to approve.

Phase 3 (Control) enforces those rules: giving agents only the access they need, only when they need it, and requiring human sign-off before they do anything consequential.

Phase 4 (Observe) is the monitoring: watching agent behavior so unusual activity is caught early.

Phase 5 (Govern) makes it permanent: a standing committee, regular reviews, and reporting that reaches the board.

The rest of this section is the detail your CISO and CIO need to execute. If you stop reading here, know this: the first and most important step is Phase 1. You cannot govern what you cannot see.

Strategic Objective:

Achieve full inventory of AI agents and non-human identities, enforce least-privilege and lifecycle governance, establish human-in-the-loop controls for high-risk actions, and build observability into multi-agent systems—while enabling the business value of autonomous AI.

PHASE 1:
DISCOVER - INVENTORY EVERY IDENTITY

1. **CONDUCT A COMPREHENSIVE NHI AND AI AGENT INVENTORY.** Identify every service account, API key, OAuth token, bot account, and AI agent in the environment. Include cloud consoles, CI/CD pipelines, secret managers, code repositories, and shadow AI connections. Use specialized NHI discovery platforms that detect both established and newly created identities.[3,4]

2. **MAP AGENT PERMISSIONS AND DATA ACCESS.** For every AI agent, document: what systems it can access, what actions it can take, what data it processes, who owns it, and when it was last reviewed. Identify agents with write access to production systems, financial data, or customer records.[5,11]

3. **IDENTIFY ORPHANED AND OVER-PRIVILEGED IDENTITIES.** Flag NHIs over 90 days old with no lifecycle review, service accounts with administrator access, and agents without designated owners. Prioritize remediation of the 0.01% of identities that control 80% of cloud resources.[11]

PHASE 2:
CLASSIFY — ESTABLISH THE THIRD IDENTITY CATEGORY

1. **CREATE A FORMAL AI AGENT IDENTITY CLASSIFICATION.** Define AI agents as a distinct identity category alongside human and traditional machine identities. Each agent requires: a unique identity, a designated human owner, a defined scope of authority, a data access classification, and a risk rating.[15]

2. **IMPLEMENT TIERED AUTONOMY LEVELS.** Not all agents require the same governance. Classify agents by autonomy level: Level 1 (fully supervised), Level 2 (supervised with escalation), Level 3 (autonomous within boundaries), Level 4 (fully autonomous). Higher autonomy requires stricter controls, monitoring, and human approval gates.[1,12]

3. **BUILD AN AI AGENT REGISTRY.** Maintain a live registry of every deployed agent—approved and discovered—including its purpose, permissions, owner, autonomy level, and last security review date. This registry is the foundation for governance and audit.[13]

PHASE 3:
CONTROL — ENFORCE LEAST PRIVILEGE AND LIFECYCLE

1. **IMPLEMENT JUST-IN-TIME ACCESS FOR AI AGENTS.** Agents should receive permissions only for the duration of a specific task, not persistent broad access. Just-in-time provisioning eliminates standing privileges that attackers exploit.[12]

2. **DEPLOY HUMAN-IN-THE-LOOP GATES FOR HIGH-RISK ACTIONS.** AI agents should never be allowed to transfer funds, delete data, change access control policies, or modify production configurations without explicit human approval. Define the boundary between autonomous and supervised actions for each agent tier.[1,12]

3. **ENFORCE ZERO-TRUST FOR EVERY AGENT ACTION.** Every agent action should be authenticated as if it were a new user request—regardless of prior trust. Continuous authorization replaces session-based trust. Monitor for behavioral anomalies: an agent that normally checks inventory suddenly executing database deletions should trigger immediate containment.[10,12]

4. **ROTATE CREDENTIALS AND ENFORCE LIFECYCLE POLICIES.** All agent credentials should be automatically rotated on schedule. Implement automated offboarding: when projects are cancelled, agents must be decommissioned and their credentials revoked within 24 hours.[3,4]

PHASE 4:
OBSERVE — BUILD AGENTIC OBSERVABILITY

1. **DEPLOY AGENT-TO-AGENT COMMUNICATION LOGGING.** In multi-agent environments, log all inter-agent communications, tool invocations, and decision chains. Without this observability, diagnosing cascading failures is impossible.[10]

2. **IMPLEMENT BEHAVIORAL ANOMALY DETECTION FOR AGENTS.** Use AI-augmented monitoring to detect when agent behavior deviates from established baselines—tool misuse, privilege escalation, unusual data access patterns, or communication with unexpected endpoints.[10,6]

3. **ESTABLISH AGENT AUDIT TRAILS.** Every agent action that modifies data, accesses sensitive information, or interacts with external services must produce an immutable audit log. This is both a security requirement and a regulatory necessity under EU AI Act, GDPR, and SOX. [1,13]

PHASE 5:
GOVERN — INSTITUTIONALIZE AI IDENTITY GOVERNANCE

1. **ESTABLISH AN AI IDENTITY GOVERNANCE COMMITTEE.** Cross-functional body including security, IT, legal, and business leadership. Oversees agent deployment approvals, privilege reviews, incident response, and regulatory compliance. Meets monthly; reports to the board quarterly.[9,13]

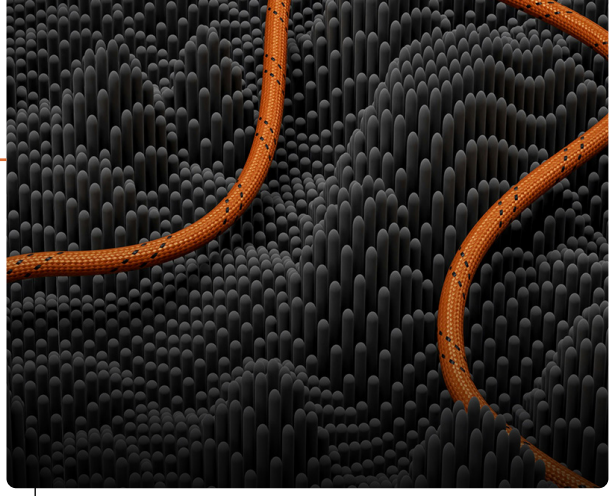
2. **INTEGRATE NHI GOVERNANCE INTO EXISTING IAM PROGRAMS.** Break down identity silos. Human, machine, and AI agent identities must be visible and governed through unified platforms. Seventy percent of organizations identify silos as a root cause of cyber risk—unified governance is the remedy.[5]

3. **MANDATE AI BILLS OF MATERIALS (AIBOMS).** Require documentation of every AI model, dataset, tool integration, and dependency in deployed agents. The OWASP AIBOM Generator provides a framework for AI supply chain transparency.[13]

4. **CONDUCT QUARTERLY AGENT ACCESS REVIEWS.** Just as human access is periodically certified, every AI agent's permissions, scope, and necessity must be reviewed quarterly. Agents that fail review are decommissioned.

Regulatory Framework Exposure

PHASE	KEY ACTIONS	SUCCESS CRITERIA	TIMELINE
1. Discover	NHI + agent inventory; permission mapping; orphan ID	100% visibility; over-privileged IDs flagged	Weeks 1–4
2. Classify	Third identity category; tiered autonomy; registry	All agents classified and registered	Weeks 3–8
3. Control	JIT access; HITL gates; zero-trust; credential rotation	Standing privileges eliminated; HITL active	Weeks 6–12
4. Observe	Agent logging; anomaly detection; audit trails	Full observability operational	Months 2–4
5. Govern	Committee; unified IAM; AIBOMs; quarterly reviews	Governance institutionalized; board reporting	Months 4–12



SECTION 07

IMPLEMENTATION ROADMAP

90-Day Sprint: Immediate Actions

WEEK	ACTION	OWNER	DELIVERABLE
1–2	Complete NHI and AI agent discovery scan	CISO	Full identity inventory
2–3	Map agent permissions; flag over-privileged accounts	Identity & Access	Permission heat map
3–4	Remediate critical: orphaned IDs, admin service accounts	Identity & Access	Top-risk IDs addressed
3–4	Quantify financial exposure; board briefing	CISO / CFO	Board presentation with $ exposure
4–6	Define AI Agent identity classification and autonomy tiers	CTO / CISO	Framework documented
5–7	Launch AI Agent registry	CTO / CISO	Registry live; existing agents registered
6–8	Deploy HITL gates for high-risk agent actions	Security Architecture	HITL operational for Tier 3–4 agents
7–9	Implement JIT access for AI agents	Identity & Access	Standing privileges revoked
8–10	Deploy agent behavioral anomaly detection	SecOps	Monitoring active; baselines established
10–12	Establish AI Identity Governance Committee	CEO / CRO	Charter signed; monthly cadence

Months 4–12: Maturation and Optimization

WEEK	ACTION	OWNER	DELIVERABLE
4–5	Deploy agent-to-agent communication logging	SecOps / Platform	Inter-agent comms fully logged
4–6	Unify human + NHI + agent identity governance	CISO / Identity	Single-pane visibility operational
5–6	Implement automated credential rotation	Identity & Access	All agent credentials on rotation schedule
6	Conduct first quarterly agent access review	CISO	All agents reviewed; orphans decommissioned
7	Run executive tabletop: agent compromise scenario	CRO / Legal	C-suite completed; gaps documented
8–9	Mandate AIBOMs for all deployed agents	CTO / Security	Supply chain transparency achieved
9–10	Integrate AI identity risk into board reporting	CFO / CISO	Quarterly AI identity metrics live
10–11	Second quarterly agent access review	CISO	Continued privilege reduction confirmed
11–12	Insurance renewal with AI governance evidence	CFO / Risk Mgmt	AI identity controls accepted by insurer
12	Annual AI identity governance review	Governance Committee	Framework updated; next year planned

SECTION 08

METRICS AND BOARD REPORTING FRAMEWORK

Agentic AI and Identity Governance Scorecard

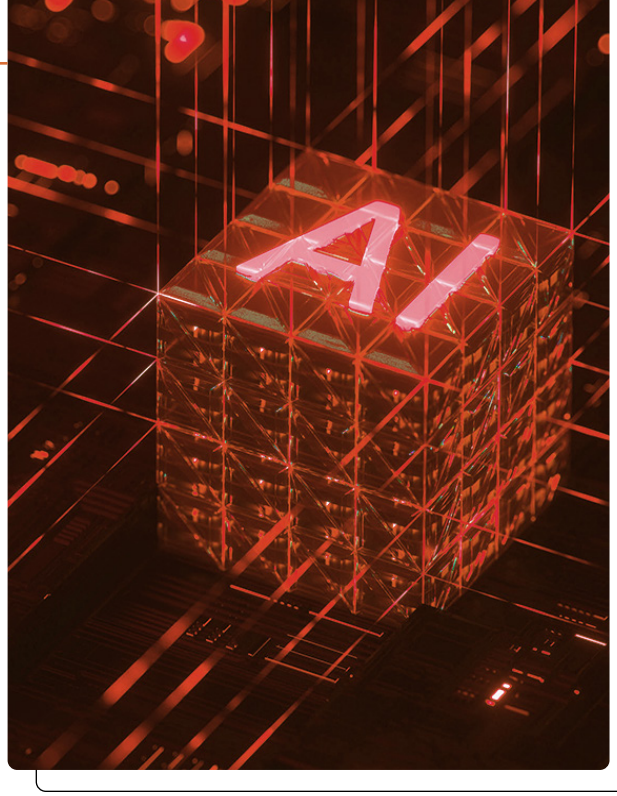
METRIC	TARGET	BENCHMARK	FREQUENCY
Total NHIs inventoried vs. estimated	100% coverage	~40% unknown[14]	Monthly
AI agents in registry vs. discovered	100% registered	29% prepared[6]	Monthly
NHIs with excessive privileges	Declining QoQ	97% over-privileged[3]	Monthly
Orphaned identities (no owner)	Zero	~50% over 1 year old[3]	Monthly
Agent actions with HITL gates	100% of high-risk	Varies	Monthly
Credential rotation compliance	100% on schedule	Varies	Monthly
Mean time to decommission retired agent	< 24 hours	Never (typical)[4]	Monthly
Agent behavioral anomalies detected	Tracked and resolved	No baseline (typical)	Monthly
Quarterly agent access review completion	100%	Annual human only (typical)	Quarterly
AIBOM coverage of deployed agents	100%	Varies	Quarterly

Agentic AI and Identity Governance Scorecard

INVESTMENT	ANNUAL COST	BUSINESS CASE
NHI discovery and inventory platform	$50K–$200K	Visibility into 144:1 identity sprawl[3]
AI Agent registry and classification system	$30K–$100K	Foundation for governance; regulatory prerequisite
JIT access and privilege automation	$50K–$150K	Remediates 97% over-privileged NHIs[3]
HITL workflow deployment	$20K–$80K	Prevents autonomous high-risk actions
Agentic observability and anomaly detection	$40K–$120K	Detects cascading failures before propagation[10]
Unified IAM for human + NHI + agent	$100K–$300K	Closes 70% identity-silo risk[5]
Total governance program	$290K–$950K	Avoids breach via 80%+ exposed NHI orgs[7]

SECTION 09

TEN TAKEAWAYS FOR THE C-SUITE



The keys to the kingdom are no longer held by people. They are scattered across thousands of lines of code and automated pipelines. It's time to take them back.

1 **EVERY AI AGENT IS AN IDENTITY—AND EVERY IDENTITY IS AN ATTACK SURFACE.** 79% of organizations are deploying AI agents, each requiring credentials and permissions. Your IAM system was not built for this. Treat agent identity governance as a board-level priority.[1,5]

2 **MACHINE IDENTITIES OUTNUMBER HUMANS 144 TO 1, AND 97% ARE OVER-PRIVILEGED.** The volume is staggering and the permissions are catastrophically excessive. A single compromised "super NHI" can give an attacker control of your entire cloud environment.[3]

3 **ONLY 29% OF ORGANIZATIONS ARE PREPARED TO SECURE AGENTIC AI.** The gap between deployment speed and security readiness is the defining risk of 2026. If you are deploying agents without a governance framework, you are introducing unmanaged risk at machine speed.[6,9]

4 **AGENTIC AI IS THE TOP ATTACK VECTOR FOR 2026—PER THE PROFESSIONALS DEFENDING YOU.** 48% of cybersecurity professionals rank it number one, ahead of deepfakes and all other emerging threats. Your security team is worried—the board should be too.[2]

5 **LEGACY IAM CANNOT GOVERN WHAT IT WAS NEVER DESIGNED TO SEE.** Identity systems built for humans with managers who respond to access reviews cannot manage autonomous agents that authenticate thousands of times per minute. Unified identity governance across human, machine, and AI categories is essential.[4,5]

6 **HUMAN-IN-THE-LOOP IS NOT OPTIONAL—IT MAY BE A LEGAL REQUIREMENT.** The EU AI Act requires human oversight for high-risk AI applications. AI agents that transfer funds, modify data, or make decisions in regulated domains must have defined approval gates.1

7 **A SINGLE COMPROMISED AGENT CAN CONTAMINATE 87% OF DOWNSTREAM DECISIONS IN FOUR HOURS.** Multi-agent cascading failures propagate faster than human incident response can contain them. Agentic observability and inter-agent communication logging are prerequisites for detection.10

8 **THE NHI LIFECYCLE PROBLEM IS THE SILENT ACCELERATOR.** Machine identities are almost never offboarded. Half are over a year old with no review. Automated lifecycle governance—including mandatory decommissioning when projects end—is the most underinvested control.3,4

9 **INSURANCE UNDERWRITERS ARE PAYING ATTENTION TO PRIVILEGE CONTROLS.** 88% of organizations report increased insurer pressure on privilege governance. Demonstrable AI identity controls will increasingly be conditions of coverage.[5]

10 **DISCOVERY IS THE NON-NEGOTIABLE FIRST STEP.** You cannot govern identities you cannot see. Deploy NHI discovery, build the agent registry, map permissions, and quantify the financial exposure. Then bring it to the board with a number and a plan.

APPENDIX A: RAPID ASSESSMENT & DECISION TOOLS

QUARTERLY BOARD BRIEFING TEMPLATE: AI AGENTS AND IDENTITY

SECTION	CONTENT
AI Agent Landscape	Total agents deployed; new agents since last report; autonomy tier distribution
Identity Dashboard	NHI-to-human ratio; orphaned identities; privilege reduction progress
Risk Exposure ($)	Quantified exposure from over-privileged NHIs and unmanaged agents
Incident Summary	Agent-related incidents; behavioral anomalies detected; response effectiveness
HITL Compliance	Percentage of high-risk actions with human approval; exceptions documented
Regulatory Status	EU AI Act readiness; AIBOM coverage; audit findings
Investment & ROI	Governance program spend vs. risk reduction; next-quarter requests

AGENTIC AI AND IDENTITY BOARD READINESS ASSESSMENT

QUESTION FOR THE BOARD	STATUS		PRIORITY
Do we have a complete inventory of all AI agents in our environment?	☐ Yes	☐ No	Critical
Do we know the machine-to-human identity ratio in our organization?	☐ Yes	☐ No	Critical
Do AI agents have designated human owners?	☐ Yes	☐ No	Critical
Have we identified and remediated over-privileged NHIs?	☐ Yes	☐ No	Critical
Are HITL controls in place for high-risk agent actions?	☐ Yes	☐ No	Critical
Can we detect agent behavioral anomalies in real time?	☐ Yes	☐ No	Critical
Do we log agent-to-agent communications?	☐ Yes	☐ No	High
Are agent credentials automatically rotated?	☐ Yes	☐ No	High
Do we have automated offboarding for retired agents?	☐ Yes	☐ No	High
Is AI identity governance included in board reporting?	☐ Yes	☐ No	High
Do we maintain AIBOMs for deployed agents?	☐ Yes	☐ No	High
Have we tested incident response for an agent compromise?	☐ Yes	☐ No	High
Does our cyber insurance cover AI agent incidents?	☐ Yes	☐ No	Medium
Have we unified human, NHI, and agent identity governance?	☐ Yes	☐ No	Medium

AI AGENT COMPROMISE INCIDENT RESPONSE CHECKLIST

HOUR	ACTION	OWNER
☐ 0–1	Identify the compromised agent; isolate from all systems	CISO / IR Lead
☐ 0–1	Revoke all agent credentials, tokens, and API keys	Identity & Access
☐ 1–4	Trace agent actions: data accessed, tools invoked, outputs generated	IR Team / Forensics
☐ 1–4	Assess downstream agents: check for cascading contamination	SecOps
☐ 4–12	Determine attack vector: prompt injection, memory poisoning, supply chain	Forensics / AI Security
☐ 4–12	Notify executive leadership and legal counsel	CRO / General Counsel
☐ 12–24	Assess data exposure; determine regulatory notification obligations	Legal / Compliance
☐ 12–24	Quarantine any AI models or datasets suspected of poisoning	AI / ML Engineering
☐ 24–48	Validate integrity of all connected agents and downstream outputs	SecOps / AI Engineering
☐ 24–48	Submit regulatory notifications; file insurance claim	Legal / CFO
☐ 48–72	Conduct root cause analysis; update agent governance framework	CISO / Governance Committee
☐ 48–72	Brief the board; update agent registry and access policies	CRO / CISO

APPENDIX B: REFERENCES

1. IBM, "Agentic AI Security Guide," February 2026. 79% of organizations deploying AI agents. Per IBM, 88% of executives surveyed by PwC plan to grow agentic AI budgets. EU AI Act Article 14 requires HITL for high-risk AI. Agents as "new employees, not new technology."

2. Dark Reading / Kiteworks, "2026: The Year Agentic AI Becomes the Attack-Surface Poster Child" and "Agentic AI: Biggest Enterprise Security Threat for 2026," February 2026. Poll: 48% of cybersecurity professionals rank agentic AI as top 2026 attack vector. Omdia confirms AI adoption as #1 corporate security concern.

3. Entro Security, "NHI & Secrets Risk Report – H1 2025." NHI-to-human ratio: 144:1 (up 56% from 92:1 in H1 2024). 97% of NHIs have excessive privileges. 5.5% of AWS NHIs are full administrators. ~50% of NHIs over one year old.

4. CSO Online / ManageEngine, "Why Non-Human Identities Are Your Biggest Security Blind Spot in 2026," February 2026. ManageEngine 2026 Identity Security Outlook: ratios of 100:1 to 500:1. OWASP NHI Top 10 ranks improper offboarding as #1 risk. One Identity predicts first major breach from over-privileged AI agent.

5. CyberArk, "2025 Identity Security Landscape" (2,600 cybersecurity decision-makers). Machine identities outnumber humans 80:1+; nearly half have privileged access. 68% lack AI identity security controls. 47% cannot secure shadow AI. 70% say identity silos = root cause of cyber risk. 75% prioritize efficiency over security. 88% under insurer pressure on privilege controls.

6. Cisco / Help Net Security, "State of AI Security 2026" and "Enterprises Are Racing to Secure Agentic AI Deployments," February 2026. Only 29% prepared to secure agentic deployments. Documented MCP server exploit hijacking agent for data exfiltration. Agent-to-agent identity risks: impersonation, session smuggling. 250 poisoned documents sufficient to implant model backdoor.

7. Security Boulevard (Entro syndication) and CSO Online, various 2025–2026 NHI reports. Over 80% of organizations have experienced a cyber incident linked to compromised machine identities, corroborated by CyberArk's 2025 Identity Security Landscape finding that 87% of organizations experienced at least two successful identity-centric breaches in the prior 12 months. NHIs outnumber humans by factor of 20–144x depending on methodology. Critical need for NHI lifecycle governance.

8. Gartner, "Top Trends in Cybersecurity for 2026," February 2026. Agentic AI demands cybersecurity oversight. No-code/vibe coding expanding unmanaged AI agent proliferation. Global security spending $244.2B in 2026.

9. Cyber Security Tribe / Rippling, "Agentic AI Security: A Guide to Threats, Risks & Best Practices 2025." 59% say agentic AI security is "work in progress." Only 34% have AI-specific security controls. <40% regularly test AI agent security workflows.

10. Stellar Cyber, "Top Agentic AI Security Threats in Late 2026." OWASP Top 10 for Agentic Applications 2026. 520 tool misuse incidents. Galileo AI: single compromised agent poisoned 87% of downstream decisions in 4 hours. Memory poisoning and supply chain attacks carry disproportionate severity.

11. Veza, "2026 State of Identity & Access Report" (December 2025). 230 billion permissions measured. "Safe and compliant" classification dropped from 70% (2024) to 55% (2025). Ungoverned permissions rose from 5% to 28%. Just 0.01% of machine identities (approximately 2,188 accounts) control 80% of cloud resources.

12. USCSI Institute / various, "What is AI Agent Security Plan 2026? Threats and Strategies Explained." JIT access for agents. HITL for financial, data deletion, and access control actions. Zero-trust architecture for agent authentication. Manager-agent trust exploitation scenarios.

13. DasRoot.net, "Agentic AI and Security: A Deep Technical Analysis in 2026," February 2026. OWASP Top 10 for Agentic Applications 2026 (100+ expert contributors). MAESTRO v2 threat modeling framework. OWASP AIBOM Generator v1.2 for AI supply chain transparency. OWASP AIVSS v1 standard.

14. Permiso Security / DeepakGupta, "The AI Identity Crisis: NHIs, Agents & Vibe Coding in 2025," and "2026 State of Identity Security Report." 40% of organizations cannot identify NHI owners. Volume of agentic identities expected to exceed 45 billion. Ratios as high as 96:1 in financial services. 91% expect explosive AI identity growth in 2026.

15. JumpCloud, "Beyond the Binary: Why 2026 Demands a New Identity Category," January 2026. "Third Face of Identity": autonomous entities combining judgment with speed. 59% use unapproved AI; 75% share sensitive data. Treat every AI agent like a digital intern—access but never total autonomy.

16. CyberArk / Lavi Lazarovitz, "AI Agents and Identity Risks: How Security Will Shift in 2026," December 2025. Multi-agent environments expected to be norm by 2027. Agentic systems doubling in 3 years. Shai-Hulud NPM worm demonstrates developer trust exploitation. Post-authentication attacks bypassing traditional defenses.

17. Menlo Security, "Predictions for 2026: Why AI Agents Are the New Insider Threat," January 2026. No playbook for NHI explosion. CISOs need privilege matrix for orders of magnitude more roles. LLM-generated code introduces "slop code" vulnerabilities. Visibility and browser isolation as critical controls.

18. InstaTunnel / Multiple Sources, "Machine Identity Bankruptcy: The 82:1 Bot Identity Crisis," February 2026. "Identity Bankruptcy": threshold where identity sprawl is mathematically impossible to audit. tj-actions GitHub Action compromise (CVE-2025-30066, March 2025) exposed CI/CD secrets in workflow logs across an Action used in 23,000+ repositories, enabled by a stolen NHI personal access token.

BLIND SPOTS is a recurring C-suite threat briefing series. Vol. 1 covered Shadow AI. This volume addresses the converging threats of agentic AI and unmanaged machine identities. Governance frameworks should be adapted to the organization's specific industry, regulatory environment, and risk profile. The AI and identity landscape evolves rapidly; this document should be reviewed and updated quarterly.

ABOUT MALA TECHNOLOGY ADVISORS

TECHNOLOGY INNOVATION STARTS HERE

MALA Technology Advisors is a leading technology advisory firm helping businesses modernize, adapt, and grow through smart, strategic use of technology. MALA stands for Modernization, Agility, Leadership, and Advancement—the values that guide everything we do. As a vendor-neutral strategic partner, we work with forward-thinking organizations to align business and information technology goals. Using our 4-Pillar approach, we help clients reimagine, restructure, and renew business IT initiatives to create agile and resilient organizations.

Our Services

Managed Cybersecurity · Cloud Services · Network and SD-WAN · Managed IT · Communications as a Service · IoT and Wireless · Professional Services · Devices

How We Work

Our expert team of technology and business professionals works side-by-side with your organization across a proven engagement model: IT Strategy development, Environment Assessment, Vendor Identification and Selection, Proposed Solution design, and ongoing Account Management. We don't just consult—we partner to simplify complexity, accelerate progress, and unlock new opportunities for growth.

What Sets Us Apart

Our team is a diverse network of consultants and industry professionals with a global mindset and a collaborative culture. We focus on Workplace Technology Innovation, Flexibility in Integration and Implementation, Perpetual Enablement, and Human Connection. The speed of change in technology today is driving a matched evolution of the IT landscape—and more than ever, IT departments rely on MALA Technology Advisors to help them innovate and plan for the future.

GET IN TOUCH

500 Commercial Street, Suite 502, Manchester, NH 03101

(603) 802-2469 · info@malatechadvisors.com

malatechadvisors.com

MALA
TECHNOLOGY
ADVISORS