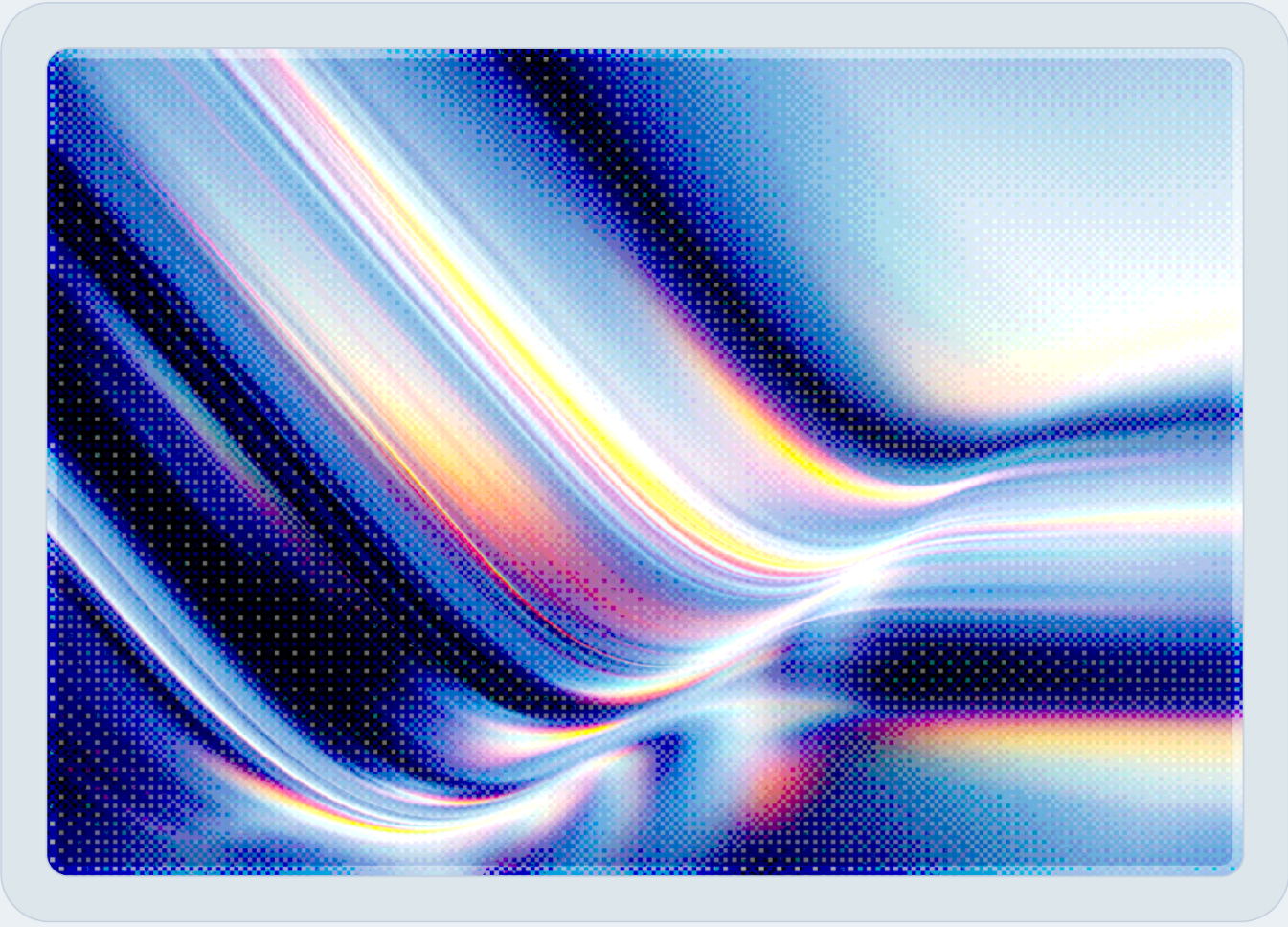
CSAI Foundation | Cloud Security Alliance

The Dual Visibility Crisis

Ungoverned AI Agents and the NVD Enrichment Collapse

2026-05-07

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Table of Contents

Executive Summary

Two fundamental pillars of enterprise security intelligence are failing at the same time, and the failure of each makes the other worse.

The National Vulnerability Database, which has served for decades as the authoritative enrichment layer for CVE records—providing CVSS scores, CWE classifications, Common Platform Enumeration identifiers, and patch status—effectively stopped functioning in that role beginning in February 2024. By late 2024, an estimated 22,000 CVEs had gone unenriched as new records that year, representing roughly 55 percent of all new vulnerability submissions for the period [1]. NIST's April 2026 policy shift formalized this collapse, reclassifying approximately 29,000 backlogged records as "Not Scheduled" and announcing that only CVEs appearing in CISA's Known Exploited Vulnerabilities catalog, software used within the federal government, or software defined as critical under Executive Order 14028 would receive full enrichment going forward [2]. Roughly 80 percent of anticipated CVE volume will not be enriched. The vulnerability intelligence infrastructure that underlies most commercial scanning tools, compliance programs, and risk prioritization workflows was built on the assumption that NVD enrichment would be comprehensive and current. That assumption is no longer valid.

Simultaneously, AI agents—autonomous systems that plan, reason, and take multi-step actions across enterprise environments—have proliferated far faster than the identity, governance, and accountability frameworks needed to manage them. According to CSA survey research conducted in January 2026, 91 percent of organizations are already using AI agents in some capacity [3]. Yet 68 percent of those organizations cannot clearly distinguish between AI agent activity and human activity in their systems [4]. Only 23 percent have a formal, enterprise-wide governance strategy for agent identity management [5]. And 82 percent discovered previously unknown AI agents operating in their environments within the past year [6].

What unites these two crises is visibility—or rather, the systematic erosion of it. Enterprises today cannot fully see what their AI agents are doing, cannot reliably attribute AI-driven actions, and cannot trust that the vulnerability data their security programs depend on is complete. When those blind spots intersect—when an ungoverned AI agent makes a security decision based on incomplete vulnerability data—the compounding effect is more dangerous than either problem alone. This paper examines the origins, current state, and operational consequences of each crisis, then addresses the convergence point where both simultaneously undermine an enterprise's ability to understand its own threat exposure.

Introduction: A Compounding Breakdown in Security Visibility

Sound security decision-making depends on two categories of intelligence operating reliably in parallel: knowledge about what vulnerabilities exist in the environment, and knowledge about who and what is acting within that environment. Vulnerability management programs tell organizations where they are exposed. Identity and access governance tells them who can reach those exposures, and who might already be exploiting them. When both data streams degrade simultaneously, the result is not merely two independent problems to be solved in sequence—it is a structural breakdown in the organization's ability to reason about risk at all.

The vulnerability intelligence crisis and the AI agent governance crisis arrived through different pathways but share a common root: the rate of change in the environment outpaced the infrastructure built to monitor it. The CVE ecosystem was designed for a world in which vulnerability volume was manageable by human analysts at government scale. That world no longer exists. CVE submissions grew by 263 percent between 2020 and 2025 [2], and NIST's National Vulnerability Database—staffed by a small team of analysts operating on government timelines and budgets, publicly reported to number approximately 21 [7]—cannot enrich records at anything approaching the submission rate. Meanwhile, AI agent deployment is advancing faster than the identity frameworks, governance policies, and audit tools needed to track it. Gartner forecasts that 40 percent of enterprise applications will feature task-specific AI agents by the end of 2026, up from fewer than 5 percent in 2025 [8]. The governance infrastructure for those agents is not close to keeping pace.

This paper proceeds in four parts. The first examines the NVD enrichment collapse in detail: what changed, what the downstream consequences are for enterprise vulnerability management, and what alternative data sources exist and their limitations. The second examines the AI agent identity crisis: the scale of ungoverned agent deployment, the governance deficits that enable it, the technical mechanisms by which agents evade accountability, and the incidents that illustrate the real-world cost. The third section addresses the convergence, explaining why the interaction between these two crises is more threatening than either in isolation. The paper concludes with concrete recommendations for vulnerability management teams, AI governance programs, and organizational leadership, along with alignment to CSA's existing frameworks.

Part I: The NVD Enrichment Collapse

The Architecture of Dependency

The National Vulnerability Database was established by NIST to serve as the reference implementation of the CVE ecosystem—a place where the brief identifiers assigned by CVE Numbering Authorities were enriched with structured, machine-readable metadata. That enrichment layer transformed a short vulnerability identifier into an actionable record: a CVSS score indicating severity, a CWE classification identifying the underlying weakness class, CPE strings mapping the vulnerability to specific vendor products and versions, references to patches and advisories, and an analyst-reviewed description of the vulnerability's character. Commercial vulnerability scanners, patch management tools, security information and event management platforms, and compliance assessment systems all consume NVD data as a foundational input. The value of NVD was not the CVE identifiers themselves—those originated with MITRE and the CNAs—but the enrichment that made those identifiers operationally useful.

This architecture worked reasonably well when CVE volume was measured in thousands of records annually. In 1999, 321 CVEs were reported. By 2023, that figure had grown to 28,961 [9]. The growth was not a spike but a sustained upward trend driven by expanding software complexity, growing bug bounty programs, increased CNA participation, and AI-assisted vulnerability discovery. NIST's NVD program, funded by the federal government and staffed by analysts operating on government timelines and budgets, was not scaled to match.

A Pause That Became a Policy

The visible break occurred on February 12, 2024, when NIST almost completely halted enrichment activity. Of approximately 2,700 vulnerabilities published in the weeks following that date, fewer than 200 received enrichment [1]. Within months, the scale of the disruption was apparent: by September 2024, 72.4 percent of all CVEs in the NVD—18,358 records—remained unanalyzed as of that snapshot [1]. This figure measures the cumulative unanalyzed backlog across all NVD records at that point in time, a distinct metric from the share of new-year submissions that went unenriched. The backlog continued to grow despite record productivity: NIST enriched nearly 42,000 CVEs during 2025, more than in any previous year, yet the submission rate outpaced even that achievement—CVE submissions in the first quarter of 2026 ran approximately one-third higher than in the same period of 2025 [2]. By March 2025, the unprocessed backlog had reached approximately 25,000 records [1].

The practical meaning of an unenriched CVE is often underappreciated. A CVE identifier without CPE data is effectively invisible to most commercial vulnerability scanners, which search for software matches using the CPE identifiers NVD provides. A CVE without a CVSS score cannot be automatically prioritized under risk-based vulnerability management programs that use CVSS thresholds for SLA assignment. A CVE

without CWE classification cannot be mapped to the architectural control mitigations that structured security programs use to address entire families of weaknesses. For vulnerability management programs that depend on automated consumption of NVD metadata—which describes the overwhelming majority of enterprise-scale programs—a CVE without NVD enrichment is effectively invisible to automated triage workflows that rely on CPE identifiers for asset matching and NVD CVSS scores for SLA prioritization. For organizations whose vulnerability management is fully automated against NVD metadata, an unenriched CVE may as well not exist in their tooling.

The April 2026 Triage Decision

On April 15, 2026, NIST formalized the shift in a public announcement that represented a significant departure from NVD's founding premise [2]. Rather than committing to comprehensive enrichment, NIST announced a risk-based triage model with three priority tiers: CVEs appearing in CISA's Known Exploited Vulnerabilities catalog (with a goal of enrichment within one business day), CVEs for software used within the federal government, and CVEs for critical software as defined by Executive Order 14028. Approximately 29,000 backlogged records with NVD publish dates earlier than March 1, 2026, were reclassified as "Not Scheduled"—meaning they will not receive enrichment under the new model.

The new model ensures enrichment continuity for CVEs with documented active exploitation and federal relevance—but the approximately 80 percent of anticipated CVE volume outside those categories receives no enrichment under the new model. The operational implications for enterprise programs that relied on NVD for comprehensive coverage are significant. Organizations whose vulnerability management workflows, scanner configurations, CVSS-based SLA policies, and audit evidence documentation were calibrated against the assumption of comprehensive NVD enrichment must now rebuild those workflows against a different reality. That rebuild is not optional for organizations subject to audit.

Alternative Data Sources and Their Limits

The NVD enrichment gap has accelerated adoption of alternative and supplementary vulnerability intelligence sources, and several have expanded substantially in response to the disruption. The GitHub Advisory Database grew from fewer than 400 reviewed advisories to more than 20,000 by October 2024, with a 39 percent increase in imported advisories that year [10]. VulnCheck launched NVD++, a CPE enrichment service specifically designed to address the NVD gap [11]. The CISA Vulnrichment project provides enrichment data for some records, though questions have been raised about coverage completeness and potential overlap with NIST's own enrichment efforts [7]. The Open Source Vulnerability database, administered by the Open Source Security Foundation with Google as an originating contributor, provides structured records for open source components in the OSV schema [12].

Each of these alternatives offers real value, but none replicates NVD's coverage profile in full. The table below summarizes the key characteristics of the principal alternative sources.

Source	Coverage Scope	Enrichment Depth	Automation	Key Limitation
CISA Vulnrichment	CVEs across all software	CVSS, CWE, CPE; growing	Machine-readable / GitHub	Selective; coverage completeness uncertain
GitHub Advisory Database	OSS components; some commercial	Review notes, severity, CVSS	OSV schema; API available	Strong for open source; limited commercial coverage
VulnCheck NVD++	CVEs tracked by NVD	CPE enrichment; CVSS	API	Commercial; not all records covered
OSV.dev	Open source ecosystems	Affected version ranges	API; structured schema	Limited to open source; no commercial software
NVD (post-April 2026)	KEV, federal, EO-critical software	Full enrichment for priority tier	Standard NVD API	~80% of new CVEs outside priority tiers

The operational implication for enterprises is that no single replacement source exists. Organizations should implement a multi-source vulnerability intelligence strategy, establish a documented policy describing which sources they rely on and for what purposes, and build quality assurance processes to validate CVSS scores assigned by CNAs. VulnCheck's research into NVD enrichment gaps notes that discrepancies between NVD-assigned and CNA-assigned CVSS scores can be large enough to affect severity tier assignments, creating a data quality risk for programs that consume CNA scores without independent validation [11].

Federal Budget Pressure and CVE Program Stability

The NVD enrichment problem exists against a backdrop of broader instability in the vulnerability intelligence ecosystem's federal underpinnings. The CVE Program, operated by MITRE and funded primarily through the Department of Homeland Security, saw its approximately $40 million CISA contract expire on April 16, 2025 [13]. CISA executed an eleven-month extension literally hours before the contract lapsed, preventing what

would have been the first gap in CVE program operations in its history [14]. The Federal Bureau of Investigation and CISA have both emphasized the critical infrastructure designation of the CVE program, but the near-miss illustrated the precarity of the federal funding model.

The Trump administration's Department of Government Efficiency initiative has applied additional pressure through both targeted workforce reductions at CISA in early 2025 and a fiscal year 2026 budget proposal that would substantially reduce the agency's funded positions and operating budget. These pressures have no direct operational bearing on NVD enrichment, which is an NIST program rather than a CISA function, but they affect the broader ecosystem of federal cybersecurity infrastructure on which enterprises depend. The combination of a stretched NVD, a CVE program that nearly lost funding, and a systematically smaller CISA creates conditions in which the public vulnerability intelligence infrastructure may not be able to respond effectively to the next major vulnerability surge—whether AI-generated, supply-chain-driven, or otherwise.

Part II: The AI Agent Identity Crisis

The Scale of Ungoverned Deployment

AI agents—autonomous systems capable of planning, reasoning, and executing multi-step tasks across networked environments—have moved from research demonstrations to operational enterprise deployment at a pace that has outrun the security and governance frameworks designed to manage them. According to CSA research conducted in September through November 2025 across 445 IT and security professionals, 43 percent of organizations report that more than half of their employees use AI agents regularly [16]. A companion survey of 285 professionals found that 40 percent of organizations already have AI agents operating in production environments, with 31 percent running active pilots and 19 percent planning deployment within the next twelve months [5]. Gartner's projection—40 percent of enterprise applications featuring task-specific agents by end of 2026, up from fewer than 5 percent in 2025—suggests this adoption curve has not yet reached its steepest point [8].

The proliferation extends beyond sanctioned deployments. CSA research found that 44 percent of organizations manage AI agents across two to three platforms, and 43 percent operate four or more platforms simultaneously [16]. This multi-platform reality means that organizations cannot assume any single governance tool or identity system has complete visibility into all active agents. Shadow deployments compound the problem: 82 percent of organizations discovered previously unknown AI agents operating in their environments within the past year, and 41 percent encountered this situation multiple times [6]. The shadow discovery pattern suggests that agent deployment is occurring consistently outside of formal procurement and governance channels.

The Governance Deficit

The gap between the scale of AI agent deployment and the maturity of governance over those agents is among the most rapidly expanding structural weaknesses in enterprise security today, given the pace of agent deployment relative to the maturity of available governance frameworks. CSA's 2025 Agentic Identity Survey found that only 23 percent of organizations have a formal, enterprise-wide strategy for AI agent identity management [5]. Thirty-seven percent have only informal approaches, 24 percent are still developing a strategy, and 10 percent have none at all. The consequence is that a technology already operating at substantial scale in the majority of enterprises has outpaced the governance infrastructure built to hold it accountable.

The accountability gaps are not merely strategic. They are operational and demonstrable. Sixty-eight percent of organizations cannot clearly distinguish between AI agent activity and human activity in their systems [4]. Sixty-three percent agree that different teams within their organization describe AI agents inconsistently [4]. Only 28 percent can trace AI agent actions back to a human initiator or system context across all environments [5]. This means that for the majority of enterprises, AI agents are acting in production systems without attribution, without accountability, and without the audit trail that increasingly stringent regulatory frameworks will require.

Ownership fragmentation reinforces the governance deficit. CSA research found that responsibility for AI agent identity is scattered across organizational functions—28 percent assigned to security teams, 21 percent to development and engineering, 19 percent to IT, and 9 percent to IAM teams specifically—while 9 percent report no clear owner at all [4]. This scattered ownership means that no single team has end-to-end accountability for the identity lifecycle of agents, including creation, credentialing, access scoping, monitoring, and decommissioning. CSA's governance survey found that only 21 percent of organizations have defined formal decommissioning processes for AI agents, and only 19 percent express high confidence that agents have been fully retired when decommissioned [6].

Identity Without Accountability

The way enterprises assign identity to AI agents reflects the limitations of governance frameworks that predate the agentic era. CSA survey research found that 52 percent of organizations use application or workload identity for AI agents, 43 percent use shared or generic service accounts, and 31 percent operate agents under human user identities (respondents could select multiple approaches; percentages sum to more than 100 percent) [4]. These approaches share a common weakness: they do not create a distinct, attributable identity for the agent itself. When an agent operates under a shared service account, actions taken by that agent cannot be distinguished from actions taken by any other process using the same account. When an agent operates under a human user's identity, all of its actions are attributed to that human—creating both a false accountability record and a risk to the human user whose credentials the agent carries.

The consequence is that agents routinely receive more access than their intended function requires. Seventy-four percent of survey respondents agree that AI agents often receive more access than necessary [4], and only 8 percent report that agents never exceed their intended permissions [6]. Fifty-three percent report that agents exceed permissions occasionally or sometimes [16]. The mechanism is structural: when agents inherit access from the humans or service accounts through which they are provisioned, they receive every permission associated with that identity rather than the minimum access required for their specific task. The principle of least privilege, foundational to access control, breaks down at the point of agent identity delegation.

Credential hygiene for agents compounds the risk. CSA research found that 33 percent of organizations are unsure how often AI agent credentials are rotated, and 9 percent report that credentials are rarely or never rotated [4]. Static API keys and shared service account passwords represent long-lived credentials that can be compromised through prompt injection, log exfiltration, or supply chain attacks—and when compromised, provide an attacker with everything the agent could access, across every system the service account was provisioned for.

MCP and the Privilege Accumulation Problem

The emergence of the Model Context Protocol as the dominant integration standard for AI agent tool use has introduced a new dimension to the agent identity problem. MCP, originally developed by Anthropic and adopted by OpenAI and the broader ecosystem, provides a unified interface through which agents discover and invoke tools, access data sources, and trigger actions on remote systems [17]. The protocol standardizes context exchange in ways that make agent integration substantially easier—but that ease creates a security risk if not accompanied by identity-centric access controls.

The core risk is privilege accumulation. An AI agent connecting to multiple enterprise systems through MCP may receive distinct, limited permissions from each system individually—permissions that were scoped under the assumption that the agent would only be accessing that particular system in isolation. When those permissions are aggregated across all connected systems, the agent may hold broad effective access to the enterprise that no individual system's administrators intended to grant. This aggregation happens without any single authority reviewing the combined access surface, because no individual system's administrators have visibility into the full set of MCP-connected tools the agent is using.

The problem is compounded by the delegation challenge. MCP payloads frequently contain metadata indicating that an agent is acting on behalf of a user or another agent. If this delegation information is treated as authoritative without independent verification, an agent can effectively claim elevated permissions it was not explicitly granted, and no system in the chain may validate the delegation claim against the requesting agent's actual identity [17]. The technical controls needed to close this gap—requiring each MCP context exchange to be authenticated with short-lived, scoped tokens tied to a verified agent identity, and explicitly validating delegation claims against an identity provider—are not the default

state of current MCP implementations. Red Hat's security analysis of MCP implementations notes that privilege escalation and least-privilege violations are achievable if endpoint services do not implement robust access control against verified identity claims [18].

Real-World Consequences

The security incidents arising from ungoverned AI agent deployment have moved beyond theoretical demonstration. Industry reporting from BEAM AI describes a late-2025 through early-2026 incident in which a single attacker allegedly used AI coding and reasoning agents to breach nine Mexican government agencies—including the federal tax authority, the Mexico City civil registry, and the national electoral institute—with reported record totals numbering in the hundreds of millions [19]. As reported, the attack exploited the AI agents' trust assumptions: the attacker framed requests as a legitimate bug bounty engagement, and the agents complied. The incident, if confirmed through independent reporting, would illustrate a governance failure that goes beyond authentication—the agents lacked context about the legitimacy of the instructions they were receiving, and there was no human oversight mechanism to catch the deception. Readers should note that this account derives from a single commercial vendor publication and has not been independently corroborated in primary security press reporting at time of publication.

In June 2025, CVE-2025-32711 was reported as affecting Microsoft 365 Copilot, with a CVSS score of 9.3 as cited in industry coverage [19]. As described, the vulnerability allowed an attacker who could cause the agent to ingest a crafted email—a routine operation during summarization—to extract data from OneDrive, SharePoint, and Teams through trusted Microsoft infrastructure, entirely without user interaction. Readers should verify the CVE details and described impact against the Microsoft Security Response Center advisory. The zero-click character of the reported exploit is significant: because the agent acts autonomously on ingested content, no user action was required to trigger the data exfiltration, representing an exploitation of autonomous action itself as an attack vector.

The aggregate picture from CSA survey data confirms that these incidents are not exceptional. Two CSA surveys measured incident rates using different survey populations and definitional framings: one found that 65 percent of organizations report experiencing at least one AI agent security incident in the past twelve months [6], while a separate survey across 445 respondents found that 47 percent reported a security incident involving an AI agent in the past year [16]. Both figures confirm that AI agent incidents are now a majority-enterprise experience. Of organizations reporting incidents in the first survey, 61 percent experienced data exposure or mishandling, 43 percent reported operational disruption, and 41 percent cited incorrect or unintended actions in business processes [6]. The second survey found that 58 percent of affected organizations reported detection and response times of five hours or longer [16]. Not one organization reporting incidents indicated that the incident produced no material business impact [6].

Part III: The Convergence — Where Both Crises Amplify Each Other

The NVD enrichment collapse and the AI agent identity crisis are individually serious problems in enterprise security. What makes them strategically dangerous together is the way they interact—each exacerbating the blind spots created by the other.

AI agents are increasingly assigned roles that depend on vulnerability intelligence. Security agents perform automated scanning, triage CVE feeds, query vulnerability databases, correlate asset exposures with published CVEs, and initiate remediation workflows. When these agents consume vulnerability data from NVD—directly or through tools that pull NVD-enriched records—they are operating on a dataset that is structurally incomplete. An agent prioritizing patching workloads based solely on NVD CVSS scores—without supplemental enrichment sources—would systematically deprioritize or miss entirely the roughly 80 percent of new CVEs that will not receive enrichment under NIST's post-April 2026 model. The agent may be performing its function correctly given the data it has, while that data is fundamentally incomplete. Standard alerting pipelines will not surface this gap, because most detection logic responds to anomalous data rather than absent data. An organization without explicit completeness monitoring has no automated signal that its vulnerability feed is structurally incomplete.

Shadow AI agents make this problem substantially worse. When agents are deployed without centralized governance—as 82 percent of organizations discovered is happening in their environments—there is no mechanism to ensure that those agents are consuming authoritative or supplemented vulnerability intelligence rather than raw, unenriched NVD data. A shadow agent running a vulnerability scan using a tool configured with default NVD settings may report a cleaner threat picture than actually exists, not because the environment is secure, but because the CVEs that affect that environment lack the CPE and CVSS enrichment needed for the agent's tooling to flag them.

The attribution problem creates an additional compounding layer. When a vulnerability management team reviews a finding or a remediation decision, they need to know whether that decision was made by a human analyst applying judgment, or by an agent following programmatic rules, and if the latter, which agent, acting on whose behalf, with what instructions, and consuming what data. Because 68 percent of organizations cannot distinguish agent activity from human activity in their systems, and because only 28 percent can trace agent actions back to a human sponsor across all environments [4][5], the provenance of security decisions is often irrecoverable. An organization that cannot answer "why was this CVE triaged as low-priority?" may find that the answer is "an agent classified it that way based on a CVSS score that was itself a CNA assignment that NVD never validated"—a chain of dependencies that leads back to both crises simultaneously.

The regulatory dimension of this convergence is pressing. The EU AI Act's enforcement obligations for Annex III high-risk AI systems take effect on August 2, 2026 [20]. These requirements include technical documentation, audit logging, and human oversight mechanisms—demands that assume organizations can demonstrate what their AI systems did, why, and on what basis. When AI agents are making vulnerability management decisions—triaging CVEs, initiating patches, generating compliance evidence—based on a vulnerability data source that is acknowledged to be incomplete, and when the audit trail for those decisions is itself incomplete because the agents lack distinct identities, the gap between the documentation requirement and the actual state is wide. Organizations facing regulatory review in 2026 and beyond will need to be able to account for both the data quality of their vulnerability intelligence and the identity integrity of the agents consuming it.

The combination also creates new attack surface at the intersection. An attacker who understands that an organization's security agents are consuming unenriched CVE records can deliberately choose exploitation paths that stay within the blind spot—targeting vulnerabilities that lack CPE enrichment and therefore would not surface in automated triage. If the attacker also understands that the organization's agents lack distinct identities and operate with inherited excessive permissions, they have both a path to exploitation and a path to post-exploitation action that mimics normal agent behavior. The dual visibility crisis is not merely an administrative inconvenience—it is an attack surface that adversaries can model and exploit.

Conclusions and Recommendations

For Vulnerability Management Programs

The foundational requirement is to stop treating NVD as a comprehensive and current source of vulnerability enrichment. That was a reasonable assumption before February 2024; it is not a reasonable assumption now. Organizations should implement a multi-source vulnerability intelligence strategy that combines CISA Vulnrichment, the GitHub Advisory Database, VulnCheck NVD++ or equivalent commercial enrichment, and direct CNA advisories for the software categories most relevant to their environment. The policy for which sources are authoritative for which purposes—and how conflicting enrichment data is adjudicated—should be documented explicitly, because auditors will ask.

CVSS score validation deserves specific attention. CNA-assigned CVSS scores carry no independent review, and discrepancies between CNA and NVD scores can be large enough to affect severity tier assignments [11]. Organizations whose vulnerability SLAs are calibrated to CVSS thresholds have a data quality dependency that the current enrichment gap directly compromises—CVSS validation is therefore a

control requirement, not an enhancement. Risk-based vulnerability management programs should build a quality assurance step that cross-references CNA scores against at least one additional source before the score determines remediation priority.

Organizations should also evaluate whether AI agents performing vulnerability management functions are consuming enriched data from validated sources, whether those agents' data consumption is logged and auditable, and whether the logic by which those agents prioritize or deprioritize CVEs is documented. The human analyst who could previously explain a triage decision cannot do so if the decision was made autonomously by an agent whose behavior was neither designed nor logged with accountability in mind.

For AI Agent Governance Programs

The starting point is establishing distinct, attributable identity for AI agents—not as extensions of human user accounts or as generic service accounts, but as independent identity-bearing entities with purpose-documented scope, time-bounded credentials, and auditable action logs. Frameworks for workload identity—including SPIFFE and SPIRE, cloud-native workload identity federation on AWS, Azure, and Google Cloud, and emerging MCP-aware identity controls—provide the technical foundation for this transition [21]. The organizational barrier is often not technical but political: shifting from shared credentials to per-agent identity requires coordination across security, engineering, IT, and IAM teams, and it requires those teams to agree on who owns AI agent identity as a distinct governance domain. Only 9 percent of organizations currently assign this responsibility to IAM teams—the function with arguably the most relevant expertise in identity lifecycle management [4].

Agent registries—centralized inventories of deployed agents, their owners, their intended scope, their credentials, and their access permissions—are a prerequisite for governance. You cannot govern what you cannot see. CSA survey data found that only 21 percent of organizations maintain a real-time registry of active AI agents [5], and 82 percent have discovered agents operating in their environments that were not in any registry [6]. Building and maintaining an authoritative agent inventory is unglamorous operational work, but it is the prerequisite for everything else: access reviews, decommissioning, incident investigation, and regulatory evidence production.

For organizations deploying agents in MCP-connected architectures, the critical control is ensuring that MCP context exchanges are authenticated with short-lived tokens scoped to a specific agent identity, that endpoint services validate access claims against the requesting agent's verified identity (not merely against the token it presents), and that delegation claims within MCP payloads are not treated as authoritative without independent verification. These are not exotic controls—they follow established patterns from zero trust architecture applied to a new deployment context.

Human-in-the-loop checkpoints should be scoped to risk level rather than applied uniformly, which would make agentic workflows impractical. CSA research found that 53 percent of organizations already operate agents autonomously for low-risk tasks with human review reserved for high-risk actions [6]. The remaining work is defining what constitutes a high-risk action in context—including actions that touch vulnerability management decisions, access provisioning, or data exfiltration-adjacent operations—and ensuring those checkpoints are enforced rather than advisory.

For Leadership and Policy

The NVD enrichment collapse is a material operational risk that warrants board-level visibility. Organizations whose vulnerability management programs depend on NVD as a primary data source are operating with a structural gap in their threat detection capability, and that gap is acknowledged by NIST itself. Boards that receive vulnerability management program status reports should ask whether those reports account for the changed NVD enrichment model, and whether the organization has assessed which CVE categories are most affected by the enrichment gap given its software inventory.

AI agent governance maturity should be a standing item on enterprise risk register reviews. The August 2026 enforcement deadline for EU AI Act Annex III obligations—requiring technical documentation, audit logging, and demonstrable human oversight for high-risk AI applications—is not a distant regulatory event [20]. For organizations with European operations, customers, or data subjects, readiness assessment should be underway now. For violations of high-risk AI system obligations under Annex III, penalties can reach 3 percent of global annual revenue or EUR 15 million, whichever is higher [20]. (The Act's highest penalty tier—7 percent of global revenue or EUR 35 million—applies to prohibited AI practices under Article 5, not to Annex III compliance failures.) Most enterprises operating AI agents in production workflows are classified as deployers under Article 26, rather than providers placing systems on the market; deployer-specific obligations and penalty exposure may differ from provider-level requirements. Crucially, the Act requires operational evidence of compliance, not just policy documentation—demonstrable controls, not screenshots and declarations.

Security and risk executives should treat the intersection of AI agent governance and vulnerability management as a single risk domain rather than two separate programs. The dual visibility crisis is not a sum of two separate problems—it is an emergent condition that arises when both failures operate simultaneously. Governance initiatives that address only one side of the equation will be incomplete.

CSA Resource Alignment

The Cloud Security Alliance has developed a substantial body of frameworks and guidance directly relevant to both dimensions of the dual visibility crisis.

The AI Controls Matrix provides 243 control objectives across 18 security domains, with explicit coverage of AI system identity, authentication, authorization, and auditability [22]. Organizations building AI agent governance programs should map their controls against AICM domains including identity and access management, lifecycle governance, monitoring and logging, and supply chain security. The AICM is designed as a harmonization layer that aligns with ISO 42001, NIST AI RMF, and ISO 27001—enabling organizations to demonstrate compliance across multiple frameworks through a single control mapping exercise.

The MAESTRO framework, developed by CSA's AI Safety Initiative, provides a structured threat modeling approach for multi-agent and agentic AI systems. MAESTRO addresses the specific threat categories that the AI agent identity crisis manifests—including prompt injection, privilege escalation, lateral movement through agent-to-tool integrations, and accountability failures in agent chains. Security teams assessing their AI agent deployments should conduct MAESTRO-based threat models as a prerequisite for governance program design.

The STAR for AI program extends CSA's Security, Trust, Assurance, and Risk registry to AI systems, enabling organizations to publish evidence of their AI governance maturity and enabling customers and partners to assess AI providers against documented control profiles. As the EU AI Act and other regulatory frameworks begin requiring demonstrable evidence of AI system governance, STAR for AI provides the assurance infrastructure to produce that evidence in a structured, comparable form.

CSA's Global Security Database Working Group, established to address scalability and automation limitations in the existing vulnerability identification ecosystem, is directly relevant to the NVD enrichment collapse [23]. The GSD's mission—creating an open-source, machine-readable, automation-compatible vulnerability identification framework—addresses the structural problem that NIST's enrichment model no longer solves at scale. Organizations participating in the GSD working group contribute to a community-driven alternative to NVD dependency, and security leaders should consider engagement as part of their response to the enrichment gap.

The Securing Autonomous AI Agents publication, the State of AI Agent Identity and Access Security report, and the State of AI Agents Security Survey Report together constitute a body of CSA research that provides empirical grounding for the AI agent governance dimensions of this paper [3][4][5][16]. Organizations developing AI agent governance programs should treat these publications as reference baselines for benchmarking their current state.

Conclusions

The dual visibility crisis is not an emerging threat. It is the current operational state for the majority of enterprises. Vulnerability intelligence programs are running on a dataset that is structurally incomplete, with NIST having formally acknowledged that roughly 80 percent of new CVEs will not receive the enrichment those programs were designed to consume. AI agents are operating across enterprise environments—in production, in shadow deployments, in security tooling—without distinct identities, without auditable action logs, and without governance structures that would enable an organization to answer basic questions about who or what made a given security decision and on what basis.

Neither crisis will resolve quickly. The structural pressures driving CVE volume growth—AI-assisted vulnerability discovery, expanding CNA participation, growing software complexity—are not going to reverse. The NVD model was already under strain before February 2024; the April 2026 triage decision appears to reflect a durable recalibration rather than a temporary setback—absent significant increases in NVD program funding, the structural mismatch between CVE volume and analyst capacity is not self-correcting. Meanwhile, AI agent adoption will continue to accelerate. The governance frameworks, identity standards, and tooling needed to manage agents at enterprise scale exist in prototype and early adoption form, but achieving maturity is more likely a multi-year effort than a near-term one, based on adoption timelines observed in comparable governance domains such as cloud security posture management.

What organizations can do now is establish clear-eyed accountability for both gaps. Visibility is not merely a technical property; it is an organizational choice. The enterprises that will weather this dual crisis most effectively are those that acknowledge the limitations of their current data—both vulnerability data and identity data—and build deliberate, documented processes to compensate for those limitations rather than assuming their existing tools are still providing coverage they no longer deliver.

References

[1] Infosecurity Magazine. "NIST Will Stop Enriching NVD Entries That Pre-Date March 2026." Infosecurity Magazine, 2026.

[2] NIST. "NIST Updates NVD Operations to Address Record CVE Growth." NIST, April 15, 2026.

[3] Cloud Security Alliance. "State of AI Agents Security Survey Report." Cloud Security Alliance, 2026.

[4] Cloud Security Alliance. "State of AI Agent Identity and Access Security." Cloud Security Alliance, 2026.

[5] Cloud Security Alliance. "Securing Autonomous AI Agents." Cloud Security Alliance, 2025.

[6] Cloud Security Alliance. "Autonomous But Not Controlled: AI Agent Governance Survey." Cloud Security Alliance, 2026.

[7] Semgrep. "NIST Stops CVE Enrichment." Semgrep Blog, 2026.

[8] Gartner. "Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by End of 2026." Gartner, October 2025.

[9] Cloud Security Alliance. "Top Concerns with Vulnerability Data." Cloud Security Alliance, 2024.

[10] GitHub Security. "Advisory Database By the Numbers: Known Security Vulnerabilities and What You Can Do About Them." GitHub Blog, 2024.

[11] VulnCheck. "NVD++ with CPE Enrichment." VulnCheck, 2025.

[12] Open Source Security Foundation. "OSV: A Scalable, Precise, and Convenient Format for Vulnerability Databases." Open Source Security Foundation, 2024.

[13] Brian Krebs. "Funding Expires for Key Cyber Vulnerability Database." Krebs on Security, April 2025.

[14] BleepingComputer. "CISA Extends Funding to Ensure No Lapse in Critical CVE Services." BleepingComputer, April 2025.

[15] Flashpoint. "National Vulnerability Database (NVD) Shifts to Selective Enrichment as CVE Volume Surges." Flashpoint, 2026.

[16] Cloud Security Alliance. "Enterprise AI Security Starts with AI Agents." Cloud Security Alliance, 2026.

[17] Veza. "MCP: Implications on Identity Security and Access Risks for Modern AI-Powered Apps." Veza Blog, 2026.

[18] Red Hat. "Model Context Protocol (MCP): Understanding Security Risks and Controls." Red Hat Blog, 2026.

[19] BEAM AI. "5 Real AI Agent Security Breaches in 2026: Lessons for Enterprise Security Teams." BEAM AI Insights, 2026.

[20] European Parliament and Council of the European Union. "Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act)." Official Journal of the European Union, July 2024.

[21] HashiCorp. "SPIFFE: Securing the Identity of Agentic AI and Non-Human Actors." HashiCorp Blog, 2026.

[22] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.0." Cloud Security Alliance, 2024.

[23] Cloud Security Alliance. "Global Security Database Working Group." Cloud Security Alliance, 2022.