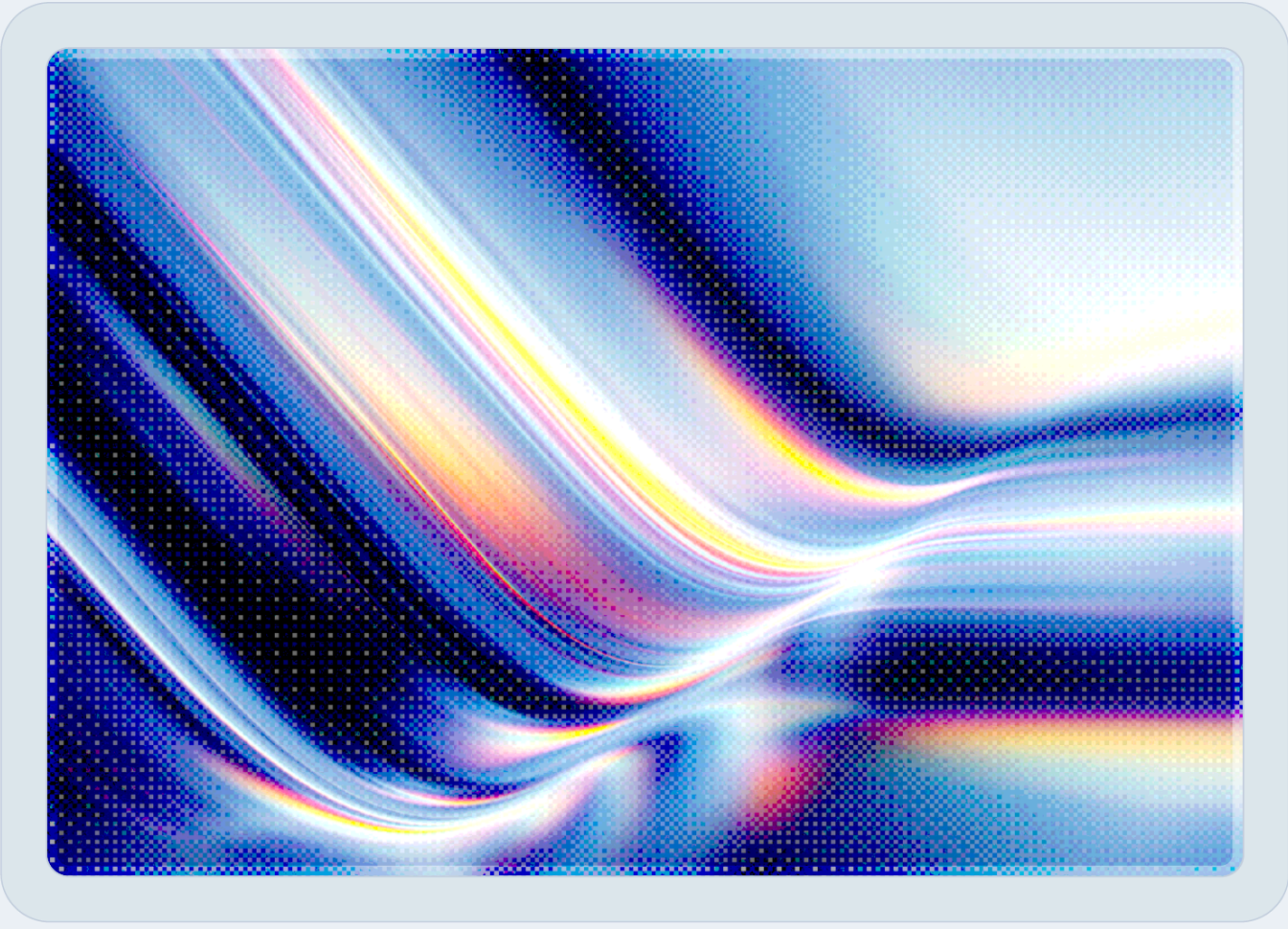
CSAI Foundation | Cloud Security Alliance

The Non-Human Identity Governance Vacuum

AI Agents and the Fastest-Growing Unmanaged Attack Surface

2026-05-20

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Table of Contents

AI Agents and the Fastest-Growing Unmanaged Attack Surface

Executive Summary

Enterprise security has spent decades hardening the human identity perimeter — enforcing multi-factor authentication, implementing privileged access management, and auditing user behavior through SIEM platforms. That investment, while substantial, addresses only the smaller half of the identity problem. In the modern enterprise, non-human identities — service accounts, API keys, OAuth tokens, machine certificates, and the credentials wielded by AI agents — now outnumber human users by an average of 45 to 1, and in cloud-native environments the ratio can reach 144 to 1 [1]. The governance frameworks applied to these identities rarely match the rigor applied to their human counterparts [2].

The emergence of agentic AI has sharpened this disparity into a crisis. AI agents are not passive credential holders; they are autonomous actors that acquire permissions dynamically at runtime, spawn sub-agents, invoke external APIs, write and execute code, and chain together actions that can span dozens of systems. Each of these behaviors expands the blast radius of any single compromised credential well beyond what a static service account could achieve. Yet a 2024 CSA survey found that only 15% of organizations feel highly confident in their ability to prevent NHI-based attacks [2], and a 2026 CSA analysis of token sprawl found that more than 16% of organizations do not track the creation of AI-related identities at all [3].

This whitepaper argues that non-human identity governance is the defining security gap of the agentic AI era. It examines the structural characteristics that make AI agent identities qualitatively different from prior NHI categories, surveys the attack landscape that exploits governance vacuums, identifies the compliance and regulatory blind spots that leave organizations exposed, and prescribes a layered governance framework built on the principles of identity lifecycle management, zero-standing privilege, and cryptographic workload attestation. Organizations that treat NHI governance as a secondary concern will find themselves structurally unprepared for the breach vectors that are already materializing in enterprise environments.

1. Introduction: The Identity Problem Doubles in Every Generation

Every major wave of enterprise technology has created a new category of non-human identity, and every generation has outpaced the governance frameworks developed for the one before it. The introduction of service accounts in Active Directory gave applications the ability to authenticate against domain resources, but created a population of long-lived, often over-permissioned accounts that rarely appeared in access reviews. The shift to cloud-native architectures and APIs generated millions of keys and tokens that existed outside the perimeter of traditional identity stores. Container orchestration platforms added workload identities to the mix. Each transition left a governance tail — identities that were created for a specific purpose, outlived their operational context, and accumulated permissions that no human owner could fully account for.

The agentic AI transition introduces qualitatively new identity governance challenges — autonomous credential acquisition, multi-agent orchestration, and runtime permission escalation — that prior generations of NHIs did not exhibit. AI agents are not simply new credential consumers; they are autonomous systems with the capacity to reason about their access needs, request new permissions, and act across systems in sequences that produce emergent behaviors their creators did not anticipate. The credential an agent holds is not merely a passive key but the principal identity of an actor that may, by virtue of its architecture, request access to systems its operators did not anticipate, chain actions across those systems, and operate outside the behavioral assumptions under which its credentials were issued.

By May 2026, the scale of this transformation has become concrete. Microsoft Copilot Studio users have collectively created more than one million AI agents, and Salesforce reported approximately $440 million in agentic AI revenue as of 2025 [4]. Gartner projects, as reported by industry analysts, that 33% of enterprise applications will incorporate agentic AI by 2028, up from less than 1% in 2024 [5]. The identity infrastructure needed to govern that population does not yet exist at most organizations, and the incidents documented in Section 4 demonstrate that adversaries have already begun operationalizing attacks against this gap.

2. What Non-Human Identities Are — and What AI Makes Them Become

2.1 The Taxonomy of Non-Human Identities

Non-human identities encompass any digital credential used by a system, workload, application, or automated process to authenticate and authorize its actions. The primary categories in contemporary enterprise environments include service accounts in directory services, API keys and bearer tokens used for REST and GraphQL integrations, OAuth 2.0 access and refresh tokens, machine TLS certificates issued to servers and containers, CI/CD pipeline identities, cloud provider IAM roles assumed by workloads, and — most recently — the credentials issued to or acquired by AI agents and multi-agent orchestration systems.

These categories differ in their lifecycle characteristics, the systems that issue and manage them, and the visibility organizations typically have over them. Service accounts are often managed (imperfectly) through directory services. API keys frequently exist outside any centralized store. OAuth tokens are scoped and theoretically short-lived but in practice persist far beyond their intended use. Machine certificates are governed by certificate management systems where they exist, but many organizations lack comprehensive PKI coverage for workload identities. And AI agent credentials occupy a new and largely ungoverned corner of the identity landscape where none of the legacy systems were designed to operate.

2.2 Scale and Growth Dynamics

The aggregate growth of non-human identities has been rapid even without the agentic AI acceleration. Entro Security research found that NHIs outnumber human identities at a ratio of 144 to 1 in cloud-native environments, up from 92 to 1 in the first half of 2024 — a 56% increase in ratio in a single year [1]. Across all enterprise environments the average ratio sits near 45 to 1, with the disparity driven by cloud adoption, microservices decomposition, and the proliferation of SaaS integrations that each require their own credential set. In one published case study, an audit of a Fortune 500 financial institution reportedly found over 4.2 million non-human identities against approximately 50,000 human user accounts [6].

Year-over-year, the NHI population across the industry grew 44% between 2024 and 2025, per Entro Labs H1 2025 research [7]. The GitGuardian State of Secrets Sprawl 2026 report found that 28.65 million hardcoded secrets were added to public GitHub repositories in 2025 alone — a 34% year-over-year increase and the largest single-year jump on record [8]. Within that figure, AI-related secrets — keys for AI APIs,

agent configuration tokens, and LLM service credentials — accounted for more than 1.27 million exposures, representing an 81% year-over-year increase and the fastest growth recorded in any single credential category [8].

2.3 How AI Agents Transform the Problem

The problem of static NHIs, while operationally challenging, is well-defined: enumerate, scope, rotate, and revoke. AI agents introduce a fundamentally different challenge because they are not static principals — they are dynamic actors.

An AI agent operating within an enterprise environment may begin a task with a defined set of permissions and then encounter a resource that requires additional access. Depending on its architecture, it may request that access from an orchestration layer, assume a different IAM role, acquire a new OAuth token, or use tool-calling capabilities to invoke a privileged API it was not expected to need at deployment time. This dynamic permission acquisition means that the blast radius of an AI agent's credential cannot be fully determined at the time the credential is issued. The attack surface is not fixed at deployment; it expands as the agent operates.

Agentic AI also introduces the pattern of multi-agent orchestration, in which a parent agent delegates sub-tasks to child agents, each of which requires its own identity and permissions. A single orchestration workflow can spawn dozens of ephemeral agents, each with distinct access requirements that change from invocation to invocation. The identity management problem that was difficult to solve for a population of static service accounts becomes an order of magnitude harder when the population is dynamic, the relationships between principals are hierarchical, and the lifecycle of each identity may be measured in seconds rather than years.

3. The Governance Gap: Where Controls Break Down

3.1 Ownership and Accountability

The most fundamental governance gap for non-human identities is the absence of clear ownership. A human identity is anchored to an individual whose employment status, role changes, and access needs can be tracked through HR systems. When a human employee departs, their identity can be deprovisioned through a defined offboarding workflow. Non-human identities lack this anchor. They are created by developers, automation systems, or AI agents themselves, and they frequently outlive the specific project, pipeline, or agent deployment that originally justified their existence.

A 2025 WEF analysis found that 51% of organizations report no clear ownership of AI identities [9]. Research published in The Hacker News in 2026 found that 8% of enterprise identities lack any HR system ownership linkage after the departure of their creator — orphaned identities that persist indefinitely with no human accountable for their continued existence [1]. Entro Security's H1 2025 research found that 47% of NHIs had gone unchanged for more than one year, suggesting that the majority of the NHI population is effectively static, unmonitored, and unmanaged [1].

3.2 Privilege Creep and Standing Access

Non-human identities consistently accumulate privileges beyond what their operational function requires. The principle of least privilege — granting only the minimum access necessary for a given task — is difficult to enforce for human users and substantially harder for automated systems that interact with many different resources under varying conditions. The result is systematic privilege creep, in which service accounts and agent credentials gradually acquire access that was never explicitly reviewed or approved.

The practical consequences are significant. Entro Security's H1 2025 research found that 1 in 20 NHIs carries full administrative privileges, even though a substantial fraction of the overall NHI population had not been active within the preceding nine months [10]. AI agents add a qualitatively new dimension to this problem because they can acquire privileges at runtime through tool use, OAuth flows, and role assumption — meaning that the standing privilege of an agent credential understates its effective access at any given moment.

3.3 Lifecycle Management Failures

The lifecycle of a non-human identity — its creation, scoping, rotation, and revocation — is the core operational discipline of NHI governance, and it is the dimension on which most organizations perform worst. The CSA State of Non-Human Identity Security survey found that only 20% of organizations have formal processes for offboarding and revoking API keys [2]. Fewer still have implemented automated rotation procedures. The CSA Token Sprawl analysis found that 24% of organizations take more than 24 hours to rotate or revoke exposed credentials after detection, while a separate 30% require more than a day to triage high-severity credential leaks — two distinct operational failures that each widen the exploitation window [3].

These delays are operationally significant because the window between credential exposure and credential exploitation is shrinking. The 2026 Verizon Data Breach Investigations Report documents that AI is accelerating attacker exploitation cycles from weeks to hours [11]. A credential exposed in a public repository at midnight may be tested against enterprise targets before morning. The rotation and revocation response times typical of current NHI governance programs are therefore misaligned with the threat environment they are designed to address.

AI agents add the specific problem of ephemeral permanence to this picture. Agent credentials are often issued for what appears to be a temporary purpose — a pilot project, a proof-of-concept deployment, or a one-time integration — but the credential persists in configuration files, environment variables, and secrets stores after the agent is decommissioned. The agent is gone; the credential is not. CSA's 2026 token sprawl analysis coined the term "persistent blast radius" to characterize this pattern: credentials that outlive their operational context and remain available for exploitation indefinitely [3].

3.4 Inventory and Discovery

An organization cannot govern identities it cannot enumerate. The discovery challenge for non-human identities is structural: they are created through code, configuration files, CI/CD pipelines, developer toolchains, and AI agent frameworks — channels that are designed for velocity, not for the governance metadata that identity management systems require. There is no equivalent of an HR onboarding form for an AI agent credential; the agent simply begins operating with whatever access it was provisioned.

The CSA token sprawl analysis found that more than 16% of organizations report they do not track the creation of new AI-related identities [3]. A 2026 NHI Reality Report reported that 78% of organizations have no documented policy for creating or removing AI identities [12]. Only 8% of respondents expressed high confidence that their legacy IAM systems can manage AI and NHI risks [3]. These figures reflect a consistent pattern: the governance infrastructure that exists was built for a world of human users and static service accounts, and it is not being extended to cover the AI agent population at meaningful scale.

4. Attack Surface Analysis: Exploiting the Governance Vacuum

4.1 The NHI Threat Landscape

The absence of governance creates exploitable conditions that adversaries are actively targeting. The 2026 Verizon Data Breach Investigations Report found that third-party and supply chain breaches now account for 48% of all breaches, up 60% year-over-year [11]. Non-human identities — particularly OAuth tokens used in SaaS integrations and credentials embedded in third-party code — represent a significant and growing vector within this category of attack, particularly through credential sprawl and insufficiently scoped integrations. Unlike phishing-based credential theft, which targets human users directly, NHI attacks exploit systemic governance failures: credentials that were never rotated, permissions that were never reviewed, and integrations that were never inventoried.

The following table summarizes the primary attack categories targeting non-human identities in agentic AI environments.

Attack Category	Description	Representative Incident
OAuth Token Supply Chain	Compromise of a third-party SaaS integration yields downstream access to customer environments via shared OAuth tokens	Salesloft-Drift (2025): attacker pivoted from a single OAuth token into more than 700 downstream customer environments [13]
API Key Exposure via LLM Tooling	AI coding assistants and LLM-integrated IDEs surface secrets embedded in code, committing them to repositories or passing them to external services	1.27 million AI-related secrets exposed in public GitHub in 2025 alone [8]
Agent Credential Harvesting	Compromise of an agent management platform yields credentials for all agents registered on the platform	Moltbook (2026): database misconfiguration exposed 1.5 million API tokens across 1.5 million connected agents [14]
Orphaned Credential	Credentials from decommissioned agent deployments remain valid and are	Commonplace in organizations without formal NHI offboarding; 47%

Attack Category	Description	Representative Incident
Reuse	discovered through scanning or code review	of NHIs unchanged for over one year [1]
Dynamic Permission Escalation	Exploiting an agent's tool-calling capabilities to acquire access beyond its original privilege scope	Documented in enterprise deployments; agents with broad tool access can achieve lateral movement without additional credential theft
Prompt Injection via NHI Context	Injecting adversarial instructions into data sources an agent reads, causing it to misuse its legitimate credentials	OpenClaw email exfiltration via prompt injection: 5 minutes from injection to data loss [14]

4.2 Supply Chain Amplification

The supply chain attack vector deserves particular emphasis because it exploits the combinatorial nature of NHI risk. In traditional IT environments, a compromised credential gives an attacker access to the resources that credential can reach. In modern SaaS and agentic architectures, a single OAuth token or API key may be the authentication mechanism for dozens or hundreds of downstream systems — because it was issued to an integration layer that itself brokers access to those systems.

The 2025 Salesloft-Drift incident, as documented by Obsidian Security, exemplifies this pattern [13]. Attackers who compromised OAuth tokens connecting Salesloft's sales engagement platform with the Drift conversational marketing platform discovered that those tokens granted access to Salesforce environments belonging to major enterprises. The blast radius of a single credential compromise was amplified across hundreds of customer environments because the integration architecture was built without adequate scope restrictions, credential compartmentalization, or monitoring for unusual downstream access patterns.

AI agents make this problem structurally worse because they are designed to integrate with multiple systems. An enterprise AI agent that automates sales workflows might hold credentials for CRM systems, email infrastructure, document management, calendar services, and external data providers simultaneously. Compromising one credential in this architecture may not yield all of these accesses directly, but it yields a foothold from which an agent's tool-calling capabilities can be exploited to traverse the remaining connections. The IBM 2025 Cost of a Data Breach report found that 97% of organizations that suffered AI-related security breaches lacked proper AI access controls [15] — a finding that suggests these combinatorial risks are not being modeled in most enterprise threat assessments.

4.3 The Shadow AI Amplifier

Shadow AI — the deployment of AI agents and integrations by employees without organizational approval or security review — substantially expands the undiscovered NHI surface. The 2026 Verizon DBIR found that employee use of unapproved AI tools tripled to affect 45% of the workforce [11]. Each unsanctioned AI agent deployment typically generates credentials that are stored in developer-controlled environments, exempt from corporate secrets management policies, never inventoried, and never rotated.

When shadow AI deployments are discovered — typically after an incident rather than through proactive monitoring — they reveal credential populations that the security team had no knowledge of. The IBM 2025 Cost of a Data Breach report found that shadow AI accounted for 20% of breaches and added approximately $670,000 to average breach costs, and that 63% of organizations that suffered AI-related breaches lacked AI governance policies entirely [15]. This creates a feedback loop: the governance vacuum drives shadow deployment, and shadow deployment deepens the governance vacuum by generating ungoverned credentials outside the organization's visibility. The absence of clear, accessible policies for sanctioned AI agent deployment pushes usage underground, where it cannot be inventoried or governed.

5. The Compliance Blind Spot

5.1 Regulatory Frameworks and the NHI Gap

The major regulatory and compliance frameworks that govern enterprise security — SOC 2, ISO 27001, PCI DSS, NIST 800-53, and the EU AI Act — each contain provisions relevant to identity and access management, but none was written with agentic AI credentials in mind. The result is a compliance gray zone in which organizations can achieve certification against frameworks while leaving their AI agent identity populations entirely unaddressed.

SOC 2 Type II audits examine access control effectiveness for the service being assessed, but the scope of "access" in most audits defaults to human user access. AI agent credentials that are never provisioned through the formal IAM system and never appear in user access reviews can pass outside the audit scope entirely. ISO 27001's Annex A.9 controls on access management similarly focus on user account management processes that were designed for human principals. PCI DSS 4.0 strengthens requirements around system component authentication, but its application to LLM-based agents and dynamic orchestration frameworks remains ambiguous in practice.

The compliance blind spot creates a risk that organizations treat NHI governance as a future requirement rather than a present obligation. When an auditor does not examine AI agent credentials, the security team has little external pressure to build the governance infrastructure needed to manage them. This dynamic is beginning to shift — multiple regulatory bodies are actively working to extend identity assurance requirements to non-human principals, and enterprise identity governance tooling has begun incorporating AI-specific controls — but the gap between current audit practice and the actual risk profile of enterprise NHI populations remains wide.

5.2 The Ownership and Accountability Vacuum

Governance frameworks consistently require that access to sensitive systems be tied to an accountable individual. The control typically looks like a matrix: every system, every dataset, and every privileged credential has a named owner who is responsible for approving access, reviewing permissions, and ensuring deprovisioning when access is no longer required. This model functions reasonably well for human user accounts because the HR system provides the accountability anchor.

For AI agent credentials, no equivalent anchor exists in most organizations. The agent was deployed by a developer who may have since moved to a different team or departed the organization. The project that justified the agent's existence may have concluded. The credentials may now be in use by an agent that was

later repurposed for a different workflow. The result is a population of credentials that are in active use — passing authentication checks, triggering transactions, writing to databases — with no one in the organization who knows who is responsible for them or whether the access they grant is still appropriate.

Research from The Hacker News's 2026 analysis found that 8% of enterprise NHIs lack HR system ownership linkage following creator departure [1]. A broader framing from multiple 2026 security reports suggests that the majority of NHI governance failures stem not from technical inability to manage credentials but from institutional inability to assign and maintain accountability for them. The governance vacuum is, at its root, an organizational problem as much as a technical one.

6. Technical Foundations for NHI Governance

6.1 Cryptographic Workload Identity: SPIFFE and SPIRE

Among the most capable open standards available for non-human identity governance, SPIFFE — the Secure Production Identity Framework for Everyone — and its reference implementation SPIRE (SPIFFE Runtime Environment) stand out for their ability to address the lifecycle and attestation problems at the core of the NHI challenge. SPIFFE provides a workload identity fabric that issues short-lived cryptographic certificates (SVIDs, or SPIFFE Verifiable Identity Documents) to compute workloads based on verifiable attestation of their execution environment, rather than static API keys or long-lived secrets [16].

For AI agent deployments, SPIFFE offers several properties that address the core governance failures described in this paper. Each agent receives a unique, cryptographic identity that is tied to verifiable properties of its deployment context — its container, its node, its code signature. The identity is short-lived and automatically renewed, eliminating the orphaned credential problem. Access control policies can be applied to the SPIFFE ID rather than to a shared API key, enabling fine-grained, workload-specific authorization. And the attestation mechanism creates an audit trail connecting every action taken by an agent to a verifiable claim about what that agent was running at the time.

SPIRE can scale to handle hundreds of thousands to billions of attestations per day [17], making it technically feasible to apply workload identity at the scale that enterprise AI deployments will require. The challenge is adoption: deploying SPIFFE/SPIRE requires changes to how agents are packaged, orchestrated, and granted access, and it requires integration with the authorization systems that evaluate the SPIFFE IDs agents present. This is non-trivial work, but it is the appropriate technical direction for organizations that want NHI governance to scale alongside agentic AI adoption.

6.2 Just-in-Time Access and Zero Standing Privilege

The principle of zero standing privilege — ensuring that no credential carries persistent elevated access, but instead requests time-limited, scope-limited permissions at the moment of need — is the architectural equivalent of the governance controls that human privileged access management systems provide for human administrators. Applied to AI agents, it means that an agent should not hold credentials that grant broad access continuously; it should request the specific access it needs for a specific task and have that access automatically expire when the task completes.

Just-in-time access for AI agents requires an authorization layer that can evaluate agent requests in real time, enforce scope and duration limits, and revoke access automatically. Attribute-Based Access Control (ABAC) and Policy-Based Access Control (PBAC) systems are architecturally better suited to this model than traditional Role-Based Access Control (RBAC), because they evaluate the agent's current task context, the data sensitivity of the resource being requested, and environmental conditions at the moment of the request — rather than applying a fixed role assignment regardless of context. RBAC can be extended through session policies to provide some contextual behavior, but its static mapping of roles to permissions is fundamentally mismatched with the dynamic, task-specific access patterns of AI agents. The trade-off is complexity: ABAC and PBAC policy management requires more sophisticated tooling than RBAC, and policy sprawl is a recognized operational risk that must be managed alongside the governance benefits [18].

The CSA Agentic AI Identity Management Approach characterizes this as ephemeral authentication — "short-lived, context-aware identities tailored to an agent's current task and operational scope" — and identifies it as a foundational principle for AI agent security that existing OAuth and SAML protocols are insufficient to provide without significant augmentation [18].

6.3 Centralized Secrets Management and Rotation Automation

For the substantial portion of NHI populations that cannot be immediately migrated to a cryptographic workload identity model, centralized secrets management combined with automated rotation provides a meaningful improvement over the status quo. Centralization means that every credential — regardless of where it is used — is issued from and tracked in a single system of record that provides discovery, ownership assignment, rotation scheduling, and revocation capability. The alternative — credentials scattered across environment variables, configuration files, developer laptops, and CI/CD platform secrets stores — makes audit, rotation, and revocation operationally infeasible at scale.

Automated rotation eliminates the reliance on manual processes for a lifecycle management task that is both critical and tedious. The CSA survey found that only 20% of organizations have formal API key offboarding processes [2]; automated rotation systems can enforce rotation on a schedule regardless of whether a human remembers to execute it. Short credential lifetimes — hours to days rather than months to years — dramatically reduce the exploitation window when credentials are exposed, even when exposure goes undetected for some time.

6.4 Behavioral Monitoring and Anomaly Detection

The identity lifecycle controls described above address the provisioning and deprovisioning of AI agent credentials. Behavioral monitoring addresses the period between those events — the ongoing operation of an agent that may have legitimate credentials but is being used in ways that deviate from expected behavior. Because AI agents can acquire permissions dynamically and operate across many systems

simultaneously, behavioral baseline is harder to establish for agents than for human users, but it is also more important: an agent that has been compromised or whose behavior has been modified by prompt injection will often exhibit detectable anomalies in the systems it accesses, the timing and volume of its requests, and the downstream actions it takes.

Continuous authorization — re-evaluating an agent's access rights at each request rather than at session initiation — is the runtime complement to JIT provisioning. It enables the authorization layer to detect mid-session deviations and revoke or constrain access without requiring the entire agent session to be terminated. Trust scoring systems that incorporate agent behavior history, combined with real-time monitoring for deviations from established baselines, provide the detection capability that pure credential management cannot offer [18].

7. Recommendations: A Governance Framework for Agentic AI Identities

7.1 Establish a Non-Human Identity Registry

Every organization that deploys AI agents should establish a centralized NHI registry — a system of record that captures, at minimum, the identity, the owning team, the business purpose, the systems accessed, the privilege scope, and the expiration or review date for every non-human credential in the environment. The registry need not be a single product; it can be implemented as a combination of secrets management systems, cloud provider identity stores, and a metadata layer that provides unified visibility across all sources.

The registry creates the accountability anchor that NHI governance requires. Once a credential exists in the registry with an assigned owner, that owner can be notified of upcoming expirations, alerted when the credential is observed outside its expected operational scope, and held accountable for reviewing whether the access it grants remains appropriate. Without this foundation, all other NHI governance practices operate in the dark.

7.2 Implement Agent Identity Lifecycle Management

AI agent identities should be subject to the same lifecycle management disciplines applied to human user accounts: creation requires documented justification and owner assignment, access should be reviewed on a defined cadence, and decommissioning should trigger automated revocation of all associated credentials. This lifecycle should be enforced through process and tooling, not left to developer discretion.

For organizations with mature DevSecOps practices, agent identity lifecycle management can be integrated into infrastructure-as-code and CI/CD pipelines — credentials are provisioned through code review processes that enforce ownership metadata, expiration dates, and least-privilege scoping, and decommissioned through the same processes that decommission the infrastructure they are associated with. The principle is that no credential should exist outside of a lifecycle management process, and no lifecycle management process should rely on human memory rather than automated enforcement.

7.3 Apply Zero Standing Privilege to AI Agents

Organizations should treat standing privileged access for AI agents as the exception rather than the rule. Where technically feasible, AI agents should be provisioned with JIT access that is scoped to a specific task, subject to a defined time limit, and automatically revoked upon completion or expiration. Where JIT access is not immediately feasible, standing access should be scoped as narrowly as possible and reviewed at short intervals — quarterly at minimum, monthly for high-privilege agents.

The technical implementations described in Section 6 — SPIFFE/SPIRE for cryptographic workload identity, ABAC/PBAC for dynamic authorization, and short-lived credentials with automated renewal — support the zero standing privilege model at different layers of the stack. Organizations should develop a roadmap toward this model even where full implementation is not immediately achievable, prioritizing agents with the broadest access and highest blast radius.

7.4 Enforce Credential Rotation and Rapid Revocation

The combination of short credential lifetimes and pre-authorized revocation automation is the most direct mitigation for the exploitation window created by exposed credentials. Organizations should set credential lifetimes for AI agent credentials at the shortest interval compatible with operational requirements — hours for ephemeral agents, days or weeks for persistent agents — and automate renewal so that rotation is not dependent on manual intervention. They should also pre-authorize revocation workflows that can execute on high-confidence exposure signals (a credential appearing in a public repository, a SIEM alert flagging anomalous use) without requiring human approval of each revocation decision.

The target is a time-to-revoke measured in minutes, not hours. The CSA token sprawl analysis found that the majority of organizations require a day or more to respond to credential exposure [3]; the threat environment documented in the 2026 DBIR requires response in a timeframe that is two to three orders of magnitude faster [11].

7.5 Extend Governance to Third-Party Agent Integrations

The supply chain attack vector documented in Section 4 cannot be addressed through governance of first-party credentials alone. Organizations should require that any third-party agent, integration, or AI tool that connects to enterprise systems operates under the same NHI governance standards as internally developed agents. This means scoped OAuth token issuance rather than broad API key grants, contractual requirements for security standards in vendor agreements, and continuous monitoring for anomalous access patterns by third-party integrations.

Third-party AI agent deployments should be inventoried as part of the NHI registry described above, with the additional metadata of the vendor identity, the contractual security obligations, and the review cadence for third-party access. The 60% year-over-year increase in supply chain breaches documented in the 2026 DBIR reflects the degree to which this vector is being actively exploited; extending NHI governance to the full integration perimeter is the structural control required to address it.

7.6 Address Shadow AI Through Policy, Not Just Detection

Shadow AI governance requires a dual approach: detection of unauthorized agent deployments through network monitoring, DLP, and identity system auditing, combined with a sanctioned AI program that makes approved agent deployment accessible enough that developers and business users have a reasonable alternative to shadow deployments. Organizations that only enforce prohibition without providing a viable sanctioned path will find that prohibition reduces visibility into shadow AI without reducing its prevalence.

The sanctioned AI program should include a self-service mechanism for requesting new agent deployments that automatically provisions credentials through the NHI registry, assigns an owner, and enforces minimum governance standards. The goal is to make the governed path the path of least resistance, so that the shadow AI population decreases through adoption of sanctioned alternatives rather than purely through detection and enforcement.

8. Conclusions

The governance vacuum surrounding non-human identities is not a new problem, but agentic AI has transformed it from a manageable debt into an active liability. AI agents are autonomous, dynamic, credential-bearing actors operating at a scale and velocity that legacy identity management systems were not designed to address. Their credentials represent the highest-leverage attack surface in the modern enterprise — a single compromised agent token may provide access to dozens of integrated systems, be amplified through multi-agent orchestration, and persist indefinitely after the agent that originally used it has been decommissioned.

The organizations best positioned to manage this risk are those that treat NHI governance as a first-class discipline rather than an afterthought to human identity programs. That means building registries, assigning owners, enforcing lifecycles, implementing zero standing privilege, and extending governance to third-party integrations — all of which require organizational commitment as much as technical investment. The market is responding: the NHI access management market is projected to grow from approximately $11.14 billion in 2025 to $27.33 billion by 2033 [19], reflecting a recognition across the industry that the current governance gap is unsustainable.

What the market response cannot substitute for is organizational urgency. The statistics in this paper — 47% of NHIs unchanged for over a year, 51% of organizations with no clear AI identity ownership, 16% not tracking new AI credential creation — describe a population of enterprises that are accruing governance debt at the same pace that they are adopting agentic AI. Given that exploitation of NHI governance failures is already documented in the 2026 DBIR and in incidents such as Moltbook and OpenClaw, the cost of continued delay is measurable, not hypothetical.

CSA Resource Alignment

This whitepaper connects directly to several active CSA programs and frameworks that provide implementation-ready guidance for the governance controls described above.

MAESTRO (Multi-Agent Environment, Security, Threat, Risk, & Outcome) is CSA's primary framework for agentic AI threat modeling [20]. Its seven-layer architecture — Foundation Models (L1), Data Operations (L2), Agent Frameworks (L3), Deployment and Infrastructure (L4), Evaluation and Observability (L5), Security and Compliance (L6), and Agent Ecosystem (L7) — maps directly to the NHI governance challenges analyzed in this paper. Identity governance is a cross-cutting concern that touches L3 (how agent frameworks issue and consume credentials), L4 (how deployment infrastructure manages workload

identities), L6 (how security controls enforce access policies), and L7 (how credentials are governed across multi-agent ecosystems). Organizations applying MAESTRO to their agentic AI threat models should treat NHI governance failures as explicit threat categories at each relevant layer.

CSA AI Controls Matrix (AICM), which extends the Cloud Controls Matrix to address AI-specific risk domains, provides the control framework most directly applicable to the NHI challenges in agentic AI environments. The AICM's treatment of AI system access management, agent authorization controls, and supply chain security for AI components is the appropriate primary reference for organizations designing NHI governance programs aligned with CSA standards. The AICM is the recommended starting point for NHI program design; the CCM described below provides the underlying baseline that the AICM extends.

CSA Cloud Controls Matrix (CCM) provides specific control objectives within the Identity and Access Management domain that establish baseline requirements for NHI governance. The IAM domain controls covering access provisioning, privileged access management, credential lifecycle management, and access reviews apply to non-human identities and should be explicitly scoped to include AI agent credentials in CCM-aligned assessments. Organizations seeking to demonstrate NHI governance maturity through a CCM lens should map their NHI program controls to the relevant IAM control IDs and document coverage explicitly.

CSA STAR (Security Trust Assurance and Risk) provides a mechanism for demonstrating NHI governance posture to enterprise customers and partners. As AI agent integrations proliferate — and as supply chain breaches via shared credentials become more common — buyers will increasingly need assurance that the agents they integrate with operate under governance standards they can evaluate. STAR-registered organizations should consider extending their STAR disclosures to address NHI governance practices, including agent identity lifecycle management, credential rotation policies, and third-party integration standards.

The CSA Token Sprawl in the AI Era blog [3] and the CSA Agentic AI Identity Management Approach [18] provide supplementary operational guidance on the specific control implementations described in this paper's recommendations section.

References

[1] Robert Kraczek (One Identity). "The Non-Human Identity Crisis: Why Your Machine Identities Are Your Biggest Governance Gap". *The Hacker News Expert Insights*, May 2026. [Data from Entro Security H1 2025 research.]

[2] Cloud Security Alliance. "The State of Non-Human Identity Security: Survey Report". *CSA*, 2024.

[3] Cloud Security Alliance. "Token Sprawl in the Age of AI". *CSA Blog*, February 2026.

[4] Oasis Security. "Oasis Security Launches Agentic Access Management, the First Identity Solution Built for AI Agents". *PR Newswire*, 2025.

[5] 1Password. "AI and the Rise of Credential Sprawl". *1Password Resources*, 2025. [Secondary citation for Gartner 2025 agentic AI adoption projections.]

[6] Protego Security. "Non-Human Identities (NHI): The Hidden Security Crisis Powering AI Agent Attacks in 2026". *Protego Security Blog*, 2026.

[7] CyberSecurity Tribe. "Research Reveals 44% Growth in NHIs from 2024 to 2025". *CyberSecurity Tribe*, 2025. [Underlying research from Entro Labs H1 2025 report.]

[8] CSO Online. "AI Coding Is Fueling a Secrets-Sprawl Crisis Few CISOs Are Containing". *CSO Online*, 2025. [Data from GitGuardian State of Secrets Sprawl 2026.]

[9] World Economic Forum. "Non-Human Identities: Agentic AI's New Frontier of Cybersecurity Risk". *WEF*, October 2025.

[10] Entro Security. "Non-Human Identity & Secrets Risk Report H1 2025". *Entro Security*, 2025.

[11] Verizon. "2026 Data Breach Investigations Report". *Verizon Business*, 2026.

[12] Cyber Strategy Institute. "2026 NHI Reality Report: 5 Critical Identity Risks". *Cyber Strategy Institute*, 2026.

[13] Obsidian Security. "What Are Non-Human Identities? The Complete Guide to NHI Security". *Obsidian Security*, 2025.

[14] SC Media. "2026 AI Reckoning: Agent Breaches, NHI Sprawl, Deepfakes". *SC Media*, 2026.

[15] IBM Security. "Cost of a Data Breach Report 2025". *IBM*, 2025.

[16] HashiCorp. "SPIFFE: Securing the Identity of Agentic AI and Non-Human Actors". *HashiCorp Blog*, 2025.

[17] Security Boulevard. "AI, SPIFFE, and the Rise of Non-Human Identity: Takeaways from Workload Identity Day 0". *Security Boulevard*, November 2025.

[18] Cloud Security Alliance. "Agentic AI Identity Management Approach". *CSA Blog*, March 2025.

[19] Grand View Research. "Non-Human Identity Access Management Market Report 2033". *Grand View Research*, 2024.

[20] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO". *CSA Blog*, February 2025.