WHITEPAPER · PRODUCT OVERVIEW

Zero-Agent Platform Overview

Zero Trust identity governance for autonomous AI agents and non-human identities — discover, govern, enforce, and remediate in one continuous loop.

Contents

01 Executive Summary

Zero-Agent AI is a Zero Trust identity governance platform purpose-built for autonomous AI agents and other non-human identities (NHIs). It discovers every agent and machine identity in an environment, assigns accountable ownership, enforces explainable Zero Trust policy on each one, and remediates risk automatically — escalating high-impact actions to human approval — while writing a tamper-evident audit trail for compliance.

Enterprises are deploying AI agents and machine identities faster than they can govern them. These identities hold credentials, take autonomous actions, and frequently have no owner, no cryptographic identity, and no oversight. Traditional IAM and IGA platforms were designed around human employees and were never built for the scale, velocity, or behavior of machine identities. The result is a fast-growing, largely invisible attack surface sitting inside organizations that believe their identity program is mature.

Zero-Agent AI closes that gap with a single governance loop — **Discover** → **Govern** → **Enforce** → **Remediate** — applied continuously to every agent, script, pipeline, and service account in the environment.

WHO THIS IS FOR

Security and IAM engineering leaders, compliance officers, and platform teams who need to bring AI agents and NHIs under the same governance discipline already applied to human identities — without slowing down AI adoption.

4	5	100%	3
AGENT CAPABILITIES	GOVERNANCE	INTEGRATED WITH	HUMAN IDENTITY
FRAMEWORKS MAPPED		COVERAGE	

02 The Problem: A Blind Spot Growing Faster Than Governance

AI agents and non-human identities already outnumber human identities in most modern environments, and the gap is widening every quarter as teams adopt agent frameworks, RPA, and automated pipelines.

Why existing identity programs miss this

- **Explosive, uncoordinated growth.** Agents are created by individual engineers and teams, not provisioned through a central identity process.
- **No accountable owner.** Most agents in the wild have no assigned human or team responsible for their lifecycle, permissions, or decommissioning.
- **Standing, long-lived credentials.** Agents commonly hold API keys, tokens, and secrets that are never rotated and often over-scoped.
- **Autonomous action with little oversight.** Agents take real actions — calling APIs, modifying infrastructure, moving data — with minimal policy enforcement at the point of action.
- **A structural blind spot in IAM/IGA tooling.** Platforms like Okta, Entra, SailPoint, and Saviynt were architected around human lifecycle events (hire, transfer, terminate) and role-based access reviews — not around ephemeral, code-spawned, machine-speed identities.

What this creates in practice

Orphaned agents with production access

Agents whose original owner has left the team or project, still holding live credentials.

Unrotated, over-privileged secrets

Long-lived API keys and tokens scoped far beyond what the agent's function requires.

No audit trail for AI actions

Actions taken by agents are frequently invisible to security and compliance teams until an incident forces a reconstruction.

Compliance gaps

SOC 2, ISO 27001, NIST AI RMF, and EU AI Act all now expect explainable, evidenced control over automated decision-making — a gap most organizations cannot currently close.

THE CORE ISSUE

You cannot govern what you cannot see, and you cannot hold accountable what has no owner. Machine identity governance has to start with discovery and ownership before policy or remediation can mean anything.

03 The Solution: One Platform, Four Capabilities

Zero-Agent AI applies a single, continuous governance loop to every agent and non-human identity it finds — regardless of framework, runtime, or origin.

1 · Discover

Continuously inventories AI agents and NHIs across source control (GitHub), container runtimes (Docker), local processes, and secret stores — detecting common agent frameworks (LangChain, LangGraph, MCP, CrewAI, RPA tooling, and custom agents) and normalizing everything into a single canonical inventory.

2 · Govern

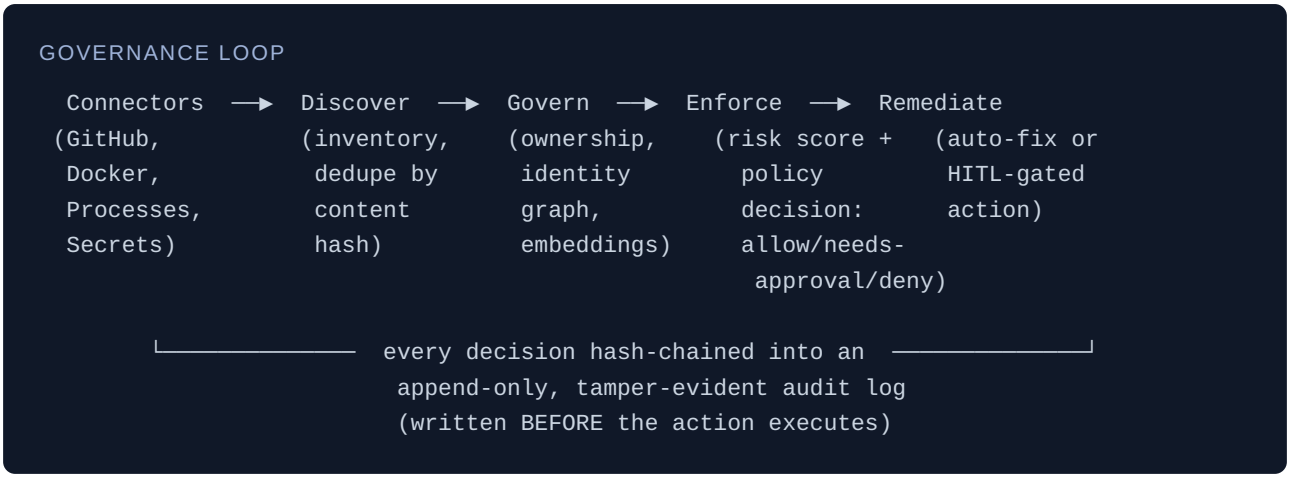
Infers ownership from converging signals (source control history, CODEOWNERS, container labels, secret-store paths, and behavioral similarity), builds an identity graph connecting owners to agents to secrets, and generates behavioral baselines for anomaly detection.

3 · Enforce

Scores risk with a transparent, explainable model and evaluates every authorization request through a policy-as-code engine (ABAC/RBAC/PBAC) to produce one of three decisions: allow, needs-approval, or deny — each with a human-readable rationale.

4 · Remediate

Automatically resolves low-blast-radius risk (rotating non-production secrets, issuing workload identities) and escalates high-blast-radius actions — production credential rotation, container termination, network isolation — to a tiered human approval queue.



04 How It Works: The Governance Pipeline

The four capabilities run as a single pipeline, triggered continuously or on demand, with the audit record written before any state-changing action executes.

Ownership inference model

Ownership is inferred from weighted signals rather than a single source of truth, because no one system reliably tracks who owns an agent end to end:

SIGNAL	WEIGHT	WHAT IT CAPTURES
Git history	35%	Commit authorship on the agent's source repository
CODEOWNERS	25%	Explicit repository ownership declarations
Docker labels	20%	Container metadata pointing to a team or service
Semantic similarity	10%	Behavioral embedding proximity to known-owned agents
Secret store path	10%	Vault/secret-manager path conventions

A composite confidence score above 0.85 auto-assigns an owner; scores between 0.60 and 0.84 are routed for human review; anything below 0.60 is flagged orphaned and escalated — orphaned status is surfaced, never silently dropped.

Risk scoring and enforcement

Every authorization request is evaluated against an explainable, additive risk model — deliberately not a black-box classifier — so that every decision can be reconstructed and defended to an auditor or regulator. Default posture is deny: an agent must present a valid workload identity, a confirmed owner, and an acceptable risk score to be allowed to act. Risk in the 0.70–0.90 band routes to human approval rather than an automatic allow or deny.

Human-in-the-Loop approval

Destructive or high-blast-radius actions — production credential rotation, stopping a container, revoking a secret, isolating an agent from the network — are gated behind a tiered approval chain (Security Engineer → SOC Lead → Platform Admin) rather than executed automatically, preserving both velocity and control.

05 Trust and Compliance: Evidence by Design

Governance is only useful if it can be proven after the fact. Zero-Agent AI treats the audit trail as a first-class output of every decision, not an afterthought.

- **Write-before-act.** The audit record for a decision is written before the corresponding action executes — including for actions later approved through HITL.
- **Tamper-evident.** Each audit entry is cryptographically chained (SHA-256) to the previous entry, so any modification to historical records is detectable end to end.
- **Append-only.** The audit store supports no update or delete path; it is architected as a write-ahead log for durability and integrity.
- **Framework-mapped.** Governance events are tagged to the control frameworks security and compliance teams already report against.

Frameworks referenced

FRAMEWORK	RELEVANCE
SOC 2 Type II	Access control, change management, and monitoring evidence
ISO/IEC 27001:2022	Information security controls for identity and access
NIST AI Risk Management Framework	Governance, mapping, and measurement of AI system risk
EU AI Act	Explainability and human-oversight requirements for automated decisions
PCI-DSS v4.0	Credential and secret handling for systems touching payment data

> **NOTE**
>
> Framework mapping indicates where Zero-Agent AI's evidence is relevant to these standards' control objectives. It is a tool to support a compliance program, not a certification or an assessment of any organization's overall compliance posture.

06 Validated in a Live End-to-End Run

The current build has been demonstrated running the full governance loop against a realistic mixed environment on a single machine, with no cloud dependencies.

In this run, a discovery scan identified nine agents across the configured connectors; ownership inference auto-assigned two with high confidence and correctly flagged the remainder for review or as orphaned; policy evaluation correctly denied a high-risk orphaned automation agent; a production credential rotation was routed to and approved through the HITL queue; and the audit chain was independently re-verified as intact.

9	2	7	1
AGENTS DISCOVERED	OWNERS AUTO-ASSIGNED	RISK FINDINGS	HITL APPROVAL PROCESSED

> **DEMO CONDITIONS**
>
> This validation ran entirely locally, with zero cloud cost and no external API keys — every integration point degraded to a deterministic local implementation. See Section 9 for what changes on the path to a production deployment.

07 What Makes Zero-Agent Different

Built for NHIs and AI agents, not retrofitted

The data model, ownership logic, and policy engine assume machine-speed, code-spawned identities from the ground up rather than adapting a human-identity system after the fact.

Explainable by design

An additive, transparent risk model replaces black-box scoring, so every decision has a traceable rationale — a requirement that is becoming non-negotiable under the EU AI Act and similar regimes.

Evidence-first

The write-before-act, hash-chained audit design treats compliance evidence as a core system property, not a downstream report generated after the fact.

Human-in-the-loop where it matters

Automation handles low-risk remediation; irreversible or high-blast-radius actions always route to a human, preserving both speed and accountability.

Open standards

Built on OPA for policy-as-code, SPIFFE/SPIRE for workload identity, and standard protocols (OAuth2/OIDC, SCIM, MCP) rather than proprietary formats — reducing integration lock-in.

08 Business Outcomes

OUTCOME	HOW ZERO-AGENT DELIVERS IT
Reduce risk	Eliminates orphaned agents and standing, over-privileged credentials by making them visible and enforcing default-deny.
Prove compliance	Produces continuous, audit-ready evidence mapped to the frameworks your compliance team already reports against.
Move faster on AI adoption	Gives security a way to govern AI agent rollout rather than blocking it outright.
Establish accountability	Every governed agent has both a human owner and a cryptographic workload identity.
Lower cost to evaluate	The current build runs locally with no cloud spend required to pilot the governance model.
Lower adoption risk	The architecture is designed so that moving from a local pilot to a cloud production deployment is a configuration change, not a rewrite (see Section 9).

09 Platform Maturity and Roadmap

In the interest of giving security and compliance readers an accurate picture, this section is explicit about what is running today versus what is on the roadmap.

Where the platform is today

The current release is a fully functional vertical slice of the governance model: a real backend, four working agent pipelines, a policy engine, an audit writer, and a dashboard, running end to end. To make that possible at zero cost and with no external dependencies, every external integration — the LLM, the vector embedding model, the secret store, workload identity issuance, the policy engine, the event bus, and notifications — runs behind an interface with a deterministic local implementation standing in for the production service. Real providers (a production LLM, HashiCorp Vault, SPIFFE/SPIRE, an OPA server, Redis, Slack) can be swapped in through configuration, without changing application code.

> **IMPORTANT**
>
> The local build's identity, secret, and cryptography providers are reference implementations, not production-hardened services. Authentication, cryptography, IAM integration, and remediation actions should undergo a full security review before any deployment that touches real credentials, production systems, or regulated data.

Roadmap

PHASE	FOCUS
Phase 1 — Foundation (current)	Four-capability governance loop, policy engine, HITL approvals, hash-chained audit, dashboard — running locally.
Phase 2 — Enterprise integration	Live connectors to SailPoint, Saviynt, Okta, and Entra; CyberArk PAM integration; production Vault, SPIRE, and OPA deployments; distributed task workers; cloud deployment.
Phase 3 — Scale	Zscaler ZIA/ZPA integration, multi-tenant SaaS delivery, advanced ML-based anomaly detection, partner marketplace, expanded analytics.

10 Next Steps

Zero-Agent AI is built to let security teams see and govern their AI agent footprint without slowing down the teams building on top of it.

- **See a live walkthrough.** Request a demo of the discover-govern-enforce-remediate loop against a representative environment.
- **Scope a Phase 2 pilot.** Identify a first enterprise connector (IGA, PAM, or Zero Trust network) to bring into a production evaluation.
- **Talk to your compliance team.** Map your current SOC 2 / ISO 27001 / NIST AI RMF / EU AI Act obligations against the evidence model in Section 5.

> **CONTACT**
>
> Visit zero-agent.ai to schedule a demo or request the companion Technical Architecture and Governance Best Practices whitepapers.