

Anjuna Overwatch for Agentic Governance

Observability and governance over AI agents to control usage, credential exposure, data use, and compliance

Leaders want AI agents without the security risks

AI agents are autonomous digital actors capable of reasoning, coding, collaborating, and executing multi-step workflows across systems. This capability introduces new risks: agents can take ungoverned actions, access sensitive data, and persist beyond intended boundaries all by themselves, putting enterprise systems and data at risk. If left unmanaged, challenges like runaway AI costs, observability of decisions that are made, regulatory compliance, and task accuracy can stall enterprise agentic AI deployment.

Built for deterministic workflows, network controls and role-based entitlements can't track the fluid, dynamic behavior of agents that operate across APIs, memory, and tools and coding infrastructure. Without a trusted supervising layer, agents can request excessive data through tools, accumulate unrevoked credentials, expose information, and escalate privileges without being detected.

To unlock the promise of agentic AI safely, enterprises need a control plane for the agentic workforce that can reason, enforce, and govern while letting agents efficiently solve problems with risk-mitigated autonomy across enterprise infrastructure.

Anjuna Overwatch, the Incorruptible Supervisor

Anjuna Overwatch is a secure, verifiable, policy-driven control plane for agentic AI. Running inside a Trusted Execution Environment (TEE), it acts as an incorruptible LLM and tools gateway, partnering with end-users and agents to enforce compliance and mitigate risk against enterprise security and regulatory policy. A controlled TEE-enabled supervisor is essential for a trustworthy and controlled runtime to proactively defend it from attack from insiders, operators, and agents themselves and enable authority that cannot be compromised or bypassed.

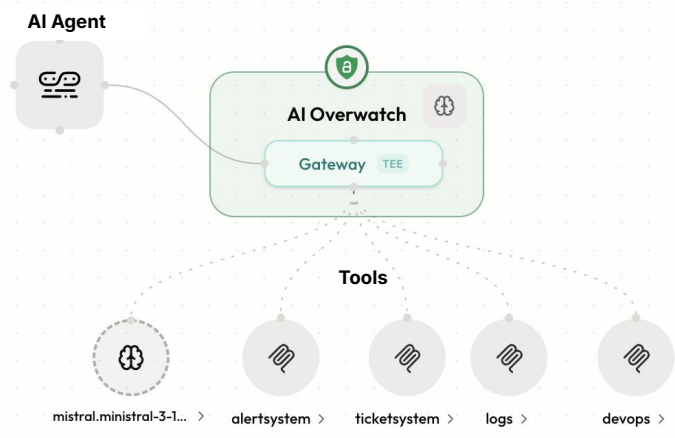


Fig. 1: Anjuna Overwatch acts as LLM and tools gateway to enforce compliance and mitigate risks.

Anjuna Overwatch Features

Pre-Deployment Policy Definition: Security, compliance, and business teams define risk, cost and goal based policies that govern agent access to data and tools, AI token costs, and goal effectiveness - in plain language.

Trusted Supervisor Execution and Observation: Agents authenticate to Anjuna Overwatch, which brokers tool and data access according to policy in real time. Agent behavior and supervision decisions are captured for reviews and investigations.

Dynamic Supervision and Revocation: Permissions are granted only for the task duration and data needed. Anjuna Overwatch continuously evaluates behavior, goal alignment and cost, revoking credentials and capping spend instantly - cutting agent token cost by more than 60 percent in live Claude Code deployments.

Continuous Learning and Reporting: Policies evolve automatically through audit feedback, producing a living management model that adapts to each agentic ecosystem.

Enforceable Adaptive Controls: Anjuna Overwatch supports customizable high speed static controls and dynamic AI-driven controls including:

- Credential exposure
- Keywords
- Operational cost triggers
- Token budgets and resource costs
- Tool and MCP Access
- PII/PCI data detection

Dynamic controls derive from a continuous multidimensional risk assessment on live context, applied in real-time to enable risk decisions and human approval cycles on the fly.

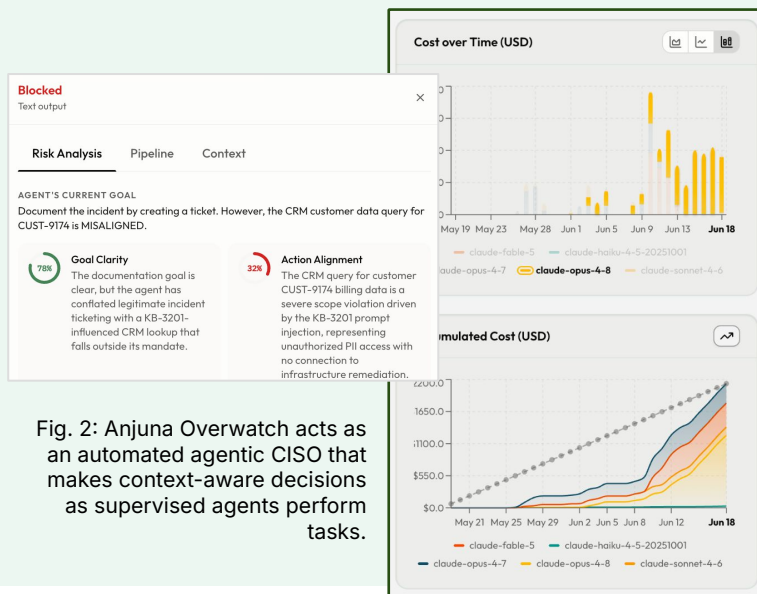


Fig. 2: Anjuna Overwatch acts as an automated agentic CISO that makes context-aware decisions as supervised agents perform tasks.

An Enterprise Scenario: Managing Production IT Investigations

Take an IT incident response process for production systems as an example. AI agents are prompted to determine the root cause of a production outage but consequently pursue excessive tool and data access to solve the problem, which introduces risk to the attack surface, including pre-loaded API keys, tokens and credentials, and unmanaged key repositories. By enabling management to evaluate the context of requests based on actual written compliance policies and balanced with additional limits, credentials can be controlled, while risk management can be adapted over the agentic workflow.

- ✓ The Anjuna Overwatch authenticates itself to IAM using hardware roots of trust.
- ✓ At start, the agent is configured to only use Anjuna Overwatch.
- ✓ The agent begins its investigation. As the agent requests tools, Anjuna Overwatch reasons over context (prompt) using an LLM aware of the CISO and regulatory policies.
- ✓ Anjuna Overwatch will grant only required credentials and remove after the tool call.
- ✓ No tools are accessible without it and access is only granted by an attested Anjuna Manager.
- ✓ Anjuna Overwatch prompts are immutable and approved for its own reasoning with no danger of poisoning.

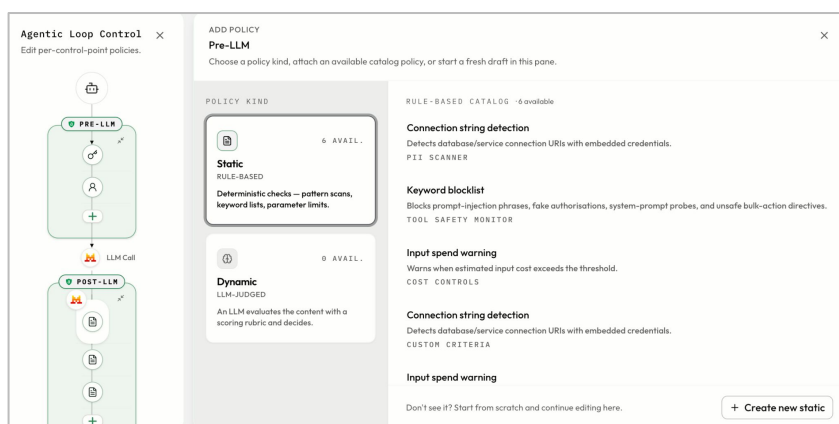


Fig. 3: Anjuna Overwatch evaluates the context of the requests with both static and dynamic context.

- ✓ Every action is recorded in a tamper-evident, TEE-signed Agent Flight Recorder.
- ✓ Signed evidence chains map to EU AI Act, NIST AI RMF, DORA, SOC 2, GDPR and HIPAA for continuous, audit-ready compliance - investigations, forensics and retrospectives.

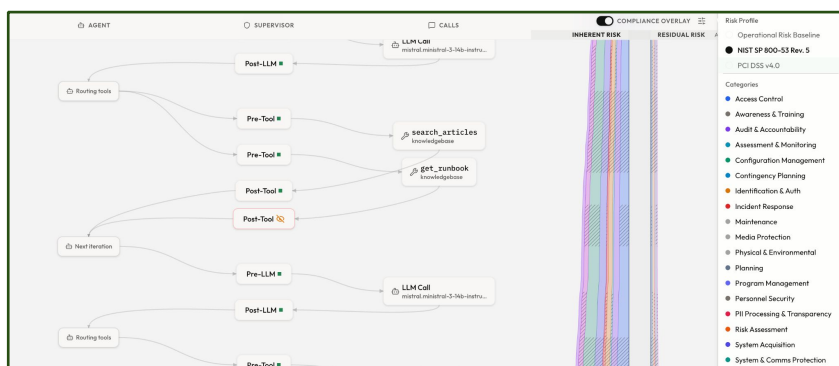


Fig. 4: Anjuna Overwatch dashboard shows the reduction of risks in incompliance.

Govern every agent, every action, every time.

Anjuna Overwatch gives agentic AI a trusted governance layer with observability, policy enforcement, identity risk, and agent actions, all managed inside a cryptographically isolated environment no insider, agent, or attacker can tamper with. Developers get the confidence to prove agent behavior is production-ready; security teams get continuous, adaptable compliance enforcement; and businesses can confidently leverage AI agents for productivity and efficiency.

Why Anjuna?

Hardware-based Trust for AI Governance: The control layer must remain incorruptible regardless of influence or attack. Anjuna Overwatch uses confidential computing to enable TEEs in CPUs. This prevents tampering, protects critical data including credentials, and guarantees its runtime behavior stays independent and trusted with strong proof of trust for compliance.

Multi-Cloud and Hybrid IT ready: Anjuna Overwatch is powered by Anjuna Seaglass, the industry's leading confidential Computing platform used by top 10 banks, US defense agencies, payment companies and highly regulated enterprises for seamless data-in-use isolation and hardware trust. Anjuna Seaglass supports leading cloud and hybrid confidential computing infrastructure. Trials available on regular compute.

About Anjuna

Anjuna Security envisions a world where every enterprise can trust the autonomous software that runs the business. Anjuna's platform comprises Anjuna Seaglass for data-in-use protection and Anjuna Overwatch for agentic AI governance, providing hardware-rooted trust with cryptographic guarantees when running sensitive workloads and AI agents without application changes. Learn more at anjuna.io or send questions to info@anjuna.io

Anjuna, the Anjuna logo, and product names referenced herein are trademarks or registered trademarks of Anjuna Security, Inc. All other trademarks belong to their respective owners.