AI RUNTIME ASSURANCE GLOBAL ENTERPRISE · MULTI-CLOUD

Deep error analysis and remediation for AI agents

By operationalizing runtime feedback and structural governance through a Guardian Agent control plane, we empowered enterprise security teams with instant, deep error analysis for their autonomous AI systems. The solution eliminated manual log-diving, automatically categorized invisible AI failures, and enabled secure, consistent, and policy-aligned agent operations at scale.

01 CLIENT & NEED

The client is a global enterprise running a multi-cloud environment with diverse autonomous AI systems — **Decisioning Agents**, **Interaction Agents**, and **Workflow Agents**. As those tools assumed autonomous operational roles, the primary challenge shifted from model quality to **runtime control**. The organization needed a structural shift in its security architecture to govern a sprawling AI estate and ensure agents behaved reliably and safely over time.

- **Invisible failures.** Agents weren't crashing — they were drifting. Decisioning Agents returned approve, deny, and rank for the same input without raising a single code error.
- **Schema & policy violations.** Non-deterministic anomalies: malformed JSON (`STRUCT-01`), invalid entity tags (`TAG-01`), and geography-skewed logic (`GEO-02`).
- **Forensic overhead.** Engineering and SRE teams burned cycles on manual log-diving to investigate agent failures and root causes.
- **Siloed observability.** Every agent drifted independently — compliance blind spots, inconsistent policy enforcement, no shared learning across the AI estate.

[T-0]

BEFORE IMPLEMENTATION

Autonomous agents silently deviated from expected logic, often returning different decisions for identical inputs. Security and engineering teams relied on manual, time-consuming log-diving to investigate behavioral anomalies. Every agent drifted independently — producing fragmented governance, compliance blind spots, and a reactive AI safety posture that made scaling risky and expensive.

[T+200 days]

AFTER IMPLEMENTATION

A unified Guardian Agent control plane sits in front of every model. It monitors agent interactions and contextual metadata in real time, blocking semantic drift and policy violations before they reach the customer. Anomalies are auto-structured into canonical failure modes for pattern recognition, feeding a continuous loop that improves model alignment and keeps every agent consistent at scale.

02 RESULTS

Small difference. Large impact.

91%
Risk mitigated
High-impact behavioral anomalies and policy violations blocked before reaching the customer.

74%
Faster investigations
Average incident investigation time cut; TTR down by **90%** end to end.

60%
Audit prep saved
Through automated behavioral history and real-time guardrail verification.

2.3×
Model-improvement velocity
Pipeline accelerated, with human-in-the-loop involvement reduced **3.2×**.