Agent Resilience in Production

Agents In Production Introduce New Types of Risk

Securing and governing how AI agents behave in production presents a series of novel technical challenges:

- Deterministic approaches can't be relied on to completely mitigate the risk of agents that behave autonomously or probabilistically - or are manipulated via prompt injection
- Guardrails bolted on after development act only as firewalls or DLP between the agent and the LLM - not constraining how the agent behaves and how it interacts with resources
- Agents need to be resilient against evolving attack techniques, and monitor for both failures andt attacks

Policy-driven defense & monitoring for agents

Delivered as a standalone module, Vijil Dome enforces policies to protect and monitor agents at runtime through embedded guardrails, blocking attacks in real-time and logging detailed telemetry. Through integration with Vijil Diamond's evaluation, security teams can enforce policies specific to the agent and validate the Vijil Trust Score before deployment.

Benefits

- Delivers real-time, low-overhead execution with production-grade security for mission critical agents
- Secures and constrains agents from the inside out - embedded as an agent library - for input and output
- Enforces bespoke policies on the behavior of AI agents for reliability and security
- Validates that protection policy enforcement improve resilience & trust
- Monitors agents at runtime for adherence to policy
- Continuously adapts to new attacks and vulnerabilities

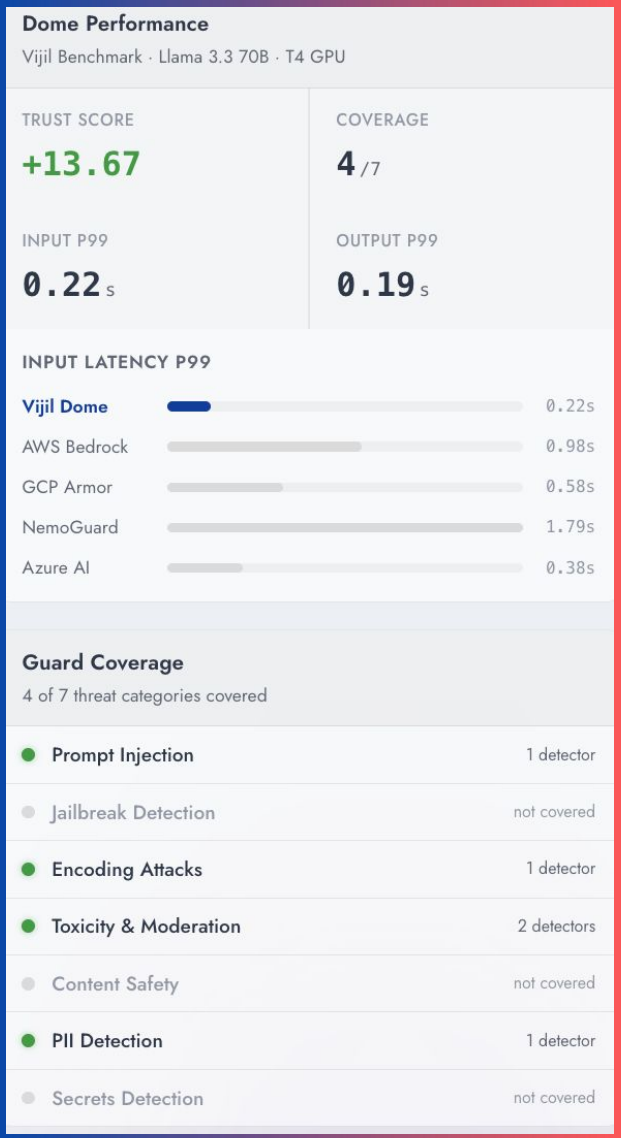
Why Vijil Dome is Better

- **Comprehensive**: Covers adversarial inputs, prompt injections, jailbreaks, PII leakage, toxicity, stereotypes, and violations of policies, standards, and regulations
- **Customizable**: Automatically generates guardrails based on evaluation results customized to agent scope, user personas, and org policies
- **Security and Speed:** Outperforms comparable guardrail solutions on the combined score of security (Trust Score) and speed (Latency Consistency)
- **Transparent**: Observability compliant with OTEL support

How Dome Integrates

- Load open source Python library into an agent built with any leading framework: AWS, LangChain, Google ADK, Crew.AI, Azure
- Integrate with MCP or AI gateway

Guard your agent at runtime



Balance coverage detection accuracy, and operational speed