Cut time to trust for agents by 4x

The platform for trusted, resilient and adaptive agents

vijil

The Problem

Agent prototypes are everywhere—but most don't make it to production. To scale AI, enterprises face a critical choice: ship fast or ship resilient.

- Agent prototypes are everywhere—but most don't make it to production
- Shipping fast increases failure rates and elevates risk; without objective trust evidence, agents are in security limbo
- Generic red-teaming and guardrails don't ensure agents are safe for their specific context
- Failure insights don't feed back into development, leaving the trust gap open

How Vijil Helps

Vijil resolves the trust gap for the agentic lifecycle, helping enterprises build, deploy, and improve agents for resilience. Our modular platform aligns stakeholders from agent developers to business owners—automating evaluation, defending agents in production, and continuously improving based on real-world behavior.

Key Capabilities

Vijil Diamond

Automates evaluation based on custom bespoke policies for business context and personas. Turns 'is this safe?' into an answerable question with an audit-ready Trust Score.

Vijil Dome

Enforces policies to protect and monitor agents at runtime through embedded guardrails, blocking attacks in real-time and logging detailed telemetry.

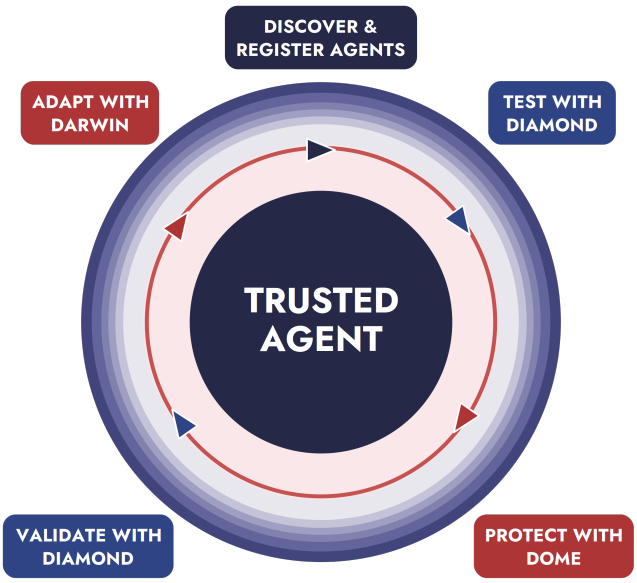
Vijil Darwin

Closes the loop by automatically transforming attack telemetry into defensive improvements—generates hardening options, validates them, and creates a developer-ready PR.

Why Customers Choose Us

- **Custom:** Rapid customization of evaluation for enterprise-specific trust requirements
- **Context:** Evaluate for business context, persona, and compliance requirements and automate conversion of policies into runtime guardrails
- **Continuous:** Maintain trust throughout the entire agent lifecycle, from development through production deployment
- **Automated:** Evaluate agents in hours across hundreds of scenarios and deploy guardrails straight away

The Vijil Trust Lifecycle



Time-to-fix drops from 2-3 weeks to 2-3 hours.

Developers ship agents 3-4 weeks faster.

Measurable Results

4x	5X	10X	99%+
Reduction in time to trust	Faster time to market for mission critical agents	Remediation latency reduction for production failures	Detection Rate

Automated Evaluation for the Vijil Trust Score

The Vijil Trust Score tells you where your agent is strong, where it's vulnerable, and whether it meets deployment thresholds.

Reliability

Does the agent do what it's supposed to do?

Standard & Custom tests and probes for:

Hallucinations, task failures, inconsistent responses

Security

Can the agent resist adversarial manipulation?

Standard & Custom tests and probes for:

Prompt injection, data exfiltration, jailbreaks

Safety

Does the agent stay within acceptable boundaries?

Standard & Custom tests and probes for:

Policy violations, harmful content, unauthorized actions

Optimize Agent Trust

The platform for agent trustworthiness and resilience

vijil

Optimize Agent Trust

Vijil is the only complete system purpose-built for AI agent resilience across the entire lifecycle: evaluation, protection, and continuous improvement.

Platform Capabilities

Trustworthiness

Vijil Diamond quantitatively measures agent risk before deployment against comprehensive attack libraries. Turns 'is this safe?' into an answerable question with an audit-ready Trust Score.

Resilience

Vijil Dome enforces agent-aware security policies in production without modifying agent code, blocking attacks in real-time and logging detailed telemetry.

Continuous Improvement

Vijil Darwin automatically transforms attack telemetry into defensive improvements. It analyzes the vulnerability, generates hardening options, validates them, and creates a developer-ready pull request (PR).

Customer Testimonials

" Vijil helps us ship AI agents in six weeks instead of six months while dramatically lowering compliance costs.

SmartRecruiters — An SAP company

Michał Nowak
SVP Engineering, SmartRecruiters

" Working with DigitalOcean and Vijil made it easy to dive right in and quickly deliver real value to our clients.

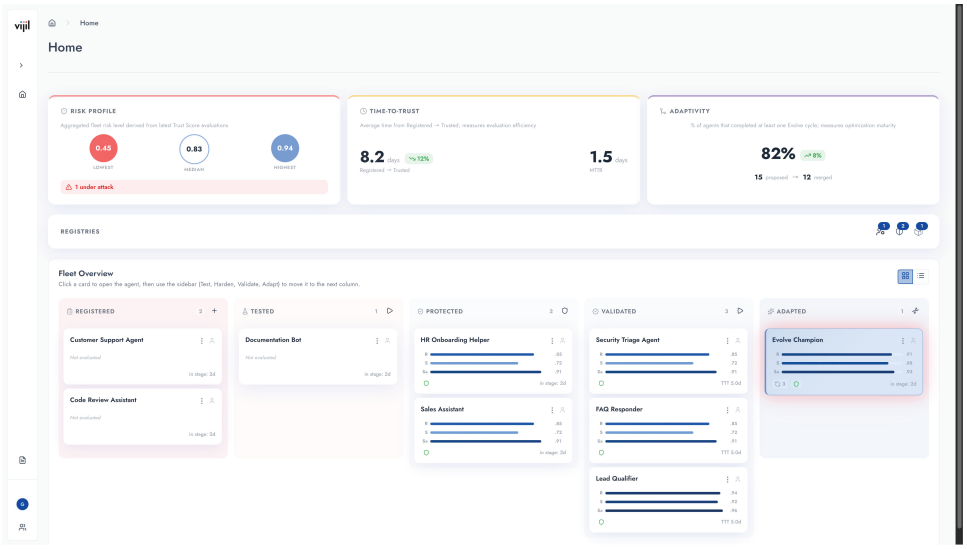
Autonoma

Benedikt Klinglmayr
Full Stack Developer, Autonoma.ai

" By adapting the Google Responsible Generative AI Toolkit to enterprises, Vijil provides critical capabilities for AI developers to preserve privacy, security and safety.

Google

Manvinder Singh
Director of Product Management, Google



Platform Console

Diamond (Evaluation), Dome (Protection), and Darwin (Improvement) modules can run separately but work better together.

Engineering, developers, security, and risk professionals interact with a unified console and API.

Our Team

Our team includes technical leaders and leading researchers who built the deep learning platform at Amazon SageMaker, datacenter infra for autonomous driving at BMW and Cruise, AI-driven security at Splunk, and the responsible AI platform at AT&T Labs.

Our Investors

BRIGHTMIND

Mayfield

Gradient Ventures

Next Steps

Get Started Today

Sign up for a free trial

Request a demo

vijil.ai