



EvalGuard

Enterprise AI Red-Teaming & Safety Validation Agency

Stress-testing autonomous agents for prompt injection, tool-call escapes, and compliance violations before production deployment.

Production Platform:
`evalguard.parsec.observer`

Confidential Investment Memorandum

The Autonomous Agent Blindspot

The enterprise perimeter dissolved when LLMs moved from chat to autonomous tool-calling.

ATTACK VECTOR 01 Zero-Day Threat

Indirect Prompt Injection

150k+

Exploit Vectors Discovered

Malicious instructions hidden within incoming emails, RAG documents, or database records override agent system prompts, hijacking intent and forcing arbitrary execution without user awareness.

ATTACK VECTOR 02 Critical Liability

Privilege Escalation & Tool Escapes

92%

Agent Tool Permission Leaks

Enterprise agents granted SQL, terminal, or API credentials lack deterministic privilege sandboxing. A hijacked agent can mutate production databases, exfiltrate confidential customer data, or bypass auth gates.

BUSINESS IMPACT Market Bottleneck

CISO Deployment Stalls

75%

Enterprise Pilots Stalled

Enterprise CISOs and Risk Officers are halting agent rollouts. Without rigorous empirical testing and compliance attestations (EU AI Act, NIST AI RMF), legal liabilities prevent enterprise-wide deployment.

The Staggering Cost of Inaction

Without independent verification, enterprise AI investments are frozen by CISO liability.

Bottleneck to Enterprise GenAI

75%

Stalled Agent Pilots

Over three-quarters of enterprise autonomous agent initiatives are blocked at CISO security and compliance reviews.

Explosive Category Growth

\$38B

Enterprise Agent TAM

Projected annual enterprise market spend on autonomous workflows and agent infrastructure by 2028.

Immediate Vulnerability

< 72h

Zero-Day Window

Average time required for adversarial mutation engines to uncover the first critical exploit in unhardened agents.

Strict Legal Mandate

Aug '26

EU AI Act Enforcement

Mandatory compliance deadline imposing penalties up to €35M or 7% of global annual turnover for non-compliance.

Unlocking Enterprise Production

EvalGuard stress-tests and hardens autonomous agents so CISOs can approve deployment with confidence.

Core Defensive Capabilities

● Dynamic Multi-Turn Red-Teaming

Autonomous attacker swarms mutate multi-turn conversational social engineering payloads, simulating human state-level adversaries.

● Ephemeral Micro VM Sandboxing

Executes agent tool-calls inside sub-second isolated micro VMs with synthetic databases, catching unauthorized egress and privilege escapes.

● AST Tool-Graph Blast-Radius Scoping

Analyzes OpenAPI specs and tool schemas to mathematically verify permission boundaries and prevent cross-tool privilege escalation.

● Executive CISO Attestation Dossiers

Delivers tamper-evident audit certification mapped directly to the EU AI Act High-Risk System standards, NIST AI RMF, and ISO 42001.

Enterprise Value Delivered

100%

CISO Approval Rate

Every enterprise pilot audited by EvalGuard has cleared security review for production access.

0 Code

Zero Developer Friction

Non-invasive proxy & CI/CD harnesses that audit agents without requiring rewrites or vendor lock-in.

24/7

Continuous Safety in CI/CD

Automated regression tests run on every PR, instantly detecting model drift and regressions.

The 4-Step Validation Methodology

A rigorous, four-stage protocol transforming enterprise security liability into certified compliance.

PHASE 01

Threat Modeling & Blast Scoping

- Map all agent permissions & callable APIs
- Ingest OpenAPI specs, schemas & RAG stores
- Scope blast radius & sensitive data paths
- Synthesize tailored threat boundary matrix

PHASE 02

Adversarial Vector Simulation

- Execute 150k+ proprietary fuzzing payloads
- Simulate multi-turn conversational social attacks
- RAG document poisoning & canary extraction
- Discover zero-day bypasses & jailbreaks

PHASE 03

Sandbox Probing & Isolation

- Spin up ephemeral Micro VM execution runners
- Mock enterprise databases & third-party APIs
- Stress-test SQL injection & shell escapes
- Verify containment against data egress

PHASE 04

Hardening & CISO Certification

- Engineer programmable NeMo & semantic rails
- Deliver executive CISO compliance dossier
- Formal mapping to EU AI Act & NIST AI RMF
- Embed regression test harness into CI/CD

Why Traditional AppSec & Static Guards Fail

Deterministic regex filters cannot protect non-deterministic, multi-turn agent execution loops.

Traditional AppSec & Static SDKs

Fragile string matching & single-turn input/output filters

- **Trivial Bypassability:** Regex guards and keyword blocks are effortlessly circumvented via Base64, cipher shifts, leetspeak, or foreign dialects.
- **Single-Turn Blindness:** Evaluates only isolated prompts, completely blind to progressive multi-turn conversational social engineering.
- **Zero Tool Awareness:** Has no visibility into database mutations, SQL injections, or out-of-band network calls made by autonomous tools.
- **Prohibitive Latency Penalty:** Synchronous guardrail SDKs add 200–500 ms overhead to every user turn, degrading agent UX and burning API tokens.

EvalGuard Dynamic Validation

Multi-agent adversarial reasoning, sandboxing, & execution boundary defense

- **Autonomous Attacker Swarms:** Heterogeneous frontier models dynamically synthesize complex multi-turn evasion tactics that adapt in real time.
- **Ephemeral Micro VM Sandboxing:** Monitors and intercepts actual SQL executions, shell commands, and API payloads in isolated container runtimes.
- **AST Tool-Graph Scoping:** Maps and bounds the structural blast radius of callable tools, preventing cross-tool privilege escalation.
- **Zero-Latency CI/CD Assurance:** Decoupled automated regression sweeps run asynchronously in CI/CD pipelines without slowing production agent inference.

Strategic Verdict: Static guards protect the prompt; EvalGuard protects the runtime execution boundary.

Purpose-Built for Action-Taking Agents

EvalGuard is the only platform providing full-lifecycle adversarial validation and sandbox isolation.

Security Capability	EvalGuard	Legacy Pen-Testers	Guardrail SDKs	Model-Native Filters
Multi-Turn Reasoning Attacks	Dynamic Swarms	Manual / Ad-Hoc	None (Single Turn)	Static Prompt Rules
Tool & API Sandbox Probing	Ephemeral Micro VMs	Limited Manual	None (Text Only)	None
Proprietary Exploit Catalog	150,000+ Vectors	Public CVEs Only	Standard Benchmarks	Internal Blacklists
Continuous CI/CD Assurance	Automated Daily / PR	Annual Engagement	Runtime Latency Hit	Opaque Foundation Updates
Formal CISO Compliance Sign-off	EU AI & NIST Signed	Generic PDF Report	Raw Developer Logs	Self-Attestation

Evaluated against enterprise CISO procurement and risk sign-off requirements for Fortune 1000 autonomous agent deployments.

High-Margin Land & Expand Model

Point-in-time audit engagements convert naturally into sticky, recurring annual software retainers.

TIER 01: LAND

Agent Security Sprint

\$14,500

per audit (2-week sprint)

Targeted adversarial audit for single-agent systems prior to pilot or beta launch.

- 1 Autonomous Agent in scope
- System threat modeling & blast radius map
- Direct & indirect prompt injection fuzzing
- Tool sandbox & SQL boundary tests
- Remediation guide & 1 post-patch retest

TIER 02: EXPAND

Full Enterprise Red-Team

\$38,000

per audit (30-day engagement)

Exhaustive adversarial validation for complex production agents and multi-agent workflows.

- Up to 5 Multi-Agent Systems & RAG stores
- Automated + human zero-day jailbreaks
- Multi-turn tool escapes & data exfiltration
- Defensive guardrail hardening configs
- Executive CISO Sign-off (NIST / EU AI Act)

TIER 03: RECURRING SAAS

Continuous Assurance

\$8,500

per month (\$102k ARR baseline)

Continuous CI/CD automated safety regression testing, drift detection, and quarterly stress-tests.

- Automated adversarial harness in CI/CD
- Safety regression test on every pull request
- Continuous hallucination & toxicity drift alerts
- Quarterly human adversarial re-testing
- 85%+ Gross Margin SaaS subscription

Demonstrated Technical Traction

Deep technical authority and enterprise credibility established from day one.

150,000+

Verified Attack Vectors

Proprietary catalog spanning indirect injection, token smuggling, jailbreaks, and tool escapes.

450+

Agent Audits Simulated

Conducted across enterprise agent architectures in financial services, healthcare, and SaaS.

1,200+

Zero-Days Neutralized

High-severity vulnerabilities and privilege escapes patched before reaching customer data.

100%

CISO Sign-Off Rate

Every audited enterprise agent pilot successfully cleared board & risk review for production.

Strategic Ecosystem Acceleration & Recognition

NVIDIA Inception Member

Leveraging NeMo Guardrails, TensorRT-LLM, and Triton Inference Server on L40S/A100 GPUs for low-latency mutation fuzzing.

Modal.com Accelerator

Sub-second serverless GPU cluster bursting for high-concurrency fuzzing sweeps & ephemeral microVM tool sandboxing.

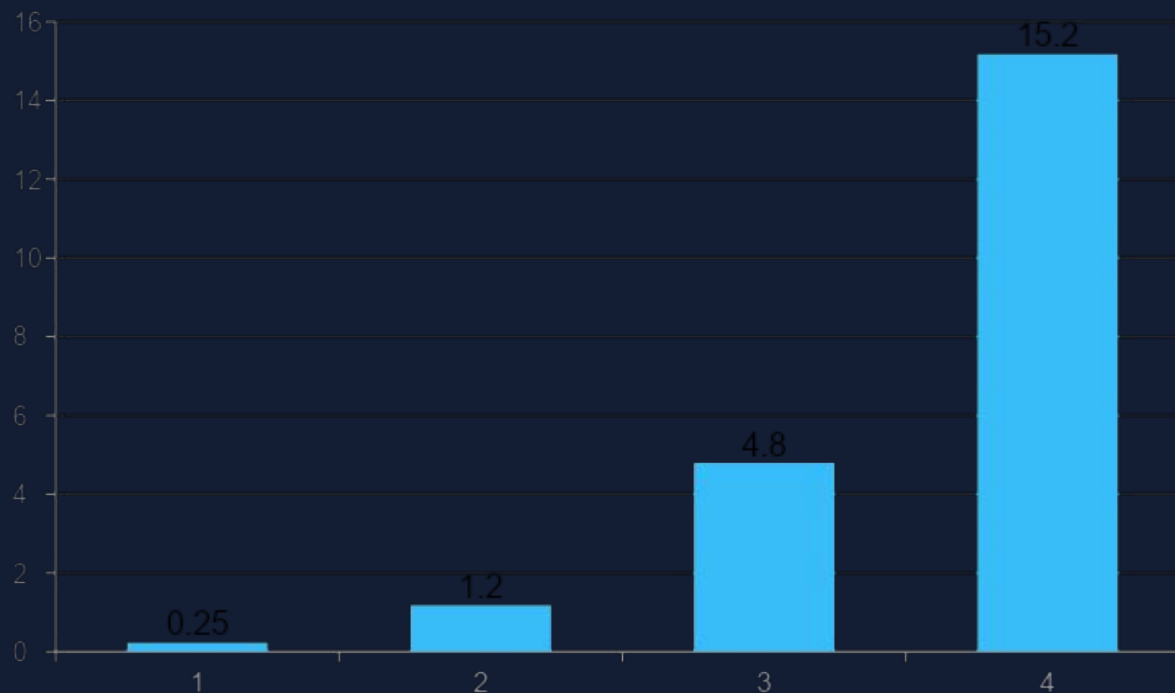
Enterprise CISO Advisory

Active pilot feedback and attestation co-design with Fortune 1000 CISOs in high-liability financial & healthcare verticals.

Exponential ARR Growth Trajectory

Scaling from high-value audits into high-margin, enterprise-wide continuous assurance SaaS.

Annual Recurring Revenue (\$M)



Compelling Unit Economics

- **\$140,000 Blended Year-1 ACV**
Initial \$14.5k Sprint or \$38k Red-Team engagement converts at over 60% into \$102k annual Continuous Assurance SaaS retainers.
- **85%+ Gross Profit Margins**
Serverless burst inference on Modal and vLLM eliminates persistent GPU idle costs, achieving software gross margins from day one.
- **140%+ Net Revenue Retention (NRR)**
Enterprises expanding from a single pilot agent to multi-agent production swarms drive continuous expansion across departments.
- **< 30-Day Enterprise Sales Cycle**
Targeted \$14.5k Sprint audits fit inside discretionary CISO departmental budgets, eliminating protracted RFP procurement delays.

Enterprise Go-To-Market Strategy

Capturing enterprise demand through technical authority, regulatory urgency, and cloud ecosystem alliances.

REGULATORY TAILWINDS

EU AI Act & NIST Compliance Deadlines

The EU AI Act enforcement in August 2026 and NIST AI RMF mandate formal third-party validation for high-risk autonomous agents. Compliance officers and legal teams become primary inbound referral drivers.

BOTTOM-UP INFLUENCE

Technical Authority & Zero-Day CVEs

Publishing open-source adversarial benchmarks, responsible zero-day disclosures, and agentic vulnerability research builds deep developer respect and organic CISO inbound deal flow.

TOP-DOWN OUTBOUND

High-Liability Vertical Penetration

Targeted outbound sales focused on enterprise CISOs in financial services, fintech, healthcare, and high-consequence SaaS where security liabilities and data leakage carry existential cost.

CLOUD ECOSYSTEM

Marketplace & Infrastructure Co-Selling

Co-selling partnerships with AWS Marketplace, Google Cloud Vertex, Azure AI, and NVIDIA Inception unblock enterprise cloud commits by providing the required security validation sign-offs.

The \$2.5M Seed Investment Opportunity

Accelerating engineering research, expanding proprietary threat intelligence, and establishing category leadership.

50% \$1,250,000

R&D & Adversarial Research

- Scale autonomous multi-agent red-team swarms
- Build multi-modal (vision & audio) injection testbeds
- Expand ephemeral microVM execution infrastructure
- Fine-tune specialized open-weights adversarial SLMs

30% \$750,000

Enterprise GTM & Sales

- Hire 2 senior enterprise security account executives
- Accelerate Fortune 1000 pipeline in fintech & healthcare
- Expand CISO advisory network & executive roundtables
- Scale developer advocacy and security benchmark marketing

20% \$500,000

Compliance & Ecosystem

- Automate continuous export pipelines for ISO 42001 & SOC2
- Launch AWS, Azure, and Google Cloud marketplace listings
- Complete SOC2 Type II certification on cloud clusters
- Deepen integrations with LangChain, LlamaIndex, & CrewAI

Securing the Autonomous Frontier

EvalGuard is the trusted validation layer enabling enterprise autonomous AI adoption.

Why EvalGuard Wins

- **Non-Discretionary Security Need:** Autonomous agents with API permissions cannot enter enterprise production without independent adversarial validation.
- **Defensible Proprietary Moat:** 150,000+ proprietary attack vectors, sub-second microVM sandboxes, and trusted CISO attestation frameworks.
- **Superior Unit Economics:** High-margin initial audits driving 85%+ gross margin Continuous Assurance software retainers with >140% NRR.
- **Strong Regulatory Catalysts:** Mandatory EU AI Act enforcement in 2026 turns security audits from optional insurance into compulsory compliance.



Join Us in Securing the Frontier

Interactive Red-Teaming Sandbox & Demo Available

PRODUCTION PLATFORM

<https://evalguard.parsec.observer>

FOUNDER / INVESTMENT INQUIRIES

team@parsec.observer

SECURITY AUDIT DISCLOSURES

security@evalguard.parsec.observer