Forge

Securing the agentic enterprise.

Enterprise AI began as a content problem: prompts in, answers out. That phase is ending. The new default is the **autonomous agent**, a long-running, stateful system that holds permissions, calls tools, accesses data, and makes decisions on the endpoint or in the cloud.

80%

of enterprises expose sensitive data through agents.

Prompt injection, lineage failures, unscoped retrieval.

93%

of organizations run AI agents with excessive permissions.

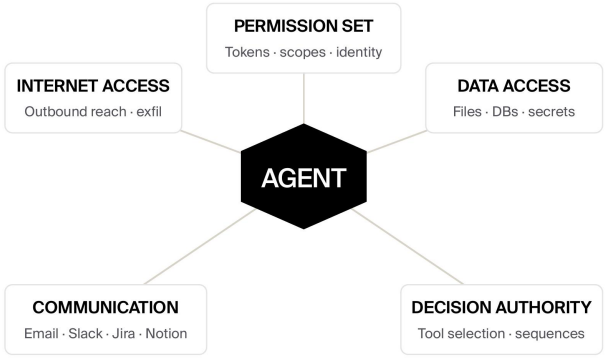
Inherited tokens. Unaudited scopes. No revocation discipline.

70%

of companies are exposed to RCE via compromised agents.

Tool misuse and untrusted input become code execution.

The risk surface, in two parts. What an agent *has*, and how an agent *fails*.



EXTERNAL ATTACK

Prompt injection
Adversarial content in retrieved docs, emails, tickets.

Social engineering
Instructions embedded in untrusted input.

AGENT ITSELF

Over-persistence
Won't stop, retries forever, escalates scope.

Tool misuse
Write where a read was intended; wrong destination.

Human error
Wrong instruction; the agent executes faithfully.

DRIFT OVER TIME

Scope drift
Permissions accumulate; nobody re-baselines.

Trust-boundary chains
Benign read + benign write = unauthorized data movement.

The static policy trap.

Today's playbook is to write a policy: block tools, disable internet, no writes to prod, require approval. It fails in three structural ways.

POLICY RULE

`tool: slack.send →` **ALLOW**

2:14 PM · #ENG
"deploy is ready, merging"
Routine team chatter.
SAFE → ALLOW

3:08 AM · DM EXTERNAL@GMAIL.COM
"[200 KB customer PII attached]"
Off-hours DM, external recipient.
UNSAFE → ALLOW

Same tool. Same verdict. **Different intent.**

01 Guesswork
Written without behavioral data. Teams over-block (kills utility) or under-block (misses risk). Either way, it's a guess.

02 Static
Agents, integrations, users, and contexts evolve continuously. Policies are frozen the moment they ship.

03 Blunt
Tool-level rules can't tell *Slack to #eng at 2pm* from *Slack DM with PII to an external account at 3am*. Same tool. Different intent.

Enterprises need intelligent, data-driven runtime protection for agents.

Grounded in how agents actually behave, not in guesswork.

Forge replaces guesswork with evidence.

01

Capture

Every prompt, tool call, data access, and output across the enterprise, normalized into one record.

02

Baseline

Behavioral profiles per person, agent, MCP, and tool. What's normal for *this* user, *this* workflow.

03

Recommend

Policy from real behavior: unused permissions, risky tool combos, gates that catch risk.

04

Enforce

Isolate agents, kill sessions, gate approvals, redact data. At the moment of action, not after.

BEHAVIORAL PROFILING

Understand what "normal" looks like.

- **PERSON** — hours, workflows, data classes
- **AGENT** — tools chosen, success patterns
- **MCP** — callers, payloads, frequency
- **TOOL** — arguments, callers, hours

Set up in minutes with the EDR your security team already runs.

CASE STUDY · DEPLOYMENT

Investment firm · $10B AUM · Procured agent platform

From blocker to enabler.

An investment firm had recently procured a popular agent platform to enable research copilots for analysts, compliance assistants, and trade-ops agents. Rollout beyond pilot stalled behind manual review. DLP, IAM logs, and secrets scanners couldn't see what the vendor's agents were actually doing across the firm's tools and data. **We gave them trace-level visibility and runtime control in one week.**

TIME TO VALUE

1 week

to first enforced policy

~600

analysts and engineers baselined by week 1

32

runtime policies enforced in week 1

30 days

to firm-wide rollout with guardrails

Team & backers

- Second-time founders with deep experience in enterprise security, AI red-teaming, and goal-driven agent systems.
- Secured systems at Meta, Google, Cloudflare, and the U.S. Department of Defense. AI red-teaming research at Harvard, Stanford, and MIT.
- Backed by $10M from Greylock Partners and exceptional angel investors.
- Early partnerships with Global 2000 enterprises across financial services, software, and infrastructure.

Rohan Kalahasty, CEO

- AI Research @ MIT, Harvard
- CEO/CTO @ Vytal.ai
- Physics @ Caltech

Sritan Motati, CTO

- AI Research @ Stanford, Harvard
- Founding Team @ Vytal.ai
- CS + Finance @ UPenn

PREVIOUSLY AT

Meta · Google · CLOUDFLARE · U.S. Department of Defense

BACKED BY

greylock · Amplify · Palantir

Anu Bharadwaj, President @ Atlassian

Frederic Kerrest, Cofounder @ Okta

Todd McKinnon, CEO @ Okta

Aaron Levie, Cofounder @ Box

Lip-Bu Tan, CEO @ Intel

Jeetu Patel, President @ Cisco

Autonomous agents are the new insider threat. We build the behavior profiling layer.

Book a demo → forge.ai/30-min-demo