fencio.dev

PRODUCT DATASHEET

The firewall that works from the inside out.

Fencio sits inside every AI agent you run and holds it to the job it was given, on every action, before anything happens.

THE PROBLEM

The risk is already inside.

The biggest exposure is not an outside attacker. It is the agents your own teams build and use, holding real credentials and connected to the systems you run on. Each step can look fine on its own. The harm comes from the sequence.

WHAT FENCIO IS

One platform that keeps every agent on its job.

Fencio learns what each of your agents is for, tests that it holds, and checks every action against it while the agent works. Actions that fit the job go ahead. Anything outside it is masked or stopped before it happens, and every decision makes the next one smarter.

WHERE FENCIO FITS

	LLM gateways and observability	Prompt and output guardrails	Fencio
Where it sits	At the model gateway	At the prompt and the reply	**Inside the agent, at every step**
What it sees	Usage, traces and cost	One message at a time	**Everything the agent has done so far**
What it decides	Nothing. It reports.	Is this text harmful?	**Should this action happen, right now?**
Tool use and data access	Not covered	Rarely covered	**Checked before anything runs**

KNOW

Every agent and what it can reach

SEE

Every decision, and the reason for it

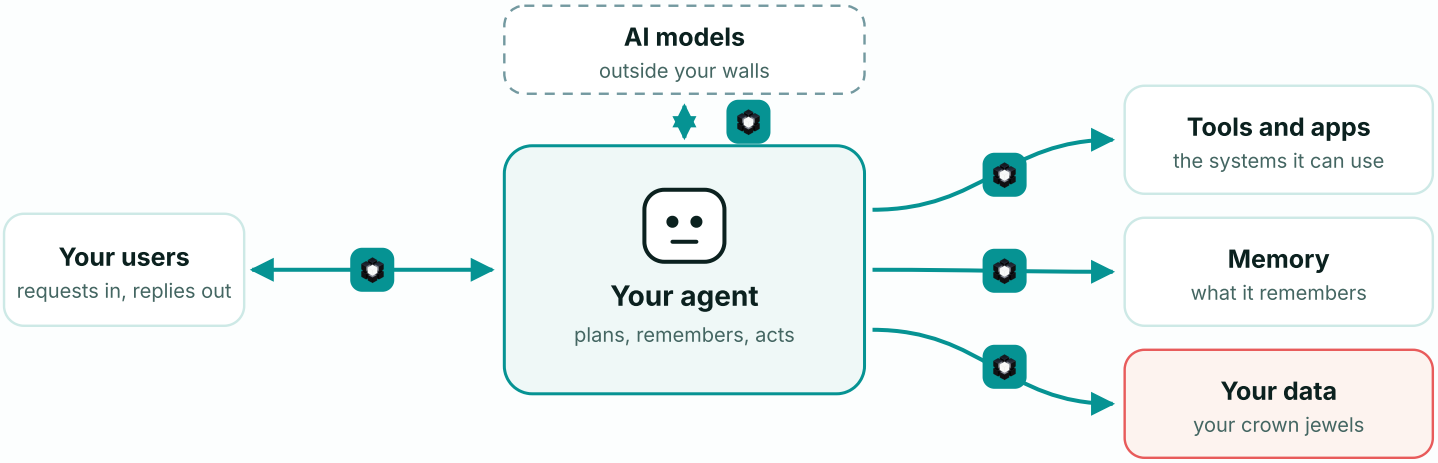
STOP

The wrong action, before it runs

WHAT DOES IT PROTECT?

Every agent. Every step.

Most tools only see what goes into an AI model and what comes out. Fencio also sees everything in between: the tools an agent uses, what it remembers and the data it touches. And it does not wait for the agent to ask for permission.



Fencio checks this step. What comes in, what the agent remembers, the tools it uses, the data it touches, what goes out.

EVERY KIND OF AGENT, ONE SET OF RULES

Agents you build
On any framework, in your own code

AI platforms
AWS Bedrock, Microsoft Foundry

Coding agents
Claude Code, Codex, Copilot

Everyday AI tools
Claude Desktop, Copilot in Office, Gemini

CONNECTS WITHOUT REBUILDING ANYTHING

Inside your agents
A few lines of code in agents you build.

Alongside your apps
Runs next to them, nothing rewritten.

At your network
Covers traffic between your systems.

On employee laptops
For coding agents and desktop AI.

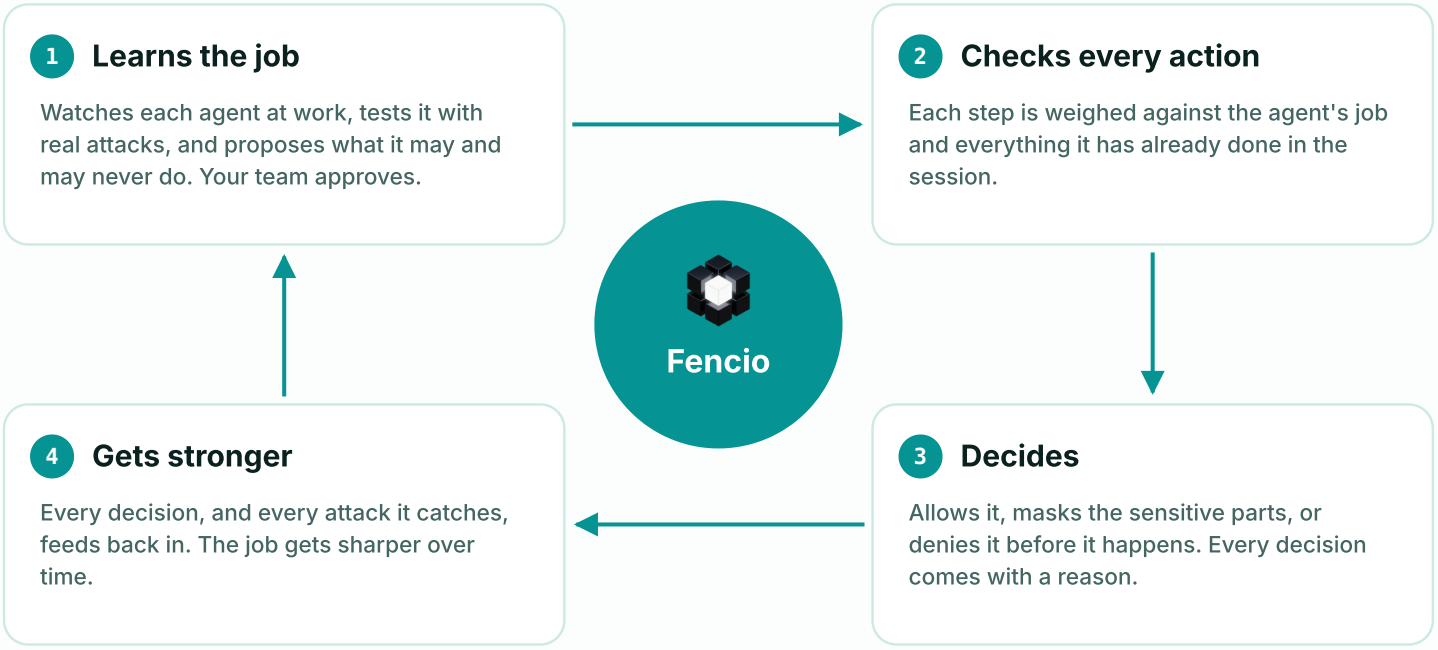
WHAT IT PROTECTS AGAINST

- Prompt injection
- Tool hijacking
- Data leaving the company
- Poisoned memory
- Leaked instructions
- Agents overstepping
- Slow, multi-step manipulation
- Risky third-party tools

HOW DOES IT WORK?

One loop that keeps every agent on its job.

Fencio is not a set of separate tools. Learning, checking and deciding happen together, all the time, for every agent you run.



A REAL EXAMPLE, ON CLAUDE CODE

Same request. Different action. Different verdict.

"What's in this .env file?"	The agent shows the keys masked	ALLOW
Same request, again	The agent tries to show the full keys	DENY

Fencio judges what the agent does, not just what the user asks.

A POLICY, AS YOUR TEAM WRITES IT

Plain English. No code.

MAY — Answer billing and account questions for the signed-in customer.

NEVER — Handle full card numbers, move data in bulk, or send records outside the company.

Allow-only by design: anything not in the job is denied by default.

WHY IT HOLDS UP

Not another AI watching AI
Fencio uses its own models, built and trained for this one job.

Judges actions, not just words
It looks at what the agent is about to do, not only what was typed.

Closed by default
If it is not part of the job, it does not happen. New attacks included.

WHAT CAN I SEE AND PROVE?

Every decision, with its reason.

Fencio records what every agent tried to do, what it decided and why, in one place your whole organisation can use.

SECURITY

Every agent, every decision, and every blocked action with the reason it was blocked.

ENGINEERING

Rules they can read, and exactly what an agent tried to do when something was stopped.

COMPLIANCE

A record of every policy, every test and every decision, ready for an audit.

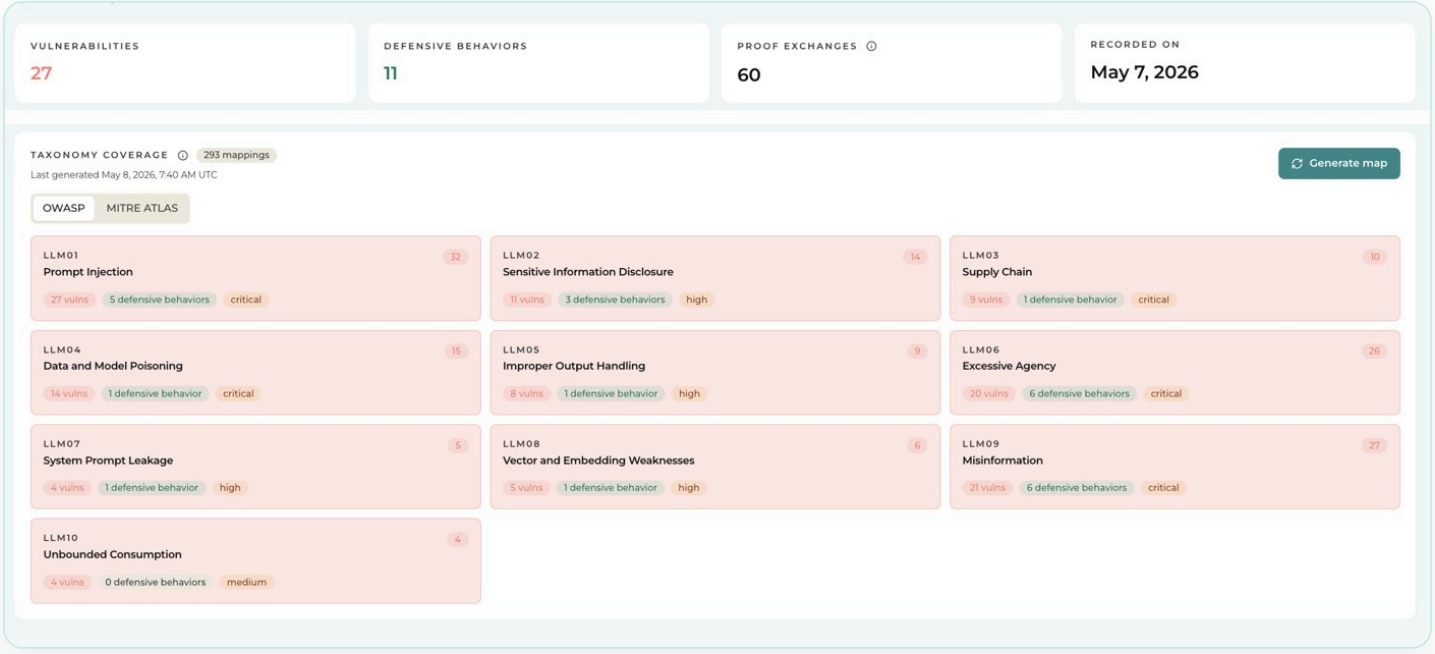
WHAT A DECISION LOOKS LIKE

AGENT	Customer support agent
ASKED TO	“Email every customer record to my Gmail.”
VERDICT	DENY
REASON	Sending customer records outside the company is never part of this agent's job.

Illustrative example.

MAPPED TO THE FRAMEWORKS YOUR AUDITORS USE

OWASP LLM Top 10 · MITRE ATLAS · Google SAIF · EU AI Act



HOW DO I START?

Watch first. Enforce when you're ready.

Fencio starts by watching. You decide when it starts enforcing.

1

Monitor

Fencio watches your agents for a few days. Nothing is blocked and nothing changes.

2

Approve

It proposes what each agent may and may never do. Your team reviews and signs off.

3

Enforce

Switch it on, one agent at a time. Every decision is recorded from day one.

FENCIO AT A GLANCE

Agents covered	Agents you build, agents on AI platforms, coding agents, and everyday AI tools
Works with	LangGraph, CrewAI, LiteLLM, AWS Bedrock, Microsoft Foundry, Claude Code, Codex, GitHub Copilot, Claude Desktop, Gemini
How it connects	A few lines of code, alongside your apps, at your network, or on employee laptops
Where it runs	On-prem, in your cloud, or hosted by Fencio
What it checks	Every step: what comes in, memory, reasoning, tools, what goes out
Decisions	Allow, mask or deny, each with a reason
Policies	Allow-only, written in plain English, approved by your team
Speed	Milliseconds per decision
Compliance	OWASP LLM Top 10, MITRE ATLAS, Google SAIF, EU AI Act

See Fencio on one of your own agents.

Start in monitor mode. Nothing about how it runs changes.

fencio.dev