DoCoreAI – Product Overview

Privacy-First GenAI Cost Observability & Governance

1) What DoCoreAI is

DoCoreAI is a privacy-first observability and governance runtime that provides cost visibility and enforces policy controls in production AI systems.

It sits inside your AI request path, monitors each call in real time, and applies usage policies —without storing prompts or responses.

DoCoreAI collects only anonymous runtime metrics such as token usage, cost, model, and latency, and displays them in built-in dashboards.

2) The problem it solves

AI features work well in demos and early launches.
The real problems appear once usage grows in production.

Most AI teams discover these issues only after costs start rising:

Financial

- AI costs grow faster than expected as usage scales
- A small number of users or environments create huge bills
- Margins shrink as usage increases
- Monthly spend becomes unpredictable

Operational (Governance)

- No guardrails once AI features are live
- Hard to detect risky or non-compliant usage in real time
- No clear cost per request or per environment

Leadership

- Pressure to control AI spending
- Fear of sudden cost spikes
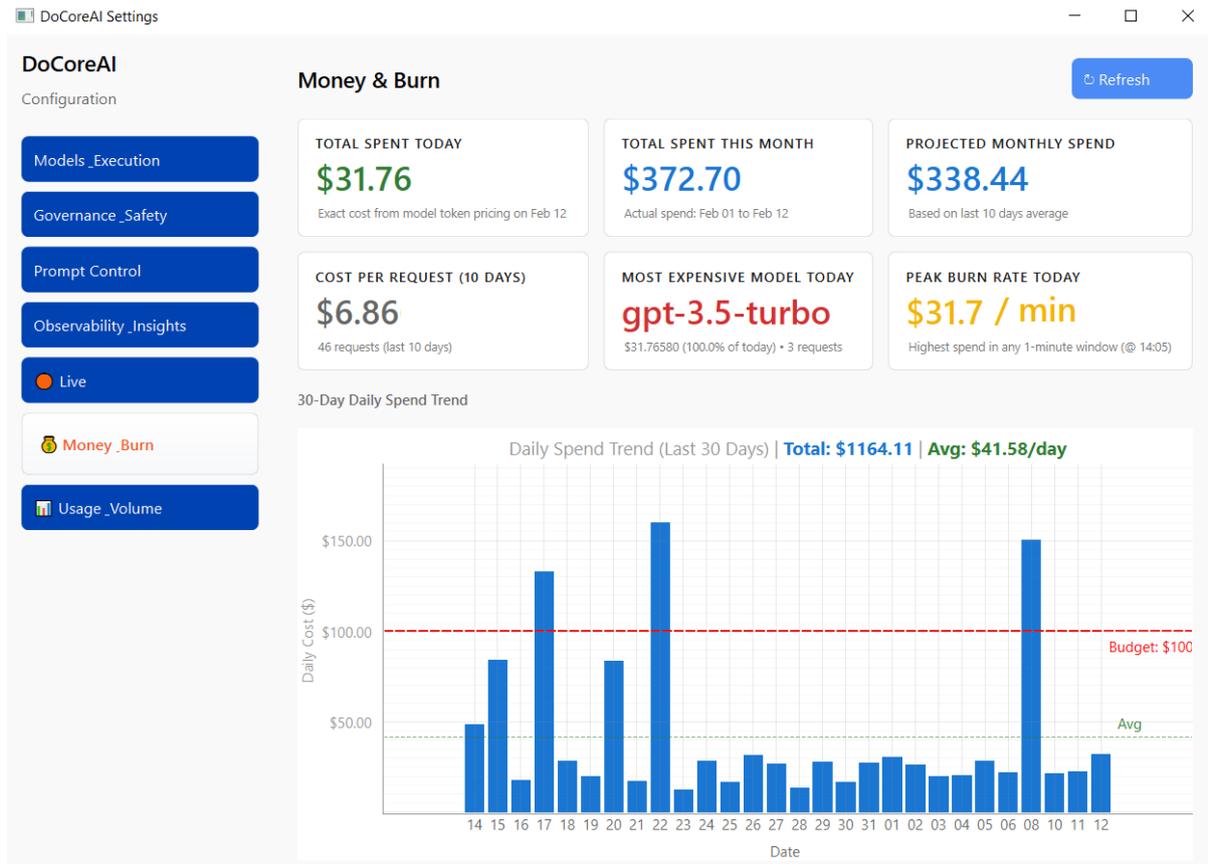- Lack of confidence in scaling AI safely

DoCoreAI – Product Overview

Privacy-First GenAI Cost Observability & Governance

Result:

AI becomes harder to scale, not easier.

AI Cost Trend (Last 30 Days) and Budget Breaches

---

3) How it works (high-level)

DoCoreAI sits inside your AI system and monitors every request.

Simple flow

1. A user sends an AI request.

2. The request is monitored by DoCoreAI.

3. DoCoreAI checks cost and policies.

4. The system:

- allows the request, or

- blocks it based on defined rules.

5. Anonymous usage metrics appear in the dashboard.

Privacy-first by design

- No prompts stored

- No responses stored

- Only anonymous runtime signals are collected

- Metrics are identified by a DoCoreAI token ID and sent to the server dashboards.

No sensitive data leaves your system.

DoCoreAI Runtime Control Flow in Production

Privacy-First LLM Observability
See cost, latency, and behavior — without storing prompts
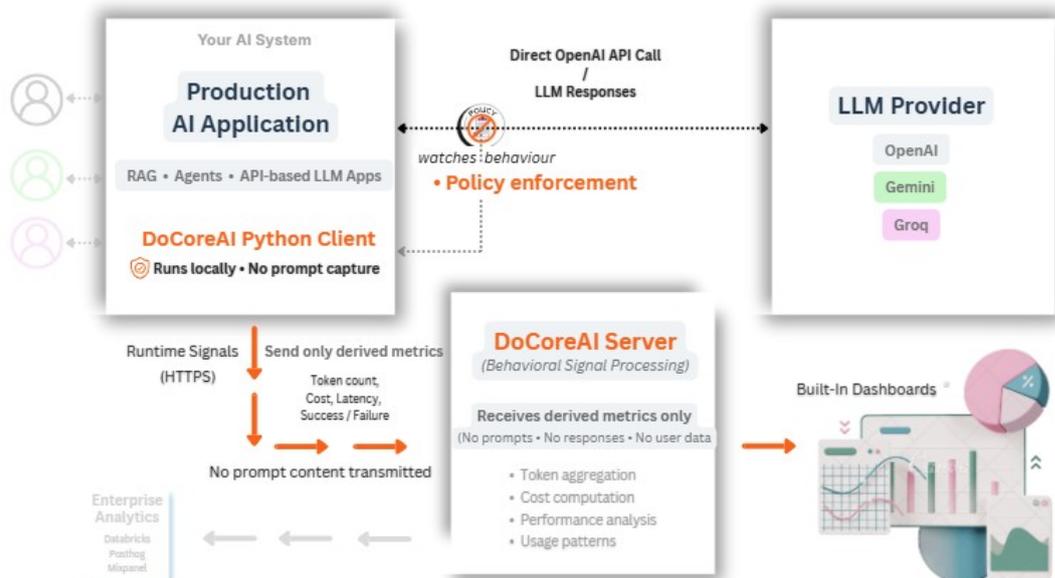
4) Key benefits

Control AI spending

- See cost per request in real time

- Detect sudden usage spikes

- Prevent surprise monthly bills

Add production guardrails

- Apply policies to live AI requests

- Block risky or non-compliant usage

- Bring governance into live production systems
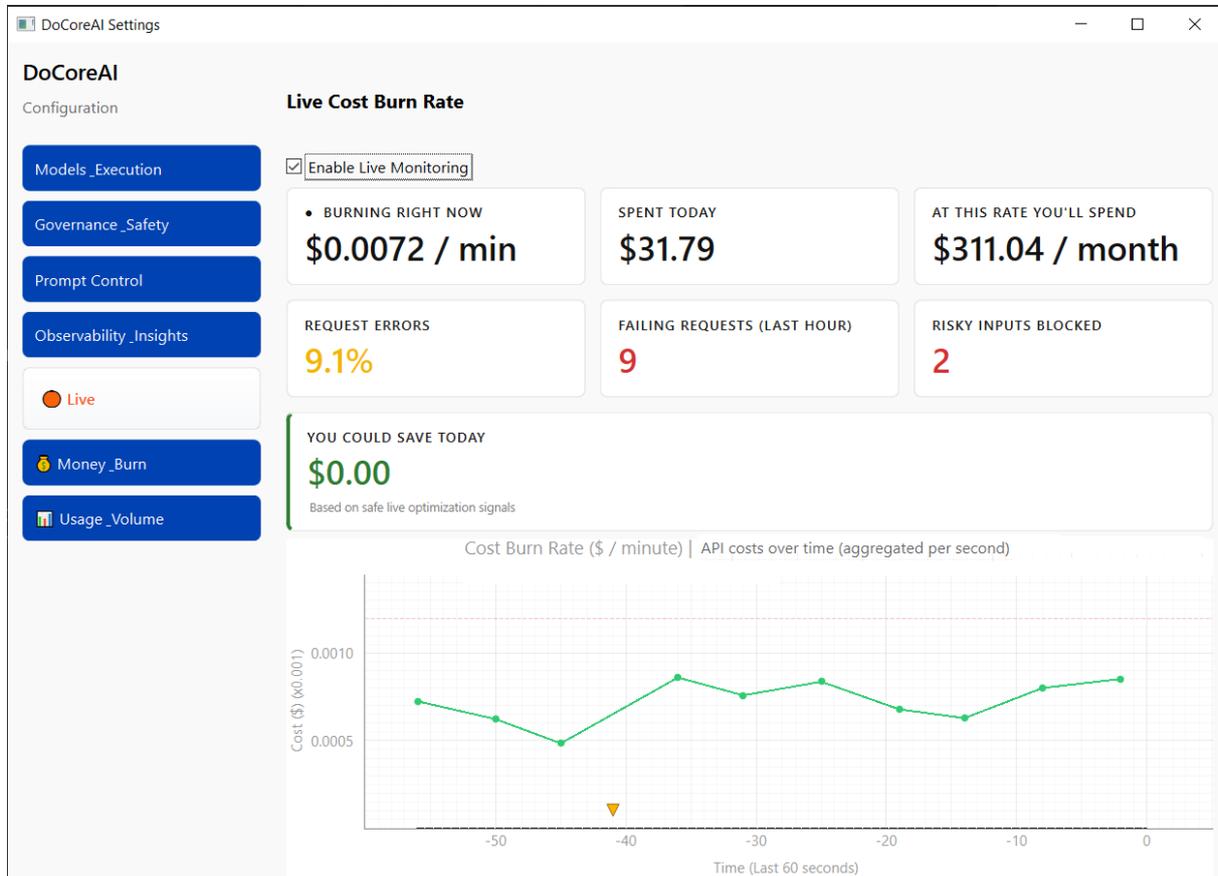
Improve operational visibility

- Understand where AI spend comes from

- Identify expensive users / environments

- Monitor usage as the system scales

DoCoreAI – Product Overview

Privacy-First GenAI Cost Observability & Governance

Live AI Spend and Policy Enforcement Signals in Production

---

5) Who it's for

DoCoreAI is built for teams running **LLM-based features** in real production products.

Ideal companies

- SaaS products with LLM-powered features

- AI support or automation platforms using LLMs

- AI content or developer tools built on LLM APIs

- Any product running LLM based features in production

Typical profile

- 10–300 employees
- AI features already live in production
- $2k+ per month in AI spend

Typical buyers

- Founder
- CTO
- Head of AI
- VP Engineering

---

6) Next step: Pilot or demo

We recommend starting with a short pilot.

Pilot approach

- 7–14 day pilot
- Installed in a single AI environment or service
- No code or architecture changes required

After the pilot, you'll see:

- Real cost per environment
- Where spend spikes occur
- How policies can control AI usage
- A clear view of cost impact before full rollout

Next step: **Schedule** a 20-minute technical walkthrough with a live runtime demo
info@docoreai.com
https://docoreai.com