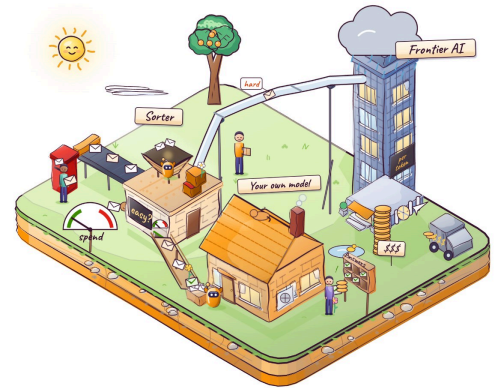


See, control and prove every AI call on AWS, from one gateway.



Every other cost has a unit and an owner; the AI line arrived with neither. **The AI Cost Control Tower Workshop** is one day on your own AI traffic, read-only. You leave with a **baseline of what your AI costs today**, and a **costed plan to put every call on a team**.

What the day leaves you with

Morning

We put one number on today.

- ✓ Every AI invoice, and AI subscriptions on cards
- ✓ **Who is using AI:** spend by key, AI agent, team
- ✓ **Unspent AWS commitment**

Afternoon

We design the target.

- ✓ A task and an outcome for each AI agent
- ✓ **Cost per outcome** and labels
- ✓ **Budgets**, each with an owner

End of day

A costed roadmap.

- ✓ **Your baseline**, and the Control Tower's plan for your accounts
- ✓ **A rollout in phases**, read-only first

Ours, the engineers who build it

AI Cost Control Tower Architect Engineer

Yours, with finance in the room

AI engineering lead Finance Security

The value, in one question

One question to an online shop's AI agent: 14 model calls.

Example

Today

● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●

One shared key

Team

Job

Outcome

?

a ticket resolved

With the AI Cost Control Tower

● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●

One gateway, in your account

Team **Support**

Job **"Where's my order?"**

Outcome **Resolved**

31¢

a ticket resolved

Stopped at the gateway, **not found on the bill.**

Every model call goes through Amazon Bedrock AgentCore Gateway, single-tenant in your own AWS account, and an AWS Service Control Policy refuses any call that doesn't. **Nothing is hosted by us, and prompts never leave your account.** Only metadata does. If the gateway is down, AI pauses rather than running ungoverned, with a documented break-glass path.

Each check watches first; you switch it on, one account at a time.

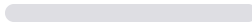
Attribution	Every call on a user, an app and a team, about a two-week build once the definitions exist; from the next month-end, the AI line arrives split by team. We count what still goes around it.	☒
Budgets	Dollar budgets per user and team, checked before the model runs. A call the gateway can't check waits.	☒
Routing and caching	A cheaper approved model where quality holds. No promised percentage: we prove the savings on your traffic first.	☒
Screening	Personal data, denied topics and prompt injection, stopped in the request path, not found in a log afterward. Every decision is logged.	☒
Evidence pack	Mapped to the EU AI Act, ISO 42001 and NIST AI RMF: an audit-ready record, built as you go. The compliance stays yours.	☒
Browser AI tools	They need a different control; we won't pretend otherwise.	☐

What it has done for clients

A U.S. AI marketing platform

48%

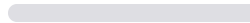
lower compute cost after we put its cloud cost on names, by workload, and right-sized it

Before 
After 

The same platform

9 → 4

AWS accounts, so billing and governance stopped blurring together

Before 
After 

An enterprise software company

//

They worked closely with our AWS team to optimize costs and minimize the financial impact of our migration.

Director CloudOps

Book the AI Cost Control Tower Workshop.

- ✓ **One day**, read-only: no prompts or outputs needed
- ✓ **Free**: AWS funding typically covers it
- ✓ Bring **last month's usage export** and the bill
- ✓ **A costed roadmap**, yours with us or without us



Scan to book

hello@armakuni.com

or ask your AWS account team