Anar Labs

Anar Gateway

One gateway. Any model.

بوابة واحدة. أي نموذج.

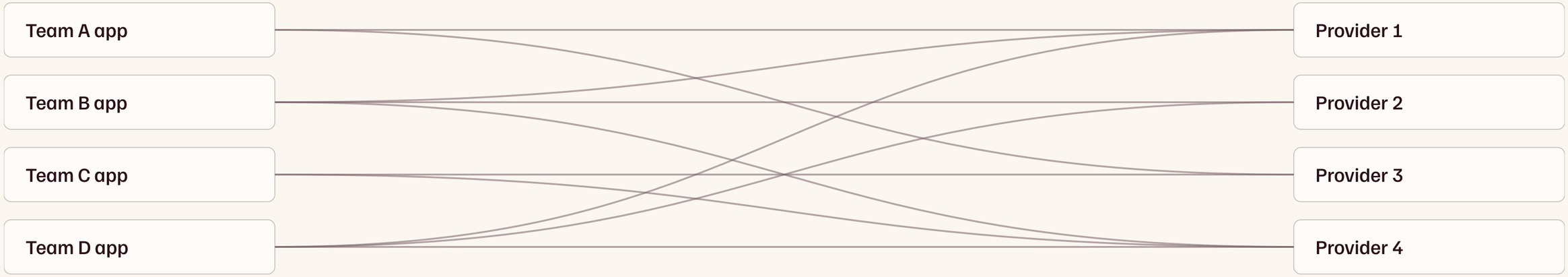
You bring the models and the provider contracts. The gateway routes, meters and checks every request inside your perimeter.

For
Security and digital leaders

Date
November 2026

The problem

Every team now calls a model. Nobody sees all of it.



01

No single record

Each team calls its own provider. Nobody can say who asked which model, and for which purpose.

02

Spend in many accounts

Costs land in separate provider accounts. Finance sees the bill after the money is spent.

03

No control of the route

A prompt can go to any region the provider chooses. Nothing stops it at the boundary.

What it is

One endpoint between your apps and your providers.

01

Your application

Calls an OpenAI-compatible endpoint. Change one base URL. Keep the code.

→

02

Anar Gateway

Identifies the caller, checks the budget and the policy, then records the request.

→

03

Your provider

The model you chose, under your contract, in the region you approve.

Anar supplies no models. The gateway governs the models you already use.

Four checks

Each request is attributed, priced, checked and recorded.

Attributed

Each application gets a virtual key that a team owns. Teams never see the provider key.

Shipped

Priced

Budgets and rate limits per customer, team and key. Cost is counted on each request.

Shipped

Checked

Routing rules on content and caller. Content scan through Anar Guard.

Rules Scan

Recorded

Each request is logged with payload, cost and latency. Admin actions go to an audit trail.

Log Audit

Shipped Partial Roadmap

Data residency

A prompt that would leave the approved region is refused.

The gateway returns an error to the caller. It also blocks failover to a provider in another region.

Shipped. Turn it on per policy.

Today the check covers chat requests that match a residency policy. Embeddings, audio and images are on the roadmap.

Response to the caller

403

`data_residency_violation`

Fallback to another region: blocked.

Deployment

It runs inside your perimeter, with no call home.

- **One static binary** with the console built in. Nothing else to install.
- **No call to Anar** at run time. No licence server.
- **Your database:** a database that you run.
- **Provider keys from your vault** or from the environment.

Item	Status
Runs without root rights	Shipped
Large payloads to your object store	Shipped
Signed offline bundle with SBOM	Roadmap

Providers

Works with the models you already run.

20

native provider integrations, plus any endpoint that speaks the OpenAI API.

Native provider integrations, November 2026.

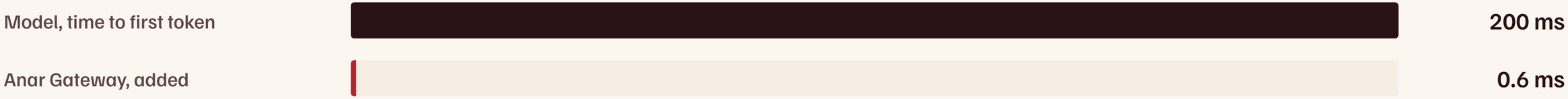
Arabic-aware routing: a rule can send Arabic prompts to the Arabic-capable model you licensed.

Anthropic	AWS Bedrock	Azure	Cerebras
Cohere	ElevenLabs	Gemini	Groq
Hugging Face	Mistral	Nebius	Ollama
OpenAI	OpenRouter	Parasail	Perplexity
Replicate	SGLang	Vertex AI	xAI

Any OpenAI-compatible endpoint, including models in your own data centre

Performance

Governance adds less than a millisecond.



142 µs
added at the median, with no optional features turned on

315 µs
added at the median, with the full governance chain

34k
requests per second on one process

0.6 ms
added to a 200 ms first token

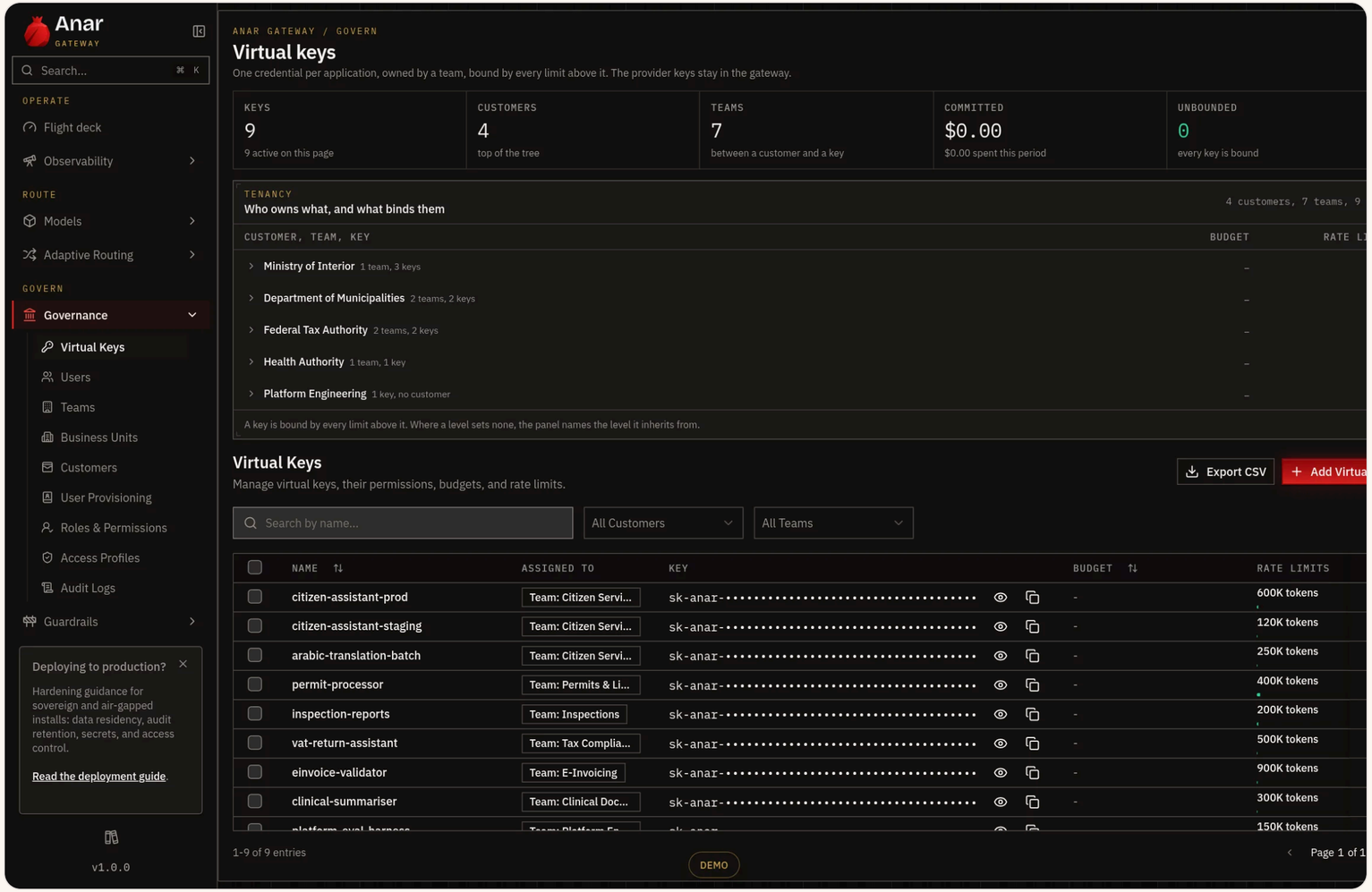
Measured September 2026 on a shared development machine: 8 cores, loopback, a mock provider, median of 3 runs. We can run the same test on your hardware.

The console

Every key has an owner and a limit.

Customers contain teams. Teams own keys. A key inherits each limit above it.

Shipped



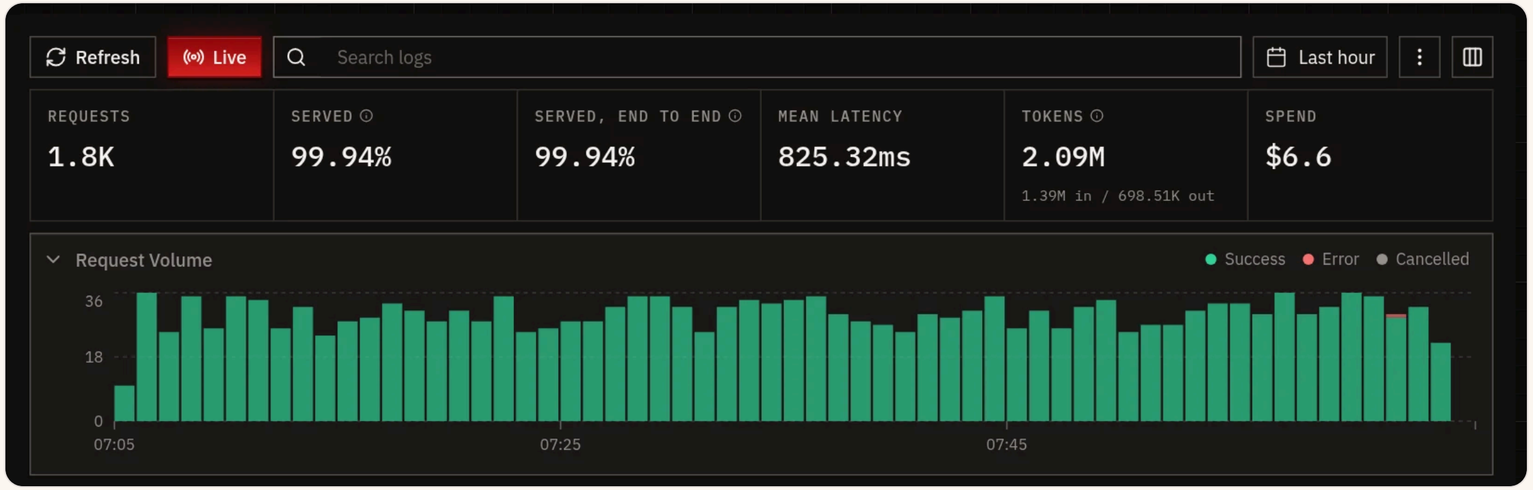
Anar Gateway console, virtual keys. Demo data.

The console

Every request is counted and priced.

See volume, success rate, latency, tokens and spend. Open any request to see who sent it and what it cost.

Shipped



Anar Gateway console, request log. Demo data.

Status today

What works now, and what does not yet.

Capability	Status	Note
Routing, budgets, keys, request log	Shipped	On in every install.
Residency refusal, cost export	Shipped	Turn on per policy.
Role-based access	Partial	Four roles set in the configuration file. Off by default.
Audit trail	Partial	Records admin actions. No export or tamper evidence yet.
Content scan	Partial	Needs Anar Guard. Passes traffic if Guard is down.
SSO, SCIM, alerts, PII redaction	Roadmap	Designed. Not built.

Roadmap

The order follows what a security review asks for.

Now

- Routing, budgets and keys
- Request log and metrics
- Residency refusal
- Runs inside your network, with no call home

Next

Roadmap

- Sign-in with your identity provider
- Roles you manage in the console
- Tamper-evident audit trail

Later

Roadmap

- Signed offline bundle with SBOM
- Arabic console
- PII redaction and alerts

The platform

The gateway is the base of the Anar platform.

Anar Guard

Scans prompts and answers for unsafe content and personal data.

Anar Comply

Scores your AI controls against GCC and EU frameworks.

Other Anar products

They can send their model calls through the gateway when you connect them.

Anar Gateway

Identity, budgets, routing policy and the request record, under every product above.

شاهدوه يعمل داخل مؤسستكم.

See it inside your perimeter.

A one-month pilot: we deploy in your environment, connect your providers and run it with one team.

anarlabs.ai/contact