The Resilience Newsletter

THALAMUS ADVISORY

ISSUE 06 SEPTEMBER 2026 READING TIME · 10 MIN SERIES · SECTOR ANALYSIS

SECTOR ANALYSIS · TECHNOLOGY, SOFTWARE & CLOUD

Selling Reliability You Must Also Have

These companies sell reliability to everyone else. Their outages come with a status page, a postmortem the industry reads, and SLA credits — which makes honest positioning harder and more necessary.

THALAMUS ADVISORY • ENTERPRISE RESILIENCE PRACTICE

01 — THE SIGNAL

The baseline is already excellent

Any argument made to a technology company about observability has to start by conceding something: they probably already have it. Google invented the SRE discipline. Netflix invented chaos engineering. Amazon operates at a scale that forces solutions nobody else needs. Pitching basic monitoring into that room is how a conversation ends.

US technology revenue runs to roughly $2.4 trillion, on top of about $1.5 trillion in annual enterprise IT spending. The reliability engineering inside these companies is the best in the economy. So the honest question is not what they lack — it is where the remaining incidents actually cost them.

02 — THE MECHANISM

What survives a mature practice

When detection is solved and instrumentation is thorough, the incidents that remain share a profile: rare, novel, cross-service, and expensive in senior attention rather than minutes.

The cost has moved. It is no longer the outage — it is the six engineers on a bridge call for two hours, each holding a different hypothesis, working sequentially through them because that is how human investigation works. The constraint is not tooling. It is the serial nature of the hypothesis loop and the seniority of the people executing it.

> *In a mature estate, the expensive part of an incident is not the downtime. It is the attention.*

Multi-tenancy adds a second pattern. Noisy-neighbour degradation is politically as well as technically hard: attributing a slowdown to a specific tenant requires evidence strong enough to have a commercial conversation about. Aggregate percentiles cannot support that.

03 — THE COST

Where it actually lands

SURFACE	HOW IT PRESENTS
SLA credits	Contractual and immediate. At hyperscale a single significant incident can run to eight figures in credits alone.
Public postmortem	The incident becomes an industry artefact. The reputational cost lands on sales cycles for quarters afterwards.
Shared fate	Where the platform is also your own dependency, an incident hits first-party services and customers simultaneously, and the investigation must separate the two before it can start.
Senior attention	The scarcest resource in the company, spent on serial hypothesis elimination that is largely mechanical.
Tenant attribution	Multi-tenant degradation that cannot be attributed with evidence becomes a support escalation instead of a commercial conversation.

04 — THE AI STAKES

Model serving broke the assumptions

Model-serving infrastructure has latency and cost characteristics unlike anything that came before it. Inference is slow relative to an API call, expensive per request, and its performance varies with input in ways that are not visible to conventional monitoring.

Worse, **model behaviour can change without a version change.** The same prompt returns different output. No error is raised, no deploy happened, and availability monitoring shows green throughout. Quality degrades and the first signal is a customer complaint weeks later.

There is also a cost failure mode with no traditional analogue: agent-driven systems become *more expensive the worse things get*. An incident storm triggers proportional agent activity, and an unbounded system can generate extraordinary spend in hours — precisely when nobody is watching the billing console.

05 — THE REMEDY

Parallelism and independence

- **Replace the serial hypothesis loop with a parallel one.** Probe latency, errors, traffic, saturation, logs, traces, upstream dependencies and change records concurrently, returning evidence with strength scores rather than opinions.
- **Verify on a different model.** Self-verification produces correlated errors. Independence is what makes automated conclusions safe to act on.
- **Instrument model endpoints as first-class dependencies** — p95/p99 latency and error rate as SLIs, plus output drift on a fixed evaluation set, because behavioural change is invisible to availability monitoring.
- **Cap agent spend structurally.** Per-run and daily ceilings enforced at the execution boundary, with alerting on rate-of-change rather than total.
- **Make tenant attribution evidence-backed.** Per-tenant SLOs and Apdex so “which tenant is affected” and “which tenant is causing it” are separable with data.

06 — IN PRACTICE

Hosted, or customized for your estate

Hosted on Thalamus AI Cloud

For most technology companies the hosted path is a proof-of-value rather than a destination. Provision an instance, point a bounded service domain at it through OTLP, and measure the diagnostic throughput difference against your existing practice on real incidents. If the parallel fan-out and the independent verifier do not measurably compress your investigation phase, you have learned that cheaply.

Customized for your enterprise

At scale, self-hosted is the norm — you already run the infrastructure and the data volumes make egress unattractive. Customization typically covers: specialist probes extended to your own internal systems and query languages; per-tenant SLO templates aligned to your commercial tiers; verifier configured against a second model from a different vendor to guarantee independence; and spend ceilings mapped to your existing budget controls. The agent fleet is configurable per role, so the boundary between what is automated and what escalates is yours to set rather than ours.

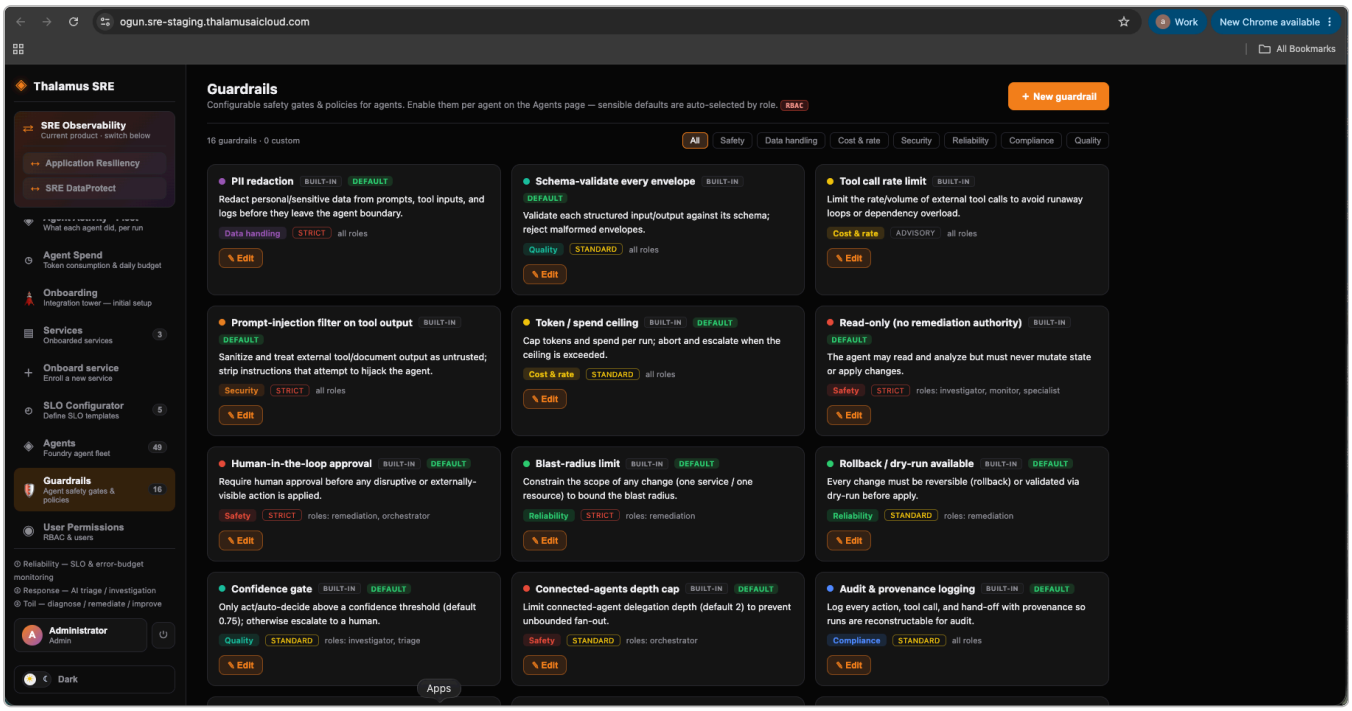

FIGURE 1 — THALAMUS SRE · GUARDRAILS

Spend and authority bounded structurally. Among the sixteen enforced policies: *token / spend ceiling* caps cost per run and aborts with an escalation when exceeded; *tool call rate limit* bounds external calls to prevent runaway loops; *connected-agents depth cap* limits delegation depth to prevent unbounded fan-out. These are the controls that make agent-driven operations financially predictable rather than an open-ended commitment.

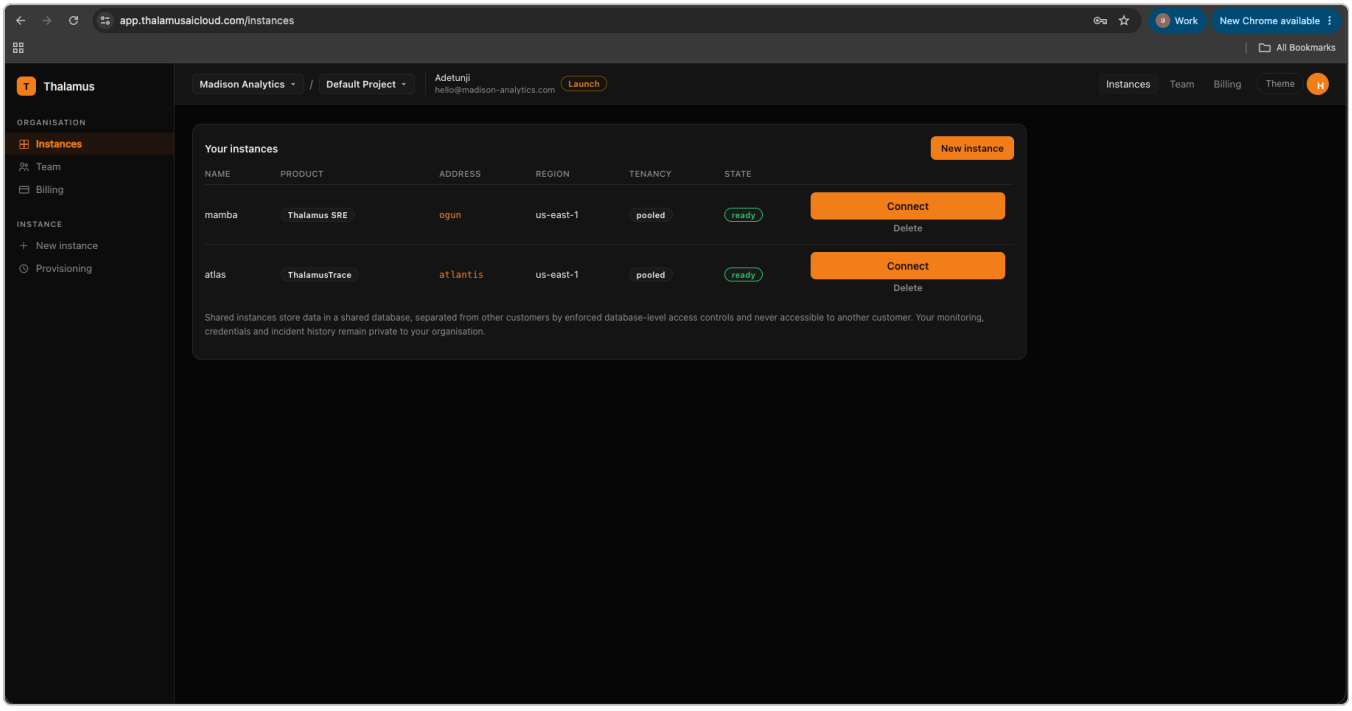

FIGURE 2 — THALAMUS AI CLOUD

Both products, provisioned and ready to connect. A bounded proof-of-value can be running against real telemetry the same day — pooled tenancy for a trial, dedicated where a shared database is not acceptable. Because ingest is OTLP-native, the instrumentation you add is the OpenTelemetry standard rather than a vendor agent, which is also what makes the data portable back out if the answer is no.

07 — GETTING THERE

How Thalamus Advisory helps

ENGAGEMENT	WHAT IT PRODUCES
Diagnostic Throughput Benchmark 4 weeks	Your current time-to-hypothesis and hypothesis-overturn rate on real incidents, measured against the same incidents run through a parallel fan-out. A defensible before-and-after rather than a claim.
Model Dependency Review 3 weeks	Every model endpoint in a critical path, its latency and error profile as an SLI, and a drift-evaluation harness for behavioural change that availability monitoring cannot see.
Agent Spend Controls 2 weeks	Ceilings, budgets and rate-of-change alerting sized to your incident distribution, so an incident storm cannot become a billing event.
Multi-Tenant SLO Design 4 weeks	Per-tenant objectives and attribution evidence strong enough to support a commercial conversation about noisy-neighbour impact.

THE COUNTER-ARGUMENT, MADE HONESTLY

If you are Google, Amazon, Meta or Netflix, you have built most of this internally and built it well. The claim we will defend is narrow: diagnostic throughput through parallel probing, and reduced confidently-wrong conclusions through independent verification. That is a throughput argument, not a capability argument.

We would not suggest replacing your observability stack, and anyone who does is not looking at what you already run. The interesting question for you is whether your investigation phase is parallel — and for most organisations, however sophisticated, it still is not.

ONE NUMBER

Six. Senior engineers on a typical major-incident bridge at a large technology company, working through hypotheses in sequence.

The specialists probe in parallel and return evidence with strength scores. The arithmetic of that difference is the entire argument, and it does not require your monitoring to be inadequate.

Where to start on Monday

Take your last major incident. Write down the hypotheses that were considered, in order, with the time each one took to eliminate.

Then ask: how many of those could have been tested simultaneously? That number, multiplied by the people on the call, is what parallel investigation is worth to you.

Thalamus Advisory

Enterprise resilience for the age of autonomous systems. We help organisations build operational practice that stays trustworthy when the systems making decisions are no longer entirely human.

THE RESILIENCE NEWSLETTER · ISSUE 06 · SEPTEMBER 2026
Every figure we print carries its assumptions. Where we model, we say so.
Written for engineering leaders. Forward it to one.