The Resilience Newsletter

THALAMUS ADVISORY

ISSUE 03 SEPTEMBER 2026 READING TIME · 10 MIN SERIES · SECTOR ANALYSIS

SECTOR ANALYSIS · HEALTHCARE & LIFE SCIENCES

When Downtime Is Clinical

A retailer's outage costs a basket. A hospital's outage diverts an ambulance. Healthcare is the one sector where the right answer to a low-confidence diagnosis is to refuse to act.

THALAMUS ADVISORY • ENTERPRISE RESILIENCE PRACTICE

01 — THE SIGNAL

Paper charting is the disaster recovery plan

Ask a hospital CIO what happens when the EHR degrades and you will not hear about failover. You will hear about downtime boxes — printed forms, kept in a cupboard on every floor, for writing orders by hand. That is the recovery posture for a $4.9 trillion sector, and it exists because the alternative is worse.

US national health expenditure is approaching 17.6% of GDP. The estate that runs it is unusually heterogeneous: HL7v2 interfaces written decades ago sitting beside modern FHIR APIs, claims adjudication on mainframes, pharmacy benefit platforms answering in real time while a pharmacist waits at a counter. All of it load-bearing. None of it replaceable on a convenient schedule.

02 — THE MECHANISM

Multi-hop paths with no obvious owner

The characteristic healthcare failure is not a service falling over. It is a latency regression three hops deep in an eligibility check that surfaces as a provider-facing timeout with no obvious owner.

A prior-authorization request crosses an integration engine, a payer API, a benefits platform and back. When it slows, four teams are paged: the portal team because the symptom is theirs, the integration team because they are always paged, the payer-interface team because it might be upstream, and the platform team because it might be infrastructure. Three of those four spend the next hour proving it is not them.

> *Four teams paged for one fault is not an alerting problem. It is a correlation problem wearing an alerting problem's clothes.*

Two properties make this worse in healthcare than elsewhere. Interfaces are often organisational boundaries — the other side belongs to a payer, a clearinghouse, a lab network — so half the trace is outside your estate. And the systems are validated: changing application code to add instrumentation triggers a revalidation cost that means, in practice, it does not happen.

03 — THE COST

Where it actually lands

SURFACE	HOW IT PRESENTS
Clinical diversion	EHR degradation forces paper charting and ambulance diversion. The cost denominator includes patient outcomes, which is why operational-only figures always understate it.
Reportable events	HIPAA and HITECH make an availability incident simultaneously a potential disclosure event. The investigation has two audiences, one of them a regulator.
Pharmacy queue	Adjudication latency is felt locally and immediately — a person waiting at a counter — and is reputationally acute in a way that back-office downtime is not.
Seasonal concentration	Flu season, vaccination campaigns and January formulary changes create predictable annual stress the estate must absorb on a known date.
Validated-system paralysis	"Never touch it" and "keep it reliable" are in direct tension, and observability is often the only intervention that does not trigger revalidation.

04 — THE AI STAKES

Automation meets a clinical blast radius

Healthcare is adopting AI quickly — ambient documentation, coding assistance, triage support, imaging pre-reads. Each addition places a probabilistic component in a path where the consequence of being wrong is clinical.

For reliability specifically, the risk is automated remediation. An agent that confidently restarts the wrong service during an EHR degradation has not saved an hour; it has extended the outage and contaminated the evidence. In a sector where the failure mode is a patient safety event, an automation that acts on a weakly-supported conclusion is worse than no automation at all.

There is also a data-governance dimension that healthcare feels more sharply than most. Traces and logs in a clinical estate contain PHI. The moment an investigating agent reads those logs, that content is in a model context window — possibly with a third party, possibly retained. The governance boundary moves without anyone having decided to move it.

05 — THE REMEDY

Defensibility before speed

The healthcare objective is not the fastest possible diagnosis. It is a diagnosis that can be defended afterwards — to a safety committee, to a regulator, to a board.

- **Correlate before paging.** One fault, one problem, one named root-cause entity, an explicit affected-entity set. The four-teams-for-one-fault pattern is solved by correlation, not by better runbooks.
- **Instrument without touching application logic.** OTLP collection is a materially easier validation conversation than a vendor agent injected into a validated system.
- **Make escalation the expected outcome for clinical paths.** Anything touching an EHR, a medication path or a controlled substance should require human approval by policy, not by convention.
- **Redact at ingestion, not at query.** PHI should never reach an agent context. Redaction after the fact is a breach that has already happened.
- **Generate the postmortem as evidence.** You have to produce the narrative anyway. Producing it as a by-product of the investigation, with probes attached, is strictly cheaper than reconstructing it a week later.

06 — IN PRACTICE

Hosted, or customized for your estate

Both deployment models are available, and in healthcare the choice is usually decided by where PHI is allowed to live rather than by preference.

Hosted on Thalamus AI Cloud

ThalamusTrace and Thalamus SRE run as managed instances, provisioned per region and tenancy model from one console. For payer analytics, revenue cycle, scheduling and other estates where the data can be de-identified or does not contain PHI, this is the fastest route — an instance in

minutes rather than an infrastructure programme. **Dedicated tenancy** gives a separate database and a dedicated AWS account where a shared store is not acceptable to your privacy office.

Customized for your enterprise

Where PHI must not leave your boundary, the same platform deploys self-hosted inside your VPC or data centre. This is the common outcome for provider organisations and for anything touching the EHR. Customization typically covers three things: SLO definitions expressed in clinical terms rather than infrastructure ones (adjudication p95, order-entry availability, interface-engine throughput); guardrail policy mapped to your change-control tiers so clinical systems escalate by default; and redaction rules tuned to your PHI classification before telemetry reaches any agent.

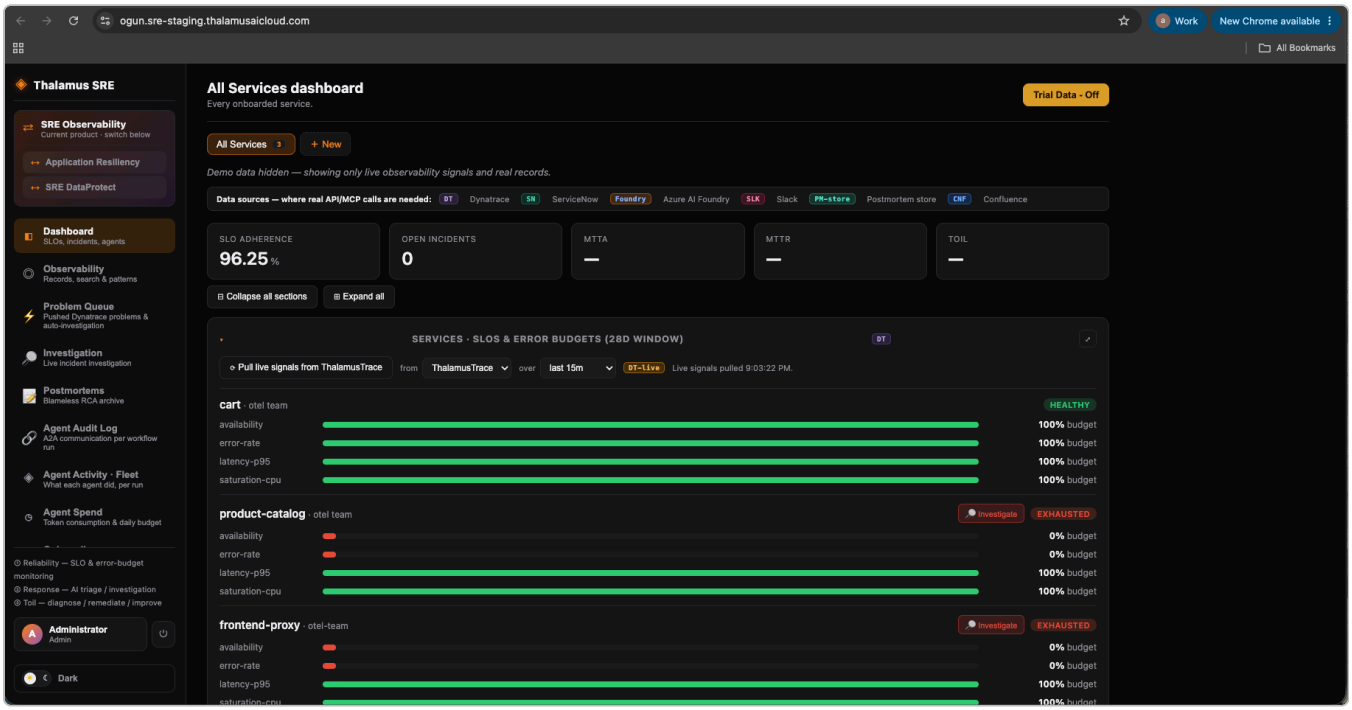

FIGURE 1 — THALAMUS SRE · SLO & ERROR BUDGETS

Reliability expressed as budget, not uptime. Each onboarded service carries availability, error-rate, latency-p95 and saturation objectives over a 28-day window, with the remaining error budget shown per objective. Two services here have exhausted theirs while their latency budgets remain intact — the distinction that tells you *which* obligation is failing, which is the question a clinical safety review will ask.

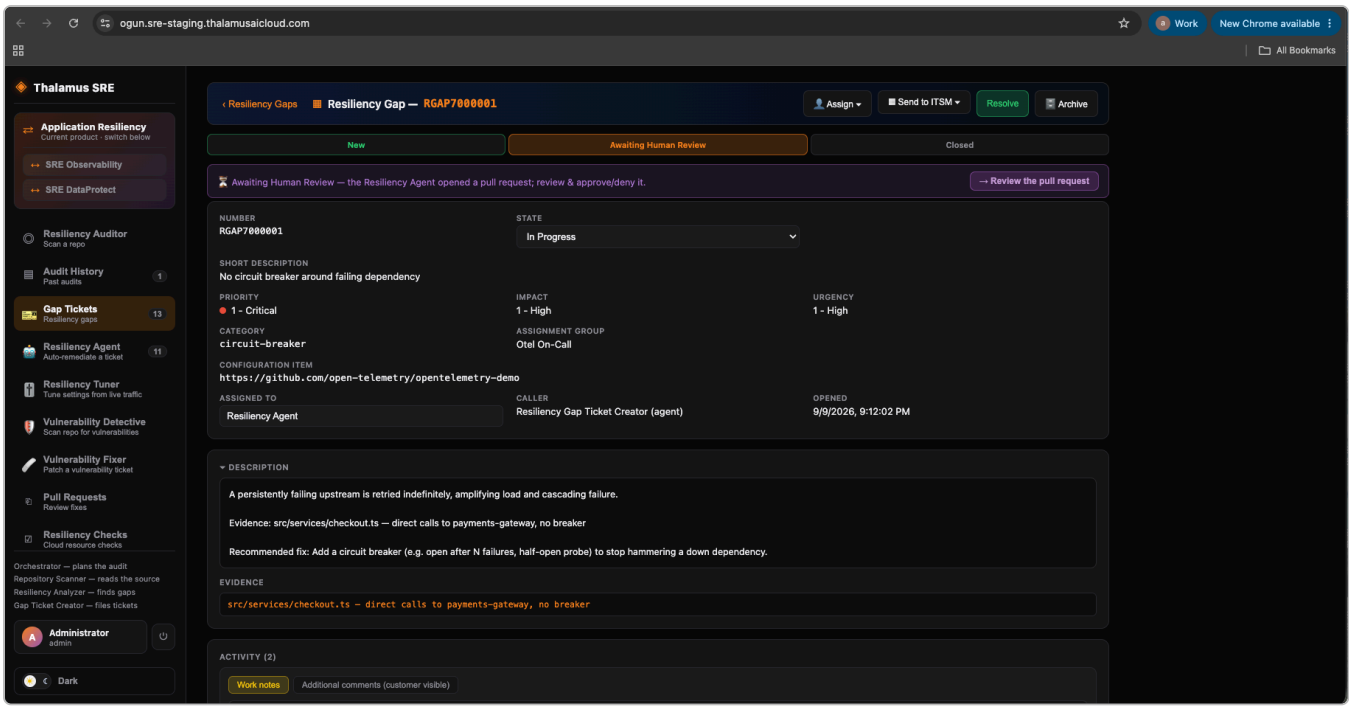

FIGURE 2 — THALAMUS SRE · HUMAN-IN-THE-LOOP

The agent stops at the boundary. A fix was diagnosed, written and raised as a pull request — and then held in *Awaiting Human Review* rather than applied. For clinical and medication paths this is the correct default: the automation does the work, a human makes the decision, and the evidence supporting it travels with the ticket.

07 — GETTING THERE

How Thalamus Advisory helps

ENGAGEMENT	WHAT IT PRODUCES
Clinical Path Mapping 3–4 weeks	Which application paths have clinical consequence, what their real dependency graph looks like, and where the trace currently breaks at an organisational boundary.
PHI Telemetry Review 2–3 weeks	Where PHI appears in traces and logs, which agents and vendors can read it, and a redaction design that preserves correlation while removing the identifiers.
Validated-Estate Instrumentation 6–8 weeks	OTLP coverage across validated systems without application change, with the documentation your validation process requires produced alongside.
Guarded Rollout 8–12 weeks	Self-hosted deployment, SLOs in clinical terms, guardrails mapped to your change tiers, and postmortems routed into your existing safety review.

THE COUNTER-ARGUMENT, MADE HONESTLY

Healthcare estates are where automated remediation is *least* appropriate, and any vendor selling you self-healing clinical systems should be treated with suspicion — us included. The realistic value here is compressed diagnosis and defensible evidence, with a human making every decision that touches care.

If your organisation wants autonomous remediation across clinical paths, we are the wrong partner. We will configure the guardrails to prevent exactly that.

ONE NUMBER

Four. Teams typically paged for a single upstream fault in a multi-hop eligibility or adjudication path. Three of them spend the first hour establishing that it is not them.

Correlation does not make the incident shorter by making engineers faster. It makes it shorter by not involving three-quarters of them in the first place.

Where to start on Monday

Take your last five multi-service incidents. For each, count how many teams were paged and how many were causally involved.

That ratio is your correlation debt. In healthcare estates we have measured, it is rarely better than three to one — and it is paid in the most expensive engineering hours you have.

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 03 · SEPTEMBER 2026

Every figure we print carries its assumptions. Where we model, we say so.

Written for engineering leaders. Forward it to one.