

SECTOR ANALYSIS • ENERGY & UTILITIES

Reliability as a Federal Obligation

Supply must match demand continuously, IT failures propagate into physical systems, and a regulator counts. In energy, the most dangerous telemetry signal is not an alarm — it is silence.

THALAMUS ADVISORY • ENTERPRISE RESILIENCE PRACTICE

01 — THE SIGNAL

No buffer, and a regulator counting

The grid is a real-time physical system with no storage in the middle. Generation and load must balance continuously, and there is no queue to absorb a mistake. That physical fact shapes everything about how the sector treats software.

US energy expenditure runs to roughly \$1.5 trillion, with about \$500 billion in electric and gas utility revenue. NERC CIP makes reliability federally enforceable with financial penalties attached — one of the few places where an engineering failure has a statutory consequence rather than a commercial one.

02 — THE MECHANISM

Silence that looks like health

The characteristic energy failure is not an error. It is an absence.

A remote asset stops reporting. A SCADA historian stops writing. An AMI collector silently drops a percentage of meter reads. Nothing errors, no threshold is crossed, no latency degrades — because there are no requests to be slow. Conventional monitoring, built to detect *too much*, is structurally incapable of detecting *too little*.

Threshold alerting cannot fire on nothing. In a sector where the most common failure is a feed that stopped, that is not a gap — it is the whole problem.

The second pattern is the OT/IT boundary. It is simultaneously a security perimeter and an observability discontinuity. Telemetry stops at the DMZ by design, so the correlation between an IT fault and its operational consequence has to be made by a human, from memory, under pressure.

Multi-jurisdiction operators add a third. Six utilities, six regulators, six sets of performance metrics — and aggregate reporting that hides a single jurisdiction’s failure inside a corporate average that looks acceptable.

03 – THE COST

Where it actually lands

SURFACE	HOW IT PRESENTS
NERC CIP penalties	Financial and public. Reliability is a compliance obligation with a named enforcer, not an internal quality target.
Market settlement	RTO participation has financially-settled deadlines. A late or wrong submission has an immediate, quantifiable cost.
Storm response	Outage management is stressed exactly when it is most needed, and restoration performance is reviewed publicly afterwards.
Silent data gaps	Missing meter reads or historian writes are discovered in billing or in an audit, months later, when remediation is expensive and the evidence is gone.
Jurisdictional masking	One utility’s degradation averaged into a corporate figure that satisfies nobody once a state commission asks the question directly.

04 – THE AI STAKES

Critical infrastructure as a target

Energy has been an explicit target of state-aligned intrusion for years. What has changed is cost: reconnaissance, lateral movement and the patience to run a long campaign have all become substantially cheaper to automate.

That matters more here than in most sectors because dwell time is the adversary’s main asset and detection latency is the defender’s main weakness. An intrusion that manifests as *slightly fewer* telemetry events, or a historian that occasionally misses a write, looks exactly like the ordinary

silent failures the sector already tolerates. The attack hides inside the noise the estate has learned to ignore.

Meanwhile the generation mix is changing faster than the operational practice around it. Renewables are weather-coupled and distributed; DER management and battery dispatch are new system classes arriving faster than anyone can build monitoring conventions for them.

05 — THE REMEDY

Make absence a breach

- **Throughput SLOs with a floor.** Expected events per window, where falling below is a breach. This single change converts the sector's dominant failure mode from invisible to alerting.
- **Baseline-relative detection.** Compare each event class against its own prior baseline rather than a static threshold, so a class that quietly disappears is a signal.
- **Observe across the OT/IT boundary without weakening it.** Collect on the IT side of the DMZ and correlate with OT events, rather than maintaining two disconnected monitoring estates and a human in the middle.
- **Scope objectives per operating entity.** Each regulated utility measured against its own commitments, not a corporate average.
- **Generate restoration evidence automatically.** Rate cases and regulatory reviews need the narrative regardless; produce it as a by-product.

06 — IN PRACTICE

Hosted, or customized for your estate

Hosted on Thalamus AI Cloud

Appropriate for corporate IT, customer information and billing, market-facing analytics and the enterprise estate generally — anywhere outside the CIP-scoped perimeter. Managed instances with dedicated tenancy remove the infrastructure lead time, which matters for a sector where standing up a new internal platform is itself a multi-quarter programme.

Customized for your enterprise

For anything inside the CIP boundary, self-hosted is the only viable model, and the platform is designed for it: the whole stack runs inside your own network with no egress to a vendor SaaS. Customization for utilities usually means throughput SLO templates per asset class with floors derived from expected reporting intervals; per-entity tenancy so each regulated utility is measured

separately; collectors positioned on the IT side of the DMZ with correlation rules mapping IT faults to operational consequence; and postmortem output shaped to the restoration-performance format your commissions actually ask for.

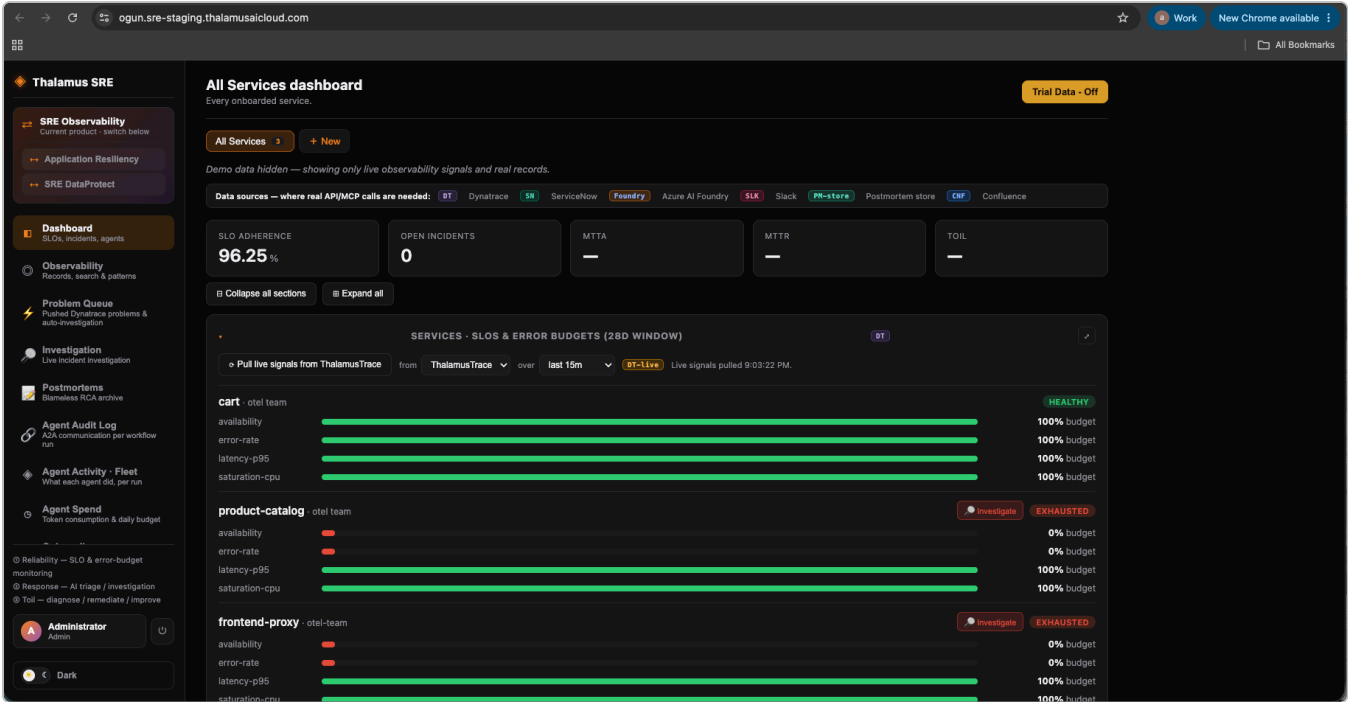


FIGURE 1 – THALAMUS SRE • SLO & ERROR BUDGETS

Objectives that can express a floor. Availability, error-rate, latency and saturation objectives tracked per service over a 28-day window with remaining error budget. The same mechanism carries a *throughput* objective — expected events per window — which is how a feed that has silently stopped becomes a breach rather than a quiet gap discovered in next month's billing reconciliation.

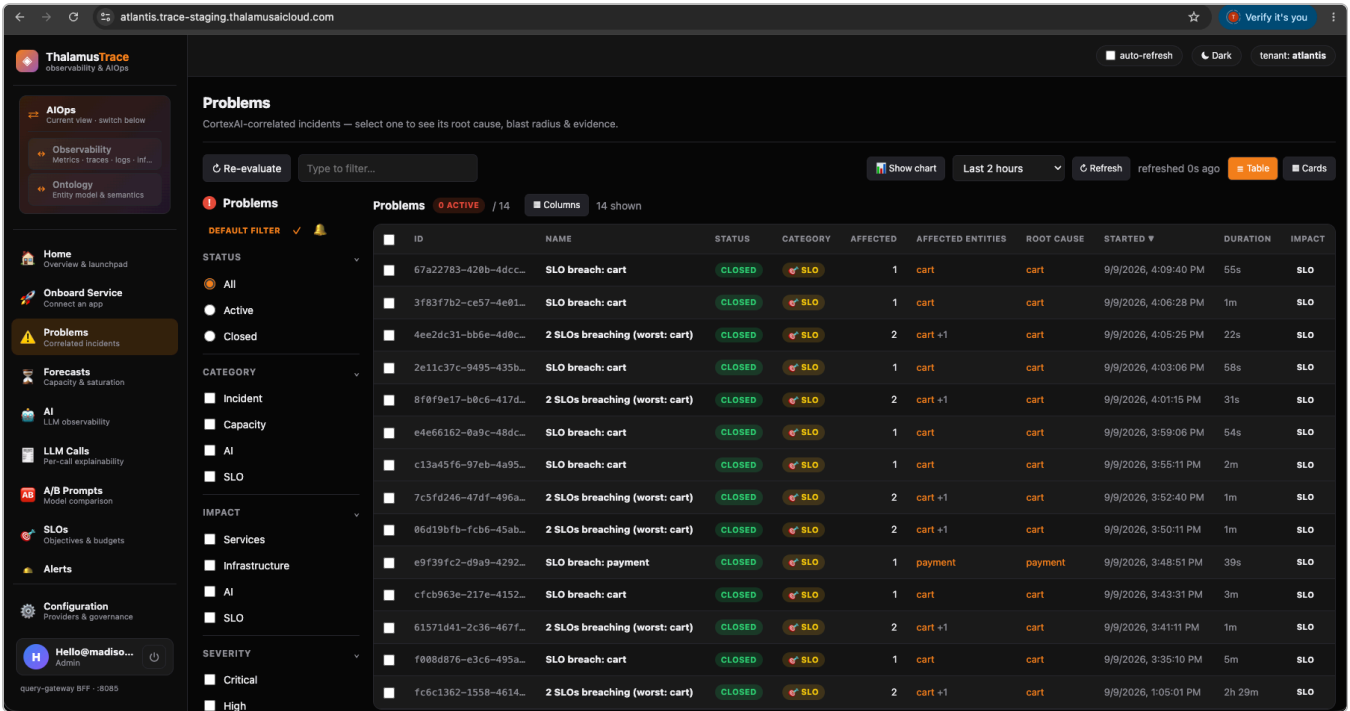


FIGURE 2 – THALAMUSTRACE • CORRELATED PROBLEMS

One cause, named, with its blast radius. Concurrent breaches collapse into a single problem carrying a root-cause entity, the affected-entity set, and the onset timestamp. For a multi-jurisdiction operator the same mechanism scopes per entity, so one utility's degradation surfaces as its own problem rather than disappearing into a corporate average.

07 – GETTING THERE

How Thalamus Advisory helps

ENGAGEMENT

WHAT IT PRODUCES

Silent Failure Audit
3–4 weeks

Every critical data flow, whether you would know within an hour if it stopped, and the throughput floors that would make each one alert. For most utilities the honest answer for most flows is currently no.

OT/IT Boundary
Observability
6–8 weeks

Collection architecture that spans the DMZ without weakening it, reviewed against your CIP controls before anything is deployed.

Per-Entity SLO Design
4 weeks

Objectives scoped per regulated utility and mapped to the commitments each commission actually measures.

Restoration Evidence
3 weeks

Postmortem and timeline output shaped to your regulatory review format, generated from the incident record rather than reconstructed.

THE COUNTER-ARGUMENT, MADE HONESTLY

CIP-scoped environments have change control for good reasons, and introducing any new component into them is a serious undertaking that should not be rushed by a vendor timeline.

The pragmatic sequence is to start outside the perimeter — corporate IT, customer systems, market analytics — prove the throughput-SLO approach where the change cost is low, and only then make the case for the regulated estate with evidence in hand. Anyone proposing to start inside the CIP boundary has not worked in this sector.



ONE NUMBER

Zero. Alerts generated by a data feed that stops delivering entirely, under threshold-based monitoring.

No errors, no latency, no saturation — because there are no requests. The estate reports perfect health right up until someone reconciles the billing.

Where to start on Monday

List your ten most critical data flows. For each, answer one question: **would we know within an hour if this stopped?**

Every “no” is a throughput SLO waiting to be written, and it is the cheapest reliability improvement available to a utility this year.

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 07 · SEPTEMBER 2026

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

Written for engineering leaders. Forward it to one.