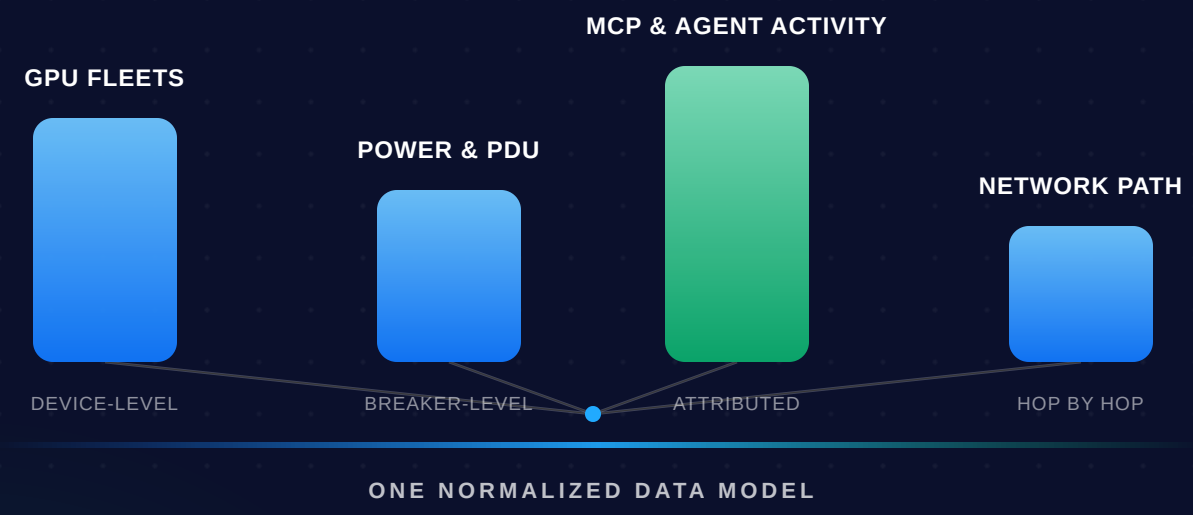
parlon

AI INFRASTRUCTURE OBSERVABILITY · CORPORATE OVERVIEW

The infrastructure layer every AI workload runs on.

GPU fleets, the power feeding them, the agents now operating inside them, and the network path between all three. Seen, correlated, and governed in one platform. Legacy observability was never built to reach any of it.



~70%
Reduction in alert noise with Alert Auto-Tune™

Hours
Time to first value, synthetics live same day

250+
Legacy vendor certifications carried over, not re-earned

3–8×
Lower three-year TCO than legacy platforms

WHY NOW

Infrastructure changed. The tools watching it did not.

Three pressures are converging on infrastructure teams at once. None of them is waiting for the other two, and together they define the window for modernizing observability now rather than at the next renewal.

01

AI created new blind spots

A thermal-throttled GPU, a rack approaching breaker capacity, an agent reading production telemetry without attribution, a congested east-west path. None of these appear in an application trace, and none of them wait for a dashboard to be reconciled.

47.7%
of enterprises already run AI training or inference workloads

35%
believe their tools are ready to observe them

EMA Network Management Megatrends 2026, n=352

02

Buyers are already moving

This is not a hypothetical replacement cycle. It is already open, and it is accelerating year over year as teams lose confidence in tools built for a different era of infrastructure.

73%
likely to replace an observability tool within two years

26% → 32.7%
reporting “very likely,” 2024 to 2026

EMA Network Management Megatrends 2026

03

Incumbent economics are breaking down

Renewal costs are climbing for reasons that have nothing to do with value delivered: private-equity pricing plays, forced bundling, and per-device economics that punish scale.

200–300%
renewal cost increases reported after PE acquisition

$200/dev
list pricing at some incumbents, before services and headcount

SolarWinds pricing analysis, 2025; IBM SevOne internal pricing, 2023

These are not features incumbents forgot to build. They are architectural limits set before cloud and AI existed, and closing them now requires a ground-up rebuild, not a roadmap item.

THE PLATFORM

One architecture. Four structural bets.

Every solution in this document, including the AI infrastructure layer, runs on the same foundation: built together from the ground up, not stitched together after the fact.

NORM

Normalization at ingest

Every source, any vendor, any location, is mapped into a unified schema the moment it enters Parlon. Vendor-specific detail is preserved. Correlation is immediate, not reconstructed after an incident.

SYN

Synthetics and telemetry, unified

Active synthetic testing and continuous telemetry collection run natively in one system, including LLM-aware workflow checks. No modules to stitch together. No swivel chair.

TUNE

Alert Auto-Tune™

A human-approved workflow that recommends thresholds with the evidence behind them, learns system behavior, and correlates related alerts. When Parlon alerts, it is worth acting on. Cuts alert noise by approximately 70%.

DEPLOY

Deploy anywhere

SaaS, on-premises, hybrid, or fully air-gapped, with customer-controlled boundaries. Every deployment mode is feature-equivalent and in production today, including HIPAA-aligned, air-gapped environments.

THE MODEL

Separate collection and later reconciliation is a data problem, not a tooling gap.

Adding another point tool does not fix a data model built to collect first and understand later. Parlon collects with context, normalizes at ingest, validates actively, and interprets coherently, so operators decide with complete context instead of reconstructing it under time pressure.

30–50%
Reduction in MTTR across customer deployments

<4 hrs
Time to first value; synthetics live same day

200+
Vendors mapped into the Normalization Engine out of the box

0
Modules to stitch together across synthetics and telemetry

PRIMARY ROUTE

Observability built for AI infrastructure.

AI workloads run on physical and operational layers most observability platforms do not reach: the GPU fleet, the power feeding it, the agents now operating inside it, and the network moving data between all three. Most platforms approach AI observability from the software side and stop at the metric. Parlon starts at the silicon and the fabric.

GPU Fleets

Fleet-wide GPU health, utilization, and waste, down to a single device. A fleet averaging a healthy-looking 45% utilization can still be half maxed out and half sitting idle. Catches thermal throttling and ECC/XID errors that a cluster average hides.

60–120s early warning before frameworks like PyTorch DDP notice a node failure

Power & PDU

Live rack power, thermal conditions, and cost, down to a single PDU. Ranks the rack closest to its breaker limit before it trips, and flags whether the A/B feed can carry the load if one side fails.

$1M+ per hour: cost of downtime reported by 41% of enterprises (ITIC, 2024)

MCP & Agent Activity

Every AI agent read tied to a known identity, human or agent, and audit-logged. Governance is included in the platform, not sold as a separate add-on module.

31% of organizations prioritize MCP support for agentic access (EMA 2026)

Network Path

East-west traffic and fabric health across the AI cluster, hop by hop, running on the same normalized data model as GPU, power, and agent telemetry above it.

One data model. No separate console for the fabric layer.

GPU Fleets and Power & PDU each ship today with a dedicated solution guide and a live console. MCP & Agent Activity governance is included in the platform now. Network Path runs on the same data model as the other three. This is the ground legacy platforms don't own, because almost none of them treat GPU fabric, power, and agent governance as first-class products.

WHERE TO START

Three routes in. One platform underneath.

Start wherever the problem is today. Nothing here is a module, a bolt-on, or a second console; every route runs on the same normalized data model described on the previous pages.

PRIMARY

AI Infrastructure Observability

GPU fleets, power and PDU capacity, agent activity, and the network path between them: the parts of the AI buildout that do not show up in an application trace, correlated on one data model.

Observability Modernization

The direct replacement path off legacy network and infrastructure monitoring platforms, with a 90-day proof of value built in, not bolted on after the sale.

Active Synthetic Monitoring

Outside-in and inside-out synthetic testing, unified with telemetry in one system instead of two panes of glass. Ten-plus check types, from HTTP/S to LLM-aware MCP checks.

WHO WE SERVE

Built for the teams that can't tolerate blind spots.

Platform + SRE Teams

Reliability, MTTR, end-to-end visibility, and AI workload readiness across a hybrid estate.

Enterprise Businesses

Tool consolidation, compliance, and cost control across a multi-vendor infrastructure footprint.

Managed Service Providers

Multi-tenant noise reduction, white-labeling, and tenant isolation from a single deployment.

Regulated Industries

Traceability and auditability, with air-gapped and HIPAA-aligned deployment options.

BY INDUSTRY

Built for environments that can't tolerate downtime or exposure.

Healthcare

HIPAA-aligned, air-gapped deployment for networks with 1,000+ locations.

Financial Services

Low-latency visibility for trading, banking, and payments, with on-premise and air-gapped compliance support.

Government & Defense

Classified-environment deployment with zero external connectivity.

Telecom & MSP

Multi-tenant deployment with tenant isolation and white-label options.

WHAT COMES NEXT

See the infrastructure first. Govern the AI that acts on it next.

The environments that matter most, regulated, sovereign, and air-gapped, cannot let telemetry, credentials, or write access leave the boundary. They need AI's help the most and cannot get it by loosening that boundary. Governance has to be part of the substrate, not a policy layered on top.

Live today

- ✓ Every read tied to a known identity, human or agent
- ✓ Scoped permissions and a full audit trail
- ✓ Usage metering and quotas to control AI spend
- ✓ Bring your own model. Nothing leaves the boundary

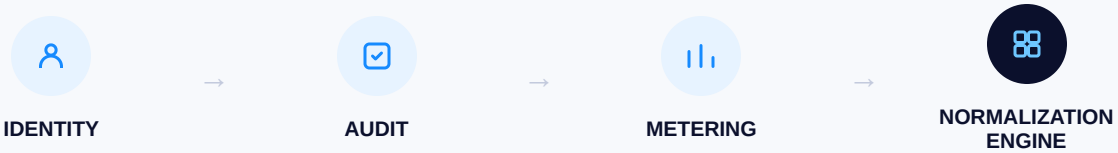
Building next

- → Write actions inheriting the same identity model
- → Required approvals before execution
- → A record of what was touched, and rollback
- → Policy that persists across model generations

WHY IT'S STRUCTURAL

Not a policy add-on. A property of the data model.

Because every read and write already flows through the same Normalization Engine that powers alerting and correlation, identity and audit attach at the same layer instead of a separate compliance tool bolted on afterward. That is what makes governance credible in an air-gapped or fully sovereign environment: nothing has to leave the boundary to be observed, attributed, or audited.



The same pipeline that normalizes telemetry for alerting also carries identity, audit, and metering, inside the boundary, by construction.

EVIDENCE, WITH THE RECORD ATTACHED

One deployment. What it does, and does not, establish.

ENTERPRISE HEALTHCARE · AIR-GAPPED & HIPAA-COMPLIANT

A national health system, consolidated into one deployment.

A multi-vendor stack across more than 1,000 clinic locations, datacenters, and remote branches, consolidated into one data model, one console, one contract. Synthetic tests were active within hours, inside a fully air-gapped environment with strict data-residency requirements.

~50%	2–3 → <0.5	2 → 1	Hours
lower observability spend than the prior multi-tool footprint	FTE of operational effort to administer	legacy tools replaced by one platform, one contract	from deployment to active synthetic tests

> **“We found Parlon's capabilities beyond parity with the legacy vendors, and the simplicity of deployment and the cost were a significant value in themselves.”**
>
> Network Infrastructure Lead, enterprise healthcare provider

Where observed: one production enterprise healthcare deployment. **How measured:** against the documented cost and staffing of the replaced multi-tool footprint. **Configuration:** on-premises, fully air-gapped, HIPAA-compliant, device monitoring and synthetic path testing. **What it does not establish:** a universal payback period, or a result in a cloud-first or GPU-dense environment.

THE SAME PATTERN, MODELED

Three-year TCO, same environment.



Deploys in hours. Proves value in 30 to 60 days, running alongside what you already operate. Nothing is decommissioned to find out, and every outcome, including a null result, counts as a valid conclusion.

ABOUT PARLON

Built by people who know the incumbents from the inside.

Parlon is the infrastructure layer every AI workload runs on: GPU fleets, the power feeding them, the AI agents now operating inside them, and the network path between all three, normalized, correlated, and governed in one platform. Founded in 2026 by three operators who sold, deployed, and architected the observability platforms enterprises are now replacing, at SevOne, IBM, Pliant, CA, Akamai, and Verizon.

3	30+	4	MA
Co-founders	Years in enterprise observability	Wedges, one platform	Headquartered in Cambridge

MG

Matt Goldberg

CEO + Co-founder

Led global strategic solutions at a major network performance management company and co-founded an IT automation platform acquired by IBM.

CR

Chris Rohter

CRO + Co-founder

Career in enterprise go-to-market across Akamai, SevOne, AppNeta, and Pliant, spanning sales, marketing, and alliances. Knows the enterprise observability buyer from years spent serving them.

SN

Soumo Nandi

Chief Architect + Co-founder

Path ran from the Verizon Wireless Network Leadership Development Program through solutions architecture at observability and DNS infrastructure companies, including SevOne and Pliant.

PRINCIPLES

The principles behind how we build.

— **Architecture over feature lists.** The right foundation makes everything else possible.

— **No lock-in, ever.** Open APIs, streaming export, vendor-agnostic collection.

— **Time to value is a product decision.** First value on day one is the expectation, not the exception.

— **Signal over noise is the whole point.** When Parlon alerts, it matters.

parlon

See your infrastructure the way Parlon does.

A bounded proof, run alongside what you already have, on one question that matters. No estate-wide commitment, and no assumption about the result.

REQUEST A DEMO →

parlon.io · **hello@parlon.io**

One Mifflin Place, Suite 400 – #610, Cambridge, MA 02138 · © 2026 Parlon, Inc. All rights reserved.