

— A ROLE, NOT A PERSON

Takes production incidents **end-to-end.**

The SRE Orchestrator turns noisy alerts and tickets into structured investigations — forms and tests hypotheses, finds root cause across every domain, remediates under your guardrails, verifies recovery, and learns. **Always-on across the whole estate, with no personal inbox.**

Detect

Diagnose

Remediate

Verify

— THE CHALLENGE

On-call is a role assigned to a person

Every production estate generates more signal than any team can read. The response is to put a human in front of it — paged at 3am, triaging alerts that mostly do not matter, reconstructing context that already exists somewhere in the logs. It does not scale, and it is not where engineering time should go.



Noise arrives faster than judgement

A single underlying fault produces dozens of alerts across monitoring, ticketing, and chat. Someone has to work out that they are one problem before anyone can start solving it.



Investigation is the expensive part

The fix is often minutes of work. Establishing what broke, why, and whether it is safe to act consumes hours — repeatedly, by whoever happens to be on call.



Context is scattered across domains

A network symptom is often a storage problem. Tracing the causal chain means crossing tools, teams, and dashboards that were never designed to answer one question together.



Knowledge leaves with the engineer

Each incident teaches the organisation something, and most of it stays in one person's head. The next occurrence starts from close to zero.

From raw signal to one structured investigation

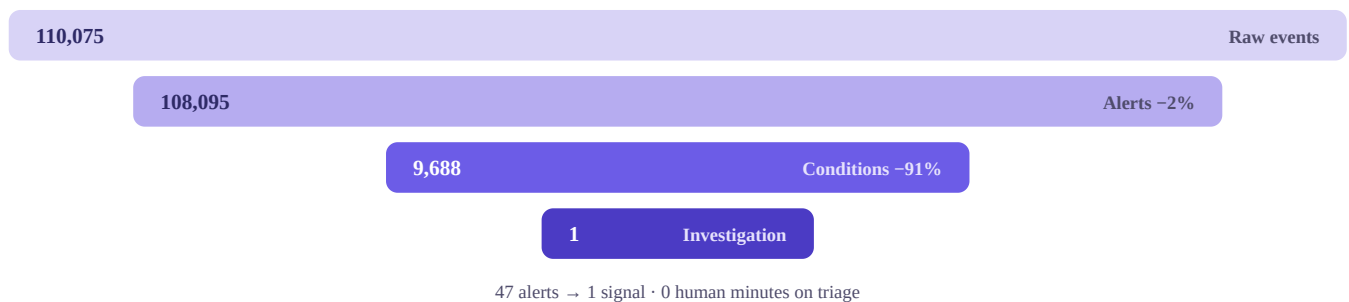


FIGURE — NOISE REDUCTION IN A 30-DAY WINDOW

THE SHIFT

Reliability should be an always-on organisational function — not a rota, an inbox, and a person who happens to be awake.

— THE INCIDENT LIFECYCLE

Detect. Investigate. Diagnose. Remediate. Verify. Learn.

The orchestrator runs the entire loop as one coherent workflow — done in minutes, at estate scale, without a human carrying the triage.

One coherent workflow — the toil a human on-call engineer would otherwise carry

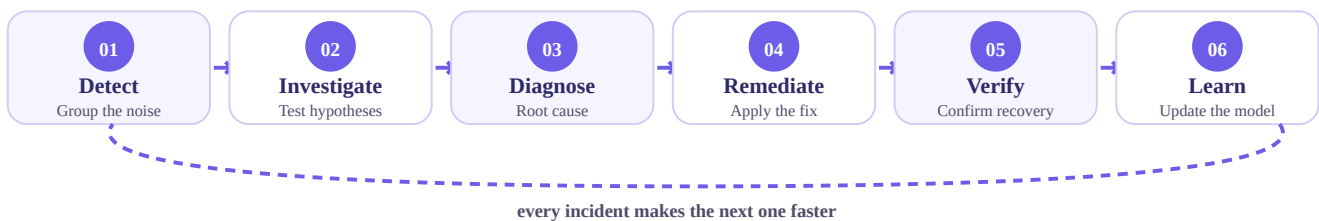
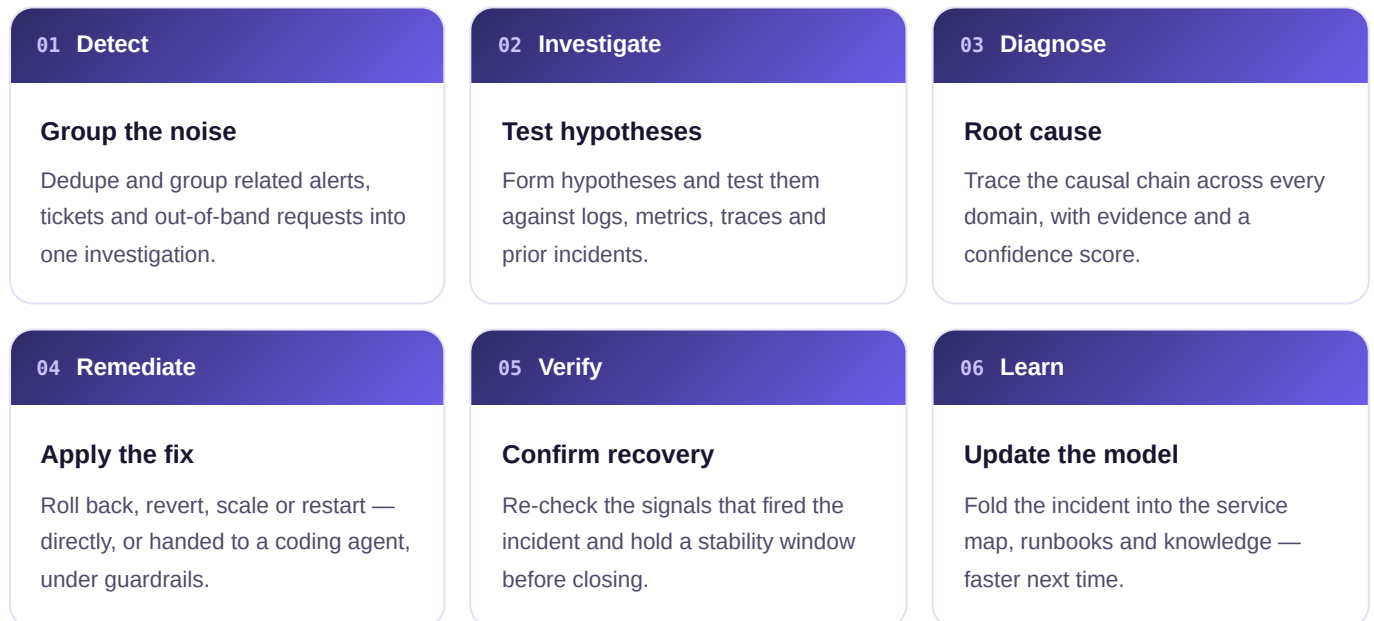


FIGURE — THE INCIDENT LIFECYCLE



— INVESTIGATE & DIAGNOSE

Evidence-backed root cause, **not guesses**

Specialist Coworkers investigate in parallel across code, infrastructure and telemetry — proposing and ruling out hypotheses against the evidence, then synthesising one explainable root cause.

- Parallel **multi-agent investigation** across domains
- Hypotheses tested against logs, metrics, traces and prior incidents
- **Cross-domain RCA** — a network symptom that is really a storage problem
- Explainable, evidence-linked summary with a **confidence score**

LUMI — INVESTIGATE

```
ask "why is payments-api p99 spiking in eu-west?"
```

```
# parallel agents build a live model & test hypotheses
investigate net · compute · storage · db
  → storage-fleet IO saturation 92%
  → replica lag cascading to payments-api
```

```
remediate --guardrails
  → rebalance storage-fleet (+3)
  → rollback #2291 · verify p99 recovered
```

```
✓ resolved 3m12s · stable 15m · 0 human minutes
```

INVESTIGATING · PAYMENTS-API P99 · 3
HYPOTHESES

Storage IO saturation → replica lag	92%
Deploy regression b7f3a2c	41%
DB connection pool exhaustion	18%

✓ Root cause: storage-fleet IO saturation · evidence:
6 signals across net · storage · db

— GOVERNED ACTIONS

It doesn't just tell you — it fixes it

The orchestrator executes real remediation through your tools, inside the guardrails and approval policies you set. Nothing runs outside the lines you draw.

Silence & route alerts

Scale & rebalance

Restart service

Post to Slack / ticket

Roll back deploy

Revert commit / flag

Open a PR

Run GitHub workflow

- auto within guardrails
- requires approval
- steer every action from the Workbench

— CENTRALIZED CONTROL

Autonomy you can govern

The SRE Orchestrator is the control layer over Lumi's autonomous operations. Every capability below exists so that AI action stays bounded, attributable, and reversible — autonomy with a hand on it.



Autonomy management

Set how far the orchestrator may act on its own, per service, per action class, and per environment.



Continuous monitoring

Estate-wide watch across every connected signal, with pre-alert investigation before a page is raised.



AI-driven remediation

Decisions become governed actions through your existing tools — rollback, scale, restart, revert, or a pull request.



Execution guardrails

Hard boundaries on what may run, where, and when. Nothing executes outside the lines you draw.



Approvals & handoffs

Consequential actions gate on a human decision; routine code fixes hand off to your coding agents with full context.



Learning & auditability

Every run logged, gated, and attributed — and folded back into the service map and runbooks.

— VERIFY & LEARN

Verify recovery

Re-checks the exact production signals that triggered the incident and holds a stability window before it calls the incident resolved — with a full audit trail and timestamps.

Learn & improve

Every investigation updates the service map, runbooks and knowledge base — so the next one starts with more context, higher accuracy, and less time to resolution.

— PREVENTION

Change intelligence

Correlates deploys, flag flips and config changes to risk — so a bad change is caught, not chased.

Pre-alert investigation

Watches telemetry and opens an investigation before the alert fires — buying you time, not surprises.

Auto-generated PRs

Surfaces pre-production risks and drafts the fix as a pull request for your team to review.

— ONE COORDINATED TEAM

Owens the incident. Pulls in **the right specialists.**

The orchestrator directs a bench of specialist Coworkers — and hands routine code fixes to your coding agents with full context.

**SRE Orchestrator — the lead**

Owens the incident end-to-end and coordinates the specialists. A role, not a person — always-on, company-wide, no personal inbox. Hands off routine fixes to coding agents such as Claude Code or Cursor with full context.

Network

Connectivity & routing

Compute

Virtualization & hosts

Storage

Data & volumes

Database

Data services

Security

Security operations

Cloud Ops

Cloud estate & FinOps

— A LIVING MODEL OF YOUR ESTATE

The orchestrator reasons on a current model of production — service map, dependencies, deploy patterns, failure modes and ownership — that updates continuously.

Service & dependency map

Auto-discovers topology, traffic flows, hidden dependencies and ownership across the estate.

Deep integration

Connect natively to your platforms for a rich, always-current model — Kubernetes-native, read-only in-cluster agent.

Light discovery

No deep integration required — Lumi rebuilds a knowledge graph from the incidents and signals you already emit. Value in days.

— UNIFIED COST DATA LAYER

The **CERNE™** control plane unifies DCIM, AIOps, HCMP, FinOps and GreenOps, so you can ask cross-domain questions no cost-only tool can answer — relating a GPU cluster's power cost to its token spend in one question, one answer.

— WORKS WITH YOUR STACK

Turns up on the tools **you already run**

50+ integrations across observability, incident, cloud, code, knowledge and collaboration — plus on-call and background agents that never sleep.

OBSERVABILITY

Datadog · Prometheus · Grafana ·
CloudWatch · Splunk

INCIDENT & ON-CALL

PagerDuty · Opsgenie · incident.io

CLOUD & KUBERNETES

AWS · Azure · GCP · OCI ·
Kubernetes

CODE & CI

GitHub · GitLab · Terraform

COLLABORATION

Slack · Teams · Jira

KNOWLEDGE

Runbooks · Confluence · Notion

On-call agent

Triages every alert and posts findings before you are paged — silences noise and routes to the right team.

Background agents

Deployment monitoring, operational reports and resource or cost optimization — on a schedule or a trigger.

Coding-agent handoff

Hands routine fixes to Claude Code, Cursor or your own agents with full context — traces, logs, service map.

— UNDER THE HOOD

**Private LLM**

In-boundary reasoning on your environment — never the public internet. Your telemetry, logs and code context stay inside the perimeter you control.

**Model orchestration**

Frontier plus domain-specialized models, the best model per task, continuously evaluated — with an agentic execution engine that turns a decision into a governed fix.

47 → 1

Alerts to signal

92%

Root-cause confidence

500+

Curated skills

50+

Integrations

— OUTCOMES

Replace hours of toil with a \$25 resolution

A resolution that takes an engineer roughly four hours comes back as a \$25 investigation. It is labour arbitrage — validate it, and take the toil out of the loop.

UP TO

70%

Lower MTTR — the investigation loop, gone

UP TO

90%

Less alert noise from grouping & triage

DAILY OPS

2–3_x

Engineer productivity

PER RUN

\$25

Per investigation vs a ~4-hour human effort

— PRICED ON THE ESTATE IT RUNS

An annual platform fee by estate size, then a per-investigation fee — with steep discounts as you commit volume.

SMALL

\$50K

platform fee · per year

MID

\$100K

platform fee · per year

LARGE

\$200K

platform fee · per year

Then per investigation: **\$25** pay-as-you-go · **\$20 / \$16 / \$12** at 10K / 50K / 200K committed per year · **\$8** at volume, 1M+ per year.

BOTTOM LINE

An always-on reliability function for your whole estate — detect to learn, under your guardrails.

Turn up your SRE Orchestrator.

An always-on reliability function for your whole estate — value-proving POC in about two weeks.

LUMIOps AI™

LUMIOPS.AI



GET IN TOUCH

lumiops.ai/contact-us

POC

~2 weeks to value

LINKEDIN

linkedin.com/company/lumiops