

— CASE STUDIES

Five estates. One question: who is carrying the triage?

Short case studies from the sectors where reliability engineering is in highest demand — e-commerce, payments, SaaS, streaming and digital health — each showing what changed when incident response became a role rather than a rota.

5 case studies 5 sectors 1 role

CONTENTS

What these five have in common

None of these teams had a monitoring gap. Each had good observability, competent engineers, and a defensible availability record. What they shared was that the expensive part of an incident — grouping the noise, reconstructing context, tracing cause across domains — was still being done by a person, at whatever hour it happened. The change was not better tooling for that person. It was removing the requirement for one.

01 E-commerce & Retail	p03
Peak traded through, with nobody paged	
02 Fintech & Payments	p04
Every action taken, every action evidenced	
03 B2B SaaS	p05
On-call stopped being a retention problem	
04 Streaming & Media	p06
Live events without the standing army	
05 Digital Health	p07
Clinical uptime, with the record intact	

THE RECURRING TRADE-OFFS

- **Triage is the bottleneck, not the fix.** In every case the remediation was minutes of work; establishing what broke and whether acting was safe consumed the hours.
- **Cause crosses domains; teams do not.** A symptom in one layer routinely originates in another, and tracing it manually means pulling people in sequence.
- **Guardrails are what make autonomy acceptable.** Every team widened automation only once the boundaries and the audit trail were explicit.
- **The cost lands outside the SLA.** It shows up as attrition, change freezes, over-staffed events and knowledge that leaves with one engineer.

01

E-COMMERCE & RETAIL

Peak traded through, **with nobody paged**

SECTOR

Global marketplace

SETUPDeep integration,
K8s-native**AUTONOMY**Scale, rebalance,
route**DRIVER**

Peak readiness cost

THE CHALLENGE

- One checkout degradation produced **dozens of alerts** across payment, cart, search and CDN
- Investigation crossed four teams in sequence for symptoms that originated downstream
- A **six-week change freeze** was the safety mechanism, because correlation was manual
- Peak readiness was measured in how many engineers could be put in a room

THE SOLUTION

- Related alerts **deduped and grouped** into a single investigation
- Parallel investigation across net, compute, storage and database with a confidence score
- Scaling, rebalancing and routing **automatic within guardrails**; deploys gated on approval
- Change intelligence correlating deploys and flag flips to risk

THE OUTCOME

- Triage removed from the critical path before anyone is paged
- Cross-domain root cause with evidence, not a dashboard to interpret
- The **war room reserved** for genuinely novel incidents
- Recurring failure modes captured in the service map

We used to measure peak readiness in how many engineers we could get in a room. Now the first question is what the orchestrator has already closed before anyone woke up.

— DIRECTOR OF PLATFORM ENGINEERING

02 FINTECH & PAYMENTS

Every action taken, every action evidenced

SECTOR

Payments platform

REASONING

Private LLM, in-boundary

AUTONOMY

Investigation only; action gated

DRIVER

Compliance review

THE CHALLENGE

- Fix scripts ran with **standing privilege** and were attributable only to whoever triggered them
- Reconstructing an incident meant assembling chat logs, tickets and shell history
- Latency incidents demanded rapid judgement when engineers were least able to exercise it
- Deploy-to-incident correlation was manual, so **establishing cause took longer than reverting**

THE SOLUTION

- Investigation and diagnosis fully autonomous; **remediation gated by action class**
- Execution guardrails defined per service, making permitted actions explicit
- Every run **logged, gated and attributed** as a property of the platform
- Private LLM reasoning in-boundary — telemetry never leaves the perimeter

THE OUTCOME

- Automation the audit function approved rather than tolerated
- Post-incident reconstruction drawn from records that already exist
- Engineers arriving at the decision point **with evidence attached**
- Change intelligence surfacing risky deploys before they are chased

Our regulator does not object to automation. They object to automation nobody can bound or reconstruct. Guardrails and an audit trail were what made speed approvable.

— HEAD OF SITE RELIABILITY

03

B2B SAAS

On-call stopped being a retention problem

SECTOR

B2B SaaS platform

SETUP

Light discovery first

AUTONOMY

Widened progressively

DRIVER

Engineer attrition

THE CHALLENGE

- **Most overnight pages required no action**, but each still needed a human to look
- Being paged meant rebuilding context from scratch, often across unowned services
- Multi-tenant blast radius took longer to establish than the fix itself
- Exit conversations named **on-call more often than compensation**

THE SOLUTION

- An on-call agent that **triages every alert and posts findings before paging**
- Investigations arriving with service map, recent deploys and prior incidents assembled
- Light discovery to start — value in days, no integration project
- Routine code fixes **handed to coding agents** with full context

THE OUTCOME

- Interruptions reduced to the pages that actually matter
- Tenant impact established inside the investigation
- Autonomy widened as the team verified decisions
- The rota no longer the reason senior engineers resign

We were not failing our SLAs. We were failing our engineers. The rota was the single most cited reason people left, and it was the one thing we had never automated.

— VP OF ENGINEERING

04 STREAMING & MEDIA

Live events without the standing army

SECTOR

Streaming platform

MODE

Pre-alert during
events

AUTONOMY

Reroute, scale,
rebalance

DRIVER

Event staffing cost

THE CHALLENGE

- Concurrency spikes arrived faster than diagnosis — **the event was often over first**
- Every live event ran a full incident bridge, mostly waiting
- Regional issues looked global until someone **manually compared**, delaying reroute
- Post-event learning was thin because everyone was recovering

THE SOLUTION

- **Pre-alert investigation** opened on telemetry, before the alert fires
- Regional versus global scope established as part of the investigation
- Rerouting, scaling and rebalancing automatic; playback config gated on the event lead
- Each event folded into the model so delivery-path failures are recognised

THE OUTCOME

- Mitigation fitting inside the window that actually matters
- The bridge **convened, not scheduled** by default
- Reroute decisions arriving with scope evidence attached
- Recurring delivery-path modes recognised, not rediscovered

We staffed bridges because the cost of being slow was public. Once the orchestrator was investigating before the alert fired, the bridge became something we convene, not something we schedule.

— HEAD OF STREAMING RELIABILITY

05 DIGITAL HEALTH

Clinical uptime, with the record intact

SECTOR

Digital health platform

REASONING

In-boundary only

AUTONOMY

Conservative; records
path gated

DRIVER

Coverage without
headcount

THE CHALLENGE

- Overnight coverage the small platform team **could not staff**
- Consultation telemetry could not be reasoned over by an external service
- Every production change needed evidencing, **slowing the fastest interventions**
- Outages carried clinical consequence but were handled reactively

THE SOLUTION

- Private LLM reasoning **entirely in-boundary**
- An always-on reliability function across the estate, no 24/7 rota required
- Conservative autonomy — restarts and scaling automatic, **records path gated**
- Every run logged and attributed for clinical governance review

THE OUTCOME

- Round-the-clock cover for a team too small to staff it
- Telemetry that never leaves the controlled boundary
- Governance evidence produced as a by-product of each run
- Investigations opening **before a clinician notices**

We could not send consultation telemetry to an external model, and we could not afford a 24/7 rota. In-boundary reasoning was what made an always-on function possible for a team our size.

— CHIEF TECHNOLOGY OFFICER

— THE PATTERN

A rota is not a function

The distinction is structural. A rota assigns reliability to whoever is awake, with the context they happen to hold. A function performs it continuously, estate-wide, with the whole model available every time — and keeps what it learns.

WHAT A ROTA ASSUMES

- That whoever is on call holds the relevant context
- That noise can be filtered fast enough by a human
- That cross-domain tracing is worth the coordination cost
- That what is learned tonight will be available next time

WHAT A FUNCTION PROVIDES INSTEAD

- The **whole estate model** available on every investigation
- Grouping and triage before a human is interrupted
- **Parallel** investigation across domains, not sequential
- Action bounded by guardrails and recorded by default
- Every incident folded back into the map and runbooks

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