Qevlar AI

for SOC leaders

Choosing Your AI SOC Path

BUILD VS BUY

Summary

Security teams are building their own AI SOC capabilities, and many are getting real results.

In 2026, Qevlar AI spoke with dozens of MSSPs and enterprise SOC teams that are building their own AI SOC capabilities. Many are getting real results. A North American MSSP running a private LLM on its own hardware cut ticket handling from 35 to 45 minutes down to two for most of its queue. A European software group's security team told us its system already does about 90% of what a vendor demo showed.

So the useful question is what it takes to get past the first working use cases. In our conversations, the same five obstacles came up again and again:

1. Getting from validation into production
2. Automating the hardest remaining alerts
3. Depending on one or two builders
4. Paying for upkeep that never ends
5. Scaling from a handful of use cases to full coverage

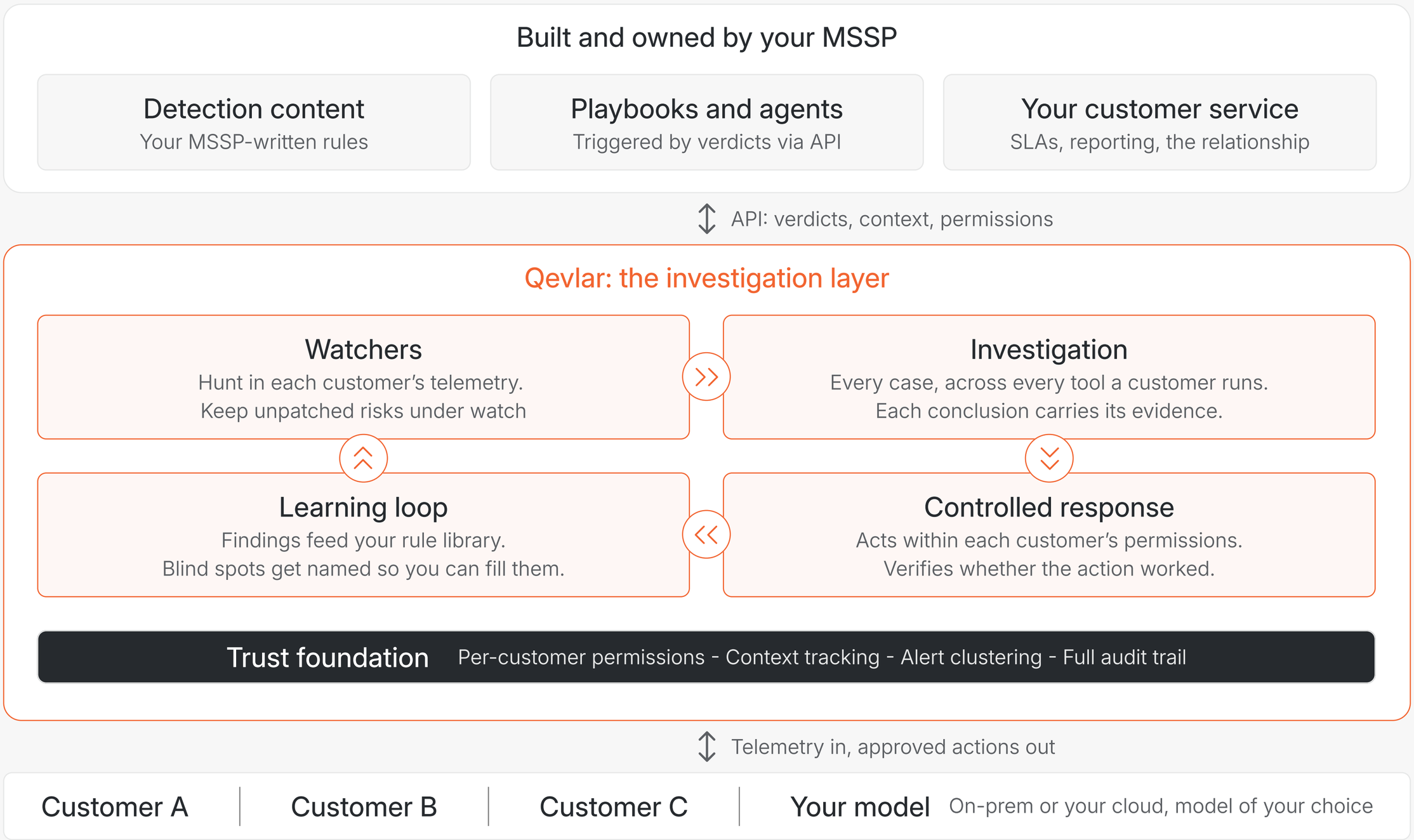
This guide covers why teams build, what their builds already do well, where they stall, and what even a working build doesn't cover.

At a Glance

Where in-house builds stall, and how Qevlar handles each point:

Where builds stall	What it takes in-house	How Qevlar AI handles it
Getting into production	Engineering a consistency layer around the LLM, then building your own way to measure analyst agreement	The platform plans each investigation and LLMs handle specific steps. Sampling measures analyst agreement on live alerts.
The hardest remaining alerts	Gathering context by hand, alert by alert and customer by customer	Investigations follow new evidence as it appears, correlate with past and grouped alerts, and pull per-customer context from tickets and analyst feedback.
Depending on one or two builders	Logic and context that live in a few people's prompts and code	Logic, context and decision history live in the platform, visible to the whole team.
Upkeep that never ends	Retesting after every model change, maintaining integrations, and forecasting token or GPU costs	Qevlar takes on model changes and integration maintenance.
Scaling to full coverage	A new build for every use case and every vendor integration	Investigation plans are built from each alert's entities, with no playbook per alert type. MSSPs get a separate scope, context and feedback loop per customer, plus custom alerts.

Own the service, run the investigation layer on Qevlar AI



Why Teams Build

The reasons are sound, and they deserve to be taken seriously.

Intellectual property

When AI capability is part of what a company sells, it wants that IP built inside its own walls.

Data control

Some organisations won't let a third party handle their customers' data, so on-premises deployment becomes a requirement. Customer contracts can also rule out third-party processing.

Business model

For MSSPs, analyst capacity is part of what they sell. Whoever automates that work owns that part of the business.

Build-first culture

Engineering-led organisations often build any tool expected to carry a large licence cost.

Capable models

Frontier models now give consistent results on many alerts, and open-weight models running on in-house GPUs can bring the cost per investigation down to pennies. Quality still depends on the stack: models struggle with query languages that are poorly documented.

What In-House Builds Already Do Well

The teams we spoke to have built things that work in production:

Faster handling	Combining automation and AI can cut ticket handling from the better part of an hour to a few minutes.
Growth without hiring	Teams can absorb significant growth in alert volume or customers without adding analysts.
Engineering around unpredictability	Deterministic scripts handle the steps that must be reliable, and AI handles enrichment.
Results without an AI background	Security teams with no AI specialists can learn as they go and still ship use cases that work.
Grounded in internal context	Agents trained on internal context can cover hours the SOC no longer staffs, such as weekends.
Dedicated AI engineers	The most advanced builders employ AI engineers specifically for this work.
Tailored to one environment	A build can be shaped tightly around one organisation's context, tools and desired outcomes.

Where Builds Stall

01 Getting from validation into production

In-house build

Analysts act on a verdict when they can reproduce it and check the reasoning. In Qevlar's study of 18,000 LLM investigations, one alert run 100 times produced 90 different investigation paths, simple alerts followed their most common path in at most about 75% of runs, and identical alerts sometimes got different verdicts. Teams can engineer around this with deterministic scripts, but they also have to build a way to measure agreement with analysts, then maintain both.

With Qevlar AI

The platform plans each investigation, and LLMs handle specific steps such as parsing a log or reading an email, so the same alert follows the same path to the same verdict. Sampling, now with early customers, seals a share of live alerts so analysts investigate without seeing the verdict, then measures how often they agree with Qevlar.

02 The hardest remaining alerts

Builds win quickly on high-volume, repetitive alerts. The hardest ones resist. These alerts usually need context the system doesn't have: who the user is, what is normal for this customer, whether a pentest is running.

In-house build

Much of that context lives in email threads and analysts' memories. Asking analysts to type it into a tool tends to work for a week before usage drops off.

With Qevlar AI

Investigations are iterative: when new evidence turns up, Qevlar investigates it in turn. Each case is correlated with past and grouped alerts, and per-customer context is pulled from tickets and analyst feedback, so an alert from an admin account and one from a standard user are judged differently.

03 Depending on one or two builders

In-house build

The logic of an in-house build lives in prompts, code and the heads of the people who wrote them. Those builders are usually some of the SOC's most capable engineers, so every hour they spend maintaining the system is an hour away from security work.

With Qevlar AI

The investigation logic, the context it has learned and the history of every decision live in the platform, where the whole team can see and audit them. Nothing depends on prompts one engineer wrote and has to keep maintaining.

04 Upkeep that never ends

In-house build

Every model change means retesting every use case, and 40 model versions across OpenAI, Anthropic and Google carry a 2026 retirement date. Models also drift: the further attacker techniques move from a model's training data, the less reliable its judgement becomes. Cost per investigation can be very low today, especially on self-hosted open-weight models, but it is hard to forecast nine months out, after new models, token prices and GPU refreshes. CFOs notice. Tech debt compounds too: each new agent adds maintenance for every one already running.

With Qevlar AI

Qevlar takes on model changes and integration maintenance, so you budget for a platform instead of forecasting token prices and GPU refreshes.



05 Scaling to full coverage

In-house build

Each new use case brings its own data sources, prompts and tests, so coverage grows one use case at a time. **MSSPs feel this hardest: a build needs an integration for each security product in each customer environment, maintained every time that product changes.** Coverage needs keep expanding too. Mandiant's M-Trends 2026 puts the mean time to exploit at minus seven days, which adds exploitation hunting and continuous CVE monitoring to the list.

With Qevlar AI

Each investigation plan is built from the alert's own entities, such as users, hosts, IP addresses and files, with no playbook per alert type, so coverage grows without a new build per use case. Connectors are vendor-agnostic, so Qevlar keeps working through a SIEM migration, such as Splunk to Sentinel, or an acquisition. For MSSPs, each end customer gets its own scope, context and feedback loop, and custom alerts are supported.

What a build doesn't cover: beyond the investigation

- **Seeing what your rules don't cover.** Qevlar Watchers keep investigating continuously in the telemetry you already collect. Point one at a critical vulnerability you can't patch yet, and the risk stays under the SOC's watch until the fix lands.

- **A defense that evolves.** In a triage build, an investigation ends in a closed ticket. Qevlar turns what it learns from each investigation into new detections and automations, and names the blind spots it finds so your team can fill them.

- **AI you can govern and prove.** Every action follows your permissions and change process, and every conclusion shows its evidence, including what the AI is unsure about. Analysts check a sample of AI decisions blind, every piece of learned context is traced to the decisions it changed, and autonomy is granted one alert type at a time.

For MSSPs: own the service, add a service line

Qevlar is live in production at 1,500+ organisations, including Fortune Global 500 enterprises and MSSPs. For MSSPs, the two most common reasons to build are intellectual property and owning the automation. Both stay with you.

01

Your IP stays yours.

Your detection rules, playbooks, customer context and the service built on them remain your own. Qevlar investigates the rules you write without taking ownership of them. It can run in your cloud or fully on-premises, on the model of your choice, so customer data stays on infrastructure you control.

02

You own the automation.

Qevlar's API exposes verdicts, context and permissions, so your SOAR playbooks, custom agents and customer-facing tools run on top of it. The service, the SLAs and the customer relationship stay with you, and Qevlar adds the investigation capacity to take on more customers without breaking SLAs or adding headcount.

03

Each customer's rules, enforced.

Every customer has its own permissions and change process, especially in co-managed setups. Qevlar acts within them and verifies whether each action worked. Automated remediation is limited to alert types a customer has approved, with confidence thresholds you set, so a cautious customer never gets automation it didn't sign off.

04

Proof you can show each customer.

Sampling benchmarks Qevlar against your analysts and detects drift in production. Context tracking links every piece of context to its source and to the decisions it changed, and every investigation shows the logs, queries and steps behind it. Escalations reach the customer with the evidence already attached, so they get fewer follow-up questions.

05

A new service line to sell.

Watchers put customers' unpatched vulnerabilities under continuous watch, with detection rules Qevlar writes and maintains. That gives you a new exposure-monitoring offer to sell, with little maintenance to weigh on margin. What each customer's investigations reveal also feeds back into your own detection content.

What this looks like in practice: real-world examples

VANCORD

Vancord started building, then stopped.

Vancord runs a SOC for around 125 customers. Before choosing Qevlar, the team had started building its own AI investigation system and stopped because of cost at the scale of one company, constantly changing models, and the need for ML engineers to backtest everything. Three to four months into production, analysts sit down to a complete investigation within five to ten minutes, work that used to take half an hour or more.

5-10 min

per investigation (down from 30 min)

"Our goal is to secure customers, not build systems. We would have just ended up being a developer shop."



Matthew Fusaro
Head of Security Operations

Almond

Almond kept its own platforms and built around Qevlar.

Almond, an MSSP with 450 experts, has spent years building its own technology: ITERA for managed-as-code detection and M&NTIS for attack simulation. It added Qevlar to that stack alongside its SOAR, SIEM and XDR tools. Qevlar now handles over 80% of cases autonomously, and mean time to remediation is under five minutes, with no new analysts added.

>80%
Cases handled autonomously

<5 mins
Mean Time to Remediation

nomios

Nomios scaled AI investigation across a large SOC.

Nomios serves more than 5,000 customers across Europe with 750 certified engineers. Qevlar was integrated in a few days and now runs its L2 investigations in under three minutes, saving about 30 minutes per investigation. That freed 60 analysts to focus on threat hunting and detection tuning.

"With Qevlar AI, our SOC analysts are now 'augmented analysts', capable of accelerating response times while maintaining quality."



Eric Bohec
Group CTO

3 min
to investigate

≈ 30 min
saved per investigation

60 FTEs freed
for proactive defense

Live in production in 1,500 organizations globally

Orange Cyberdefense · Atos · sodexo · Global 500 · ECI · MediaMarkt SATURN · i-TRACING Cybersecurity · stoïk · GlobalConnect · nomios · Almond · PERFORMANTA · Deloitte · NYSE · La Poste

See how it works →