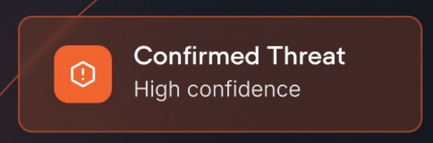
Qevlar AI



The SOC Survival Guide for the Frontier Model Era

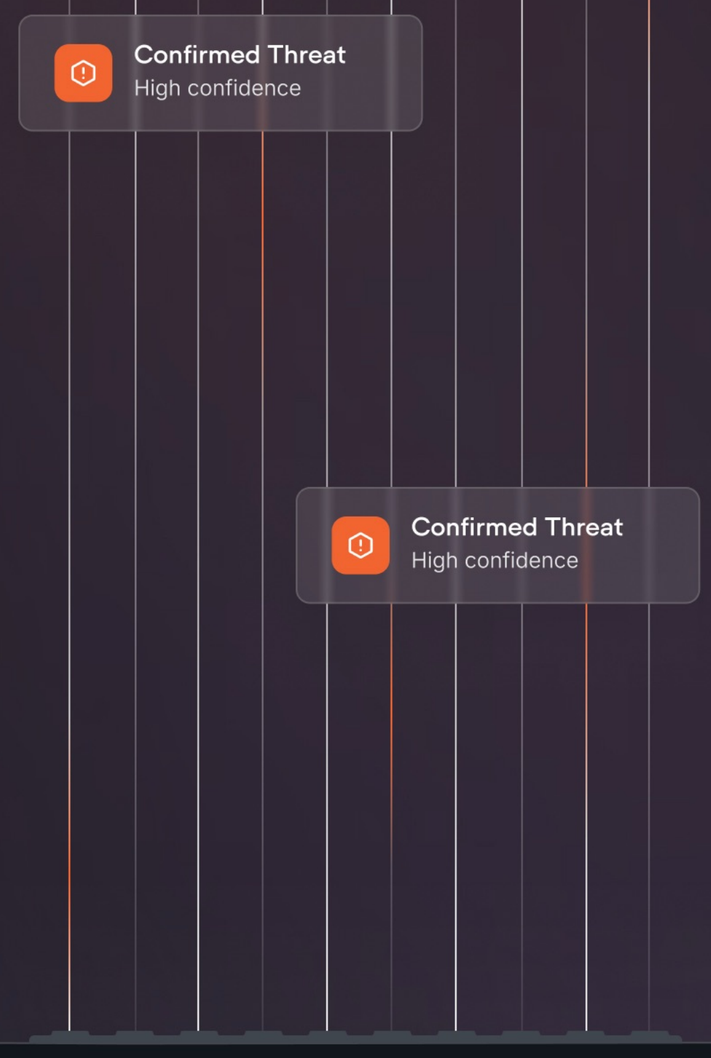
4 risks Claude Mythos and similar models bring and how to solve them

Intro

The arrival of Claude Mythos has forced every security leader to recalibrate. The headlines have been loud. The actual operational implications for SOC teams have been far less clear.

Most of the coverage has focused on the capabilities of the model itself. This guide focuses on something more useful: what Mythos (and similar models like Open AI Daybreak) means for the way your SOC operates, which risks deserve your attention, and what your team should be doing differently as a result.

By the end, you will have a clear answer to the question every CISO is asking right now: are we exposed in ways we haven't accounted for, and where do we start?



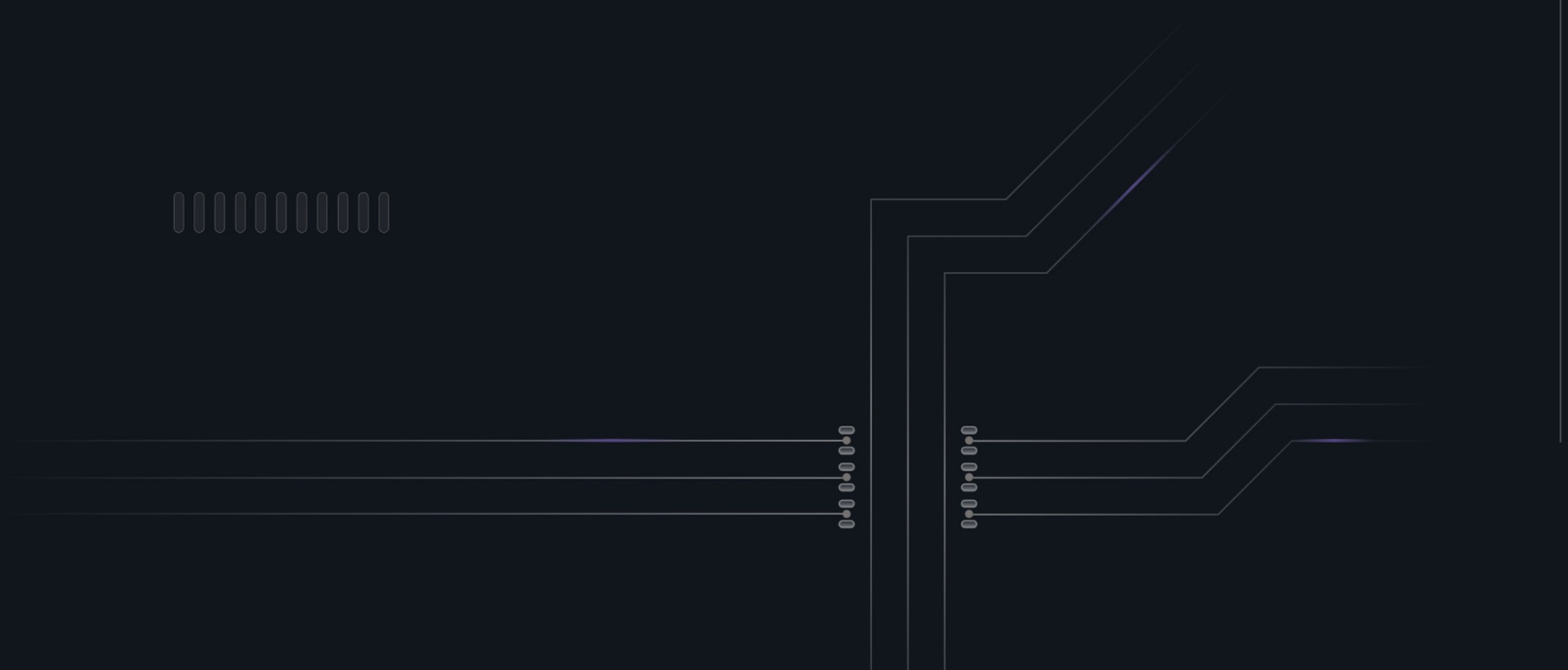
What actually happened with Mythos

On April 8, 2026, Anthropic announced Claude Mythos Preview alongside Project Glasswing — a restricted consortium of roughly 40 organizations given access to the model to find and patch critical vulnerabilities before adversaries could exploit them. The announcement was framed as a watershed moment.

Here is what actually happened, separated from the media.

The capabilities are real, but Anthropic chose what to show you

Mythos Preview is capable of identifying and exploiting zero-day vulnerabilities in every major operating system and every major web browser, including bugs that are ten or twenty years old. The oldest found so far is a now-patched 27-year-old flaw in OpenBSD which is an operating system known primarily for its security. In pre-release testing, it reproduced vulnerabilities and developed working exploits on the first attempt in over 83% of cases. Those numbers are significant. (Source: Anthropic)



The economics are muddier than the headlines suggest

When Anthropic published their technical findings, one number became the center of the debate: $20,000. That was the total cost of running roughly 1,000 automated scaffold runs to discover a 27-year-old vulnerability in OpenBSD, one of the most hardened operating systems in the world. An untargeted sweep across thousands of attempts on a deliberately difficult target. Most coverage treated it as the cost of an attack.
It wasn't.

A separate exploit chain on a known Linux kernel vulnerability, starting from a CVE identifier, completed in under a day at a cost under $2,000. The headlines collapsed that distinction entirely, producing both the "this is financially prohibitive" dismissal and the "AI will hack everyone" panic, neither of which reflects the actual cost structure.

For context on where the cost curve is heading: GPT-5.5 solved a reverse-engineering challenge, reconstructing a custom virtual machine's instruction set and recovering a cryptographic password in 10 minutes and 22 seconds at an API cost of $1.73. A professional analyst using standard tools took 12 hours to solve the same problem.

The economics of a sustained, large-scale agentic attack are still genuinely unclear. But the direction of travel is unambiguous, and security investment decisions anchored to any specific cost figure today are likely to be wrong within 12 months.

The 3 risks that actually matter for your SOC

The Mythos coverage focused almost entirely on what the model can do. The more useful question for security operations leaders is: what does it expose about the way your SOC currently works?
The following three risks were real before Mythos, but the model makes each of them more urgent.

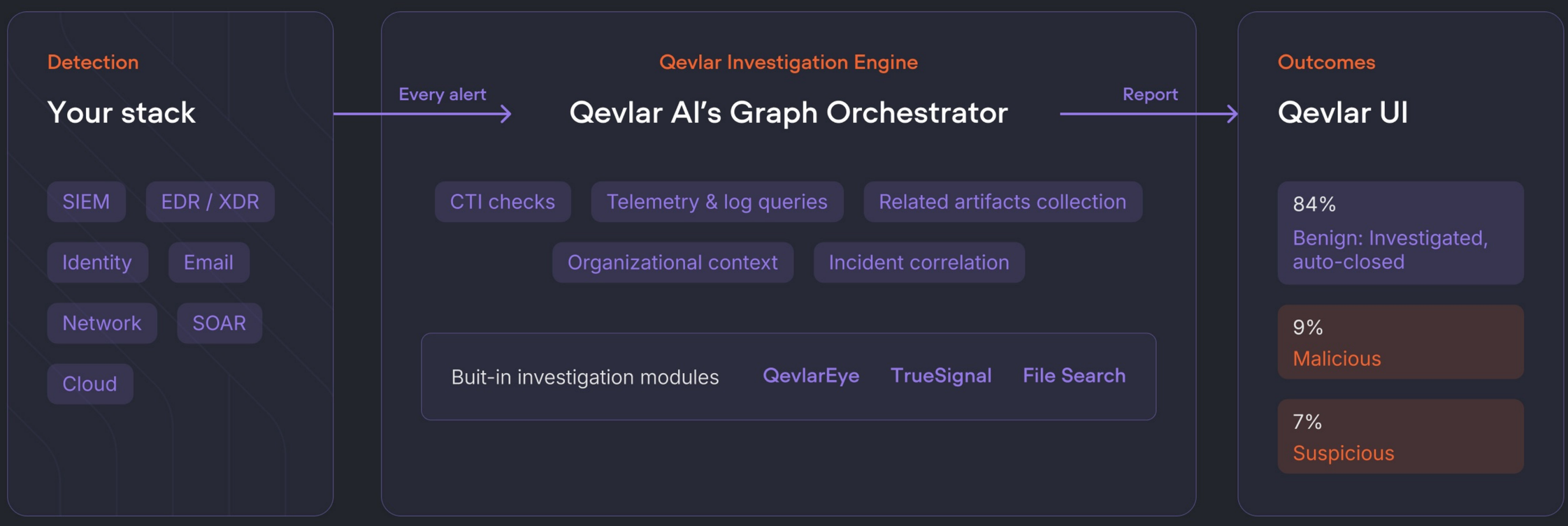
Risk 1:
Your response window is collapsing

What good looks like: every alert gets a full investigation, automatically, with consistent depth regardless of volume, time of day, or which analyst is on shift.

Where most teams fall short: SOCs receive thousands of alerts per day, and analysts can manually triage fewer than half. That was a sustainability problem before Mythos. With AI compressing the time between vulnerability discovery and working exploit from weeks to hours, it is now a material risk. A manual triage process that investigates 40% of incoming alerts at adequate depth is making an implicit resource allocation decision about the other 60%.

What closes the gap: Qevlar investigates 100% of alerts in under 3 minutes. Each investigation follows a defined, validated path with built-in self-checks, collecting evidence across your entire stack without requiring playbooks or pre-built rules. Every verdict is evidence-backed and auditable. Analysts receive confirmed threats with a full investigation context ready for response.

> "We not only reduced operational costs significantly but also improved our operational security efficiency and excellence."
>
> 
>
> Abdelhalim ELMOUADAN
> Head of Global Operational Security, Sodexo
>
> sodexo



Risk 2:
Your SOC architecture was not built for this volume

What good looks like: low severity alerts are surfaced, connected to related signals across your stack, and read as part of a broader pattern. The SOC that catches multi-stage attacks is the one that can read those patterns at scale, across every shift, without depending on an analyst who happens to remember seeing something similar last week.

Where most teams fall short: in many SOCs, detection engineers constrain the rules they deploy because the analyst team cannot absorb the resulting alert volume. This creates a ceiling on detection coverage: the SOC can only detect what it can investigate. AI-driven attack chains are designed to exploit exactly this kind of blind spot.

What closes the gap: Rather than treating each alert in isolation, Qevlar correlates activity across entire security stack and surfaces, mapping the full blast radius of an incident, including lateral movement, credential anomalies, and asset exposure. When Qevlar confirms a false positive, it automatically generates a detection tuning request so the bad rule gets fixed. The detection ceiling lifts.

> "With Qevlar AI, we've achieved the depth and consistency of investigation we've always aimed for. Every alert, no matter how subtle, is analyzed and documented in minutes. It has given us confidence, clarity, and the ability to scale without compromising quality."
>
> 
>
> Daniel Aldstam
> Chief Security Officer, GlobalConnect
>
> GlobalConnect

From alert to containment in under five minutes

0:18	**3 EDR alerts** correlated to 2 endpoints
1:15	PowerShell + LSASS access on **2 hosts**
2:08	Correlated to single campaign **4 identities affected**
5:00	Containment actionned **2 compromised accounts**

Identifies affected 03

Compromised

Anders Holm VIP

a.holm@metrix-retail.com

Escalated to SOC

Compromised

Ilia Petrov

i.petrov@metrix-retail.com

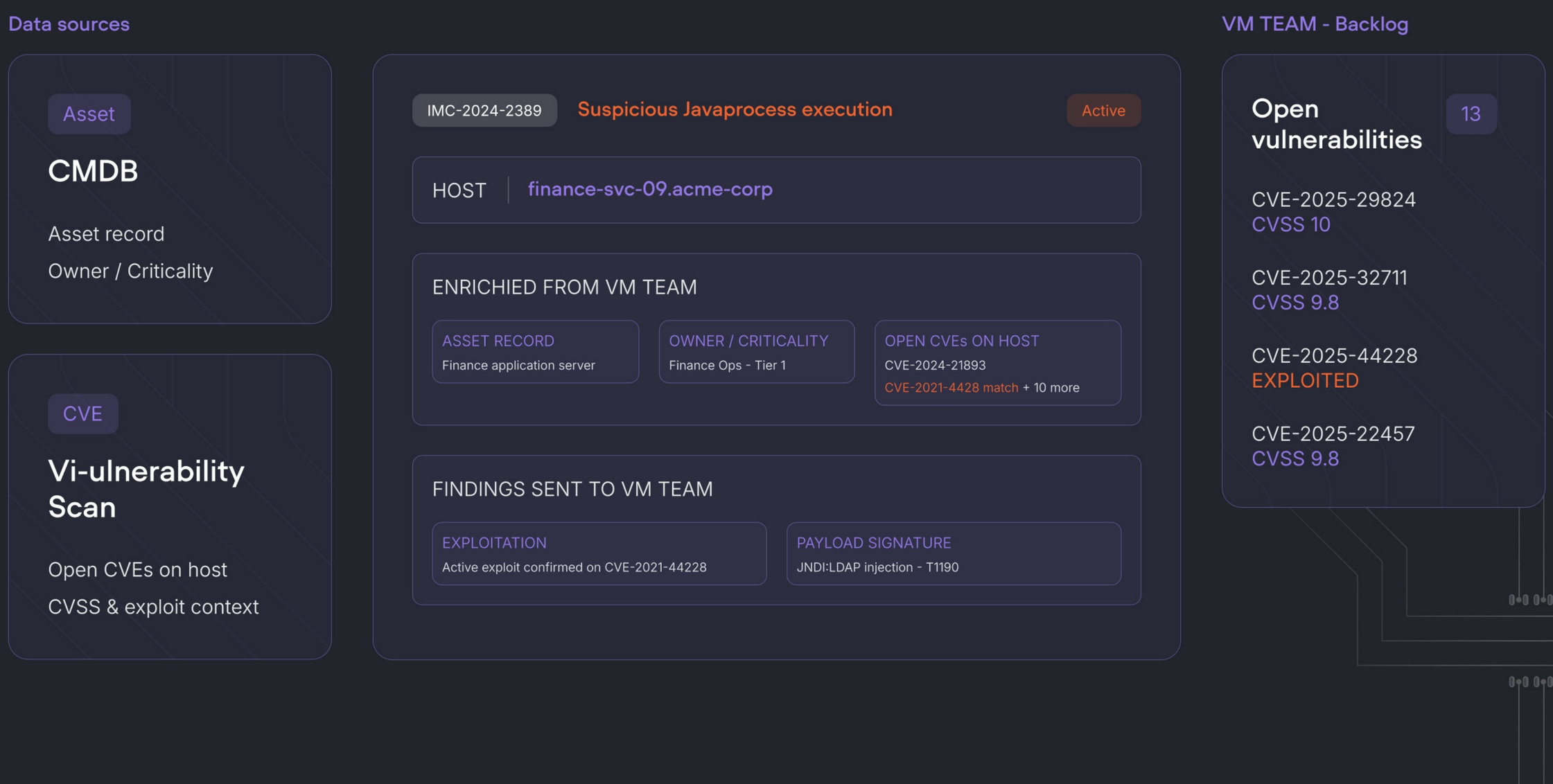
Account disabled Sessions revoked

Risk 3:
Your SOC and vulnerability teams are flying blind to each other

What good looks like: the SOC and vulnerability operations function as a single intelligence loop, not two separate teams sharing spreadsheets on a quarterly cadence. When the SOC confirms active exploitation of a CVE, that finding immediately elevates it in the vulnerability queue. When a critical unpatched exposure sits on a crown jewel asset, it becomes an active hunt hypothesis in the SOC. The two teams work from the same picture, in real time.

Where most teams fall short: in most organizations, SOC and vulnerability management are separate functions with separate tools, separate priorities, and ad hoc collaboration. The result is that the most dangerous exposures stay invisible to the people who could act on them fastest.

What closes the gap: Qevlar creates a live intelligence loop between SOC and vulnerability operations in a vendor-neutral way that works across your existing stack. When Qevlar's investigation confirms active exploitation of a CVE in your environment, that signal flows directly to VM prioritisation. In the other direction, when an alerted device carries high-EPSS unpatched vulnerabilities, that exposure context enriches the SOC investigation in real time.



Most organisations find out they have a problem when it is already expensive to fix. These three risks are detectable in advance, but only if you are looking at the right signals in the right way.

14-day pilot on your live alerts, plus a custom SOC readiness assessment. Zero lift from your team.

For qualified SOC teams, until June 30.

See how it works →

qevlar.com