White paper

The CISO's guide to building a Mythos-ready security program

When AI is the attacker, the network becomes the first line of defense

EXECUTIVE SUMMARY

In April 2026, Anthropic released a limited-access cybersecurity-specific LLM that shook the cybersecurity world. As we struggle to figure out how to respond to the changes in the security landscape Mythos and similar models introduce, it's important to consider the broader shifts rather than focusing solely on patching the most recently announced vulnerabilities. Mythos-class models represent an epoch in the development of language models, a structural shift in what AI can do to software security, not a one-time event tied to a single model. Other models will reach comparable capabilities within 6 to 12 months, and some will eventually exceed what current models have demonstrated. As a result, the right response is structural, not Mythos-specific. This guide lays out four pillars for building a resilient security program that addresses risk from Mythos and every capability jump that follows. It closes with a 90-day action plan that can be executed starting today.

What changed—and what didn't

Corelight Asset Classification analyzes protocol fingerprints captured in network traffic to continuously identify and categorize every device on the network. Built on Zeek®-based network evidence, it passively classifies assets using DHCP and HTTP user-agents—emitting a dedicated asset_classification.log that enriches existing Zeek logs with device context. No additional hardware, agents, or active polling is required.

CONTENT

What has changed recently is the economics and speed of vulnerability discovery. Yes, adversaries have exploited long-unpatched infrastructure before; for example, the Volt Typhoon campaign built on vulnerabilities that organizations had years to remediate but didn't. However, Mythos automates and scales that pattern: time-to-exploit has fallen from 2.3 years in 2018 to under 10 hours in 2026, and more than 99% of the vulnerabilities Mythos has found remain unpatched. So the threat model is not exactly new, but it's operating now at greater speed and scale.[2]

The 2026 Verizon Data Breach Investigations Report underscores the urgency: software vulnerabilities now account for 31% of breach entry points, overtaking stolen credentials as the leading attack vector for the first time in the report's history.

What isn't changing, even in the age of Mythos, are the fundamental tenets of effective defense. Segmentation, network visibility, behavioral detection, assume-breach architecture—these controls were right before Mythos, and they are even more urgently important now. But in the age of Mythos, it's critical to prioritize speed.

The same controls that were best practices before Mythos are even more relevant now, including complete asset visibility, behavioral detection at the network layer, AI-augmented defense at machine speed, and vulnerability prioritization grounded in what is actively being exploited. Organizations that built this infrastructure are positioned to absorb the next wave on their own terms.

NOT ONE MODEL. A NEW ERA.

In April 2026, six of the eight vulnerabilities disclosed for FreeBSD were found using AI or heavy AI assistance, even before the release of the Mythos model. Although much attention has focused on Mythos, it's important to recognize that advanced models from many vendors - and soon, open-source initiatives - will democratize the power of vulnerability discovery. We are entering a new epoch, and regardless of which model is dominant next year, organizations need to respond structurally to increase resiliency. For defenders, this means thinking beyond the goal of quickly patching new vulnerabilities, which will be necessary but not sufficient. Fortunately, AI helps greatly on the defensive side as well. The combination of agentic automation, sophisticated anomaly detection, and high-quality real-time data will help organizations weather coming storms.

What CISOs should expect: Three phases

Technology-driven disruption in security often follows a recognizable arc. Understanding that pattern will help orient investment against the overall narrative, rather than responding to each stage in isolation.

Phase 1: Disruption (now): high noise, rapidly evolving attack techniques, and significant variance in defender outcomes. Exploit windows are measured in hours. Patch cycles are structurally outpaced. Organizations that fare worst in this phase are those that have deferred network visibility, behavioral detection, and assume-breach architecture.

> **"As humans, we can't defend ourselves perfectly from every bacteria and virus, but we do have a flexible immune system that monitors what's happening, and responds in real time to new threats. Sophisticated security teams have been taking a similar approach for years, and the pattern needs to become much more common in the age of Mythos."**
>
> - Greg Bell, Co-Founder, Corelight

Phase 2: Adaptation (12 to 24 months): Defenders absorb the new threat model. AI-powered defensive tooling matures. Updates to patching and vulnerability management tools and processes. Best practices for AI SOC architecture emerge. Equilibrium begins to form around organizations that have built an immune system infrastructure in Phase 1.

Phase 3: Equilibrium: The new normal. Organizations that invested in Phase 1 operate with a structural advantage. Even if they can't predict future vulnerability waves, they've deployed systems to maximize resiliency.

The most important strategic question is not whether an organization is 'safe' in a binary sense from future vulnerability waves. The question is whether that organization is investing now to maximize resiliency, as the 'arms race' of vulnerability detection, exploitation, and defense continues to escalate.

The cost variable buys limited time

Token costs for frontier models remain high, which creates an economic disincentive for certain attacker classes, while nation-states and well-resourced criminal groups will spend whatever it takes. The mid-tier attacker's economics are still shifting, but that window will not stay open. The

cost of Large Language Model (LLM) inference has declined rapidly, and the capabilities are proliferating into open-weight models that cost almost nothing to run.

THREE HARD TRUTHS FOR THE MYTHOS ERA

These truths will hold, regardless of which model is most capable of discovering vulnerabilities or automating attacks next year.

Hard truth 1: The discover-prioritize-patch cycle is broken

CVEs logged in the first 100 days of 2026 already equal all of 2025 combined. Each new model will uncover another set of previously unknown vulnerabilities—producing recurring waves of exposure that overwhelm patching programs and create a new status quo, a steady state of dynamically shifting software vulnerability. The vulnerability backlog is no longer simply 'technical debt'. It becomes active, compounding operational risk. Quarterly patch cycles and CVSS-based prioritization were designed for a world that no longer exists.

The '90-day cliff' is real. Statistically speaking, findings that are not closed within 90 days are very likely to be open permanently.[3] Given an ever-growing backlog of unpatched bugs, the rational operating model going forward is triage based on real-time data and dynamic risk factors, including reachability and exploitability, that go beyond a simple severity score.

Breaches will occur, and the question is whether organizations can detect and contain them in time to limit damage.

There is a deeper problem that CVSS-based prioritization cannot solve. Mythos excels at discovering vulnerabilities that have no signature, no patch, and no public record yet. Vulnerability management tools can only scan for, prioritize, and remediate known CVEs. The assumption that keeping software up to date ensures safety no longer holds—attackers will have access to a large, ever-growing supply of novel entry points that defenders cannot yet see.

Hard truth 2: Breach prevention is insufficient as a primary strategy

You cannot prevent every breach when exploit windows are measured in hours, and vulnerability discovery greatly outpaces the capacity to patch. Operating assumptions must shift: Breaches will occur, and the question is whether organizations can detect and contain them in time to limit damage. No enterprise IT department allows a tool to patch critical databases or production infrastructure without testing automatically—doing so risks catastrophic downtime. There will always be a human-enforced delay and long exposure windows. The program must be built to defend what cannot yet be patched.

That requires a different kind of infrastructure, one that watches everything on the network, not just known endpoints. Endpoint tools are a critical defensive layer, but cannot be universally deployed and are vulnerable to new attacks, just as all software is. However, attackers cannot avoid leaving network evidence as they breach systems and move laterally towards their ultimate targets.

Hard truth 3: Your AI-layer is an attack surface you haven't inventoried

Every coding agent, CI runner, and MCP server in your organization carries developer credentials and, in some cases, repo write access. A compromised coding agent can commit code, call external tools, and exfiltrate data at machine speed, outside business hours.

Most security programs have only a limited inventory of which models their developers use, what IDE plugins are installed, which MCP servers are installed, or which agents have repo write access. The consequences of that gap become very significant during an active breach.

Defenders need to treat the code factory (including agents, MCP servers, developer workstations, CI runners) as a production attack surface. Every agent needs a named owner, scoped permissions, and other safety guardrails. In practice, such environments often fall short of meeting strict security requirements.

FOUR PILLARS OF A MYTHOS-READY SECURITY PROGRAM

Each pillar below addresses a specific structural gap. Together they contribute to an immune-system-style defense that does not depend on knowing every vulnerability in advance.

Pillar 1: Complete asset visibility

See everything AI will find—before attackers do

Attackers prefer to target lightly managed or unmanaged devices, which are often unknown to many asset management tools. As the attack-defense cycle shortens and automates, having accurate, real-time data on current asset composition and exposure is critical.

What good looks like

- Continuous, network-based asset identification and classification that covers managed and unmanaged systems, in addition to vulnerability and asset management.
- Device type, OS, manufacturer, role, and network behavior continuously updated as traffic is observed, not through scheduled scans or quarterly audits.
- Proprietary and industry-specific protocol parsing for ICS/ OT in critical infrastructure such as power plants and manufacturing facilities.
- Program to adhere to the standard deadline for agencies to patch Common Vulnerabilities and Exposures (CVEs) that are posted to the CISA Known Exploited Vulnerabilities (KEV) catalog.
- Complete AI-layer inventory: shadow AI detection, every LLM endpoint, MCP server, and coding agent with repo access, with a named owner and defined blast radius.

How Corelight helps

Corelight continuously classifies and analyzes every asset that communicates on the network, including the unmanaged, unknown, and unpatched systems that AI-driven exploits might target first. This is particularly important in ICS and OT environments where patching is slow, difficult, or operationally infeasible, because assets that cannot be patched are precisely those that require the greatest visibility. This asset intelligence from Corelight then feeds your SIEM/XDR, AI detection pipelines, and agentic SOC platform. Corelight's shadow AI detection also highlights the use of unauthorized, unmonitored AI tools that compromise security, compliance, and business outcomes.

Pillar 2: Detection and containment

Build for the breach that will happen.

Mythos changes how an attacker gets into the network, but it cannot change what an attacker must do once inside. To achieve objectives such as ransomware deployment or data exfiltration, attackers must leave a

The time-to-exploit window that security programs were built for—measured in months—has collapsed to hours.

trail of network evidence behind. Endpoint-only detection cannot cover the full attack surface. Unmanaged systems, unknown devices, lateral movement through network infrastructure, and compromised endpoints with disabled agents are all invisible to endpoint tools. Without network-level visibility, a breach may be confirmed - but the attack chain will remain obscure, leading to incomplete investigations and containment, even when using AI tools.

What good looks like

- AI-powered network-level detection as the non-negotiable baseline: it's the only layer that sees every asset, every session, and every exploit attempt.
- Behavioral detection correlated to asset context and service identity, not just IP addresses.
- Deception tech that can provide high-fidelity alerts on specific malicious behaviors, many of which are difficult to detect with log analysis or a SIEM alone. Canary tokens as high-confidence breach signals: fake credentials, decoy MCP endpoints. AI attackers enumerate exhaustively, so a single canary hit is a signal rather than noise.
- Detection-as-code pipeline grounded in strong network telemetry that an analyst can inspect, a Detection Engineer can write against, and a regulator can audit.

How Corelight helps

Corelight delivers industry-leading threat and anomaly detection built on the richest structured network metadata available. When a Mythos-class exploit chain runs in your environment, Corelight detects it at the network layer, regardless of whether endpoint tools are present or already compromised.

Corelight's AI-powered threat detection includes unsupervised ML models and behavioral detections that build behavioral baselines for every entity in the network and surface anomalies that lack a signature. This technique can surface zero-day and novel exploit behaviors before any CVE exists.

In addition, the platform has supervised models and signatures, developed and maintained by Corelight Labs, that are trained on curated threat data to identify specific adversarial TTPs with high confidence across more than 100 MITRE ATT&CK techniques—including lateral movement, credential harvesting, and C2 over standard protocols—the predictable post-exploitation behaviors that follow any initial access method. Every detection comes with the specific network context that underpins it, so analysts can validate findings, tune thresholds, and carry verdicts into board reporting. Finally, Corelight's behavioral detections deliver the same multi-attribute coverage and detect threats that signatures miss, with significantly lower false-positive rates, because they analyze protocol-level semantics deterministically rather than generating probabilistic scores. Together, these multiple analytical approaches ensure that a detection program covers unknown zero-days, known TTPs, and confirmed active exploits—without requiring defenders to know in advance which category an attack falls into. Black-box detections are outdated in the age of Mythos. Corelight's ML models are backed by network evidence you can inspect, tune, and write code against.

Detection is only half the equation, though. Corelight's response integrations close the loop without adding a new console. From Investigator, analysts can isolate a compromised host through an EDR provider, block a port on a firewall with a single click, force a logout or password reset through an identity provider, and open a case in ServiceNow or Jira—all from the same investigation view. In a Mythos-era environment where attacks move at machine speed, the gap between detection and containment needs to be as small as possible, and some organizations may prefer an automated response. Corelight's integrations with endpoint and vulnerability management leaders, such as CrowdStrike, SentinelOne, Microsoft, Tenable, and major identity and firewall platforms, help organizations to drive confident response based on trusted data.

Pillar 3: AI-augmented defense

Fight at machine speed—deploy the same class of capability defensively

You cannot outrun machine-speed threats with human-speed processes. The volume of Mythos-class vulnerability waves will exceed any human team's capacity to absorb. The answer is not more headcount, but rather deploying advanced AI capability on the defensive side.

What good looks like

- AI-powered vulnerability scanning of internal codebases and infrastructure to find what attackers will discover before they can exploit it.
- Agentic triage that autonomously correlates alerts, surfaces high-risk entity activity, and delivers investigation-ready findings rather than alert queues.
- Automated response thresholds are defined in advance, setting clear criteria for when the SOC moves from assisted triage to auto-containment, with human review of the outcome.
- Evidence quality is treated as a strategic investment because AI defensive tools perform only as well as the underlying data they operate on.

How Corelight helps

Corelight's network metadata is purpose-built for AI-driven analysis. All commercial and open-source language models already understand Zeek-based metadata, because they were trained on the open Internet. In Corelight's solutions, that same high-quality telemetry powers AI detection models, IR workflows, and NetOps baselines. For your AI SOC to triage and respond at machine speed, it requires access to structured, enriched, correlated network evidence - not raw packet dumps or incomplete flow records. In fact, Corelight research has shown that data quality sets a hard ceiling on AI inference and, in turn, on SOC performance. Even the most advanced models could not overcome limitations imposed by low-quality network data. This finding has important implications for the architecture of SOCs, as efficiency improvements compound in an increasingly automated environment. Assuming that agentic automation will boost SOC efficiency over the next year or two, the choice to use better data and raise investigative success rates from 30% to 90% will result in a dramatic increase in overall efficiency.

Critically, this does not require displacing the tools your team already uses. Corelight evidence and multi-layered detection flows natively integrate with popular SIEM and XDR platforms such as Splunk, Microsoft Sentinel, and Elastic via built-in connectors, so your analysts can work in the platform they prefer rather than in a separate vendor console competing for attention. In addition, Corelight's Investigator offering features Agentic Triage functionality that addresses the problem of adversarial attack automation. Rather than delivering discrete alerts with assigned priorities, it intelligently groups detections into entity-centric incidents, with visible AI reasoning at every layer, allowing forensic pivots to raw evidence at any

point. Entity-centric consolidation eliminates repetitive review steps, enabling 10x faster triage and more informed analyst decisions. As threat volume escalates, organizations can define predetermined thresholds for moving from assisted triage to auto-containment.

Pillar 4: Intelligent vulnerability prioritization

Triage at machine speed—based on what is actually being exploited.

Over the next year, Mythos class models will drive waves of vulnerability disclosure, some of which will not be valuable, actionable, or relevant. A surge in vulnerability disclosures has already been reported in open source projects, creating issues of prioritization and work planning. In this environment, part of a rational prioritization process is knowing what to ignore and what to prioritize. And what is true for developers is also true for defenders. When AI floods defensive environments by exploiting newly discovered vulnerabilities, triage must involve dynamic prioritization based on asset criticality, along with real-time telemetry to identify which attacks (among the hundreds available) are actually unfolding. The question shifts from what is vulnerable to what is actively being exploited now, and which systems and services are mission-critical. Triage based solely on CVSS scores will produce a backlog, but not a priority list.

What good looks like

- Real-time flagging of reconnaissance and exploitation attempts against known vulnerabilities, providing the signal that a CVE has moved from theoretical to active in your environment.
- Correlated data across network, host, identity, and vulnerability sources that surfaces post-exploitation activity: lateral movement, suspicious credential use, data staging before exfiltration.
- MTTD, MTTR, and coverage that are tracked against baselines, ensuring prioritization decisions are grounded in measurable risk reduction.
- Preemptive threat hunting that leverages network telemetry and insights, allowing hunters to search for threats that may have bypassed other security measures.

How Corelight helps

Corelight's network evidence context is deepened through Corelight's ecosystem integrations. Connectors for CrowdStrike, SentinelOne, Microsoft, and other integration partners enrich relevant host, identity, and CVE data directly into Corelight sensor logs and/or detections, ensuring that live network activity is correlated with current host and identity data to prioritize alerts based on actual enterprise risks. Threat hunting is also more effective with real-time context that helps prioritize threats before they escalate into an incident that can disrupt business and cause enormous damage. Corelight provides tools for threat hunting, including behavioral analysis, anomaly detection, threat intelligence integration, detailed search capabilities, and mapping to the MITRE ATT&CK framework. Having host, vulnerability, and identity correlated with network activity effectively consolidates alerts from multiple systems into a single alert. That is how the prioritization pillar becomes operational rather than theoretical.

Software vulnerabilities now account for

31% of breach entry points,

48% and of all breaches entry involve ransomware.

THE BOARD CONVERSATION

Boards are actively inquiring about the implications of Mythos and other advanced models, exploring the evolving landscape of risk, and developing programs and investments to improve the resiliency of their organizations.

Your board needs to know that the economics of cyberattacks have changed permanently. The time-to-exploit window that security programs were built for—measured in months—has collapsed to hours. That compression will not reverse. As model capabilities improve and token costs decline, attacker economics will extend down-market from nation-states to organized criminal groups to opportunistic attackers.

The board's most important strategic question is not whether an organization is 'safe' from vulnerabilities that Mythos and similar models will discover. Instead, the most important discussion should focus on the optimal mix of investments to improve resiliency before a disruption phase begins.

The 2026 Verizon DBIR is a reality check. Software vulnerabilities now account for 31% of breach entry points, and 48% of all breaches involve ransomware.

This is the baseline that Mythos-class models will make much worse.

Every AI defense tool in the stack—including detection models, agentic triage, investigation workflows, and agentic pipelines—is dependent on the quality of the underlying evidence available to it. Corelight's own research shows that data quality sets a hard ceiling on AI SOC performance: even the most capable frontier LLMs cannot overcome the limitations imposed by low-quality data. Organizations that raise their investigative success rate from 30% to 90% by improving evidence quality will see that improvement compound dramatically as agentic automation scales.

The three questions every board is asking—and how to answer them

Are we exposed? The honest answer is yes, because every organization is exposed. But some organizations will have much greater resilience against future attacks than others. Those that gather high-quality data about the state of vulnerability and exploitation in real time, across every asset (including unmanaged and unknown systems), stand a much greater chance of weathering the storm successfully. If a response to Mythos starts and ends with endpoint agents being installed and active, then Mythos-class tooling will eventually find and exploit the structural gap. Yes, EDR controls are critical, as are vulnerability management and patching. But they are not sufficient for the era we are entering, one of permanent and shifting vulnerability. Network visibility and detection can close that gap by providing the raw fuel for AI automation and detecting anomalous network activity correlated with vulnerabilities that have not yet been patched.

How quickly can we respond? If we face a compromise, how swiftly can we recover and ensure comprehensive remediation of the issue? This is where operational metrics such as Mean Time to Detect (MTTD) and Mean Time to Respond (MTTR) provide genuine insights. They measure the duration from detection to containment for the most recent confirmed incident. Additionally, they reflect whether agent-based triage is employed and what the implications of a tenfold increase in triage speed are for analyst capacity liberation. Governing bodies rely on these metrics. CISOs who can effectively interpret this data hold the strategic narrative.

You cannot prevent every breach when exploit windows are measured in hours.

Are we ready for the next one? The right answer is structural, not focused on a specific model or on the vulnerabilities making headlines today. With thoughtful investment in the components that make it possible to automate the SOC using the same tools and models that attackers use, defenders have a good chance of turning the tables eventually. These tools include basic cyber hygiene practices (including patching) that have been underfunded, as well as next-generation capabilities such as AI-powered detection, machine-speed triage, and agentic workflows driven by exceptional data. Such infrastructure does not need to be re-architected for each new model, because it's designed to survive successive waves of disruption and keep organizations focused on their core missions.

YOUR 90-DAY ACTION PLAN

The four pillars above are not a sequential roadmap; they are parallel workstreams that reinforce each other from day one. The 90-day action plan below provides perspective getting started on building a sustainable foundation for Mythos defense.

Today

Understand your current approach to network visibility. Does it cover unmanaged assets and east-west traffic? Does it provide modern, high-fidelity data that can power AI workflows? Can it detect shadow (as well as legitimate) AI services? Most organizations discover significant gaps here, and Mythos-class tooling will find and exploit these. Simultaneously, run an AI-layer inventory: identify every LLM endpoint, MCP server, coding agent, and IDE plugin that has access to your development environment or production systems, and confirm each has a named owner and scoped permissions. When it comes to vulnerability prioritization, stop triaging solely by CVSS score and establish a working protocol that factors in network reachability, active exploitation signals, and the 90-day cliff beyond which findings are effectively permanent.

30 days

Deploy sensors and confirm network evidence is flowing into your SIEM or XDR platform, with host, CVE, and identity enrichment integrations active. Validate that supervised ML detections are tuned to your environment's traffic patterns, and that shadow AI detection is flagging unauthorized model usage. Run the first Corelight-backed tabletop against a lateral movement scenario to establish your MTTD/MTTR baseline.

90 days

Assure that agentic triage is enabled, with entity-centric incidents replacing per-alert analyst review, for a 10x boost in human investigative productivity. Make sure that response integrations are active, including one-click host isolation via EDR, firewall blocks, and identity actions, all available from a single investigation view. Consider a second tabletop run, this time designed for simultaneous multi-incident scenarios, which is the operational reality AI-scale attacks create. By this point, you have the four structural capabilities in place: real-time asset intelligence, AI-powered behavioral detection with explainable outputs, agentic triage at machine speed, and vulnerability prioritization grounded in what is actually being exploited on your network right now.

WHERE CORELIGHT FITS

Corelight is the real-time evidence and detection layer that enables the 90-day plan, powering agentic triage and other highly efficient defensive workflows. Network-based asset classification closes the visibility gap on day one. Supervised and unsupervised ML detections—backed by the richest structured network metadata available and correlated with host, CVE, and identity context—give the detection and containment pillar its signal quality. The agentic triage capability in Investigator operationalizes AI-augmented defense without requiring your analysts to learn a new console. And the vulnerability prioritization pillar only becomes real when grounded in live network telemetry showing what is actively being exploited. Corelight is not one more tool in the stack—it is the connective tissue that makes the rest of the stack work. Without network-level ground truth, every AI detection, every triage decision, and every investigation that follows is operating on incomplete information. That gap is what Mythos-class tooling is built to find. Closing that gap is what a Mythos-ready program is built to prevent.

References

1. https://zerodayclock.com/
2. https://red.anthropic.com/2026/mythos-preview/
3. https://thehackernews.com/expert-insights/2026/03/which-code-vulnerabilities-actually-get.html

corelight

To learn more about building a Mythos-ready security program, request a demo at

https://corelight.com/contact

info@corelight.com | 888-547-9497