WHITEPAPER · MAY 2026

Post-Mythos Security

Why the AI vulnerability storm calls for a root-and-branch redesign of the detection layer.

Evan Powell and Josiah Lashley
DeepTempo

EXECUTIVE SUMMARY

The argument in brief

On April 7, 2026, Anthropic published an Alignment Risk Update for Claude Mythos Preview[1]. Their characterization is plain. Mythos Preview is “significantly more capable, and is used more autonomously and agentically, than any prior model. In particular it is very capable at software engineering and cybersecurity tasks, which makes it more capable at working around restrictions.” The model is in limited research preview, not general availability. Anthropic is holding it back precisely because of these capabilities.

The vulnerability discovery side of Mythos is the visible half of the capability. The other half is the autonomous attack orchestration that Anthropic itself documented in November 2025, when a Chinese state-sponsored group, designated GTG-1002, manipulated Claude Code into executing approximately 80 to 90 percent of the tactical workload of a multi-target espionage campaign across roughly 30 global entities[2]. Reconnaissance, vulnerability identification, exploitation, lateral movement, credential harvesting, exfiltration. That was Claude Code. Mythos is more capable along precisely the dimensions that matter for autonomous attack orchestration.

The Cloud Security Alliance, with SANS, [un]prompted, and the OWASP GenAI Project, has published a draft strategy briefing for what it calls a Mythos-ready security program[3]. It is a serious document with serious authors, and it offers useful operational guidance for a CISO walking into a Monday morning meeting. Read alongside the CMU CyLab Cyber Autonomy Initiative research manifesto[4], and the keynotes from RSAC 2026 calling for a fundamental reimagining of security in the agentic era, the picture sharpens. Each document, taken alone, addresses part of the problem. Read together, they describe a set of structural changes that the detection layer beneath the SOC has not yet absorbed.

This whitepaper makes the architectural argument that ties those documents together. Signature-based detection and bespoke per-customer anomaly detection have collapsed into the same structural failure. The path forward is a foundation model for cybersecurity that learns the language of telemetry and generalizes to behaviors it has not seen. DeepTempo has built one. It is called LogLM. It runs in production with Deutsche Telekom, BNY, Stanford, the Technology Advancement Center in OT environments, and across the Snowflake Native App ecosystem.

[1]Anthropic, "Alignment Risk Update: Claude Mythos Preview," April 7, 2026 (updated April 10, 2026). https://www.anthropic.com
[2]Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign," November 14, 2025. Threat actor designated GTG-1002.
[3]Cloud Security Alliance, SANS Institute, [un]prompted, and OWASP GenAI Security Project, "The AI Vulnerability Storm: Building a Mythos-ready Security Program," April 12, 2026 (Draft).
[4]Carnegie Mellon University CyLab Security and Privacy Institute, "Cyber Autonomy Initiative" research agenda. The CyLab manifesto argues for proactive, behavioral, system-level defenses across data-plane and control-plane infrastructure.

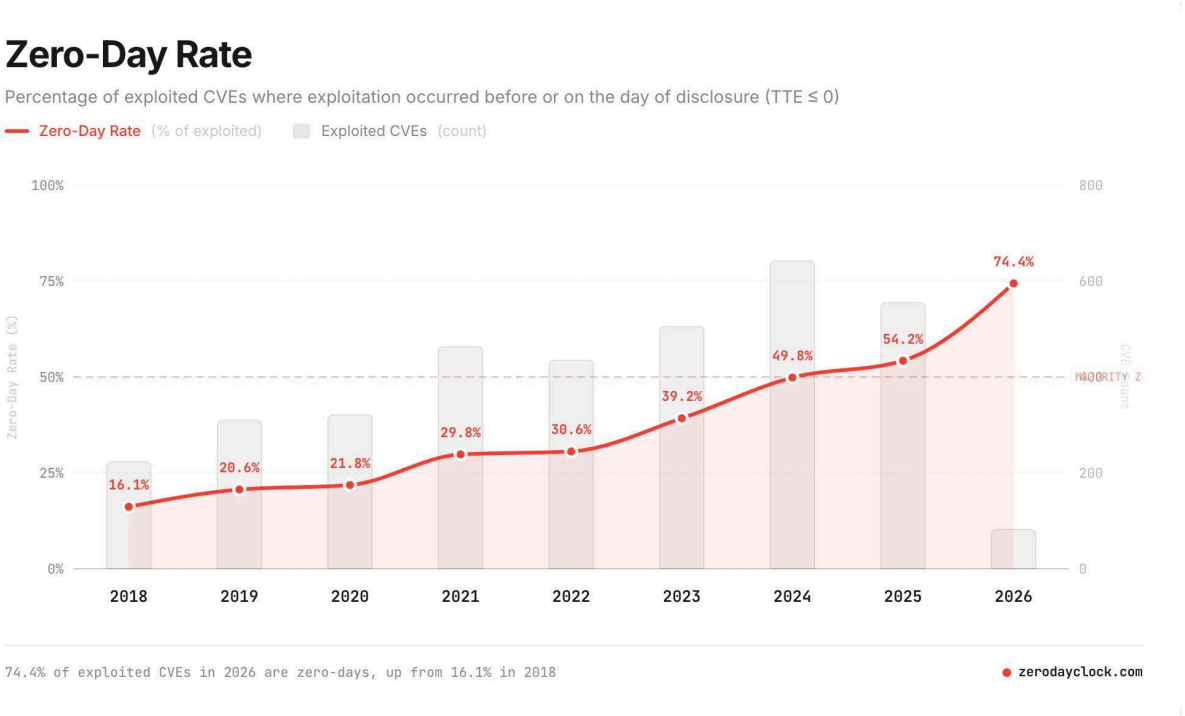
SECTION 1

The Zero Day Clock and the death of “in time”

The Zero Day Clock draws on roughly three thousand five hundred CVE-exploit pairs from CISA KEV, VulnCheck KEV, and XDB[5]. It tracks two related curves. The first is the median time from CVE disclosure to first observed exploit. The trajectory is unambiguous.

- **2018:** 771 days. Two years to patch.
- **2021:** 84 days. A nine-fold compression in three years.
- **2023:** 6 days.
- **2024:** 4 hours.
- **2025:** the majority of exploited vulnerabilities were weaponized before public disclosure.
- **2026:** median time-to-exploit under one day. The data fits an exponential decay curve.

The second curve is the share of exploited CVEs where exploitation occurred before or on the day of disclosure. The Zero Day Clock now publishes this directly, and the slope is the same.



Source: zerodayclock.com. 74.4 percent of exploited CVEs in 2026 are zero-days, up from 16.1 percent in 2018.

Sandra Joyce, Vice President of Google Threat Intelligence, made the same point at RSAC 2026: the time between initial access and hand-off to a human operator collapsed from eight hours in 2022 to

[5]Zero Day Clock, https://zerodayclock.com. CVE-exploit pairs sourced from CISA KEV, VulnCheck KEV, and XDB. Time-to-exploit and zero-day rate trajectories independently maintained.

twenty-two seconds in 2025[6]. Whether measured at the disclosure boundary or at the operational hand-off, the same compression appears.

The implications for signature-based detection follow directly. A signature is a record of something already seen. When the attacker's exploitation precedes the public record, signature matching is structurally late. The category of detection that defined commercial security tooling for two decades is, with respect to the leading edge of attacks, a record-keeping function rather than a defensive one.

The effectiveness of signature-based detection degrades significantly in an environment where exploitation precedes disclosure, because the signature does not yet exist. This is not a tuning problem. It is a sequencing problem in the definition of the technique. Absent the invention of time travel - and not even AGI may get us there - the current pace of novel attacks mean signature based detections are far behind.

[6]Sandra Joyce, Vice President of Google Threat Intelligence, RSAC 2026 keynote: "We found that the time between initial access from threat actors to the hand-off has collapsed from eight hours in 2022 to twenty-two seconds in 2025."

SECTION 2

The bespoke-anomaly trap

The traditional alternative to signatures has been per-customer anomaly detection, generally implemented as User and Entity Behavior Analytics. The promise is that an anomaly engine learns what normal looks like in your environment and flags deviation. The reality, on inspection, is less generous.

- **Hand-engineered features.** Most UEBA systems extract a fixed set of statistical features from logs: counts, rates, ratios, thresholds. The choice of features encodes prior assumptions about what attacks look like. Signature matching reappears one level up the stack, in the feature engineering rather than in the rule set.
- **Per-customer models.** A new customer requires a new model, trained on the new customer's distribution. Onboarding becomes a research project. Time to value extends from days into quarters.
- **Per-distribution models.** When the customer's environment changes, and it always changes, the model must be retrained.
- **No generalization.** These are not foundation models. They do not transfer knowledge across customers, across telemetry types, or across time. Every deployment is a clean start.
- **Alert fatigue as a structural property.** Threshold-based deviation detection produces noise on benign change. The operational response is to raise thresholds until alert volume is tolerable, which is the same as not detecting anything.

One school of recent argument has concluded that behavioral detection is structurally exhausted, and that deception is therefore the only remaining defense. That conclusion is correct about the failure of the bespoke-anomaly approach. It is wrong about the alternative.

Deception is a useful adjunct. It is not a foundation. A decoy fires when an attacker touches it, which is information, but the information is bounded by the placement and realism of the decoys, and the attacker's path to the decoy still passes through the network undetected. A defense that depends on the attacker stepping on a tripwire is a defense against the attackers who step on tripwires.

> *Hand-engineered, per-customer anomaly models are signature matching by another name. They share the structural failure mode of signatures. They cannot be the answer to ever more creative and prevalent attacks enabled by AI.*

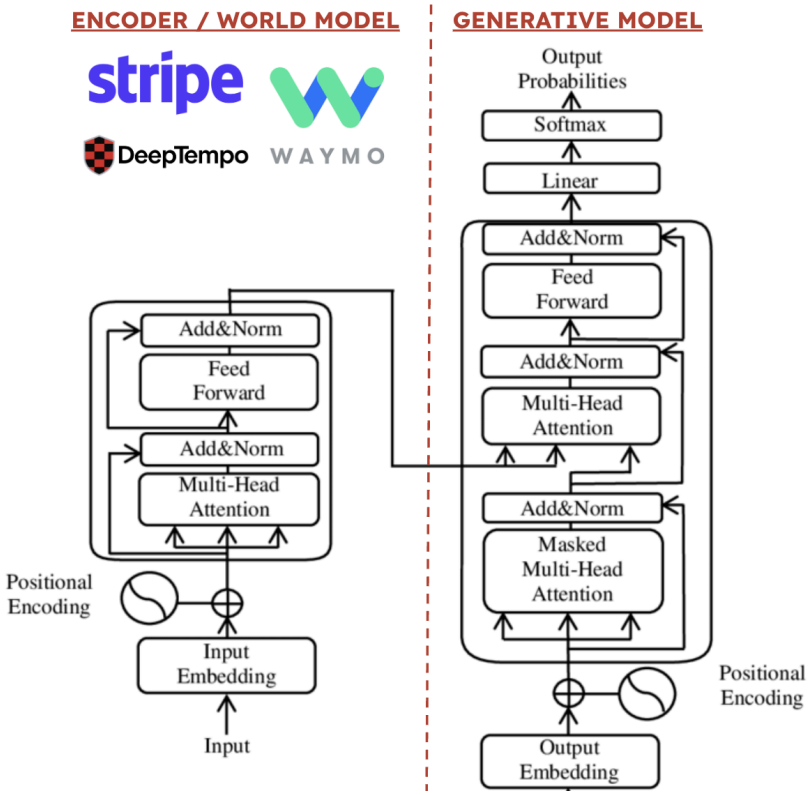
Bespoke anomaly detection, as conventionally practiced, does not generalize. Each new customer means a new model. Each new distribution of data means a new model. The architecture cannot keep pace with attackers operating across many customers using foundation-model-orchestrated tooling.

SECTION 3

Attention is all you need, and what cyber missed

In 2017, Vaswani and colleagues at Google published "Attention Is All You Need"[7]. The paper introduced the transformer and established self-attention as a sufficient mechanism for sequence modeling. Most of the field followed the decoder branch into language modeling. A smaller set of teams recognized that encoder-only transformers offered something different: representations of structured input that could be used to reason about the world rather than to generate text.

Stripe built fraud and payments models on the encoder branch. Waymo built perception and prediction on encoder-style world models. The thesis in both cases was that an encoder trained on a sufficiently broad distribution learns embeddings that generalize: that what is normal in one environment can be recognized as such in another, and that what is novel can be flagged as such regardless of whether the specific instance has been seen before.



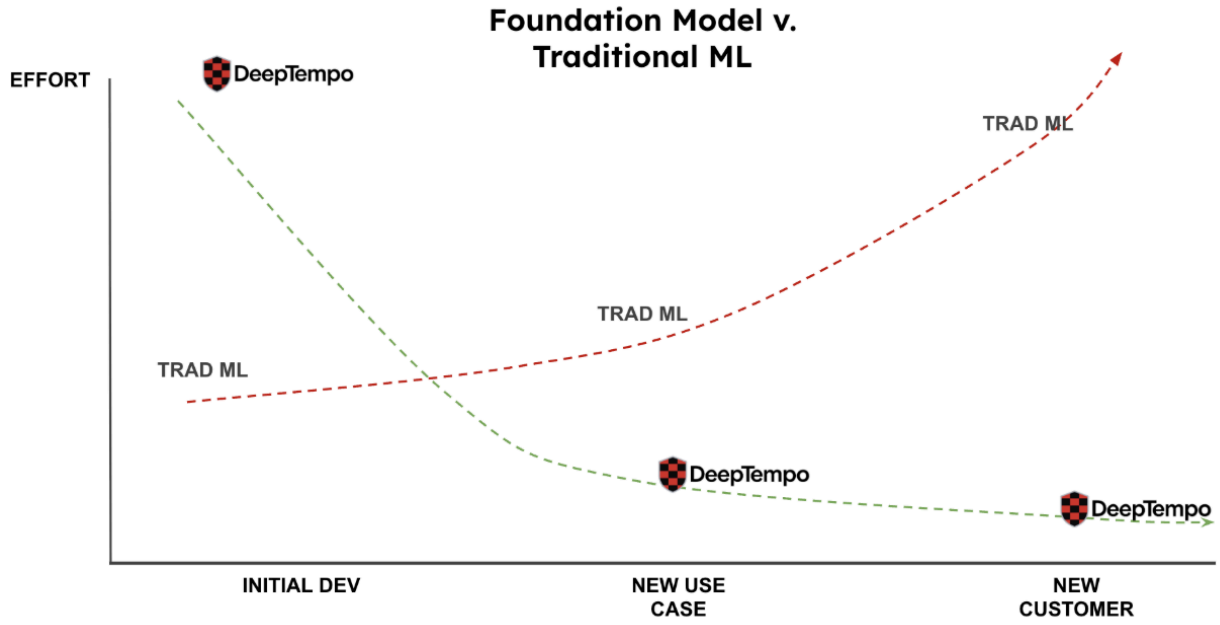
The argument resurfaced in public in November 2025, when Yann LeCun left Meta after twelve years to found AMI Labs and raise approximately one billion dollars on the thesis that autoregressive language models are a dead end[8]. LeCun's Joint Embedding Predictive Architecture, JEPA, is the encoder-side bet at frontier scale. V-JEPA encodes positions and trajectories as abstract embeddings and predicts future

[7]Vaswani et al., "Attention Is All You Need," NeurIPS 2017.
[8]Yann LeCun departed Meta in November 2025 after twelve years to co-found AMI Labs. The thesis behind JEPA and its visual variant V-JEPA is documented in Meta AI publications and in LeCun, "A Path Towards Autonomous Machine Intelligence," 2022.

embeddings rather than reconstructing pixels. The architecture trades surface fidelity for semantic generalization, which is precisely the property cyber detection requires.

How this impacts users is fairly clear. On the one hand, training and pretraining a Foundation Model is significantly more challenging, in general, than building a traditional ML model. For example, these models require substantial training runs and the expertise required to build models at the intersection of AI and cyber security is scarce. Additionally, the quality of available data is lacking, so partnerships and other techniques are required to gather a useful corpus of data.

On the other hand, once you have such a Foundation Model the ongoing costs due to adaptation are significantly lower than building traditional machine learning models per use case and per customer and often per entity within a particular customer. A Foundation Model such as the LogLM from DeepTempo that has high zero shot performance is much lower cost to maintain for any given use case or customer because it is inherently more resilient to changes in the distribution of the underlying data.



Cyber generally didn't build new foundations: what did Cyber do instead?

Cybersecurity, with rare exceptions, missed this branch of the conversation. The defensive AI conversation has been a chatbot conversation. Large language models reading alerts. Summarizing tickets. Drafting playbooks. Answering analyst questions. Useful, but downstream of the actual problem. The actual problem is that the detection layer beneath the chatbot is the same detection layer that has been failing for a decade, and pasting an LLM on top of it does not fix the underlying architecture.

THE MISS

While Waymo, Stripe, and now AMI Labs (and DeepTempo) were operationalizing the world-model thesis, cybersecurity treated AI as a chatbot category. The detection layer, where the actual defense happens, was left to the same signature-and-bespoke-anomaly architecture that has been failing on the leading edge for a decade.

SECTION 4

What Mythos *actually* means

The vulnerability discovery story is the more visible half of Mythos. Project Glasswing, Anthropic's coordinated disclosure program for Mythos-discovered vulnerabilities, has surfaced thousands of zero-days across every major operating system and browser, with a 72 percent exploit success rate, including a 27-year-old OpenBSD bug[9]. Anthropic's own characterization, in their April 7 alignment update, places vulnerability discovery alongside cybersecurity and software engineering generally as a domain where Mythos is materially more capable than its predecessors.

The same document is the clearest public characterization of the model's overall capabilities and is worth quoting at length. From their core findings:

> *Mythos Preview is significantly more capable, and is used more autonomously and agentically, than any prior model. In particular it is very capable at software engineering and cybersecurity tasks, which makes it more capable at working around restrictions.*

And from the same document's overall risk assessment: "The overall risk is very low, but higher than for previous models. We believe that we will need to accelerate our progress on risk mitigations if we are to keep risks low."[10] Mythos is in limited research preview rather than general availability. Anthropic's own assessment is that the alignment and monitoring infrastructure that contained risk for prior models is approaching, but has not yet exceeded, the capabilities of the model it is meant to contain.

This characterization changes how to read the November 2025 GTG-1002 disclosure. In that incident, Anthropic detected and disrupted a Chinese state-sponsored campaign that, in their words, manipulated Claude Code into functioning as an autonomous cyber attack agent. The model in that disclosure was Claude Code. The tactical workload executed by the model was approximately 80 to 90 percent of the campaign, including reconnaissance against multiple targets in parallel, vulnerability identification, exploitation, credential harvesting and validation, lateral movement, database querying, and exfiltration scope determination. Request rates were, in Anthropic's phrase, "physically impossible" for human operators.

Three reasonable inferences follow.

- **Inference one.** If Claude Code could orchestrate eighty to ninety percent of an espionage campaign's tactical workload, Mythos can orchestrate at least as much, more reliably, with less hand-holding. Anthropic's own framing, that Mythos is "more capable at working around restrictions," extends directly to operational restrictions imposed by defenders.
- **Inference two.** The model that finds vulnerabilities at the rate Mythos finds them is the same kind of model that can chain them. Discovery and weaponization are not different cognitive tasks. They are adjacent ones, and Anthropic's alignment report places them in the same capability cluster. Individual CVEs are often low severity in isolation. The skill in exploitation has always been chaining, finding the path from a low-privilege foothold through a sequence of vulnerabilities to full

[9]Anthropic, "Project Glasswing." Vulnerability discovery program associated with the Mythos Preview release.
[10]Anthropic, "Alignment Risk Update: Claude Mythos Preview," Section 1, "Introduction," findings 1 and 2.

compromise. What previously required a skilled human operator to spend days mapping an environment becomes an automated traversal problem.

- **Inference three.** The open source ecosystem is structurally exposed in a way it has not been before. Supply chain attacks have always been asymmetric: defenders must patch every vulnerable dependency across every version in production, while attackers need only one unpatched instance downstream. Mythos tips that asymmetry further. It can enumerate dependency graphs, identify exploitable downstream consumers, and sequence campaigns across multiple targets simultaneously. The supply chain attack surface is not growing incrementally. It is being industrialized.

The orchestration layer

One of the more immediate and not as talked about operational concerns is Mythos not as a discovery tool but as an orchestrator. Stanford's Artemis project[11] demonstrated that a multi-agent red team framework, where specialized sub-agents handle discrete segments of the kill chain coordinated by a central reasoning layer, multiplied attack effectiveness beyond the sum of its parts. The bottleneck was always the quality of the orchestrating model. Artemis demonstrated the framework. What it could not demonstrate was what happens when the reasoning layer has deep, native cybersecurity knowledge baked into its weights rather than bolted on. Mythos does not need to be prompted to think like an attacker. It already does. That changes what orchestration looks like at every level, from how sub-agents are tasked, to how failed attempts are interpreted, to how the campaign adapts in real time when a defender responds.

DeepTempo's Branch 61G research group has been tracking this convergence closely, and their assessment aligns with what the Artemis results imply: the jump from research demonstration to operational threat capability is not a question of framework design. The frameworks already exist. It is a question of reasoning quality, and Mythos potentially clears that bar. When it does, the parallelization that made GTG-1002's request rates physically impossible for human operators becomes a deliberate feature of every well-resourced campaign.

> *The vulnerability storm is the visible half of the Mythos problem. The orchestration storm is the half that has already arrived in production.*

What Modern Detection Was Never Built For

The detection stack most organizations are running today was not built badly. It was built rationally, for the threat environment that existed when it was designed. Signatures were written against known malware families because malware families were stable. Behavioral rules were tuned against human attacker patterns because attackers moved at human speed. UEBA models were trained on organizational baselines because even sophisticated attackers would eventually cross one.

That logic was sound. The assumptions underneath it are no longer valid.

[11] https://github.com/Stanford-Trinity/ARTEMIS

The entire architecture of modern detection rests on a single foundational premise: that the next attack will resemble a previous one enough to be caught by rules derived from it. Every signature, every behavioral model, every anomaly threshold encodes that premise. It is not a flaw in the design. It is the design.

Nobody built a system to contend with a model that produces novel zero-days at industrial scale, chains them faster than a human analyst can triage the first alert, and reasons explicitly about what the detection layer is looking for. The threshold of what was possible in vulnerability discovery and autonomous campaign orchestration was low enough that the pattern-matching paradigm held. Attackers found new techniques, defenders eventually caught up, signatures got written, the cycle continued.

Mythos breaks that cycle at the root. It does not produce a new variant of a known attack family. It produces attacks that have no prior instances, orchestrated by a system that has already read the same threat intelligence the defenders have. The current detection stack was built for a different game entirely.

The question for every SOC operating today is not whether their current tooling will eventually fail against this class of threat. It will. The question is whether they will know when it does.

SECTION 5

Reading the field together

Three documents, read in sequence, describe the shape of post-Mythos defense more completely than any one of them does alone. The CSA, SANS, and OWASP GenAI Project *AI Vulnerability Storm* briefing of April 12, 2026 is the operational layer: a Monday-morning playbook with eleven priority actions and a draft risk register. The CMU CyLab Cyber Autonomy Initiative paper is the architectural layer: a multi-year research agenda for behavioral detection, system-level autonomy, and proactive monitoring at the data and control planes. The RSAC 2026 keynotes are the industry signal that the architectural redesign is no longer a research question.

At RSAC 2026, Cisco's President and Chief Product Officer Jeetu Patel called for a fundamental reimagining of security in the agentic era, framed as "protect agents from the world, protect the world from agents, and detect and respond at machine speed," and announced DefenseClaw and a suite of open-source agentic security tools[12]. Microsoft's Vasu Jakkal opened the conference with "Ambient and Autonomous Security: Building Trust in the Agentic AI Era," arguing that autonomous, self-healing systems are no longer theoretical[13]. Google's Sandra Joyce documented the time-to-handoff collapse from eight hours to twenty-two seconds. The argument that the SOC stack of the last decade is structurally insufficient is no longer simply a DeepTempo vendor pitch - although this is very much our “I told you so” moment. It is the public position of the largest players in the industry.

[12]Cisco RSAC 2026 keynote, Jeetu Patel, "Reimagining Security for the Agentic Workforce," March 2026. Cisco announced DefenseClaw and a suite of open-source agentic security tools at the same event.
[13]Microsoft RSAC 2026 keynote, Vasu Jakkal, "Ambient and Autonomous Security: Building Trust in the Agentic AI Era," March 2026.

CRITERION	CSA / SANS MYTHOS-READY	CMU CYLAB CYBER AUTONOMY
THREAT CHARACTERIZATION		
Scope of threat vectors	`Narrow` Vuln discovery & exploit gen dominate. LoTL, credential abuse, insider threat essentially absent.	`Broader` Full autonomous chains — recon, lateral movement, adaptive multi-stage campaigns, novel techniques.
Attacker autonomy & speed	`Strong` TTE collapse to hours. One-shot exploit gen. Zero Day Clock data.	`Strong` Machine-timescale OODA loops. Attacks dynamic and unpredictable at human scale.
Threat model realism	`Mixed` Admits most breaches are credential/misconfig and not due to new vulnerabilities being added to an already enormous backlog.	`Realistic` Enterprise SOC scenario. Asset gaps, multi-vendor stacks, unbounded perimeters.
Proliferation timeline	`Specific` Frontier → open-weight in 6–12 months. Commodity attacker access projected.	`Implicit` Notes open vs. locked model risk. No attacker adoption timeline.
DEFENSIVE APPROACHES		
Detection philosophy	`Reactive` Detection in two vague action items. No behavioral baselines, no anomaly architecture.	`Proactive` Real-time telemetry, per-host behavioral monitoring, deception at scale, self-learning loops.
Defensive AI architecture	`Principles` "Use AI agents." No telemetry pipeline, inference layer, or orchestration model.	`Systems` Full thrust on data-plane/control-plane infra. SDN orchestration, log analytics, sim-to-real.
Specialized vs. general AI	`Generic` AI = LLM throughout. No encoder models, embeddings, or RL for detection.	`Specific` Distinguishes LLMs, RL, classifiers, formal methods. Questions LLM generalization limits.

The Mythos-ready playbook gives a CISO a vulnerability and patch action plan for this week. The CyLab Cyber Autonomy Initiative paper fills in the detection architecture and behavioral threat model the playbook does not reach. The two documents, read together, describe both halves of the post-Mythos response.

Neither document holds the full answer. The CSA briefing addresses what to do this quarter while patches and headcount are still the most actionable levers. The CyLab paper sets out how to move your environment towards machine speed intelligence. The RSAC keynotes signal that the largest infrastructure providers are committing to the same architectural direction.

The architecture exists in initial form today, with DeepTempo's LogLM running in association with a data lake or data pipeline. It is in use at financial services, telecom, higher education, OT, and on Snowflake's data-cloud platform

SECTION 6

LogLM and the world-model thesis for cyber

DeepTempo's LogLM is an encoder-only transformer trained on cybersecurity logs and telemetry broadly including network flow, Layer 7 logs, WAF data, threat intelligence feeds, application-layer behaviors. The training distribution is intentionally diverse, because we are using the transformer architecture to build a world model for cyber security across environments.

The LogLM learns to project sequences of log records into high-dimensional embeddings that capture the structure and meaning of the underlying activity. Purpose-built classifiers run on top of those embeddings and map detections to MITRE ATT&CK techniques: command and control, credential access, lateral movement, exfiltration. By identifying patterns of behavior that betray these intents, the LogLM can see GTG-1002 and others using reasoning models to research, plan and execute complex and innovative cyber attacks.

Three properties of this architecture matter for the post-Mythos environment.

- **Generalization.** An embedding trained on broad cyber telemetry recognizes the meaning and intent of a behavior, not its surface. Novel command and control infrastructure that uses unfamiliar domains, unfamiliar protocols, and unfamiliar timing still occupies a region of embedding space adjacent to known C2 activity, because the embedding has learned what C2 is.
- **Zero-shot accuracy.** LogLM operates on day one in environments it has never seen, because the encoder's representations transfer. The bespoke-model treadmill, train per customer, retrain on drift, train per distribution, does not apply.
- **Adaptation without retraining.** Where customer-specific tuning is required, it is implemented through base-layer fine-tuning and a purpose-built classifier on top of the foundation model. We are not training a new model. We are adjusting the projection at the head. Months of building a bespoke model can be collapsed into at most hours of adaptation.

Architecture

PROPERTY	LOGLM
Model class	Encoder-only transformer foundation model
Training data	Cybersecurity logs and telemetry broadly: flow, Layer 7, WAF, threat intel
Detection mechanism	High-dimensional embeddings plus purpose-built classifiers mapped to MITRE ATT&CK
Customer adaptation	Base-layer fine-tuning and head classifiers, no new foundation model per customer
Distribution drift	Continuous adaptive learning at the embedding layer

PROPERTY	LOGLM
Revision cadence	Monthly model revisions across the customer base
Deployment surface	Cloud, on-premises, OT, hybrid; agentless ingestion
Privacy posture	Federated learning supports BYOC deployments where customer data does not leave the environment

Why this generalizes when UEBA does not

A UEBA model is asked to learn what is normal for one customer from that one customer's data. Its inductive bias is bounded by the one distribution it has seen. A foundation model is asked to learn what is normal for cybersecurity telemetry across many customers and many environments, and then to recognize one customer's environment as a sample from that broader distribution. The inductive bias is bounded by the global distribution, which is much larger and much more informative.

Stripe makes this argument for payments. Waymo makes it for driving. The argument transfers to cyber for the same reason it works in those domains: the surface variety of the input is large, but the underlying behaviors that matter, fraud, collision risk, command and control, credential abuse, are stable across instances. An encoder trained at sufficient scale finds the stable structure beneath the surface variety.

SECTION 7

Built with practitioners, in stealth

After ChatGPT hit there was a great rush towards AI enabling all the things. We saw things slightly differently. As old school builders and operators in cyber - and category creators in infrastructure - we worried first about what the attackers would do. We started asking around about whether anyone was worried about LLMs being used for attackers. Generally we received blank stares or more direct negative feedback.

For whatever reason - we doubled down nonetheless. Perhaps it was because we were already hands on keyboard, familiar with the incredible power of purpose built transformer models to become foundation models. We saw that essentially a reinvention of anomaly detection would be required to counter a threat that was on the near horizon - the prevalence of novel attacks and AI orchestrated attacks that recently Mythos has made front page news.

DeepTempo did our work in stealth while the investors in the cyber security industry were mostly chasing models that wrap prompt engineering around leading models and call it a SOC.

Our design partners were enormously helpful. Today we have our LogLM working in some of the most demanding environments in the world, achieving 90%+ reductions in false positive rates while finding attacks that existing systems simply miss.

Vigil: an open-source AI SOC

Vigil is an open-source AI SOC project started by DeepTempo and released under Apache 2.0 in March 2026. It ships with thirteen specialized AI agents, more than thirty integrations, and over seven thousand two hundred detection rules across Sigma, Splunk, Elastic, and KQL formats. Vigil orchestrates investigation and response on top of LogLM detections. The architectural separation is deliberate: the foundation model below is unique and proprietary, and the orchestration layer above is open because customers should not be locked in.

> *Vigil is the first open-source AI SOC. Its purpose is to free enterprises from lock-in while unshackling the intelligence of underlying foundation models.*

SECTION 8

Three deployments, one model

The same LogLM serves three deployment models. Customers choose by data residency, regulatory posture, and how much of the security stack they want DeepTempo to host.

SaaS, free trial

Try DeepTempo without a contract. Telemetry routes to the DeepTempo cloud, detections return to the customer's SIEM or Vigil instance. The fastest path from interest to first detection.

BYOC and on-premises

LogLM runs in the customer's cloud or on-premises infrastructure. Customer data never leaves the environment. Federated learning is the architectural commitment that makes this work at the model level: the foundation model improves from aggregate patterns across the install base without any single customer's data being centralized. Required for regulated industries, sovereign environments, and OT/ICS deployments where data egress is not negotiable.

The Snowflake Native App provides the same experience for Snowflake customers, with data staying in the customer's Snowflake account.

Adaptation through purpose-built classifiers

Where a customer's environment is unusual enough to warrant tuning, adaptation happens at the head: a purpose-built classifier trained on the customer's labels, running on top of the same LogLM embeddings. We do not train a new foundation model per customer, and we do not retrain the foundation model when the customer's distribution drifts. The cost structure of adaptation is closer to fine-tuning a final layer than to training from scratch.

SECTION 9

Conclusions

Post-Mythos security accepts what the Zero Day Clock, Anthropic's own alignment and threat reporting, the CMU CyLab manifesto, and the RSAC 2026 keynotes have collectively established. Signatures are late: 74.4 percent of exploited CVEs in 2026 are zero-days. Bespoke anomaly models do not generalize. Reasoning models like Claude and similar open source models with significant cybersecurity capability are being weaponized in production today. Mythos, amongst the most capable model, is now in restricted preview rather than general release, which means the threat curve is likely even steeper than the public conversation reflects.

Doing more of the same, faster, will deliver a faster patch cycle and a larger headcount. It will not close the architectural gap.

The gap is closed by detection that generalizes. By a foundation model that understands what is normal because it has learned the language of telemetry, and that surfaces what is abnormal regardless of whether anyone has seen the specific tool before. This is what the CMU CyLab Cyber Autonomy Initiative argues for in a research register, what Cisco and Microsoft argued for in their RSAC 2026 keynotes, and what DeepTempo has built and deployed.

We must be able to see novel attacks. We must anticipate having been breached. Only systems that can recognize command and control without being thrown by the novelty of the tooling can hope to defend against AI-orchestrated campaigns that, in some cases, are already inside.

DeepTempo built one of those systems. The LogLM architecture is a vertical foundation model, the deployment surfaces meet customers where their data is, and the open-source orchestration layer prevents the lock-in that has compromised the proprietary AI SOC category. The validation comes from the customers and partners who shipped with us in stealth: Deutsche Telekom, BNY, Stanford, the Technology Advancement Center, Snowflake, and Cribl.

Next steps

- **Learn more about DeepTempo and the underlying technologies of the LogLM at www.deeptempo.ai**
- **Free trials for on premise and SaaS users are available.**