The Defender's Program for a Post-Mythos World

Exposure management, secure architectures, security operations, and cyber resilience. AI threads through all four.

Written for CISOs of regulated enterprises. The framework can be applied broadly, but the specific recommendations assume a maturity floor.

AHEAD

INTEGRATED SECURITY

In this piece

- **The threat.** The capability to discover previously undiscoverable zero-days is now shared across multiple frontier models. The responsible labs gated their releases. The ones who care less did not. The defender consequence is the same either way.
- **The asymmetry.** AI-powered attackers operate with infinite parallel agents, sub-minute speed, comprehensive knowledge, and marginal cost approaching zero. Defender teams operate with 514K unfilled positions, 30-minute median triage, six to twelve month onboarding cycles, and flat budgets.
- **The framework.** Four areas to rethink: Exposure Management, Secure Architectures, Security Operations, and Cyber Resilience. Using AI for security is no longer a nice-to-have. It is a required thread across all four.
- **Before Friday.** Schedule the tabletop with legal, audit, and CAB. Pull the top fifty internet-facing assets into a single ASM view. Stand up an AI-aided code review pilot. None of this requires Glasswing access.
- **The scorecard.** Four numbers reported quarterly to the board, one per shift.

What Changed

On April 7, 2026, Anthropic released Claude Mythos Preview, a frontier AI model so capable at finding and exploiting software vulnerabilities that the company chose not to release it publicly. Access was granted only to twelve launch partners and roughly 40 critical-infrastructure organizations through a parallel announcement: Project Glasswing. It was the first time a major AI lab restricted a model specifically because its offensive cyber capabilities were too dangerous to put in the wild.

The numbers behind that decision are the part most leadership conversations skip past. They are the part this piece is built around.

MODEL EFFECTIVENESS: THREE NUMBERS THAT REFRAME DEFENSE

Figure 1

The cost number is the leading indicator. Confirmed zero-day discovery dropped from roughly $25,000 in expert labor and infrastructure to $0.11 in compute. A 227,000-fold collapse. Compute scales infinitely; experts don't. The skill collapse is the second-order effect: an eighteen-word plain-English prompt replaced what used to require a team with years of specialist training. The time collapse is what executives feel: six weeks of expert work compressed into twenty-seven minutes. None of these numbers come from the model trying hard. They come from a single pass.

Across testing, Mythos identified thousands of zero-day vulnerabilities: a 27-year-old TCP SACK overflow in OpenBSD, a 16-year-old codec flaw in FFmpeg that survived five million automated fuzzer runs, a 17-year-old FreeBSD NFS remote-code execution that Mythos autonomously weaponized into a working exploit. Every one of those bugs had been in production for over a decade. None were caught by human review or by existing automated tools. The bugs weren't fixed; they were never found. Mythos didn't add the bugs. It removed the cover that hid them.

And the Substrate Is Growing Faster Than Anyone Is Reviewing It

Forty years of unreviewed code in OS kernels, language runtimes, and protocol stacks. Decades of hidden bugs that needed an expert team months to find and a frontier model minutes. And on top of that fixed-cost layer, a variable layer that is now growing faster than any review cycle has ever been able to keep up with.

THE THREAT SUBSTRATE — DECADES OF BAD CODE, THE ICEBERG, AND VIBE CODING

Figure 2

Vibe coding, the practice of building software primarily through AI prompts with limited formal review, has become the default mode of development inside enterprises and out. GitHub Octoverse data shows 46% of all new code is AI-generated. Veracode's 2025 GenAI Code Security Report, testing over 100 large language models against 80 OWASP-aligned coding tasks, found that 45% of AI-generated code introduces security vulnerabilities. The numbers compose: most enterprises are now writing more code than ever before, and a higher percentage of that code is shipping with flaws than human-written code did.

This is not a future concern. It is the present-day attack surface Mythos is searching. Every organization is now running thousands or millions of lines of code its security team has never seen, written by models that were not optimizing for security. The substrate Mythos finds vulnerabilities in is growing daily, faster than any traditional review cycle could ever cover. And the next generation of frontier-class models will read it just as easily as Mythos reads OpenBSD.

Over 99% of Mythos's findings remain unpatched while coordinated disclosure runs. The public Glasswing report lands in early July 2026. At that moment, the rest of the world has the same vulnerability map Anthropic's red team has been working from since April. Even containment has already partly failed. Bloomberg reported that a private Discord forum gained unauthorized access through a third-party vendor environment the same day Mythos was announced. Glasswing was defeated through the supply chain, not the model itself.

Gated, Gated, and Proliferating

The story most leadership conversations land on is Mythos. The story defenders need to plan against is broader. Frontier offensive cyber capability is now shared across multiple AI companies and is not under any single organization's control.

Three categories of provider matter. The responsible labs gated their releases. The merely commercial labs gated theirs slightly less. The third category, every other lab and every open-weights derivative, did not gate anything. **The defender consequence is the same in all three cases.**

CAPABILITY AVAILABILITY — MYTHOS, GPT-5.5, AND THE PROLIFERATION CURVE

Figure 3

MYTHOS · GATED	GPT-5.5 · GATED	EVERY OTHER LAB · PROLIFERATING
Anthropic	**OpenAI**	**Frontier-class capability**
April 7, 2026	April 23, 2026	**Google DeepMind, Meta, xAI, Mistral, open-weights**
Project Glasswing	**Trusted Access Controls**	AISLE replicated several Mythos findings using smaller open-weights models. Independent researchers confirmed: the capability is not Mythos-specific. It is a property of frontier-class AI.
First time a major lab restricted a model specifically because the offensive cyber capability was too dangerous to release broadly. Access limited to twelve launch partners and ~40 critical-infrastructure organizations.	Sixteen days after Mythos. Roughly at parity on cyber benchmarks. Access controlled, broadly available to qualified customers.	
Bloomberg: unauthorized access through a third-party vendor within 24 hours.	Two products. Both gated. Both already shipped within sixteen days of each other.	Comparable access from ungoverned sources: 6 to 18 months.

OpenAI shipped GPT-5.5 sixteen days after Mythos, roughly at parity on cyber benchmarks, broadly available to qualified customers under Trusted Access Controls. AISLE replicated several of Mythos's flagship findings using smaller open-weights models. The capability is not Mythos-specific; it is a property of frontier-class AI. Restrictive access programs slow proliferation. They do not prevent it.

Same Task. Four Eras. One Trajectory.

The capability is not a step function from nothing to Mythos. It is a curve, and Mythos is the first inflection on the curve that defenders cannot ignore. The same offensive task, run by a skilled human and three frontier-class systems, traces the trajectory.

THE CAPABILITY TRAJECTORY — SAME OFFENSIVE TASK ACROSS FOUR ERAS

Figure 4

METRIC	TRADITIONAL *Skilled human*	OPUS 4.6 *Production*	GPT-5.5 *Production*	MYTHOS *April 2026*
Cybench solve rate *35 CTF challenges*	<5%	62%	87%	100%
Time to working exploit *End-to-end zero-day*	6 weeks	14 hrs	2.5 hrs	27 min
Autonomous exploit chains *Of 20 hardened targets*	0/20	2/20	11/20	18/20

GPT-5.5 is in production today and clears 87% of CTF challenges, finds an exploit in two-and-a-half hours, and chains autonomous exploits in over half of hardened targets. Mythos at one-hundred percent and 27 minutes is the headline; GPT-5.5 at 87% and 2.5 hours is the present. The slope is the threat; the slope is also why “wait and see” is no longer an option.

WHAT THIS MEANS FOR PROGRAM DESIGN

Plan for capability availability, not for one model release. The defender’s timeline is set by the third category, not the first. Glasswing buys a quarter. The proliferation curve is the one that determines program urgency.

The Remediation Challenge: Why This Is Structurally Different

Enterprise security programs have been optimized for a specific threat model: skilled human attackers, days or weeks to chain exploits, vulnerability-to-exploit windows measured in weeks. Mythos breaks the model on six operational axes simultaneously.

THE REMEDIATION CHALLENGE — AI-POWERED ATTACK VS. OUR TEAMS ACROSS SIX OPERATIONAL AXES

Figure 5

CATEGORY	AI-POWERED ATTACK	OUR TEAMS
Talent	Infinite parallel agents. No fatigue, no burnout, no onboarding cycles.	514K unfilled US positions. 4.8M global gap. 19% YoY shortage growth.
Speed	22 seconds from compromise to ransomware operator on keyboard.	Median SOC alert triage: 30+ minutes. Critical-vuln MTTR: weeks.
Knowledge	Infinite parallel agents. No fatigue, no burnout, no onboarding cycles.	514K unfilled US positions. 4.8M global gap. 19% YoY shortage growth.
Volume	22 seconds from compromise to ransomware operator on keyboard.	Median SOC alert triage: 30+ minutes. Mean time to remediate critical vulns: weeks.
Prioritization	Reasons across full attack chains. Identifies the 18 of 20 hardened targets where exploitation actually succeeds.	Patches in disclosed-CVE order. Treats CVSS as priority. Loses cycles on findings that don't matter while missing ones that do.
Cost	$0.11 per zero-day. Marginal cost approaches zero at scale.	$200K+ per analyst. Tool stack growing 30% YoY. Budgets flat.

This is the case for why the response cannot be "work harder." On every axis above, the defender cannot match the attacker by adding more analyst hours, larger triage queues, or faster patch cycles within the existing operating model. The asymmetry is structural, not operational. Defenders close it by changing how the SOC operates: AI threaded through detection, triage, and response; compensating controls credited automatically; remediation routed agentically across change windows; and recovery validated continuously rather than confirmed in tabletop.

That is what the four shifts below describe. AI for security is the operational thread that makes them possible at the cadence the threat now requires.

The Four Shifts

Autonomous exploit development compresses the vulnerability-to-exploit window from days to hours, and frontier-class capability finds zero-days directly without waiting for any disclosure cycle to begin. That forces a reset across four interlocking disciplines, with AI for security threaded through all of them.

THE FOUR SHIFTS FRAMEWORK — OUTCOMES, SLA TARGETS, AND AI FOR SECURITY PER SHIFT

Figure 6

SHIFT 01	SHIFT 02	SHIFT 03	SHIFT 04
Exposure Management	**Secure Architectures**	**Security Operations**	**Cyber Resilience**
This Quarter	*Multi-Year*	*Continuous*	*6 – 12 Months*
Find what matters first. Reduce critical exposures through context-aware triage and compensating controls. Detection runs ahead of disclosure.	**Compress blast radius. Defense in depth, zero trust, and identity-first design across human, non-human, and agentic principals.**	**Match attacker speed. Detect and respond in minutes, not days. AI-assisted triage, behavioral analytics, and agentic response.**	**Recover faster than damage compounds. Immutable vault, isolated recovery environment, and quarterly exercises in scope.**
SLA TARGETS	SLA TARGETS	SLA TARGETS	SLA TARGETS
• ≤ 24 hr critical patch SLA on internet-facing assets • ≥ 70% compensating-control coverage on critical CVEs • Continuous SBOM-based exposure hunting in place	• 3 of 5 pillars at ZTMM Advanced within 12 months • Non-human identity inventory and scope-bound tokens • AI firewall, AI gateway, and AI/MCP gateway in place	• MTTD in minutes against simulated lateral movement • MTTR ≤ 72 hr on Tier 0 high-confidence detections • MITRE ATT&CK technique coverage tracked quarterly	• Tier 0 / Tier 1 service recovery within 72 hours • Vault unreachable from production domain-admin • Quarterly exercises with identity and virtualization
AI ON EXPOSURE	AI ON ARCHITECTURE	AI ON RESPONSE	AI ON RECOVERY
Find what attackers find first. Continuous autonomous AI penetration testing. AI hunters running 24 hours a day, surfacing vulnerabilities ahead of disclosure.	*Govern the agent in your stack. AI firewall and AI/MCP gateway in place. Non-human identity least privilege. Inline protection from prompt injection and excessive agency.*	*Match attacker speed. Sub-minute detection and response on high-confidence alerts. Agent-led triage. AI-generated human-in-the-loop response.*	*Validate that you can actually recover. AI-native, enterprise-resilient recovery validation. Continuous monitoring of recovery environment, security tools, and backup data.*

Exposure management is where the volume actually hits. **Secure architectures** compress blast radius so no single compromise becomes a business event. **Security operations** carries the detect-and-respond muscle. **Cyber resilience** ensures the business recovers when prevention and detection both fail.

Shift 01 Start This Quarter: Exposure Management

The wave is already visible. Daniel Stenberg, who maintains the open-source cURL project that ships in nearly every enterprise stack, watched security submission rates climb from two per week to peaks of 8× historical norms through 2025. He shut down cURL's HackerOne bug bounty program in January 2026 after the volume became unmanageable. Enterprise vulnerability management teams are next.

The starting point is Gartner's Continuous Threat Exposure Management framework. The Mythos-era refinement integrates Anthropic's four operational priorities: close the remediation gap (24 hours on internet-facing assets), prepare for 10x volume (automate triage end-to-end), build a unified inventory, and tighten mitigations (favor compensating controls).

A Four-Phase Exposure Lifecycle: AHEAD's Sightline Program

AHEAD's Sightline Program operationalizes Exposure Management as a four-phase lifecycle, run as a six-to-twelve-month engagement on your team. The pattern is visible end-to-end before any individual phase is deployed in detail. Where Figure 6 named the AI for Security capability at the shift level (AI on Exposure), Figure 7 decomposes that capability into four phase-specific applications, with the cross-platform remediation layer running underneath all of them.

THE SIGHTLINE PROGRAM — FOUR-PHASE EXPOSURE LIFECYCLE WITH CROSS-PLATFORM REMEDIATION

Figure 7

Each phase below describes the work in detail. The phases run continuously and reinforce one another: detection feeds triage, triage drives compensating-control activation, validation tests whether everything actually holds. Phase One is where most programs sit at the start of an engagement; Phase Four is where mature programs catch the failures the first three phases missed.

CONTINUOUS DETECTION AHEAD OF CVE ENUMERATION

SBOM-based hunting finds exposure before a CVE is formally published: you see the problem before the attacker does. Detection runs on a unified inventory reconciling CMDB, ASM, CSPM, and DSPM. These four sources each see a different slice of your attack surface and rarely agree. Reconcile the feeds you already pay for before investing in new ones.

Attack Surface Management is now required, not optional. *ASM continuously discovers internet-facing assets the organization does not know it has: shadow IT subdomains, abandoned cloud accounts, forgotten dev environments. Anything ASM doesn't see, the adversary will.*

ENVIRONMENTAL CONTEXT AND APPLICATION TIERING

Automated triage against runtime exposure, internet reachability, active-exploit intelligence, and compensating-control state. The output is the thirty items that matter this week, not the thirty thousand technically open.

Application tiering is the business-context layer most programs don't apply. *A Tier 0 asset (identity providers, domain controllers, crown-jewel databases) tolerates near-zero exposure window. Without this classification, a Tier 0 exposure and a commodity Tier 3 finding compete for the same triage queue.*

COMPENSATING CONTROLS: USE WHAT YOU ALREADY HAVE BOUGHT

Most enterprises already own the controls that could mitigate the exposure, but they sit in monitor mode, are misconfigured or untuned, or are shelfware with capabilities that were never operationalized. In enterprise security, compensating controls rarely have an owner. Five surfaces are worth addressing first:

- ***Infrastructure:*** *NGFW, EDR, microsegmentation, CNAPP, DNS security, cloud security groups.*
- ***Application:*** *WAF, AI firewalls, AI/MCP gateways, RASP, SCA/SAST, CI/CD pipeline gates.*
- ***Identity:*** *ITDR, PAM, conditional access across human and non-human identities.*
- ***Data:*** *Data classification, DLP, DSPM, CASB, tokenization.*
- ***Cyber Recovery:*** *Immutable backups, isolated recovery environments, cyber recovery vaults, tested restore runbooks.*

Take Log4Shell as the canonical case. Every enterprise had Java applications they could not remediate on day one. Six compensating controls, none of them remediation, reduced exposure while the patch program caught up: WAF rules blocking the ${jndi:} pattern; EDR rules flagging the Java-to-shell process chain; network egress filters blocking outbound LDAP and RMI from application servers; SCA tooling flagging the vulnerable Log4j version in the build pipeline; RASP blocking deserialization at runtime; a single JVM flag disabling the vulnerable feature in some environments.

A rule in block mode or a pipeline gate that rejects a vulnerable dependency buys days for teams to plan remediation. In some environments, especially agentic environments, these controls aren't a bridge to remediation. They are the architecture.

EBPF MITIGATION: THE NEW FRONTIER OF COMPENSATING CONTROLS

The most consequential advance in compensating controls since Log4Shell is kernel-level mitigation through eBPF. Cisco Live Protect, released on Nexus 9000-series switches in late 2025 and reaching enforce mode in early 2026, embeds the Isovalent Tetragon agent natively into NX-OS to compile CVE compensating controls into eBPF policies that execute in the switch control-plane kernel. A vulnerability-specific shield deploys without a reboot, without a software upgrade, and without a maintenance window. Cisco's own framing of the gap is direct: enterprises take roughly 45 days to patch a newly disclosed vulnerability; attackers exploit it in three. Live Protect closes that exposure window inline.

For network infrastructure, historically the slowest layer to patch because of change windows, blast radius, and uptime SLAs, this is structurally significant. The same eBPF technology that hyperscalers like Google, Meta, and Netflix run in production now executes inside the control plane of enterprise switches, applying compensating controls against privilege escalation CVEs and control-plane DDoS in real time. The mitigation runs at the kernel with full system context and minimal latency, and rolls back as cleanly as it deployed. Other vendors will follow. The pattern to track: programmatic, kernel-level, hot-deployable mitigation as a structural answer to the patch-cycle problem. This is what "use what you already have bought" looks like at the leading edge. The compensating control becomes the deployment vehicle for vulnerability response, not a holding action while you wait for the patch.

For non-patchable OT environments, the playbook inverts. When you cannot apply preventive controls, shift to detective: passive network monitoring with industrial-protocol deep packet inspection (Dragos, Claroty, Nozomi, Armis) and baseline-anomaly detection on OT protocols.

INDEPENDENT CONFIRMATION THAT IT ACTUALLY HOLDS

External Validation is outside-in. Continuous pen testing and attack-path validation platforms (Tenzai) test whether reachability claims hold under adversary conditions and whether compensating controls actually block what they claim to block.

A clean internal dashboard paired with a dirty external validation report is one of the most common patterns behind breaches that "shouldn't have happened." Run quarterly at minimum, not as a compliance exercise, as an operational one.

Cross-Platform Remediation

Where programs actually fail: *A critical web-application CVE may require a code patch from the application team, a WAF rule from the network team, a cloud security group change, an identity rotation, an EDR policy update, and a CMDB update. Each team runs different tools, different change windows, and different SLAs. This is where most enterprise vulnerability management programs break down.*

In regulated industries, compression requires legal, audit, and CAB alignment on emergency change paths. In tabletops run with regulated-enterprise clients in late 2025, the existing emergency-change path required between five and fourteen days to patch a critical internet-facing asset. Compressing that path inside 24 hours took six to twelve weeks of cross-functional work. None of those organizations had run the exercise before the CVE forced it.

30-day patch SLAs on internet-facing assets stopped being a defensible breach response somewhere between Volt Typhoon and Mythos.

Where AHEAD closes this gap. The Sightline Program is built around a force-extension bench specifically for this problem. See Where AHEAD Engages later in this piece for how the cross-platform execution work gets done in practice.

Shift 02 Multi-Year: Secure Architectures

Patching faster is necessary and not sufficient. Mythos forces defense-in-depth, the oldest principle in security architecture, quietly deprioritized across a decade of platform consolidation. Defense-in-depth is the approach. Zero trust is the posture inside it. CISA's Zero Trust Maturity Model is how to measure progress. Mythos does not change the model; it changes the cost of staying at Traditional or Initial.

Where You Need to Be: CISA Zero Trust Maturity Model v2.0

Five pillars, four maturity stages, and a two-phase post-Mythos target. Phase 1 sets the minimum: Advanced rating on at least three of the five pillars within twelve months. Phase 2 sets the destination: Optimal across the program over the multi-year horizon. Traditional and Initial are the cost centers. Every additional quarter at those stages compounds risk that compensating controls cannot fully offset.

CISA ZERO TRUST MATURITY MODEL V2.0 — POST-MYTHOS TARGET PROGRESSION (PHASE 1: ADVANCED → PHASE 2: OPTIMAL)

Figure 8

PILLAR	TRADITIONAL	INITIAL	ADVANCED ← *PHASE 1 MINIMUM*	OPTIMAL ← *PHASE 2 TARGET*
Identity	Static perms, partial MFA	Risk-based access	**Continuous validation, ITDR, NHI inventory**	**Adaptive, ML-driven**
Devices	Inventory partial, EDR uneven	EDR baseline, MDM partial	**Full posture, hardware attestation**	**Continuous trust scoring**
Networks	Network-perimeter access	Initial micro-segmentation	**Identity-aware SASE, micro-seg**	**Data-aware policies, full SASE**
Apps & Workloads	Perimeter-only security	Initial workload identity	**Runtime attestation, workload IDs**	**Continuous policy validation**
Data	Partial classification, basic DLP	DLP at boundary, CASB initial	**DSPM + automated classification**	**Continuous lineage and policy**

The pillars below describe the work each row implies in practice. None of this is new conceptually. What Mythos changes is the deadline.

Identity

Identity is human, non-human, and agentic. Non-human identities (service accounts, workload identities, API keys, OAuth tokens) outnumber human ones fifty to one in most enterprises. Tokens don't MFA; they are bearer credentials. Agentic controls that matter now: a non-human identity inventory covering agents alongside services and workloads; short-lived, scope-bound token issuance; and runtime behavioral baselines that treat the agent itself as a monitored principal.

Devices

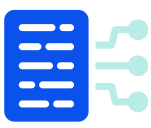
Device posture is the first access decision. Endpoint compliance signals (patch state, disk encryption, EDR active), mobile device management, and hardware-backed attestation determine whether an authenticated identity actually gets access. IoT and OT endpoints require compensating controls at the network layer.

Networks

SSE and SASE are zero trust expressed at the network layer: identity-aware access with consistent policy across a distributed workforce. Micro-segmentation contains an intrusion to the workload where it began. For OT, which runs 10–18 years and is often unpatchable, the answer is segmentation, detective controls, and ruthless isolation from IT.

Apps & Workloads

Treat each workload as a first-class identity with its own authentication, least-privilege authorization, runtime attestation, and continuous validation. Secure-by-default deployment patterns (infrastructure-as-code with policy guardrails, signed artifacts, immutable images) cost far less than retrofitting security after the fact.



Data

Encryption is table stakes. The harder work is classification at volume, DSPM instrumentation in cloud environments, and treating data flow as a security control rather than a compliance artifact. Egress controls (DLP, CASB) enforce policy at the boundary. An attacker's objective is usually your data; govern it accordingly.

Shift 03 Continuous Security Operations

Signature detection cannot keep up with AI-assisted exploitation. The frontier-era SOC has to detect on behavior, not on hashes. Mean time to detect against an attacker who can chain custom exploits in hours decides whether an incident becomes a breach. Mandiant's 2025 global median dwell time was 14 days. The Mythos-era target is minutes.

Behavioral Analytics

UEBA across human, non-human, and workload identities. A service account that suddenly opens a database connection from a new region is a stronger signal than the equivalent human anomaly; services do not improvise. Identity Threat Detection and Response (ITDR) operationalizes this and belongs in every Mythos-era SOC stack.

Cross-Domain Correlation

Endpoint, network, identity, cloud control plane, and SaaS audit logs in one detection graph mapped to MITRE ATT&CK. The technique-coverage question to track quarterly: which ATT&CK techniques relevant to your threat profile are covered by at least one high-confidence detection across at least two telemetry sources.

Detection Engineering

Detections are code: version-controlled, reviewed, tested against red-team output, retired when stale. The metric that travels to the board is technique coverage against MITRE ATT&CK relevant to your threat profile, measured quarterly against simulated lateral movement, not against inbox alert counts. Mandiant's 2025 global median dwell time was 14 days. The post-Mythos target is minutes. Closing that gap is a SOC-modernization program, not a tooling refresh.

Response Automation

Isolate endpoint, revoke token, block egress, disable account, kill workload. Five response actions that should run inside minutes on high-confidence detection without human approval. Defining those predicates honestly, in tabletop, with legal, with the engineering teams whose workloads might be killed, is the real work.

Shift 04 6–12 Months: Cyber Resilience

> ***The attackers are targeting the tools you bought to stop them.***
>
> *Mandiant M-Trends 2025*

Prevention never reaches zero. In a world where frontier-class offensive AI is broadly available, the planning assumption is not if but when. Cyber resilience makes recovery a measurable capability rather than a stated intention: immutable storage no in-band credential can reach, a standby production environment sized to run the business at a Minimum Viable Company level, an isolated environment to bring Tier 0 services back, a clean room for forensics, and a recovery cadence exercised quarterly with identity and virtualization planes in scope.

TIER 0 DEPENDENCY MAPPING

Most enterprises know which applications are Tier 1. Few have an honest map of the Tier 0 services those applications depend on (Active Directory, DNS, certificate authorities, virtualization control planes, identity providers, secrets managers) and the order they have to come back in. The mapping is the first artifact of a credible recovery program.

IMMUTABLE VAULT AND ISOLATED RECOVERY ENVIRONMENT

The vault holds copies that cannot be altered or deleted by an in-band credential, including domain-admin. The Isolated Recovery Environment (IRE) is the segregated infrastructure where those copies become running services. If a domain-admin credential in production can reach either, you do not have a recovery capability; you have a backup.

STANDBY PRODUCTION ENVIRONMENT (MINIMUM VIABLE COMPANY)

The vault and IRE answer what to restore. The standby production environment answers where to run the business while production recovers. If production is ransomed and the IRE is your only target, you are recovering into the crime scene. A fully isolated standby environment, scoped to the Minimum Viable Company (the critical processes, customer-facing services, and revenue functions that cannot go dark) is what makes recovery a business-continuity event rather than a business interruption. Full parity is a years-long project; MVC is a scoping decision that can be built and tested in a quarter.

CLEAN ROOM

Forensics, malware analysis, and recovery validation happen in a clean room: a separate sterile environment where compromised artifacts can be inspected without re-introducing them to production. Standing one up after the incident is too late. It is a two-quarter build project before one is needed.

RECOVERY EXERCISE DISCIPLINE

Quarterly exercises against named ransomware scenarios, with identity services and virtualization planes in scope, not just application data. The metric that travels to the board is the trailing four-quarter recovery exercise success rate, not the existence of a runbook. Cyber-insurance carriers increasingly ask for it directly.

RANSOMWARE NOW TARGETS YOUR RECOVERY STACK

Per Mandiant M-Trends 2026, the three assets sophisticated ransomware operators target first are (1) identity services, (2) virtualization management planes, and (3) backup infrastructure, the same three you would use to recover from an attack. Resilience planning has to assume all three may be compromised simultaneously.

ARCHITECTURAL CONSEQUENCE

If your IRE and your production identity provider share any credential, any control plane, or any human operator, you do not have an IRE. Treat the recovery stack as adversary-aware infrastructure, not as backup.

A Four-Number Scorecard for the Board

Four metrics, one per shift, reported quarterly.
These are lag indicators that aggregate the operational work underneath.

THE FOUR-NUMBER SCORECARD — ONE LAG INDICATOR PER SHIFT

Figure 9

EXPOSURE	ARCHITECTURE	OPERATIONS	RESILIENCE
≤24 hr	**3 of 5**	**Minutes**	**<72 hr**
Critical patch SLA on internet-facing assets	ZTMM Advanced rating across pillars in 12 months	MTTD against simulated lateral movement	Tier 0 / Tier 1 recovery, exercised quarterly

Exposure finds it first.

Architecture denies reach.

Operations catch the rest.

Resilience holds the line.

ONE QUESTION FOR MONDAY'S LEADERSHIP MEETING

If a frontier-class capability was pointed at your environment this week, what would fail first, and what would fail next?

That answer is where your program starts.

Before Friday: What to Stand Up This Week

None of what follows requires Glasswing access, a budget cycle, or a tooling decision. They are the lowest-cost moves that close the largest exposure gaps, and they are the four items that should be on the program-management calendar before this week ends.

Schedule the cross-functional tabletop. Legal, audit, change advisory board, and the security leadership team in one room, walking through a single critical CVE on an internet-facing asset. Target a 24-hour patch path. The first time you run this exercise should not be the first time a CVE forces it.

Pull the top fifty internet-facing assets into a single ASM view. Reconcile what CMDB says you have against what your attack-surface management tool sees. The delta is the work. Anything ASM does not see, the adversary will.

Stand up an AI-aided code review pilot. On one repository, against one team, with one set of acceptance criteria. The point is to learn the failure modes of AI-generated code review against your codebase before the substrate problem becomes the breach.

Inventory non-human identities and agentic workloads. Service accounts, workload identities, API keys, OAuth tokens, deployed agents, and what data and tools each can reach. The agents you deploy on defense need the same governance discipline as the ones an adversary deploys on offense.

Each of these maps to a shift. Each surfaces a measurable result inside thirty days. Each is the entry point to one of the AHEAD engagements that follow.

Where AHEAD Engages

AI for security threads through all four engagements below as the operational execution layer. The outcomes show up in faster remediation on Exposure Management and faster detection on Security Operations, with supervised-autonomous remediation as the longer horizon.

WHERE AHEAD ENGAGES — PRACTICE AREAS MAPPED TO THE FOUR SHIFTS

Figure 10

EXPOSURE MGMT	ARCHITECTURES	OPERATIONS	RESILIENCE
Sightline Program	**Zero Trust + Secure AI Architecture**	**SOC Modernization**	**Cyber Recovery**
Six-to-twelve-month engagement running the four-phase exposure lifecycle on your team. Cross-domain remediation bench. Discipline gets built on your staff, not parked on ours.	**CISA ZTMM current-to-target roadmap across five pillars, paired with Secure Runtime and Secure AI Architecture: AI gateway, AI firewall, and AI/MCP gateway selection and implementation.**	**Inventory the existing SIEM, SOAR, XDR, and threat-intel stack against capability requirements. Behavioral analytics and ATT&CK-aligned correlation come in next, not first.**	**Tier 0 dependency mapping that most enterprises do not have. Immutable vault and IRE built around those services. Quarterly recovery exercises with identity and virtualization in scope.**
OUTCOMES TRACKED	OUTCOMES TRACKED	OUTCOMES TRACKED	OUTCOMES TRACKED
• Remediation velocity ≤ 24 hours on internet-facing assets • Compensating-control coverage above 70% on critical CVEs • Reachability-validated finding ratio: signal versus noise	• Advanced ZTMM rating on 3 or more pillars in 12 months • AI gateway, AI firewall, and AI/MCP gateway in place • Lateral-movement containment time in tabletop simulation	• MTTD in minutes against simulated lateral movement • MTTR on high-confidence detections within shift SLA • MITRE ATT&CK technique coverage tracked quarterly	• Tier 0 / Tier 1 service recovery within 72 hours • Vault unreachable from production domain-admin credential • Recovery-exercise success rate over trailing four quarters

Sources

[1] Anthropic, Claude Mythos Preview red team report (red.anthropic.com/2026/mythos-preview/)

[2] Anthropic, Claude Mythos Preview system card (projected 244-page document)

[3] Anthropic, Project Glasswing announcement (anthropic.com/glasswing)

[4] OpenAI, Introducing GPT-5.5, April 23, 2026

[5] AISLE, AI Cybersecurity After Mythos: The Jagged Frontier (Stanislav Fort, April 8, 2026)

[6] UC Berkeley, Wang et al., CyberGym: Evaluating AI Agents' Real-World Cybersecurity Capabilities at Scale (arXiv:2506.02548)

[7] Cloud Security Alliance, SANS Institute, OWASP Gen AI Security Project, Building a Mythos-ready Security Program (April 12, 2026)

[8] Bloomberg, Anthropic's Mythos model accessed by unauthorized users

[9] Daniel Stenberg, daniel.haxx.se, cURL submission-rate data and bug bounty program closure

[10] Mandiant, M-Trends 2025 and M-Trends 2026 annual reports

[11] Synopsys, 2026 Open Source Security and Risk Analysis

[12] Veracode, 2025 GenAI Code Security Report (100+ LLMs tested across 80 OWASP-aligned tasks)

[13] GitHub Octoverse 2025 (AI-generated code adoption metrics)

[14] Cisco, NX-OS Live Protect feature documentation (Nexus 9000-series, Isovalent Tetragon agent)

[15] Google Threat Intelligence Group, 2025 zero-day exploitation analysis

[16] CISA, Zero Trust Maturity Model v2.0

[17] ISC2 / CyberSeek, cybersecurity workforce supply/demand data (514,000 unfilled U.S. positions)

About the Author

Felix Vargas

SENIOR DIRECTOR, SPECIALIST SOLUTION ENGINEERING

Felix Vargas is a senior leader in AHEAD's security practice, advising enterprises on modernizing their defenses across cloud, data center, and AI-powered environments. He specializes in zero trust architecture, secure-by-design infrastructure, and aligning cybersecurity strategy with business outcomes, and frequently works with customers building on NVIDIA technologies. Felix has more than 15 years of experience helping organizations reduce risk while accelerating innovation, and is a regular speaker on emerging threats, AI security, and the future of cyber resilience.

AHEAD

Combining cloud-native capabilities in software and data engineering with an unparalleled track record of modernizing infrastructure, we're uniquely positioned to help accelerate the promise of digital transformation.

Visit us at [ahead.com.](ahead.com)

National Hubs

CHICAGO
444 W. Lake Street
Suite 3000
Chicago, IL 60606

NEW YORK
500 5th Avenue
Floor 17
New York, NY 10010

ATLANTA
1117 Perimeter Center
W406
Atlanta, GA 30338

SAN FRANCISCO
2000 Crow Canyon Place
Suite 250
San Ramon, CA 94583