AIUC-1

After Mythos: Defending at Machine Speed

A whitepaper presented by the AIUC-1 Consortium

Co-authors

Mandy Andress, CISO, Elastic
Neil Bennett, CISO, Post Office Ltd
Dr. Manfred Boudreaux-Dehmer, Former CIO & CISO, NATO
David Campbell, Head of AI Security Research, Scale AI
Jason Clinton, personal capacity
Brett Cumming, F500 CISO
Rajiv Dattani, Co-founder, AIUC
Jen Easterly, CEO, RSAC & Former Director, CISA
Gadi Evron, CEO, Knostic, CISO-in-Residence for AI, Cloud Security Alliance
Lars Falch, Partner, Kopenhagen Konsulting & ex-CISO, Novo Nordisk
Christian Gorke, Group CISO, Deutsche Börse Group
Bil Harmer, CISO, Supabase
Erik Hart, Cushman & Wakefield
Jimmy Heschl, CISO, Red Bull
Matt Hillary, SVP, Security & CISO, Drata
Mark Hillick, CISO, Brex
Heather Hinton, PhD, CISO, Sitecore
Simon Hodgkinson, Strategic Advisor and Former CISO, BP
John Israel, Global CISO, KPMG
Amyn Jan, Chief AI Architect, CDAO, US Department of War
Katie Jenkins, EVP & CISO, Liberty Mutual
Omar Khawaja, Field CISO, Databricks
Santosh Kumar, Chief Security Architect, Cisco
Rune Kvist, Co-founder, AIUC
Louise McElvogue, Board Director, Fluency
Hardik Mehta, Head of Payments, Contractual, Regulatory Risk Assessments, JPMorganChase
Xabier Muruaga, Global Head of AI & Data, Iberdrola
David Mussington PhD, University of Maryland
Daniel Nunez, CISO, NYC Employees' Retirement System
Kumar Palaniappan, CISO, Cloud Software Group
Algirde Pipikaite, Head of GRC, CrowdStrike
Kevin R. Powers, Faculty Director & Lecturer-in-Law, MLS in Cybersecurity, Risk & Governance, Boston College Law School
Jim Reavis, CEO, Cloud Security Alliance & CSAI Foundation
Scott Roberts, CISO, UiPath
Adeel Saeed, SVP, CTO Cyber, Kyndryl
Indu Sajeev, CISO
Chris Sandulow, CISO, Confluent
Rahat Sethi, Sr Director TechGRC, Adobe
Rinki Sethi, CISO & CSO, Upwind Security
Alex Shamsutdinov, VP CISO, Upwork
Gagandeep Singh, VP, Global Compliance & Certifications, Salesforce
Rajeev Singh, Group SVP, NTT Data
Lena Smart, Ex-CISO, MongoDB & Ambassador, AIUC-1
Ravi Soin, CIO & CISO, Smartsheet
Phil Venables, Partner, Ballistic Ventures
Dan Walsh, CISO, Datavant
Nancy Wang, CTO, 1Password
Craig Weatherhead, SVP IT Infrastructure & Security, Fastenal Company
Min Xu, Senior Principal Security Architect, Navy Federal Credit Union
Tom Zick, Berkman Klein Center for Internet and Society, Harvard University
Editors: Abby Shen & Emil Lassen, AIUC-1

About the AIUC-1 Consortium

The AIUC-1 consortium unites 200+ CISOs, GRC leaders, practitioners, and academics from leading organizations around the world including Google Cloud, Cisco, Stanford, MITRE, MongoDB and more. Together, the Consortium has developed AIUC-1, the industry-led comprehensive certification standard for AI agents, addressing top enterprise risks.

Thank you to Apostol Vassilev (NIST) for valuable comments on this paper. We note that all the listed co-authors and contributors represent only themselves, and not their employer(s).

Contributors

Trent Adams, Director, Ecosystem Security, Proofpoint
Sanjeev Agarwal, Founder & CISO, Bidodi InfoSec
Abhi Arikapudi, Senior Director for Security Engineering, Databricks
Hammad Atta, Lead Advisor, AI Security Research and Consulting, Qorvex Consulting
Tushar Badlani, Security Specialist - Customer Trust and Third-Party Risk, Figma
Neil Barlow, Advisor & CSO, UnGovr
Jerrad Bartczak, IT Audit Manager - AI & Special Projects, Advantage Partners
Sagar Behere, Head of Third Party Risk Management, Circle Internet Financial
Unai Bermejo, Global AI Expert Engineer, Iberdrola
Richard Bird, Chief Security Officer, Singulr AI
Vidya Bodepudi, Global Director of Cybersecurity, Fuze Health
Chirayu Buch, Principal Cloud GRC, Automation Anywhere
Mihaly Bujaki, Security Manager, Deutsche Telekom
Mark Carney, CEO & Founder, Duro Advisors
Sheron Chakalakal, Head of GRC, UiPath
Julie Chatman, CEO, ResilientTech Advisors
Matthew Chiodi, Chief Strategy Officer, Cerby
Wilson Chiu, Head of Cyber Security, Allianz Australia
Thomas Clarke, Director - AI Safety & Security Lead, Faculty
Sammy Chowdhury, Co-founder & Chief Compliance Officer, Prescient Security
Adnan Dakhwe, CISO & CIO, DelphinusCyber
Drew Danner, Managing Director, Andersen Consulting
Phani Dasari, Global CISO, HGS
Chris DeNoia, CIO/CISO, DeNoia Consulting
Rahul Dubey, Vice President, Global Regulated Markets Solutions, Palo Alto Networks
Amanda Ehrhart, Senior IT Auditor, eDelta Consulting
Dr. Zachary Elewitz, PhD, MBA, AI Executive, Professor
Arvinda Gangadhararao, Director of Compliance, Veeam
Saurabh Ghelani, Global Trust Officer, Confluent
Travis Good, Co-founder & CISO, Workstreet
Shali Goradia, Head of Geo Expansion and Compliance, SailPoint
Bhavya Gupta, Information Security Officer, Stanford University
Miguel Angel Gutiérrez, Regional Lead & vCISO, Toyota Financial Services
Anthony Hannon, Head of Cyber Data Defense, Aon
Peter Holcomb, Founder, OPTIMO IT
Rasha Horn, Founder, CyberNarwhal
Ken Huang, Project Lead, OWASP Top 10 for Agentic AI, OWASP
Chris Hughes, VP, Security Strategy, Zenity
Ankit Jain, Agentic AI Strategy and Implementation Leader, EY
Jess Jeetley, AI Governance Advisor, Techstars
Keren Katz, Co-lead, Agentic Top 10, OWASP
Prateek Kalasannavar, Staff AI Security Engineer, Lenovo
Arpitha Kaushik, Senior Manager, Technical Risk, Marvell Technology
Scott Kennedy, Agentics Security Researcher, Agentics Foundation
Ayman Khalidi, Cloud Security Specialist, Bank of Montreal
Volkan Kutal, Founder & Lead AI Red Teaming Engineer, PaperToCode
Brian Levine, Executive Director, Former Gov
Robert Lowry, CISO, Tonic AI
Mark Lundin, Partner, Tech, Cloud and AI, BDO
Ryan Macfarlane, Incident Response Practice Lead, TrustedSec
Murali Mallina, CTO, Tria Federal
Danny Manimbo, Managing Principal, ISO & AI Practice Leader, Schellman
Annie Mathew, Head of Compliance, Dropbox
Tara Diana Mavrovitis, Head of GRC and Internal Audit, Collibra
Greg McCord, Founder & CISO, Lightcast
William Mendez, Director of Cyber Testing, NYC Cyber Command
Vamsee Krishna Metlapalli, Senior Manager, Tech GRC, Adobe
Chris Monson, Senior Security Architect, Atlassian
Madjid Nakhjiri, Consultant (Former Chief Security Architect), Avancez Group
Vineeth Sai Narajala, Senior AI Researcher, Cisco
Boris Sieklik, Senior Director of Security, MongoDB
Rohit Parchuri, CISO & Board Advisor, Yext
Sanat Pattanaik, Principal Security Architect, ADP
Uma Rajagopal, Devices Security, Amazon
Nicole Reineke, Chief AI Officer, N-able
Enis Sahin, Global Head of Security Architecture, Federated Hermes
Abhinav Singh, Founder, Wingback
Navrina Singh, Founder & CEO, Credo AI
Sean Todd, CISO, Coral
Chris White, Director, Cybersecurity DevRel, NVIDIA
Gill Whitehead, Visiting Policy Fellow, Oxford Internet Institute
Larry Whiteside Jr., Co-Founder & President, Confide Group
Junior Williams, Principal Enterprise Architect, BITSUMMIT
Jordan Worth, Security & GRC, Block

Executive Summary

Mythos' offensive cyber capability was not designed intentionally. It was discovered. Nation-state-grade performance emerged from training a frontier model to write code well, driven by the same scaling dynamics responsible for advances across the frontier. The implication is that this is not an Anthropic story, but an industry inevitability.

Elite offensive cyber capabilities are becoming a commodity. Frontier peers have reached parity shortly after Mythos' release. Already in late April 2026, research demonstrated how open-weight models - wrapped in purpose-built multi-agent orchestration architecture - are reaching parity with frontier models' cyber capabilities. By 2027, it is predicted that the capability that once defined elite offensive teams will be in the hands of all adversary actors - sharply increasing the volume, frequency, and sophistication of novel attacks.

The leaders who build today's programs are candid about the readiness gap. Polling 50+ CISOs and security executives across the AIUC-1 Consortium shows that organizations rate their preparedness for a Mythos-class threat environment at 4 out of 10 on average. They expect this to rise to ~6.7 out of 10 one year from now. The challenge ahead is clear.

We provide three tactical imperatives to close the gap. Each belong in every CISO's playbook and must move from roadmap to operational reality in a matter of months:

1. **Compress the patch window:** Every patch release is now an exploit roadmap for adversaries who compare binaries faster than defenders can deploy. Best practice includes reducing patch SLAs from months to hours, leveraging AI in the remediation loop, and pushing the same discipline into the supply chain.
2. **Design for breach:** Following zero trust architecture, design code and infrastructure containment from the start, so that attackers can't move once inside. Best practice includes constraining AI agents to least privilege, allowlisting what runs on endpoints, segmenting locally so a single foothold stays a single foothold, applying autonomous red teaming, and designing for recovery.
3. **Defend at machine speed:** With Mythos-grade tooling on the offense, defenders must operate on the same cadence as attackers. Best practice includes implementing continuous self-detection of breaches, and deploying AI throughout the SOC from alert prioritization, detection engineering to writing the production fix.

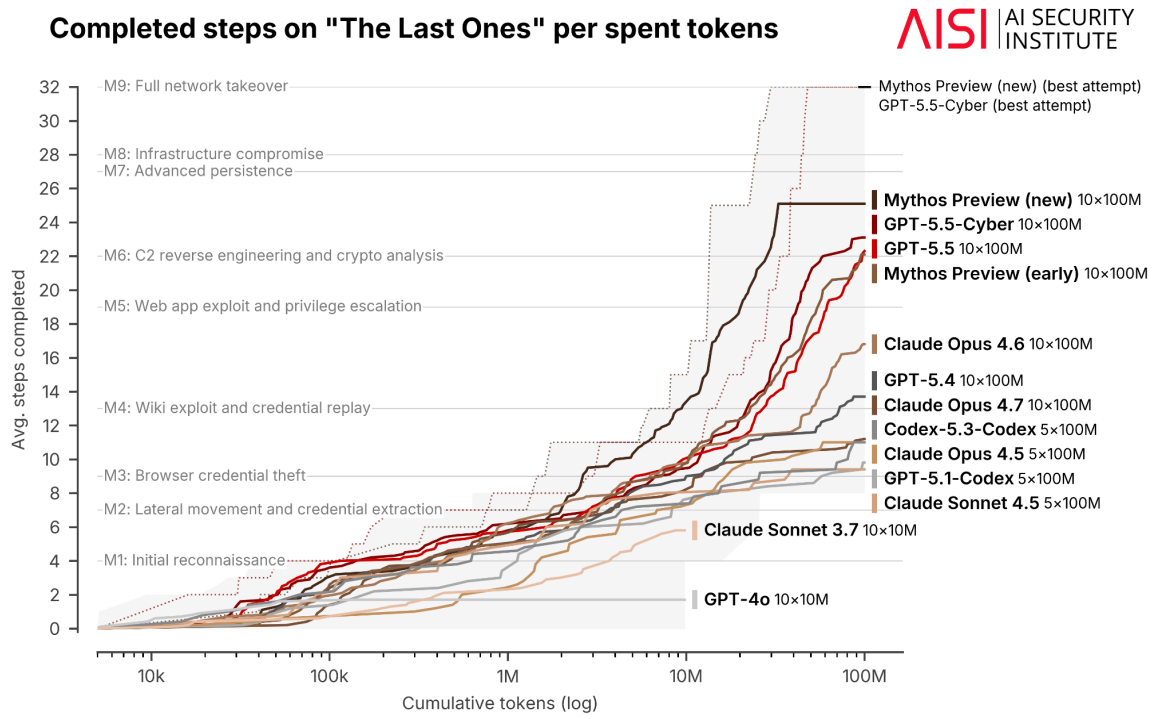
Gold-standard security organizations already run this playbook. What is needed is widespread adoption to turn the practices of a few to the standard methods of many. This paper lists the best practices already used by 120+ security leaders.

The CISO's question is no longer whether adversaries will gain access to these capabilities, but how defenders will operationalize at equivalent speed and scale.

Elite cyber capabilities will be widespread in the coming year

The exploit window has collapsed. The timeline from disclosing a vulnerability to weaponizing it via a working exploit chain has compressed from months to minutes. Already in 2025, the fastest breakout took only 27 seconds; one intrusion documented how data exfiltration began within four minutes of the initial access (CrowdStrike Global Threat Report). The breakout time is expected to decrease – not on a one-off basis but rather in large volume, as cyber offensive capabilities become commoditized.

Anthropic's Claude Mythos Preview in April 2026 solidified this shift: in its early-access cohort, the model surfaced thousands of high-severity flaws across every major operating system and web browser. Mythos' offensive cyber capabilities have been characterized as nation-state grade by the security community.



Frontier models converge on Mythos-level offensive capability as token spend scales. Source: UK AI Security Institute.

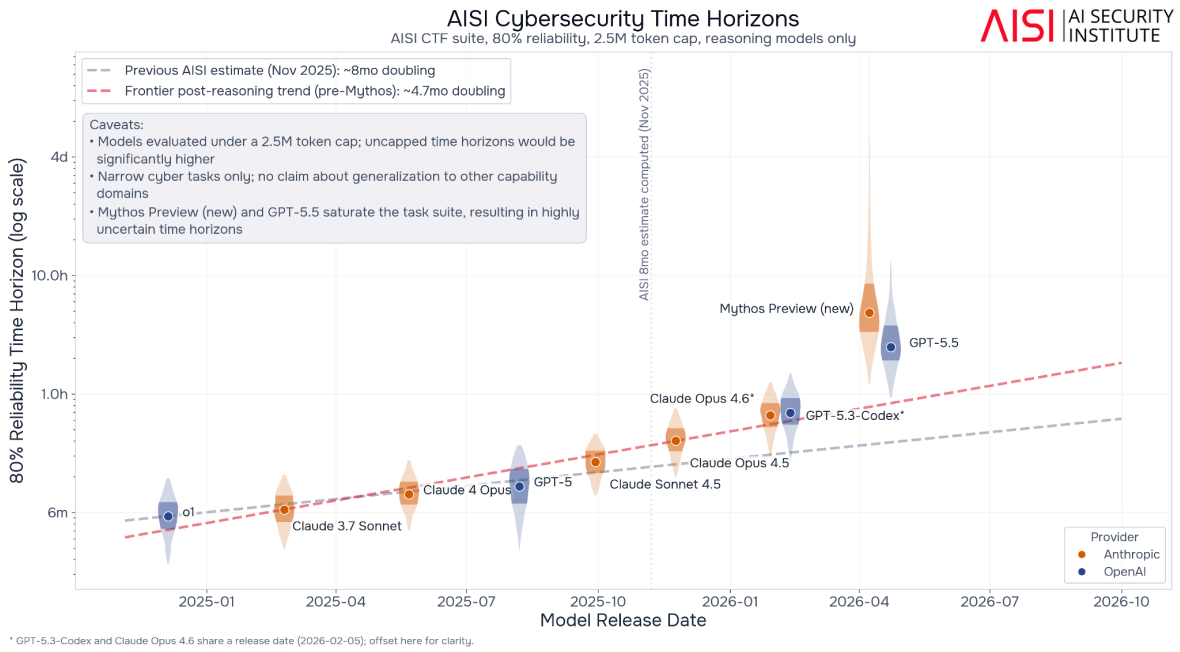
Nation-state-grade cyber capability is no longer scarce. Within weeks of Mythos, OpenAI released GPT-5.5, which the UK AI Security Institute evaluated as slightly outperforming Mythos on complex cybersecurity tasks. With offensive cyber as an emergent property of frontier coding capability, there is growing evidence that the offensive pipeline CISOs need to plan against has already become achievable with the integration of publicly available components and published architectures. Recent research 'Synthesizing Multi-Agent Harnesses for Vulnerability Discovery' (Liu et al., 2026) published in April 2026 evidenced that a mid-tier open-weight model wrapped in a purpose-built multi-agent orchestration architecture, was able to produce ten previously unknown zero-day vulnerabilities in Google Chrome, including two critical sandbox-escape vulnerabilities confirmed by Google (CVE-2026-5280 and CVE-2026-6297).

"*Open weight models with the ability to develop exploits and chain them together will scale threat actor's capabilities in a way we haven't seen before. No company is too small to target. Smaller, unknown and less-sophisticated threat actor groups will quickly ramp in capability. The only way to combat this is to engineer scalable practices that utilize AI to operate at the threat actor's scale and pace*."

Chris Sandulow
CISO, Confluent

Defenders must plan for proliferation. Once open-source model capability improvements allow Mythos-grade weights to become downloadable, the volume, frequency, and sophistication of novel attacks will rise sharply across all industries including sectors that historically relied on obscurity and low adversary interest. CrowdStrike's 2026 Global Threat Report already documents an 89% year-over-year increase in AI-enabled adversary operations.

The strategic implication is unambiguous: defenders must prepare for a world in which nation-state-grade offensive capability is available to every adversary, regardless of resources or sophistication.



Frontier cyber capability is doubling every 4.7 months, down from 8 months in late 2025. Source: UK AI Security Institute.

Every control implemented in existing security programs, such as patch cadence, detection latency, and blast radius tolerance, was calibrated for a world with scarce elite offensive capabilities. This world is ending in months and not years. The overarching question is whether defenders will be ready.

Security leaders do not feel prepared to build a Mythos-ready security program

Existing programs were built for human-constrained threats. Existing practices around enterprise vulnerability management, patching, detection, and remediation were designed for a world where finding a high-severity bug in hardened software took a skilled human days or weeks and where defenders had a reaction window between disclosure and exploitation. This model works when the rate of arrival of vulnerabilities is in line with how fast human researchers could find and disclose bugs.

"If you conduct manual configuration management, patching, and remediation, the coming onslaught of new AI-discovered vulnerabilities will quickly overwhelm the system. My advice is to automate whatever can be automated – use the machine's power to speed-up throughput"

Dr. Manfred Boudreaux-Dehmer
Ex-CIO & CISO, NATO

The defender's planning horizon is shortening. In the limited Mythos early-access cohort, the model surfaced thousands of critical and high priority findings in production software within weeks. This included zero-days in every major operating system and web browser. One example is an OpenBSD flaw that had survived 27 years of expert scrutiny and decades of automated fuzzing. Reproducibility benchmarks show that Mythos produced a working exploit on the first attempt for more than 83% of vulnerabilities as compared to a near-zero rate for the previous frontier generation. As access broadens beyond Project Glasswing, the volume of findings will grow faster than any human-paced disclosure or remediation program can absorb.

Security leaders do not believe that their organizations are prepared for a Mythos-class threat environment. We polled 52 security leaders from the AIUC-1 Consortium - a cohort that includes Fortune 500 CISOs, federal agency leaders, and security executives from sectors spanning banking, retail, healthcare, critical infrastructure, and defense. 65% of their organizations already run AI agents in production today with 1 in 5 reporting business-critical agent deployments.

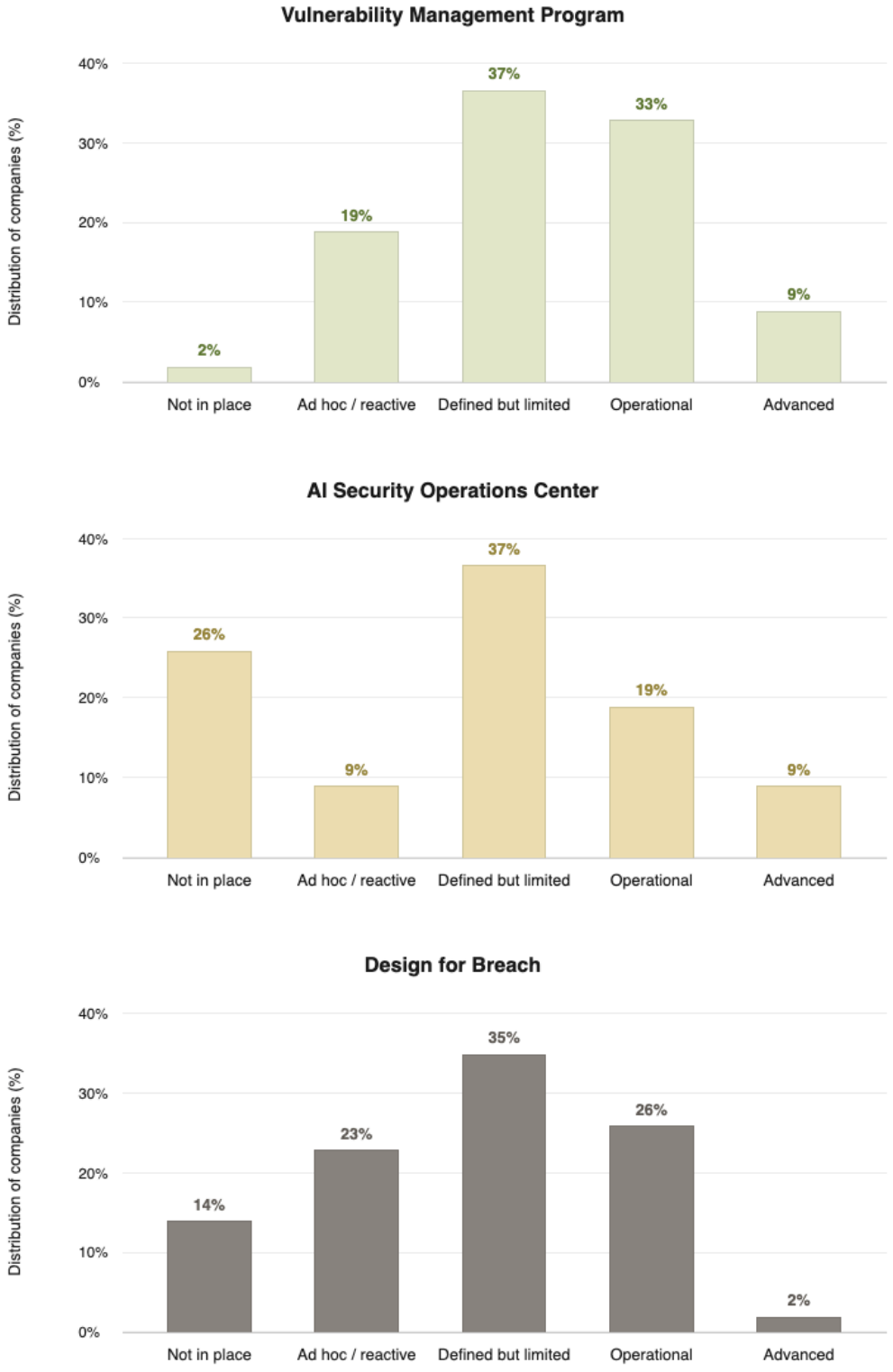
> "*The post-Mythos era won't be won by reaction. AI-enabled security operations are table stakes, but defense has to move upstream, where AI writes, reviews, tests, and pen-tests software at machine speed before it ships. Preparedness at machine speed starts where software and infrastructure are built, not where breaches are caught.*"
>
> **Alex Shamsutdinov**
> VP CISO, Upwork

The foundational defensive capabilities are not where they need to be. We asked respondents to rate the maturity of three capabilities against AI-specific threats. Most organizations are in the “defined but limited” tier for defensive capabilities around vulnerability management for AI systems, AI SOC, and design for breach capabilities.

Key takeaways from the survey:

- **On average, respondents rate their organization's preparedness for a Mythos-class threat (i.e., major cyber incident) today as 4 out of 10.** Approximately 40% rate themselves at 3 or below. Only approximately 12% rated themselves at 7 or higher and approximately 85% scored themselves at 5 or below.
- **The same group expects to reach a preparedness of 6.7 out of 10 in the next twelve months.** Only 63% expect to surpass 7 out of 10. This represents the aspirational forecast from leaders who direct their organization’s security programs, roadmaps, and budgets.

Distribution of organizations by maturity stage across 3 defensive capabilities



Survey results from 52 CISOs / Security Executives from the AIUC-1 Consortium

Three tactical imperatives to close the gap

CISOs and security organizations know what must be done: patching faster, assuming breach, and detecting at speed. The challenge is to now operationalize these concepts in months. We organize the path forward around these three imperatives:

1. Compress the patch cycle

Shipping the fix is now tantamount to vulnerability disclosure. The exposure window starts with the moment that a patch ships and not when the Common Vulnerabilities and Exposures (CVE) is published. A Mythos-class model can take a vendor binary released to patch an undisclosed CVE, compare it against the previous release, identify the changed call paths, infer what the patch was protecting against, and write a working exploit. This can be done without ever seeing the source code.

The ninety-day patch SLA written into most security policies is no longer a viable control. CISA's recent consideration in May 2026 of compressing the Known Exploited Vulnerabilities (KEV) remediation deadline from fourteen days to three is the regulatory acknowledgment of that fact.

What to do

- ☑ **Set patch SLAs that match the threat, while adopting better vulnerability management practices.** Patch SLAs should be measured in hours to days for assets that matter most, and communicated to the business as exposure windows, rather than compliance metrics.

  Leading organizations are already prioritizing actively exploited, internet-facing, and business-critical assets first, from requiring browser updates within two days to operating system updates within five days. Where patches cannot be deployed immediately, compensating controls should reduce exposure until remediation is complete.

  This rests on an operating posture where firstly, planned maintenance is treated as a rhythm rather than an exception and secondly, rollback readiness is treated as the safety net.

☑ **Automate triage and remediation with LLM-based tooling.** Severity assessment, dependency mapping, and increasingly remediation itself are all amenable to AI. Every modified line of code should pass an LLM-driven security review before it is released, and secure-by-design agents should be embedded earlier in the Secure Software Development Lifecycle (SSDLC) so the organization reduces vulnerabilities it later has to patch under pressure.

☑ **Push the same patch discipline into the supply chain.** Third-party as well as open-source components present the largest exploitable surface in most environments.

Vendors should be required to demonstrate short patch SLAs and AI-augmented discovery in their own pipelines as a part of the procurement process. This will require organizations to revisit outdated SLAs in contracts, hold vendors accountable for existing performance, and result in considerable cost of contract renegotiation that should not be underestimated. All commercial vendors that cannot provide security updates are highly recommended to be phased out.

"Secure before release: Security programs must move earlier in the software development and deployment lifecycle. In a Mythos era threat environment, it is no longer sufficient to rely primarily on post-release patching or traditional vulnerability severity models. Organisations should embed AI enabled security tooling throughout the SSDLC to identify, contextualise and prioritise vulnerabilities before software artefacts are deployed into production."

Santosh Kumar
Chief Security Architect, Cisco

---

2. Design for breach

Organizations must execute "assume breach" in practice, with zero trust architecture in mind. The model grants no implicit trust by network location, authenticates and authorizes every request against least-privilege policy, and segments the environment so a single foothold does not become free movement. As vulnerability patching cycles compress and become harder to roll out across complex environments, designing for breach containment becomes a baseline requirement, not an optional maturity goal.

Every time the organization stands up a network, endpoint, or service, the question of how to contain a breach should be posed as a design question rather than a roadmap aspiration. It is important for the design to be continuously validated. Human-led red and purple teams remain essential, but they should be augmented by autonomous adversarial simulation that continuously probes the environment for reachable paths, exposed privileges, lateral movement opportunities, and control failures.

"*Assume breach is no longer a contingency model; it is the operating model for machine-speed ecosystems.*"

Amyn Jan
Chief AI Architect, US Department of War

What to do

- ☑ **Apply least privilege to every agent.** As the fastest-growing population of credentialed actors in most enterprises, agents should have access to exactly the tools and data they need for their tasks, and nothing more. This should include segregated high-risk operations, alerts on anomalous agent behavior, and frequently audited agent permissions. Agents should default to first-class managed non-human identities, with provisioned IAM, scoped entitlements, clear ownership, and lifecycle controls.

- ☑ **Lock down endpoints with binary allowlisting.** An attacker who achieves remote code execution against an allowlisted endpoint has a much shorter list of next moves. This remains as one of the highest-leverage controls available and one of the least commonly deployed outside highly regulated environments. Ensure that usage of system tools (i.e., PowerShell) are protected.

  It remains effective against AI-augmented attacks for the same reason: an allowlist policy does not negotiate. Even if the AI model can find a vulnerability and write an

exploit at machine speed, the second-stage payload it wants to run will not execute.

- ☑ **Make compromise local by design.** Microsegmentation remains the gold standard for limiting lateral movement across workloads, but it is a multi-year program most organizations will not complete in the window this threat curve demands. The practical priority is to push as many controls as possible outside the context window (i.e., gateways, IAM, network paths, and execution environments) so that actions are enforced by architectural constraints.

- ☑ **Validate the design with autonomous red teaming.** Containment that hasn't been tested under machine-speed adversarial pressure is a hypothesis, not a control. All existing protection infrastructure needs to be independently and continuously verified with Mythos-grade capabilities: standing autonomous red-team agents should continuously probe the environment for reachable paths, privilege escalation opportunities, lateral movement, and control failures. Human red and purple teams remain essential for novel threat modeling and high-context investigation.

- ☑ **Design for recovery along with containment.** Even well-contained breaches will sometimes succeed at scale, so recovery has to be designed into the architecture: immutable and air-gapped backups, restore procedures that do not depend on the infrastructure being restored, identity systems that can be rebuilt without trusting compromised credentials, and manual operating procedures for the business functions that cannot afford an extended IT outage.

"*Assume breach can't be a quarterly ritual anymore. In an autonomous threat environment, defenders need continuous proof that compromise stays local, privileges don't cascade, and the controls they're counting on still hold when the attacker moves at machine speed.*"

David Campbell
Head of AI Security Research, Scale AI

---

3. Compress detection and response to minutes

Detection and response must operate in minutes, not days. With Mythos-class capability in adversarial hands, an attacker who lands on any machine in an organization's environment can reach the data that matters in the time it takes an existing SOC to triage a single alert. For many regulated organizations, this is also a compliance reality: incident classification, containment, evidence capture, and notification regimes assume a step change in response speed that human-paced SOC workflows cannot reliably meet.

What to do

- ☑ **Catch your own breaches first, continuously.** With handoff times now measured in seconds, self-detection must be tested continuously and benchmarked against adversarial speed.

  Organizations must shift from scheduled exercises to standing purple-team validation: detection-engineering and Incident Response (IR) capacity that runs the most probable intrusion paths against live controls on a rolling basis. This should be tracked against three measures: self-detection rate, latency to the first internal indicator, and failure modes that would defer notification to an external party.

- ☑ **Augment triage with AI.** Buying or building automated SOC analysis has become the operating standard, where building them typically involves leveraging cheap AI models for high-volume alert triage, creating an aggregation layer for correlation and prioritization, and using a frontier model at the top to perform contextual reasoning. As models improve, more of the incident lifecycle can be resolved before a human responds.

- ☑ **Remediate with the help of AI and appropriate controls in place.** Increasingly, AI can be used to write and ship the production code change that resolves the incident before a human responds.

  Practically, a remediation agent will have write access to production, and be a high-value target for adversarial manipulation. As a result, the same controls that are used to govern any high-privilege AI agent should apply here: implementing least privilege, mutual authentication, segmentation, and binary allowlisting.

"In a machine-speed threat environment, security controls cannot simply exist - they must be continuously verified under live operational conditions. A control documented in policy, validated six months ago, or marked compliant during an audit may already be ineffective against an adaptive AI-assisted adversary. The shift underway is from static assurance to continuous validation."

Min Xu
Senior Principal Security Architect, Navy Federal Credit Union

"As advanced AI systems increase the speed and scale of cyber risk, incident response is no longer merely a best practice or a compliance box to check; it is a legal imperative, requiring organizations to assess, contain, investigate, mitigate, and respond with increasing speed and precision to protect their business operations, networks, and sensitive data."

Kevin R. Powers
Faculty Director & Lecturer-in-Law, MLS in Cybersecurity, Risk & Governance, Boston College Law School

Conclusions for CISOs post-Mythos

Mythos and succeeding models will accelerate every aspect of the offensive cyber attack lifecycle. The organizations best prepared for the next few months will be the ones who compress every part of their defensive lifecycle to match this new speed pattern, and do so by building on top of solid cybersecurity foundations.

That work resolves into three imperatives:

1. Compress the patch cycle. Treat shipping the fix as the disclosure.

- ☑ Set patch SLAs in hours to days, while adopting better vulnerability management practices
- ☑ Run an LLM-driven security review on triage and remediation
- ☑ Require the same discipline from vendors as a procurement gate

2. Design for breach. Design based on zero trust architecture, and contain the blast radius before it does.

- ☑ Apply least privilege to every agent defaulted to non-human identities
- ☑ Lock down endpoints with binary allowlisting so second-stage payloads cannot run
- ☑ Segment and push controls outside the context window so a successful exploit against one workload stays local
- ☑ Run autonomous red teams against live controls, so the blast radius is verified and not assumed
- ☑ Pair containment with recovery - immutable backups, and manual fallbacks

3. Defend at machine speed. Move the SOC from backlog triage to a continuous cycle.

- ☑ Catch your own breaches first, measured by self-detection rate
- ☑ Augment triage with AI across the SOC stack, from alert prioritization upward
- ☑ Remediate with AI, governed by the same controls as any high-privilege agent

Each item is a control that gold-standard security organizations already run. CISOs should move these from quarterly reviews into operational reality this year.

The readiness gap will only be closed by accelerating the defensive lifecycle to match the offensive lifecycle. The same scaling dynamics that produced Mythos will continue to produce its successors and on a timeline that does not pause for procurement cycles or board approvals. Organizations that fail to compress defensive response times will increasingly rely on external notifications to discover compromise, and absorb the consequences from it.

Appendix: Consortium best practices

More than 120 AIUC-1 Consortium members co-authored this whitepaper, over the course of a live session and a detailed peer-review.

The Appendix captures the AIUC-1 Consortium's considerations beyond SOC controls. All converge on a single condition: the three imperatives hold only when there is a stable cybersecurity foundation - from implementing a baseline security hygiene, comprehensive third-party and supply-chain risk coverage, and practicing governance, operating models, and assurance flexible enough to keep pace with rapid change.

"*The era of reactive, low-friction security is over. In an AI-driven threat landscape, organizations that haven't invested in proactive fundamentals will find themselves structurally unable to respond. Accepting some friction now is far less costly than a crisis later.*"

Mandy Andress
CISO, Elastic

---

1. Baseline Security Hygiene

Organizations must have basic cybersecurity hygiene to mitigate existing vulnerabilities amplified by Mythos, and to implement the imperatives listed. This paper assumes organizations have implemented a baseline of fundamentals that may not have been reached yet (e.g., accurate asset inventory, patch visibility, disciplined IAM, configuration management, comprehensive logging, backup integrity, clear system ownership). A defender operating at machine speed is only as good as the environment it can see; automation layered on an incomplete inventory or stale IAM simply makes the wrong decisions faster. Legacy technical debt warrants the same urgency: forgotten, unpatched flaws are the raw material attackers chain into attack paths, accelerated by AI usage.

The April 2026 Cyber Security Breaches Survey revealed that 43% of businesses suffered a breach in the past year, yet 76% of firms rushing to adopt AI have no cybersecurity practices in place to manage the risks. In a post-Mythos world, gaps in basic cybersecurity hygiene is an existential liability.

Without basic cybersecurity fundamentals, the three imperatives are aspirations rather than actions realistically operationalized. Organizations cannot automate patching without asset inventory, cannot design for breach without identity governance, and cannot defend at machine speed without centralized telemetry.

None of this is novel; it is the cyber hygiene the industry has codified for years. The protective baseline should span:

- **Secure configuration and hardening.** Apply hardened baselines such as the CIS Benchmarks, remove default credentials and unused services, and manage configuration drift so systems do not quietly fall out of a known-good state.
- **Data protection.** Classify sensitive data, encrypt it at rest and in transit, and manage keys so that a single foothold does not automatically become a data loss.
- **Vulnerability and patch management.** Scan continuously and remediate on a defined cadence, prioritizing by real-world exploitability such as CISA's KEV catalog and the Exploit Prediction Scoring System (EPSS).
- **Comprehensive logging and monitoring.** Centralize tamper-resistant logs with enough retention to both detect an intrusion and reconstruct it afterward; a SOC cannot triage signals it never collected.

"Now is the time for good housekeeping – know your environment and know your users. If you do not yet have a complete and up-to-date registry of your assets (hardware, software) with their respective configuration and patching levels – it is high time to create one. Identity and Access Management has become critical. Ensure that you have air-tight processes in place to provision/deprovision users and frequently audit their access. Use behavioral analytics as an additional safety measure to flag anomalous user behavior."

Dr. Manfred Boudreaux-Dehmer
Former CIO & CISO, NATO

---

2. Comprehensive Third Party and Supply Chain Risk Coverage

In a Mythos-class threat environment, supplier risk should be treated as inherited exposure. Even if an organization compresses its own patch window, it remains exposed when critical suppliers operate on human-speed remediation cycles. Vendor posture is a live risk signal, not a procurement formality: critical vendors such as third-party service providers should be engaged on how they are handling machine-speed vulnerability discovery, monitored continuously, required to provide Software Bill of Materials (SBOM) and open-source dependency management, and assessed for fourth-party exposure.

The same discipline applies internally to the software supply chain. The code that applications depend on is now a primary attack vector, not a background concern. In March 2026, attackers hijacked axios (a JavaScript library with more than 100 million weekly downloads) and published two poisoned releases that installed a cross-platform remote-access trojan. The impact reached downstream organizations including OpenAI's macOS signing pipeline. A single compromised dependency delivered an attacker into the build systems of organizations that made no mistake of their own.

The discipline that contains this: SCA at every build, signed builds and code provenance, and treating AI-generated code as insecure until reviewed. Block promotion of any artifact carrying a known critical vulnerability, and require explicit review of newly introduced or materially changed libraries before they reach production. The aim is to protect the internal environment while raising the standard externally.

3. Governance, Operating Model and Assurance Designed to Keep Pace

Machine-speed defense is as much an organizational question as a technical one. Operating at such a rapid pace is not only a SOC or engineering issue, but a governance, operating model and assurance issue.

Governance grants the authority, the operating model exercises it, and assurance verifies it holds. A SOC re-tooled to run in minutes is still restricted when every new capability has to clear control mapping, vendor risk, a quarterly procurement cycle, and a board cycle before it goes live. Attackers are scaling through AI, while defenders remain constrained by approval chains, authority structures, and risk processes.

Governance is where the gap closes first. Its posture must evolve from periodic compliance validation toward continuous, threat-responsive decision-making: production-change authority for patch-cycle compression, architecture veto power for

breach-resilient design, and pre-approved business-impact authority for automated responses. Equally important is defining which decisions can be delegated to automation, which require human approval, and the escalation paths between them. Machine-speed operation requires explicit authority boundaries, not simply faster approvals.

"As risk grows and AI changes how it's managed, the structure of the security and risk organization itself comes into question. From where the CISO sits, how product security, platform security, GRC, and risk management interlock, and how the operating model handles work that increasingly spans all of them. Organizations getting this right are treating it as a deliberate redesign."

Omar Khawaja
Field CISO, Databricks

The operating model and assurance then carry that authority into practice. AI-enabled defense does not dissolve the human role; in a machine-speed SOC, the analyst becomes the system orchestrator, threat hunter, and policy validator, with autonomous agents as the supporting structure. The organizations that close the readiness gap will automate aggressively while protecting the human capacity that decides where the automation points and owns the outcome when it fails. The more the operating model delegates to autonomous agents, the less informative a periodic attestation. Assurance must also move from periodic attestation toward continuous, independent verification that controls still hold under live adversarial pressure.

AIUC-1

About AIUC-1

AIUC-1 is the standard for AI agent security, safety, and reliability. The standard is developed through the AIUC-1 Consortium with Technical Contributors from trusted institutions including Stanford, MIT, Orrick, MITRE, the Cloud Security Alliance, Google Cloud, Cisco, MongoDB, and more.

Read more and access the full standard: aiuc-1.com or contact at contact@aiuc.com