Deloitte.

Mythos: AI-driven cyber asymmetry

The new threat reality

April 2026

“Traditional security is a race against the clock. Automated resilience remove the clock entirely by ensuring the attack surface disappears before the exploit can even land.”

Michael Mosaad
Partner, Enterprise Security
Deloitte Middle East

“For decades, security models rested on an unsaid assumption: if a vulnerability wasn't widely known, it probably wasn't being exploited. Mythos broke that assumption. The gap between vulnerability discovery and weaponization has collapsed from years to hours. The organizations that treat AI-native security as a core capability rather than a pilot program will carry a structural advantage their competitors cannot close with headcount alone.”

Gary Arora
Managing Director, AI & Engineering
Deloitte US

Claude Mythos – Overview

Introduction

Independent assessments confirm that AI-driven attack capabilities now significantly exceed previous threat models. The AI Security Institute (AISI) evaluation demonstrating **Mythos executing 32-step expert-level exploits** [1] exemplifies the sophistication of current AI-powered vulnerability research and autonomous attacks. Throughout 2025 and into 2026, continuous acceleration in both research and in-the-wild attacks reveals a critical gap, with offensive AI capabilities advancing faster than defensive ones.

Anthropic's Claude Mythos (Preview), announced on April 7, 2026, represents a step change in that trajectory, autonomously finding thousands of critical vulnerabilities across every major operating system and browser, generating working exploits without human guidance, and empowering autonomous attack orchestration, all at a speed and scale that outpaces any prior capability.

What is Claude Mythos Preview?

- A new **frontier model** from Anthropic so capable at finding and exploiting vulnerabilities that it is not being released publicly.
- Claude Mythos Preview demonstrates **181 working exploits across 595+ vulnerabilities**, compared with only 2 working exploits in the previous generation, signaling a dramatic expansion of attacker capability.
- Average time-to-exploit has **fallen from 2.6 years (2018) to less than 20 hours (2026)**, turning long-tail vulnerabilities into near-real-time exposure and compressing the window for defensive response.
- Mythos achieves a **72% exploit success rate** across tested vulnerabilities, meaning most identified weaknesses can be reliably weaponized, increasing the likelihood of material business impact.

Why does it matter?

- Organizations face **elevated cyber risk** that will impact insurance premiums, mergers & acquisitions (M&A) valuations, and board liability, a structural change in the cost of doing business.
- Vulnerabilities are now discovered and weaponized faster than organizations can patch, **creating regulatory exposure** and forcing a rethink of due diligence requirements for compliance certifications.
- Organizations need to **reassess their security posture** against this new threat model.
- There are both **immediate risks** (i.e., operational disruption, regulatory and compliance breaches, liability and reputational damage) and **strategic opportunities** (i.e., competitive differentiation, operational efficiency, and cost optimization) for clients to address.

Source [1] - Our evaluation of Claude Mythos Preview's cyber capabilities | AISI Work

Project Glasswing

The scale and speed of Mythos prompted Anthropic to create **Project Glasswing,** possibly the largest multi-party vulnerability coordination effort in history. Anthropic provided selected critical infrastructure providers, industry partners, and open-source maintainers early access to Mythos so they could patch their own products, as an initiative to use Mythos defensively before attackers get similar capabilities.

Who has access to Project Glasswing?

Critical infrastructure providers

Organizations responsible for supplying essential infrastructure (CPU, GPU, servers).

Industry partners

Major technology companies (Microsoft, Apple, CrowdStrike, AWS) with security responsibilities.

Critical software infrastructure developers

A broad coalition of 40+ companies developing foundational software.

Open-source maintainers

Developers who maintain publicly available code libraries and projects that millions of applications rely on.

What is Anthropic committed to?

Anthropic is investing in the below resources to make Project Glasswing users Mythos-ready.

Defensive security usage credits ($100M)

Anthropic is providing $100 million in computational credits to support defensive security initiatives, enabling the use of AI tools for security research.

Direct funding for open-source security ($4M)

Anthropic is committing $4 million to open-source security organizations, strengthening the ecosystem of freely available security tools to protect critical infrastructure.

Why this matters for our clients?

Organizations face an unprecedented convergence of threats and opportunities that demands immediate strategic preparation to maintain their security posture.

Prepare for patch tsunami

40+ vendors will release critical security patches in rapid succession, creating an urgent response window.

Reduced defensive window

The advantage for defenders is narrowing as comparable exploit capabilities will reach attackers within months.

Strategic mindset shift

Organizations must prepare budget allocations for AI-native security investments in the following fiscal year.

AI-driven cyber asymmetry: Why the Middle East faces disproportionate risk

Rapid AI-led digitization in the Middle East is colliding with autonomous offensive AI, creating a near-term imbalance where cyber offense will outpace defense at scale.

Structural exposure

- Organizations are **aggressively adopting digital services** including but not limited to smart cities, open banking, and AI-first government.
- The **API ecosystem expansion outpaces security control maturity**, leaving organizations with an expanding attack surface and inadequate defenses.
- Security capabilities are improving at a fraction of the speed at which digital transformation is occurring, **widening the vulnerability gap**.

What changes with mythos release

- AI systems now enable **autonomous end-to-end attack execution**, automating reconnaissance, exploitation, and lateral movement across networks.
- Vulnerability discovery is accelerating faster than organizations can patch, creating an **expanding window of exploitable weaknesses** in systems.
- The **technical skill barrier** for launching sophisticated attacks has collapsed, enabling novice attackers to execute complex campaigns at scale [2].
- The **threat landscape is shifting** from individual hackers to organized AI-powered operators capable of adapting tactics in real time.

Where it hits first

- **Banking and fintech organizations** face escalating API abuse and fraud attacks, while smaller players lack the security maturity to defend against coordinated threats.
- **Critical infrastructure operators** face emerging risks as attackers pivot from IT systems to operational technology (OT) environments in energy, ports, and logistics sectors.
- **Digital identity systems** are becoming primary attack targets, with adversaries shifting focus from network compromise to identity layer exploitation.

Readiness gap

The strategic risk

- Offensive capabilities are scaling faster than defensive responses, creating a temporary but severe imbalance in attacker advantage.
- Cyber threats are evolving from episodic incidents into continuous economic pressure.
- Regulatory and security frameworks are increasingly disconnected from actual organizational capability.

What leaders are missing

- Organizations across the region are accelerating AI adoption and exposure simultaneously.
- Perimeter-based security models are fundamentally obsolete in cloud-native environments.
- Most existing defenses lack AI-native capabilities.
- Identity has become the new control plane, yet most organizations leave it critically under-protected.

Implication

- Middle East organizations must adopt AI-native, identity-first security architectures by 2027 or face systemic vulnerability across their digital infrastructure.

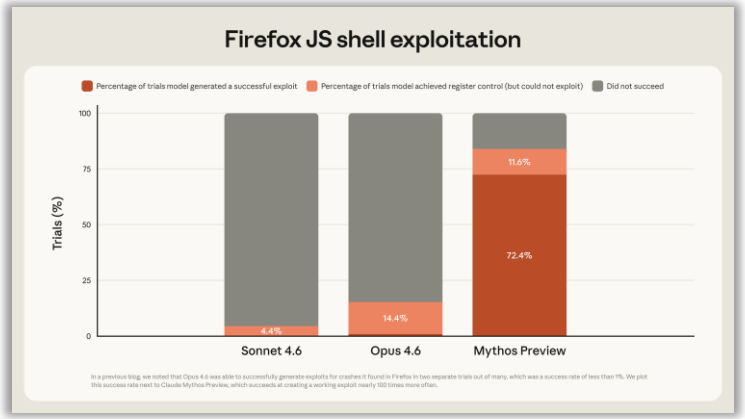
Source [2] - What AI-Driven Attack Chains Mean for CFOs and CISOs - MENA Fintech Association

Strategic impact

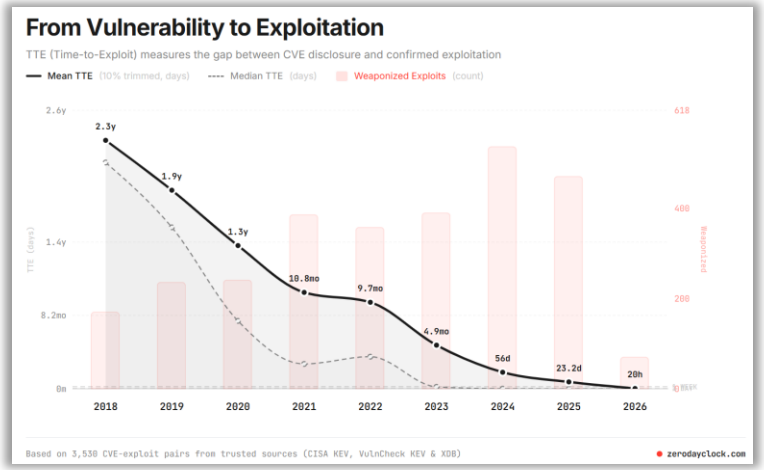
What the competitive landscape indicates today?

- **Existential risk –** Mythos compresses the time between vulnerability discovery and business disruption from weeks to hours. This demands a budget for 24/7 automated response capabilities and pre-authorized containment actions. The **cost of inaction exceeds the cost of transformation.**
- **Speed becomes competitive advantage –** Organizations deploying **AI-driven security tooling accelerate time-to-market** by automating threat detection, response, and remediation at scale. Hence, AI investment in security is not just risk mitigation; it's a competitive enabler. Organizations that move fast with confidence will outpace those lacking AI adoption.
- **Risk metrics and board reporting must evolve –** Mythos **invalidates legacy risk models** and demands board-level discussions shift from "Are we compliant?" to "Can we survive a zero-day exploit in hours?" The **UK AI Security Institute validated this threat model across 32 simulation steps** (reconnaissance to environment takeover), with superior performance against competing models, underscoring the strategic urgency.
- **Talent and capacity become the limiting factor -** The Mythos threat model requires **organizations to operate at machine speed**. This demands an increase in security staffing and automation investment over the next 12 months. Delay increases risk of losing experienced staff to burnout and creates a dangerous capability vacuum.
- **Collective defense become strategic necessity -** Mythos capabilities will be available to attackers within months. No single organization can defend alone against autonomous exploit tools. Hence, organizations should **allocate resources to participate in industry information-sharing initiatives** and collaboratively strengthen their resilience.

Claude Mythos demonstrates improved performance in vulnerability exploitation compared to previous generation models, with notably faster exploitation timelines.



Source [3] - Claude Mythos Preview \ red.anthropic.com



Source [4] - Zero Day Clock

Mythos exploit capability shift

Comparing Claude Mythos with earlier generations

Capability	Opus 4.6	Mythos Preview
Autonomous exploit development	2 working exploits	**181 working exploits**
Full control flow hijack	0 cases	**10 separate, fully patched targets**
Vulnerability discovery	150-175 cases	**595+ cases**

Security impact

- **Mythos can find and exploit vulnerabilities without human guidance**
- **Mythos can discover vulnerabilities in every major OS and web browser**
- **Mythos can find bugs dating back up to 27 years**
- **Mythos can turn publicly disclosed, patched vulnerabilities into working exploits**

Statistics

- Thousands of high/ critical-severity vulnerabilities discovered
- 72% exploit success rate across tested vulnerabilities
- 89% of manually reviewed reports had exact human severity agreement
- Cost per campaign amounts to $10,000-$50,000
- Time to exploit: Hours to days (vs. weeks/months for humans)

Source [5] - The “AI Vulnerability Storm": Building a "Mythos-Ready" Security Program

The vulnerability apocalypse

These asymmetries converge to create a systemic crisis – simultaneous discovery, accelerated weaponization, and organizational unpreparedness collide to overwhelm traditional security models.

Cost asymmetry

AI tools let attackers generate and test exploits at marginal cost, while defenders fund broad remediation programs.

Skill compression

Mythos-style systems reduce required expertise, enabling less sophisticated actors to deliver high-quality exploits at scale.

Speed advantage

Time-to-exploit has collapsed from years to hours, outpacing traditional patch and change-management cycles.

Scale democratization

AI can scan hundreds of vulnerabilities and generate many working exploits, making advanced offense widely accessible.

Volume explosion

Problem
AI-powered vulnerability discovery generates thousands of exploitable flaws simultaneously, collapsing traditional sequential triage models.

Key Metrics
Mythos discovers 595+ vulnerabilities
Traditional - 50-100/year, 6-12x acceleration

Impact
Triage teams overwhelmed, CVE prioritization impossible, Critical vulnerabilities missed in noise, and patch inventory explodes

Business
Thousands of actionable vulnerabilities exceed management capacity.

Weaponization acceleration

Problem
Exploit weaponization accelerates from weeks to hours, rendering traditional patch cycles obsolete before deployment.

Key Metrics
2018: 2.6 years, 2026: <20 hours
99.7% reduction, 72% exploit success rate

Impact
Patch cycles become liabilities, vulnerabilities weaponized before patches available, zero-day window collapses, continuous patching mandatory

Business
Organizations cannot patch faster than attackers can weaponize

Organizational unpreparedness

Problem
Enterprise security models remain optimized for sequential threats, lacking automation and processes for simultaneous, high-velocity response.

Key Metrics
Manual triage, sequential patching, limited automation, and human-driven response

Impact
Automated prioritization, AI-assisted triage, continuous deployment, autonomous response

Business
Structural gap in processes, automation, and cultural readiness

The closing defensive window

The timeline demonstrates that organizations must urgently close the gap between rapidly advancing offensive AI capabilities and their defensive readiness, with a critical deadline approaching where defensive capabilities must match offensive threats.

Offensive AI capability evolution

Emerging	Escalating	Escalating	Critical	Active	Systemic	Apocalypse	Apocalypse
Jun 24, 2025	Aug 5, 2025	Aug 8, 2025	Sep 2025	Nov 14, 2025	Feb 5, 2026	Mar 2026	Apr 7, 2026
XBOW Tops HackerOne	Google Big Sleep: 20 Zero-Days	DARPA AIxCC: 54 Vulns in 4 Hours	Singularity Warning: Autonomous Exploit ~6 Months Away	First AI-Orchestrated Espionage Campaign	Hundreds of Bugs; Autonomous Attacks Discovered	Linux Kernel Bugs Jump 2→10/Week	Claude Mythos Preview Announced

Defensive readiness & organizational response

Baseline	Awareness	Planning	Pilot	Reactive	Transform	Crisis	Critical Gap
Jun 24, 2025	Aug 5, 2025	Aug 8, 2025	Sep 2025	Nov 14, 2025	Feb 5, 2026	Mar 2026	Sep 2026
Traditional Vulnerability Management	Early Awareness; Limited Adoption	Strategic Discussions Begin	Pilot Programs Launched	Reactive Measures; Emergency Projects	Widespread Transformation Begins	Vulnerability Velocity Exceeds Capacity	Mythos Released; Window Closing

Organizations must act NOW. Every month a delay reduces preparation time

From traditional client-server failure to ephemeral, dynamic architecture

The **traditional client-server model is fundamentally incompatible with AI-driven threats**. It relies on a trusted internal network and permanently exposed entry points that AI-enabled attackers can continuously discover, analyze, and attack. To assist organizations secure their business and critical digital assets in an AI-first world, we must rethink this model from the ground up – moving beyond static infrastructure to ephemeral, constantly-evolving systems that make it difficult for attackers to maintain persistence, even if they successfully identify vulnerabilities.

Mythos exploits what it can reach; the first line of defense is **eliminating reachability.**

- **Decommission internet-facing entry points –** Remove SSL inspection gateways, VPN concentrators, and perimeter firewalls reachable from public internet.
- **Make applications undiscoverable** – Publish through cloud-native access fabric, invisible to external IP and port scanners.
- **Zero inbound connections** – No discoverable gateways for AI-driven scanners to probe; applications become invisible and unreachable by IP or port.

Attack surface

Least-privilege access **limits lateral movement**, containing credential compromise.

Application-level access only – Grant access to specific applications, never underlying network infrastructure.

- **Context-aware policy enforcement –** Broker every connection per session based on verified identity, device posture, location, and real-time risk.
- **Eliminate lateral movement –** Compromised identity reaches a single application while lateral movement requires network access.

Context access

AI-resilient protection layer underpins the Mythos-ready security model with continuous behavioral analysis.

- **Inspect all traffic inline –** All traffic is inspected in cloud for threats, data loss, and policy violations to ensure no bypass paths.
- **Cloud-scale analytics and AI –** Implement AI capabilities to detect anomalous behavior and AI-generated attacks in real time.
- **Continuous enforcement –** Enforcement across every session removes blind spots and creates AI-resilient protection layer.

Inline protection

Key architecture principle: Ephemeral infrastructure = moving target defense

Traditional architectures are static and discoverable. Mythos-ready systems are constantly changing – credentials rotate, access paths shift, infrastructure evolves.
This dynamic nature makes it exponentially harder for attackers to:

- Maintain persistence after initial compromise
- Exploit known vulnerabilities before they're patched
- Build reliable attack chains against a moving target

Client actions – High-level plan

The expected timeline is contingent upon the organization's risk appetite and operational execution capability.

Within 45 days

Immediate

1. Deploy AI-powered code security
2. Mandate AI agent adoption in security
3. Prepare for continuous patching
4. Recalibrate risk models and reporting

In 45 – 90 days

Medium-term

1. Secure your AI agents
2. Inventory and minimize attack surface
3. Strengthen defensive fundamentals

Up to 12 months

Long-term

1. Establish cross-functional innovation governance
2. Deploy deception and detection
3. Automate incident response
4. Establish permanent vulnerability operations (VulnOps)

Source [5] - The “AI Vulnerability Storm": Building a "Mythos-Ready" Security Program

Client actions

Immediate (within 45 days) – Mythos stabilization plan

01 Deploy AI-powered code security

- Organizations must **embed AI agents and LLM capabilities directly into their development workflow** to continuously scan code and dependencies for vulnerabilities.
- Start with immediate security reviews, then integrate automated scanning into CI/CD pipelines and developer tools.
- All code (human or AI-generated) must pass LLM-driven security validation before merge.
- Use tools like **Claude Code Security, Codex Security, OpenAnt, raptor, and exploitation-validator** to automate detection at scale.

02 Mandate AI agent adoption in security

- Organizations must **formalize AI agent usage across all security functions** with mandatory controls and oversight.
- While defensive AI hasn't fully caught up to offensive capabilities, adoption is non-negotiable for effectiveness.
- Make AI agent adoption in security a requirement, not an option. Cultural resistance will otherwise limit the organization's ability to execute other critical actions.

03 Prepare for continuous patching

- As Mythos becomes available to major vendors through Project Glasswing, vulnerability discovery will accelerate dramatically. Organizations must start **building triage and deployment capacity** to handle a potential flood of critical patches.
- Establish staffing, automation, and pipeline infrastructure to manage continuous vulnerability disclosures without operational disruption.
- Implement dynamic infrastructure polymorphism through **IP/port shuffling, multi-version microservice deployment, and memory randomization** to eliminate attacker reconnaissance patterns.

04 Recalibrate risk models and reporting

- **Update security risk metrics, reporting frameworks, and business risk calculations** to reflect AI-accelerated exploit timelines and attack complexity.
- Legacy assumptions about patch windows, exploit scarcity, and incident frequency are now obsolete.
- Outdated models risk underfunding controls and producing inaccurate business reporting.
- Align stakeholders on revised threat projections and resource requirements.

Source [5] - The "AI Vulnerability Storm": Building a "Mythos-Ready" Security Program

Client actions

Medium-term (within 45 - 90 days) – Foundational phase

01 Secure your AI agents

- Most critical tasks cannot be completed without agents, but agents themselves must be rigorously defended. They are not covered by existing security controls which introduces new cyber defense and supply chain risks.
- The agent harness (prompts, tool definitions, retrieval pipelines, and escalation logic) is where the most consequential failures occur.
- Audit agent configurations with the same rigor as permissions.
- Before deploying agents in or near production, **define scope boundaries, blast-radius limits, escalation logic, and human override mechanisms**.
- Establish organizational-specific governance framework based on best practices.

02 Inventory and minimize attack surface

- Organizations must **establish complete asset visibility within 90 days using AI agents** to accelerate inventory creation and enable continuous updates.
- Prioritize internet-facing critical systems first, then expand toward full coverage.
- Generate **authoritative SBOMs** to understand supply chain dependencies.
- Eliminate unnecessary exposure by decommissioning unused functionality, phasing out non-compliant suppliers, and isolating at-risk systems.
- Deploy autonomous self-healing mechanisms with automated quarantine of anomalous workloads and continuous drift correction against established baselines.

03 Strengthen defensive fundamentals

- Core security practices remain valid and should be prioritized for risks that cannot be easily mitigated.
- Implement **egress filtering** to block known exploits attacks.
- Enforce **deep network segmentation** and **zero-trust architecture** where feasible.
- Enforce maximum workload lifetime policies (e.g., 4-hour container/server expiration) on critical systems to eliminate persistent attacker access windows.
- Each defensive layer increases attacker cost and friction, hence, layered defenses remain essential, especially in an AI-accelerated threat environment.

Source [5] - The “AI Vulnerability Storm": Building a "Mythos-Ready" Security Program

Client actions

Long-term (up to 12 months) – Strategic positioning

01 Establish cross-functional innovation governance

- Actively contribute to **cross-functional security initiatives** among multiple domains including but not limited to security, legal, and engineering, to rapidly evaluate emerging offensive threats and accelerate adoption of defensive technologies.
- **Restructure decision-making frameworks** to remove bureaucratic delays and accelerate security deployment cycles as a strategic competitive advantage.

02 Deploy deception and detection

- Build **deception capabilities** that are attack-tool and vulnerability-independent, identifying threats based on attacker behavior and TTPs.
- Deploy canaries and honey tokens throughout your environment, **layer behavioral monitoring, pre-authorize containment actions, and establish response playbooks** that execute at machine speed.
- Organization must **accelerate investment in their insider threat programs** to mitigate the amplified risk of AI-empowered threats operating at unprecedented scale and speed.

03 Automate incident response

- Enhance **detection engineering and incident response** to be systemic and autonomous wherever possible.
- Implement **automated asset and user behavioral analysis, pre-authorized containment actions, and response playbooks** that execute without human delay. The goal is to match attacker speed with defender speed through automation.

04 Establish permanent vulnerability operations (VulnOps)

- Build a dedicated, staffed **Vulnerability Operations** function, similar to DevOps but for autonomous vulnerability research and remediation.
- VulnOps owns **continuous discovery of zero-day vulnerabilities across your entire software estate** (internal code and third-party dependencies) and establishes automated remediation pipelines.
- Achieve SLSA Level 3 compliance for CI/CD pipelines to ensure **supply chain integrity** and **prevent code/artifact tampering.**

Source [5] - The “AI Vulnerability Storm": Building a "Mythos-Ready" Security Program

Mythos readiness assessment

A **10-question** based approach to help you triage your security program, to reach ground truth, as well as gauge your influence on various business functions.

Questions	Context
What is your organization's current level of AI advancement and its impact on your asset protection?	Clarify whether AI use is allowed, tolerated, restricted, or undefined in your organization.
Can employees use agentic coding tools in the enterprise today?	Determine if AI coding agents are permitted, and whether security guardrails exist for their use.
Can your organization enable open-source contributions without legal ambiguity, a prerequisite for Mythos deployment?	Assess legal and IP policies governing employee open-source contributions. Unclear policies create compliance risk and limit safe Mythos adoption.
Do we have disciplined control repos, artifacts, and software, including for agentic supply chain such as MCP servers, plugins, and skills?	Evaluate whether source control, artifact management, and supply chain governance cover AI agent components.
Is there a real cooling-off point/security gate between code change and production?	Confirm whether security controls enforce a separation between development and production deployment.

Questions	Context
Is security operational, or primarily advisory?	Determine if security actively prevents incidents or mainly reviews and escalates issues.
What is the shortest period of time this company has made a security-driven production change in the last year?	Use a real example to measure your organization's actual speed of security response.
Are our critical "crown jewels" explicitly tracked and current?	Identify which systems truly matter most and ensure their inventory is maintained.
Do we know how to get urgent work prioritized by our key third parties?	Assess whether you have escalation paths and leverage with critical vendors and partners.
Does executive leadership have a working definition of urgency?	Clarify whether leadership distinguishes between routine work and genuine crises.

Source [5] - The "AI Vulnerability Storm": Building a "Mythos-Ready" Security Program

Key takeaways for clients

Deploy LLM-based vulnerability discovery

Clients should start by immediately engaging an agent for security code reviews. Build towards a **dedicated VulnOps** capability that continuously scans their codebase and dependencies for vulnerabilities using LLM tools. This is the fastest path to autonomous vulnerability detection.

Prepare for increased incident response

Clients should conduct tabletop exercises simulating multiple simultaneous high-severity incidents within a single week. Develop and **test incident playbooks** for critical scenarios. Verify and enable mitigating controls to contain blast radius when exploitation occurs.

Recalibrate risk metrics

Clients should **audit their current risk assessments** and metrics against the new threat timeline. Update assumptions about patch windows, exploit availability, and incident frequency. Communicate **revised risk profiles** and resource requirements to stakeholders and leadership.

Accelerate teams with coding agents

Clients should **mandate AI agent** adoption across security functions (incident response, GRC, patch management, audit automation). Use agents to triage patches, red team environments, automate data collection, and accelerate security operations.

Prioritize defensive fundamentals

Clients should focus on **core controls** that remain effective - network segmentation, patching known vulnerabilities, identity and access management, and defense-in-depth. Expand these efforts while building advanced capabilities.

Prepare for team burnout

Clients should **request additional headcount and budget** now to build reserve capacity. The volume and velocity of vulnerability disclosures will exceed historical norms. **Invest in automation** to reduce manual workload and prevent staff burnout during the transition period.

Evolve to a mythos-ready security program

Clients should incorporate **Mythos and autonomous exploit capabilities** into their security strategy. Assess implications for their threat model, defensive posture, and organizational readiness. Plan for a fundamentally different threat landscape.

Build collective defense

Clients should engage with sector coordinating groups, **ISACs, CERTs, and standards bodies** to share threat intelligence and coordinate response. Establish information-sharing relationships with peers. Collective defense through coordinated groups with shared tools and intelligence will outperform isolated teams.

References

1. Our evaluation of Claude Mythos Preview's cyber capabilities: AI Security Institute. Available at: https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities

2. Fintech, M. (2026) What AI-driven attack chains mean for CFOs and CISOs, MENA Fintech Association. Available at: https://mena-fintech.org/news/what-ai-driven-attack-chains-mean-for-cfos-and-cisos/

3. Assessing Claude Mythos Preview's Cybersecurity Capabilities: Claude Mythos Preview \ red.anthropic.com. Available at: https://red.anthropic.com/2026/mythos-preview/

4. Zero Day Clock. Available at: https://zerodayclock.com/

5. The 'AI Vulnerability Storm': Building a 'Mythos-Ready' Security Program (2026) CSA CISO Community, SANS, [un]prompted, the OWASP Gen AI Security Project. Available at: https://labs.cloudsecurityalliance.org/mythos-ciso/

6. Eliminating Your Attack Surface is the Best Defense Against Vulnerabilities Discovered by Anthropic's Mythos Model: Zscaler. Available at: https://www.zscaler.com/blogs/product-insights/eliminating-your-attack-surface-best-defense-against-vulnerabilities

7. Claude Mythos Proves the AI-persistent Threat Era has Arrived: Straiker. Available at: https://www.straiker.ai/blog/claude-mythos-proves-the-ai-persistent-threat-era-has-arrived

Get in touch

Michael Mosaad
Partner | Enterprise Security
Co-author
mmosaad@deloitte.com | +971529448642

Lindsay Thorburn
Director | Enterprise Security
Co-author
lthorburn@deloitte.com | +971526970479

Harshil Shah
AI Security Lead | Enterprise Security
Co-author
hshah10@deloitte.com | +971585942864

Deloitte.