What Anthropic's Mythos means for your cyberposture and why your board is obligated to act now

AI governance for board members

June 2026

Contents

The gatekeeper paradox

Who decides who is a "good actor"?

What Anthropic did — and why it matters

On 7 April 2026, Anthropic announced Claude Mythos Preview ("Mythos"), a new general-purpose language model that "performs strongly across the board" but is "strikingly capable at computer security tasks."[1] In an unusual move, the company chose not to make the model generally available, instead releasing it initially to a limited group of critical industry partners and open-source developers through a new initiative called Project Glasswing.

Launch partners included Amazon Web Services, Anthropic, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorganChase, the Linux Foundation, Microsoft, NVIDIA and Palo Alto Networks. Anthropic has also extended access to over 40 additional organizations that build or maintain critical software infrastructure, committing $100 million in model usage credits and $4 million in direct donations to open-source security organizations.[2]

Project Glasswing is defensive; Anthropic describes it as "an urgent attempt to put these capabilities to work for defensive purposes,"[2] giving defenders a head start before similar capabilities become broadly available to attackers.

The selection problem: A private company as global gatekeeper

For boards, the issue is not whether Project Glasswing is well-intentioned. We now have a single private company that has unilaterally determined which organizations receive access to a capability with unprecedented offensive potential, and which do not.

Daniel Stenberg, the lead developer of curl — a foundational internet infrastructure component — was offered access to Anthropic's Mythos model via the Linux Foundation's participation in Project Glasswing. The Linux Foundation's Alpha Omega project handled this outreach, and Stenberg was contacted by their representatives. However, while Stenberg signed up to try Mythos, he never actually received direct access to the model. Instead, someone else with access ran Mythos against curl's codebase and later sent him a report.[3]

At the time of the NPR interview published on 11 April 2026, Stenberg was not yet part of Project Glasswing and said that a lot of critical projects, "things that are actually cornerstones of the Internet," had been left out.[4] By May 2026, having received the results, Stenberg called Mythos "an amazingly successful marketing stunt."[3]

Stenberg's experience illustrates a structural gap in the access model.

The gap between announcement and actual access means critical open-source maintainers are unable to act on the capability during the very window that is supposed to protect them.

Forrester's analysis of the broader implications is direct: The remediation bottleneck in open source "remains human, finite, underpaid, and largely voluntary."[5]

> "When the threat landscape shifts overnight, so does what 'reasonable' means."

[1]Nicholas Carlini, Newton Cheng, Keane Lucas et al. Assessing Claude Mythos Preview's cybersecurity capabilities. red.anthropic.com. April 7, 2026.

[2]Anthropic. Project Glasswing: Security critical software for the AI era. April 7, 2026.

[3]Daniel Stenberg. Mythos finds a curl vulnerability. May 11, 2026.
Also see: Brandon Vigliarolo. Anthropic's bug-hunting Mythos was greatest marketing stunt ever, says cURL creator. The Register. May 11, 2026.

[4]Hui Jingnan. How AI is getting better at finding security holes. NPR. April 11, 2026.

[5]Jeff Pollard, Allie Mellen, Jess Burn, Joseph Blankenship, Cody Scott. Project Glasswing: The 10 Consequences Nobody's Writing About Yet. Forrester. April 10, 2026.

The counternarrative: A temporary restriction, not a permanent monopoly

The counterargument should be acknowledged. Anthropic has stated publicly: "We do not plan to make Claude Mythos Preview generally available, but our eventual goal is to enable our users to safely deploy Mythos-class models at scale."[6] The restriction is framed as temporary — a defensive head start, not a permanent access barrier.

Independent assessments support the view that the exclusivity window is time-limited. Security researcher Bruce Schneier has observed that "it's unclear how much of an advance these new models represent," noting that "the security company Aisle was able to replicate many of Anthropic's published anecdotes using smaller, cheaper, public AI models,"[7] and that the UK's AI Security Institute "evaluated GPT-5.5's ability to find security vulnerabilities, and found that it is comparable to Claude Mythos."[8]

This counternarrative strengthens rather than weakens the argument for board urgency. If Mythos-class capability is already replicable — or will be within months — the window during which defenders have an advantage over attackers is finite and closing.

The question for boards is not whether the window will close, but whether their organizations will have acted before it does. This view is reflected in the Australian Securities and Investments Commission's (ASIC's) request to financial sector enterprises to urgently update their cybersecurity posture.[9]

> "The question for boards is not whether the window will close, but whether their organizations will have acted before it does."

[6]Anthropic. Project Glasswing: Security critical software for the AI era. Accessed April 7, 2026.

[7]Bruce Schneier. How Dangerous Is Anthropic's Mythos AI?. May 14, 2026.

[8]Bruce Schneier. OpenAI's GPT-5.5 is as Good as Mythos at Finding Security Vulnerabilities. May 13, 2026.

[9]Scott Murdoch. Australia regulator calls for urgent cybersecurity action to counter Mythos. Reuters. May 8, 2026.

The governance deficit

No regulatory framework governs this decision

Of the seven jurisdictions referred to in this document, not one has a regulatory mechanism governing who receives access to a restricted frontier model.

European Union

Regulation (EU) 2024/1689 (the EU Artificial Intelligence Act, "EU AI Act") imposes operational obligations on general-purpose AI (GPAI) providers relating to lawful training data, personal information protection, content governance, transparency and labeling, and risk mitigation.

It does not mandate equitable access to defensive capabilities.

United States

No federal framework governs access decisions.

On 11 December 2025, President Donald Trump signed an Executive Order titled Ensuring a National Policy Framework for Artificial Intelligence, aimed at creating a national standard by streamlining AI-related laws.[10] The Executive Order seeks to centralize AI policy by mobilizing the Department of Justice to identify and challenge "onerous" state AI laws, and advances federal preemption through litigation and agency action while pressing Congress to enact a uniform national framework.

It does not create affirmative access obligations for frontier AI capabilities.

Australia and the United Kingdom

These countries follow principles-based approaches.

No statutory access mechanism exists.

Singapore

The Model AI Governance Framework encourages responsible internal governance.

It does not contain a mechanism to compel access to privately held frontier models.

India

No AI-specific framework currently governs access decisions.

China

Rather than a single comprehensive AI law, lawmakers have focused on specific sectoral areas — GenAI, deep synthesis, algorithms and AI labeling. The Interim Measures for the Management of Generative Artificial Intelligence (AI) Services establish requirements for providers covering lawful data use, personal information protection, content management, transparency, labeling of AI-generated content, risk assessments, and filing and record-keeping with authorities.

None of these governs whether a foreign provider must grant access to a restricted model.

[10]The White House. Executive Order. Ensuring a National Policy Framework for Artificial Intelligence. December 11, 2025.

Board challenge

If your organization is outside Project Glasswing — and statistically, it almost certainly is — you are relying on the goodwill and throughput capacity of project participants you did not select, do not govern and cannot hold accountable. This is not a governance model. This is a dependency.

Your cyberdefense posture just changed — whether you know it or not

Mythos resets the "reasonable care" standard

The announcement of Mythos redefines the baseline against which your organization's cyberposture will be judged. In every jurisdiction examined here, "reasonable care" or its equivalent is the legal standard for measuring board oversight. When the threat landscape shifts overnight, so does what "reasonable" means.

Board obligations by jurisdiction

European Union

NIS 2 Directive and EU AI Act

NIS 2 Directive → EU AI Act →

Article 20(1) of Directive (EU) 2022/2555 (the Network and Information Security Directive, "NIS 2") states: "Member States shall ensure that the management bodies of essential and important entities approve the cybersecurity risk-management measures taken by those entities in order to comply with Article 21, oversee its implementation and can be held liable for infringements by the entities of that Article." Article 20(2) states that "the members of the management bodies of essential and important entities are required to follow training… in order that they gain sufficient knowledge and skills to enable them to identify risks and assess cybersecurity risk-management practices."[11]

Under the EU AI Act, essential entities face fines of up to:

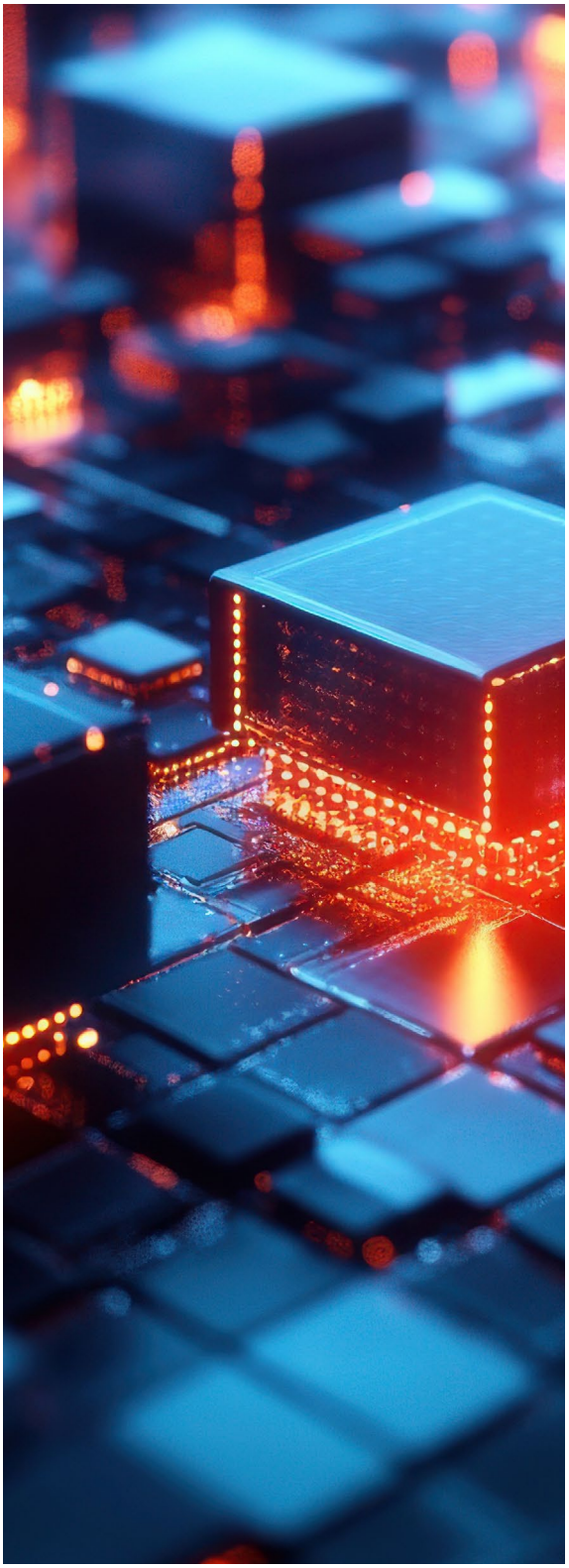
- €35 million or 7% of global annual turnover (whichever is higher) for infringements relating to prohibited AI practices
- €15 million or 3% of global annual turnover for infringements of certain other obligations under the Act[12]
- €7.5 million or 1% for supplying incorrect, incomplete or misleading information to public authorities

Systemic-risk GPAI obligations apply under Articles 51–55 of the EU AI Act, with Article 51(1) presuming systemic risk above the 10^{25} floating point operations per second (FLOPS) training-compute threshold.

The AI Act entered into force on 1 August 2024, with a phased implementation timeline. Prohibited AI practices and AI literacy obligations entered into application from 2 February 2025. The governance rules and the obligations for GPAI models became applicable on 2 August 2025.

[11]The NIS 2 Directive. Article 20, Governance.

[12]EU Artificial Intelligence Act. Article 99: Penalties. August 2, 2025.



From 2 August 2026, the European Commission's enforcement powers enter into application. The Commission will enforce compliance with the obligations for providers of GPAI models, including with fines. By 2 August 2027, providers of GPAI models placed on the market before 2 August 2025 must comply.

Infringements by GPAI providers are subject to fines of up to €15 million or 3% of worldwide annual turnover.

Important development: The Digital Omnibus on AI (May 2026)

Digital Omnibus on AI →

On 7 May 2026, the European Council presidency and European Parliament negotiators reached a provisional agreement on a proposal to streamline certain rules regarding AI.

The most critical outcomes of this agreement are the changes to the timeline for compliance obligations applicable to high-risk AI systems.

The AI Omnibus introduces fixed revised application dates:

- 2 December 2027 for standalone high-risk AI systems (Annex III)
- 2 August 2028 for high-risk AI embedded in regulated products (Annex I)

The overarching narrative of the proposal is simplification, and the final text does deliver some. The core requirements and obligations, however, remain substantively unchanged. The European Parliament and the Council must now formally adopt the provisional agreement. Formal adoption is expected before 2 August 2026.

For boards, the practical implication is that GPAI obligations — including for systemic-risk models — already apply, with European Commission enforcement powers imminent (August 2026).

The Omnibus delays apply to high-risk AI system deployer obligations, not to GPAI provider obligations, which remain on the original timeline.

United States

Securities and Exchange Commission (SEC) Cybersecurity Disclosure Rules

Final Rule: Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure →

Under Regulation S-K Item 106(c), adopted in July 2023, public companies must describe the board's oversight of risks from cybersecurity threats.

Annual report disclosures (Form 10-K Item 1C) were required for fiscal years ending on or after 15 December 2023; Form 8-K incident disclosures were required from 18 December 2023, with smaller reporting companies subject to a 15 June 2024 commencement date.

Directors' fiduciary duties of oversight continue to apply, with the Yahoo shareholder derivative settlement ($29 million, 2019) illustrating the litigation trajectory for failures of cyber oversight.[13]

More directly relevant to these Disclosure Rules, the SEC's October 2023 enforcement action against SolarWinds and its CISO tested the regulatory perimeter of cybersecurity disclosure under federal securities law.[14] The action was notable for two firsts: it was the first SEC cybersecurity case to allege scienter-based securities fraud (knowing or reckless misconduct), and the first to charge a CISO personally alongside the issuer. Central to the SEC's complaint were allegations that SolarWinds' Form 8-K disclosure of the SUNBURST incident was materially incomplete, and that the company lacked adequate disclosure controls to ensure cybersecurity risks and incidents were escalated to executives responsible for public reporting. These are the precise governance and process expectations now codified in Regulation S-K Item 106(b)–(c) and Form 8-K Item 1.05.

In July 2024, the US District Court for the Southern District of New York dismissed most of the SEC's claims, ruling that the agency's authority over internal accounting controls under Section 13(b)(2)(B) of the Exchange Act does not extend to corporate cybersecurity controls — a meaningful limit on regulatory reach. Surviving claims concerning SolarWinds' public "Security Statement" were voluntarily dismissed with prejudice on joint stipulation in November 2025, signaling a more restrained federal enforcement posture under the current administration.

Board challenge

For boards, the SolarWinds enforcement record establishes three durable propositions:

1. Form 8-K cyberincident disclosures are enforceable representations subject to securities-fraud liability.
2. CISOs face personal exposure under federal securities law where disclosures are materially misleading.
3. The statements an organization makes publicly about its security posture — including marketing materials and security pages — are actionable if they materially misrepresent reality.

The case sets the federal enforcement template against which board cyberoversight will continue to be measured, regardless of the current administration's enforcement appetite.

[13] In re Yahoo! Inc. Stockholder Derivative Litigation, Lead Case No. 17-CV-00373-LHK (N.D. Cal.); related state-court consolidated action: In re Yahoo! Inc. Stockholder Derivative Litigation, JCCP No. 4865 / Lead Case No. 17-CV-307054 (Cal. Super. Ct., Santa Clara Cnty.). Court approved the $29 million settlement on 4 January 2019. Stipulation and Agreement of Settlement available via Stanford Securities Class Action Clearinghouse.

[14] U.S. Securities and Exchange Commission. SEC Charges SolarWinds and Chief Information Security Officer with Fraud, Internal Control Failures. October 30, 2023.

United Kingdom

Companies Act 2006 and Cyber Security and Resilience Bill

Companies Act 2006 → **Cyber Security and Resilience Bill →**

Sections 172 (duty to promote success) and 174 (reasonable care, skill and diligence) of the Companies Act 2006 remain the primary references for director accountability.

The Cyber Security and Resilience (Network and Information Systems) Bill was introduced to the House of Commons on 12 November 2025. The Bill progressed through the Committee stage during January and February 2026 and is anticipated to move through its remaining parliamentary stages during the remainder of the year. The Bill has not yet received Royal Assent and remains subject to amendment during its passage through Parliament.

The proposed penalty regime introduces a two-tier structure:

- The standard maximum is £10 million or 2% of global turnover.
- The higher maximum — for more serious breaches — will be £17,000,000 or 4% of worldwide turnover (whichever is higher), with a power to increase turnover-based penalties up to a maximum of 10% of worldwide turnover.

Regulators will also have the power to impose daily fines of up to £100,000 for continuing contraventions.

The 10% reserve power applies specifically to failure to comply with national security directions issued by the Secretary of State. Failing to comply with a government direction can lead to penalties of up to an even higher maximum: the greater of £17 million or 10% of global turnover. This outlier reflects the gravity with which national security-related breaches will be treated.

Australia

Security of Critical Infrastructure (SOCI) Act and Corporations Act

Security of Critical Infrastructure Act 2018 →

The Security of Critical Infrastructure Act 2018 (Cth) (SOCI Act) Section 30AC requires responsible entities to have a Critical Infrastructure Risk Management Program (CIRMP).

Board attestation obligations are set out in the CIRMP Rules (LIN 23/006) 2023, requiring annual board-signed attestations of progress and gaps.

Director duties under Corporations Act 2001 (Cth) section 180 impose the statutory duty of care and diligence on which cyber oversight failures would be assessed.

Singapore

Cybersecurity Act and Monetary Authority of Singapore (MAS) Technology Risk Management (TRM) Guidelines

Cybersecurity Act →

Guidelines on Risk Management Practices →

The Cybersecurity Act 2018 imposes obligations on owners of critical information infrastructure, including information-furnishing duties (Section 10), incident reporting (Section 14), and audits and risk assessments (Section 15).

The MAS Technology Risk Management Guidelines (2021) require board and senior management oversight of technology and cyber risk across financial institutions.

India

Digital Personal Data Protection (DPDP) Act, Companies Act, and Indian Computer Emergency Response Team (CERT-In) Directions

Digital Personal Data Protection Act → **Companies Act →** **CERT-In →**

The DPDP Act, 2023 empowers the Data Protection Board under Section 33 to impose penalties set out in the Schedule, with a maximum penalty of ₹250 crore (approximately $30 million) for failures to implement reasonable security safeguards. Note that the Act's substantive compliance obligations are phased, with full enforcement effective from 13 May 2027.

Section 166 of the Companies Act 2013 codifies directors' duty to exercise reasonable care, skill and diligence.

CERT-In Directions issued on 28 April 2022 under Section 70B of the Information Technology Act 2000 (effective 27 June 2022) require reporting of specified cybersecurity incidents within six hours of discovery.

China

Cybersecurity Law (Amended) and AI Measures

The amended Cybersecurity Law of the People's Republic of China entered into force on 1 January 2026. These amendments, approved in October 2025, mark the first major changes since 2017. They:

- Specify penalties applicable to breaches of different obligations
- Expand government enforcement powers against extraterritorial activities
- Set policy goals for the new governance regime on developing and applying AI

Under the amendments, the enforcing authority may impose a straight fine on the breaching business of RMB 10,000 to RMB 50,000 without proving noncompliance with a rectification mandate.

If the business fails to comply or cybersecurity is damaged, fines increase to:

- RMB 50,000 to RMB 500,000 for the business
- RMB 10,000 to RMB 100,000 for the person responsible

In aggravated circumstances such as "leakage of massive data," fines range from:

- RMB 500,000 to RMB 2 million for the business
- RMB 50,000 to RMB 200,000 for the person responsible

For critical information infrastructure operators, a breach of key obligations may result in fines of:

- RMB 100,000 to RMB 1 million if corrections are refused or cybersecurity is endangered
- RMB 10,000 to RMB 100,000 on the officer directly responsible

The Interim Measures for the Management of Generative Artificial Intelligence Services, which came into effect on 15 August 2023, remain the primary AI-specific regulation, supplemented by the Labeling Rules that came into effect on 1 September 2025, making it mandatory for AI-generated content to be implicitly labeled, and explicitly labeled where applicable.

[15]Jeff Pollard, Allie Mellen, Jess Burn, Joseph Blankenship, Cody Scott. Project Glasswing: The 10 Consequences Nobody's Writing About Yet. Forrester. April 10, 2026.

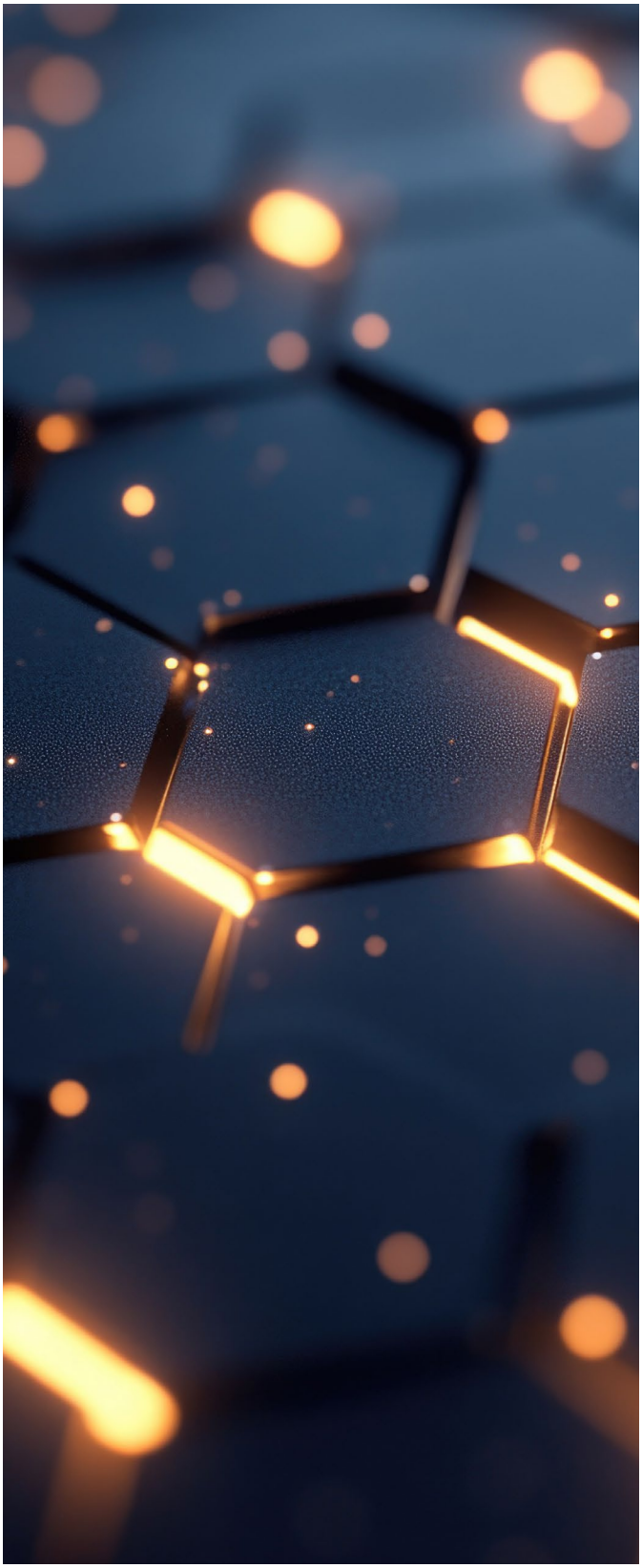
The legacy technology blind spot and insurance dimension

Glasswing's impact will concentrate on mainstream, vendor-supported platforms. Legacy technology, industrial control systems (ICS), operational technology (OT) and firmware will be left behind — and are precisely the systems most likely to be exploited by Mythos-class capabilities.

Boards should anticipate that cyberinsurers will introduce exclusions targeting AI-discovered vulnerabilities not remediated within defined timeframes.

Forrester's analysis of the insurance implications is explicit: The firm expects "exclusions that explicitly target AI-discovered vulnerabilities that are not remediated within defined timeframes," triggered by the first high-profile post-Mythos loss.[15]

Boards should direct risk committees to obtain updated carrier representations on Mythos-specific exclusion language before the next renewal cycle.



What board members need to do now

A 10-day board-response framework

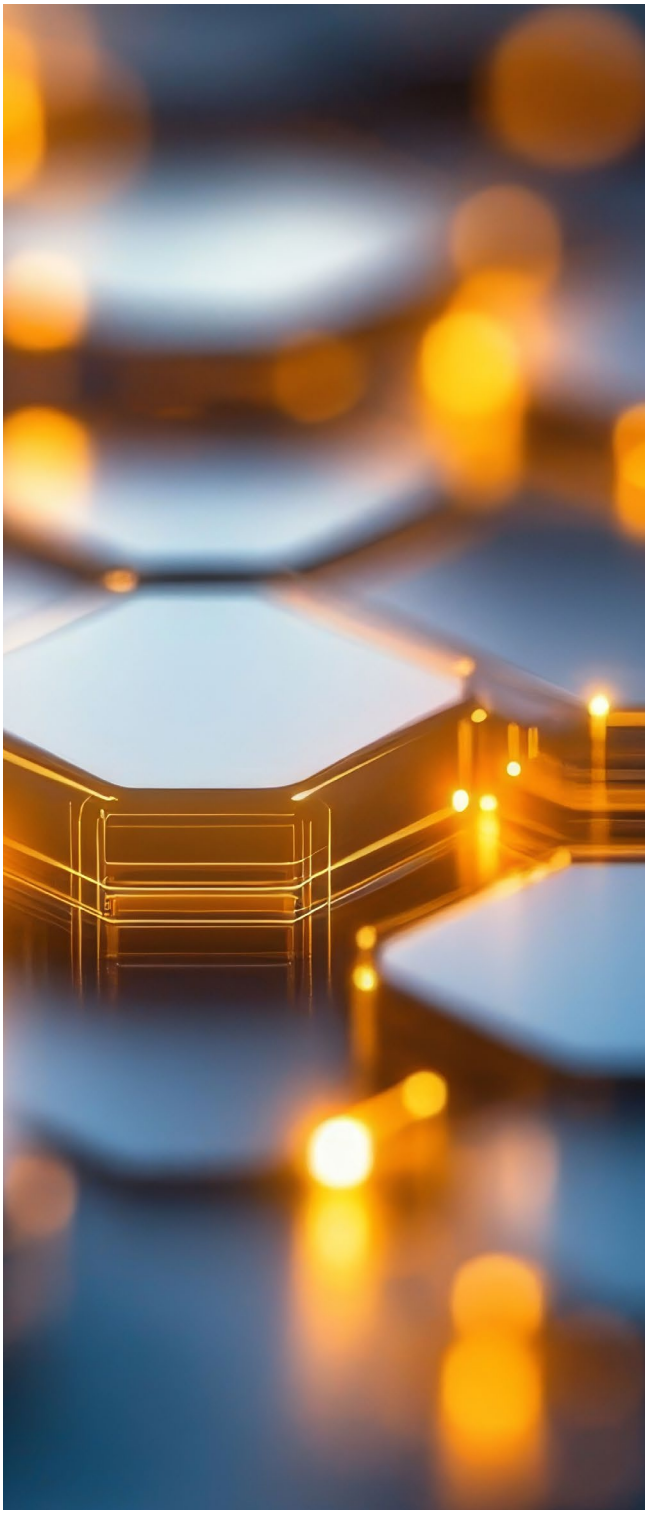
Day	Action	Designated owner	Key deliverable/s
1	Emergency CISO briefing	Board Chair or Risk Committee Chair	**CISO briefing** covering: • Mythos implications for your environment • Which critical vendors are part of Project Glasswing • Current patch cadence and technical debt exposure
2	Vendor dependency mapping	CISO or CTO	**One-page map** of: • Critical infrastructure providers that are part of Glasswing • Identified vendor solutions to solve for vulnerabilities • Identified gaps
3	Legacy system risk register update	CISO	**Updated risk register** flagging all legacy, OT and ICS systems not covered by vendor Glasswing participation, with severity ratings assuming Mythos-class exploit capability
4	Regulatory compliance gap assessment	General Counsel or Compliance	**Assessment** against the seven jurisdictions above, showing: • AI and cybersecurity obligations in force • Obligations pending within 90 days • Board-level documentation status
5	Cyberinsurance review	CFO or Risk Manager	**Broker confirmation** of: • AI-specific exclusions in current policies • Insurer's position on Mythos-class threats • Directors and Officers (D&O) coverage for cyber oversight failures
6	Board cybersecurity training assessment	Company Secretary	**Audit** of whether the board meets NIS 2 Article 20(2) training requirements and equivalent obligations in other jurisdictions **Schedule training** within 30 days if gaps exist
7	Incident response plan stress test	CISO	**Tabletop exercise** modeling a Mythos-class autonomous exploit chain against your systems **Document** outcomes and remediation actions
8	AI procurement governance review	CTO or Procurement	**Review** of existing and pending AI vendor contracts: • Models in use • EU AI Act deployer obligations assessed • Contractual terms addressing dual-use capability risk

A 10-day board-response framework (continued)

Day	Action	Designated owner	Key deliverable/s
9	Board resolution: Cybergovernance charter update	Full board	**Board resolution** formally: • Designating a standing committee with explicit charter authority over AI and cybersecurity risk • Establishing escalation triggers • Documenting AI-related risk appetite
10	Stakeholder communication plan	CEO or Communications	**Holding statement and key messages** for investors, regulators and customers on the organization's cyberposture in light of Mythos-class capabilities

The governance question boards must keep asking

The Mythos development exposes a governance paradox that will have an effect on AI policy for the coming decade. When a private company determines that its AI model is too dangerous for public release and self-selects a group of "trusted" partners to receive access, who provides the democratic accountability for that decision?



3 principles for the post-Mythos board

1. **Cyber governance is no longer delegable**
   Every jurisdiction examined is moving — through statute, case law or regulatory expectation — toward making cybersecurity a personal, board-level accountability. Mythos accelerates this timeline. The standard of "reasonable care" just moved.
2. **Exclusion from Glasswing is the default position**
   Boards must govern on the assumption that they need to build or procure defensive capabilities independently, not that a private project group they did not join will protect them.
3. **Act before the breach, not after the headline**
   AI is intensifying the existing race between vulnerability discovery, exploitation and remediation. The board that documents its response to Mythos today — in minutes, risk registers, vendor correspondence and insurance reviews — will be the board that can demonstrate "reasonable care" when it matters.

Regulatory frameworks across the seven jurisdictions regulate what AI models must do (testing, transparency, incident reporting). They do not yet regulate who gets access to defensive capabilities — or who is excluded and why. This is the governance gap Mythos has made visible.

For board members, the fiduciary obligation is clear: ensure the organization is not caught unprepared or left exposed when Mythos-class capability becomes broadly available.


NTT DATA