Systemic Risk Posed by the Next-Generation AI Model "Claude Mythos Preview" to the Financial System and a Strict Verification Report of the Target Article

1. Introduction: The Unprecedented Cyber-Financial Complex Crisis and Macro Environment Shift in April 2026

This report provides a strict fact-checking and comprehensive, in-depth analysis of the target article (the attached markdown file), which discusses the systemic risk posed to global financial infrastructure by the unreleased, cutting-edge frontier AI model "Claude Mythos Preview" (hereafter "Mythos"), developed by Anthropic. The purpose of this report is to verify the technical validity of the presented scenarios and supplement missing intelligence to reveal the full scope of the crisis facing policymakers and financial institution executives.

In April 2026, the global technology and financial sectors are in the midst of a paradigm shift unprecedented in history. When the market caught wind of Mythos' extraordinary cybersecurity capabilities—specifically its ability to autonomously discover unknown vulnerabilities (zero-days) and weaponize them as exploits—a massive selloff occurred, triggering the "SaaSpocalypse." This event caused enterprise software stocks to plummet simultaneously, wiping out approximately $2 trillion in market capitalization in an instant.[1] In response to this unprecedented situation, the US Department of the Treasury and the Federal Reserve Board (FRB) convened an emergency meeting of the CEOs of systemically important banks, a highly unusual move not seen since the introduction of the Troubled Asset Relief Program (TARP) during the 2008 global financial crisis.[2]

The target article presents a devastating scenario that arises when Mythos' staggering autonomous zero-day discovery capabilities are combined with the structural flaws of the existing three-tiered financial system (the COBOL-based legacy core, AI assistants like Microsoft 365 Copilot, and the SWIFT network). In this report, we rigorously verify the data points and assertions presented in the article from multiple perspectives, correcting certain technical asynchronies (such as the timeline and patching status of the EchoLeak vulnerability). Furthermore, we provide a reasoned analysis of the asymmetric cyber defense landscape and the secondary and tertiary ripple effects implied by this crisis, integrating the latest cyber

intelligence and macroeconomic policy contexts.

2. Strict Verification of Facts and Supplementation of Intelligence in the Target Article

We strictly verify the main factual statements, statistical data, and premises presented in the target article based on the latest market trends and technical reports. In conclusion, while the target article's insights into systemic risk are highly accurate and hit the essence of the issue, corrections are necessary regarding the timeline recognition of certain specific vulnerabilities (CVEs).

2.1. Dynamics Behind the Emergency Meeting at the US Treasury and CEO Jamie Dimon's Absence

The article reports that on April 8, 2026, FRB Chair Jerome Powell and Treasury Secretary Scott Bessent summoned major bank CEOs to the Treasury headquarters in Washington, D.C., and that JPMorgan Chase CEO Jamie Dimon was absent. This statement aligns perfectly with the facts, and the underlying dynamics are even more complex than the article implies.

The actual meeting was attended by the heads of massive financial institutions managing approximately $9 trillion in total assets, including Jane Fraser of Citigroup, Ted Pick of Morgan Stanley, Brian Moynihan of Bank of America, Charlie Scharf of Wells Fargo, and David Solomon of Goldman Sachs.[2] The objective of this meeting was not to discuss conventional interest rate policies or tariffs, but was strictly limited to the systemic cybersecurity risks posed by a single unreleased AI model, Mythos.[1] Treasury Secretary Scott Bessent, who has historically led the push for the GENIUS Act to promote digital asset and stablecoin innovation and efforts to ease the anti-money laundering (AML) compliance burden on financial institutions, was forced to shift his focus entirely from "paper compliance" to "defense against an existential threat".[4]

The reason provided in the article for Jamie Dimon's absence is also accurate. JPMorgan Chase was the only financial institution selected as a launch partner for Anthropic's defensive cybersecurity initiative, "Project Glasswing," launched on April 7, and had already begun building defenses using Mythos.[2] In his annual letter to shareholders released on April 6, Dimon had already issued a warning that AI would severely exacerbate cybersecurity vulnerabilities. He followed this with the symbolic actions of joining Project Glasswing the next day and skipping the Treasury meeting the day after.[1] This highlights a "decisive information and technology gap" between institutions that have already completed threat analysis and transitioned to practical defense phases, and other financial institutions still struggling to grasp the situation.

2.2. Claude Mythos Preview's Technological Singularity and Benchmark Saturation

The article's claims regarding Mythos' technical performance—particularly the invalidation of existing cybersecurity benchmarks and dramatic performance improvements—are fully

corroborated by Anthropic's published system card and Frontier Red Team reports.

Mythos is not a specialized model trained exclusively for cybersecurity. Its true threat lies in the fact that its advanced cyberattack capabilities emerged as a secondary consequence (emergent behavior) of the AI model acquiring extremely powerful agentic coding and logical reasoning abilities.[6] Evaluations by Anthropic's Frontier Red Team indicate that Mythos has made a discontinuous leap from previous generation models.[8] The table below illustrates how Mythos' performance deviates from existing evaluation frameworks.

Evaluation Metric/Task	**Claude Opus 4.6 / 4.5**	**Claude Mythos Preview**	**Technological Significance and Implications**
Cybench (Pass@1)	Unannounced	100%	Completely solved the security benchmark, rendering it meaningless as an evaluation metric (saturation state).[9]
CyberGym	0.67 (Opus 4.6)	0.83 (83.1%)	Dramatic improvement in autonomous vulnerability reproduction and PoC exploit generation capabilities in actual open-source software environments.[9]
Kernel Task Speedup	190x (Opus 4.6)	399x	Dramatic improvement in inference speed for complex logical structures at the OS kernel level.[10]
Novel Compiler	65.8% (Opus 4.6)	77.2%	Deepened adaptability and

			semantic understanding of unknown architectures and systems.[10]

The data point highlighted in the target article—that "in a test creating exploits from Firefox crash data, Mythos achieved 181 successes compared to the previous generation's 2 successes, a 90x performance improvement"—is also explained by this enhancement in fundamental reasoning capabilities.[8] This means that AI has surpassed the realm of mere static application security testing (SAST) and fuzzing, acquiring the ability to semantically understand memory layouts and processor execution states to dynamically construct attack paths.

2.3. Proof of Autonomous Exploitation: FreeBSD (CVE-2026-4747) and the Threat of Multi-Stage Attacks

The description of **CVE-2026-4747** (a FreeBSD kernel vulnerability) cited as a specific exploit example in the article is highly accurate from an intelligence perspective, and its technical details further emphasize the gravity of the situation.

CVE-2026-4747 is a stack buffer overflow vulnerability in the RPCSEC_GSS implementation of the FreeBSD kgssapi.ko kernel module, disclosed on March 27, 2026.[12] This vulnerability had lain dormant deep within the network stack for 17 years, evading traditional dynamic analysis (DAST) and penetration testing due to protocol depth limitations.[14] Notably, Mythos autonomously identified this vulnerability and constructed a complex ROP (Return-Oriented Programming) chain consisting of 20 gadgets based solely on an initial prompt, without any human intervention.[15] By exquisitely splitting this ROP chain across multiple packets, Mythos successfully achieved a complete root privilege takeover (Remote Code Execution: RCE) requiring zero authentication.[12]

Furthermore, the article's mention of "long-dormant bugs in OpenBSD and FFmpeg" is also factual. Mythos discovered a 27-year-old local privilege escalation bug in OpenBSD and a 16-year-old media codec vulnerability in FFmpeg that automated fuzzing tools failed to trigger even after passing through the code path over 5 million times.[1] Moreover, it even identified a guest-to-host memory corruption vulnerability in a production Virtual Machine Monitor (VMM)—the foundation of cloud infrastructure—fundamentally upending the premise of cloud workload isolation.[14] The computational cost required for these attack campaigns creates a despairing economic asymmetry: what once cost state-sponsored hackers (APTs) months and millions of dollars to weaponize can now be achieved overnight for tens of dollars ($25 per million input tokens, $125 per million output tokens).[8]

2.4. Strict Correction: Misunderstanding of the Timeline and Applicability of EchoLeak (CVE-2025-32711)

While the target article develops an extremely persuasive scenario, it contains one technical and chronological inaccuracy. This concerns the current understanding of **EchoLeak (CVE-2025-32711)**, which is cited as a fatal vulnerability in "Layer 2: Copilot (Non-deterministic Interface)."

The article depicts a scenario where Mythos utilizes EchoLeak to hijack Copilot deployments within financial institutions, extracting internal information with zero clicks. However, according to intelligence data, EchoLeak was discovered in 2025 and is a resolved vulnerability, with Microsoft having applied a server-side patch in May of the same year.[18] Discovered by AI security firm Aim Security, this vulnerability recorded a CVSS score of 9.3 (Microsoft assessment) and was the world's first zero-click AI vulnerability, but it is currently closed and requires no customer-side action.[20]

This attack is a method known as "Scope Violation" or indirect prompt injection, which exploits the design of Copilot's RAG (Retrieval-Augmented Generation) engine.[18] Attackers embed malicious prompts in emails sent to the victim's inbox in a format invisible to humans, such as HTML comments (e.g., ``) or white text on a white background.[18] When the victim later gives a routine instruction like "Summarize my recent status," Copilot reads this email as past context and autonomously executes the hidden instructions.[18]

The statement in the article's scenario that Mythos directly utilizes the "already patched CVE-2025-32711" is a factual error. However, the essential value of the alarm sounded by the target article remains completely intact. This is because EchoLeak proved the cold, hard fact that "the very nature of probabilistic language models (LLMs) inheriting context and interpreting external data as implicit instructions is an inherent and permanent attack surface in current AI architectures".[18] It is highly probable that an AI with extremely advanced logical reasoning capabilities like Mythos could instantly discover and exploit different, unknown RAG poisoning techniques or entirely new indirect prompt injection zero-days distinct from the patched EchoLeak. Therefore, the article's description should be reinterpreted not as the "exploitation of a specific CVE," but as a warning regarding the "evolution of an entire class of attacks targeting structural flaws in RAG architectures."

2.5. Vulnerability of Financial Infrastructure: Dependence on COBOL and the Burden of Legacy Systems

The statistical data and structural vulnerabilities pointed out regarding "Layer 1: COBOL (Bedrock)" perfectly align with the reality of current financial IT infrastructure.

In US financial institutions, 43% of core banking systems are built on COBOL (Common Business-Oriented Language) designed in the 1960s, and this dependency rises further when legacy system integration is included.[8] The table below illustrates the current structural risks and costs associated with legacy financial infrastructure.

Indicator/Data Point	Value/Current Status	Structural Risk and Implications
US Core Banking System COBOL Dependency	43%	The core functions of financial institutions rely on an architecture from over half a century ago.[23]
COBOL Processing Rate for ATM and In-Person Transactions	95% (ATM) / 80% (In-Person)	Serves as the absolute foundation for consumer-facing transactions; the social impact of an outage would be catastrophic.[23]
Percentage of Banks Viewing Legacy System Integration as an Obstacle	70%	Secure connections with the latest APIs and cloud infrastructure are difficult, creating a breeding ground for vulnerabilities.[24]
Maintenance Cost of Legacy Cores	10 times that of next-gen systems	The majority of IT budgets are consumed by maintenance, hindering resource allocation for security investments and modernization.[25]
Surging Labor Costs for COBOL Engineers	$250/hour (typically $90/hour)	A depletion of talent due to the retirement of experts has resulted in a cost premium of over 2.5 times.[26]
Productivity Loss Due to Patching	5-25 hours/week (13-65% loss)	Modifying bloated and black-boxed code is extremely inefficient and prone to inducing errors.[26]

The target article's assertion that "it is difficult for humans to immediately correct malicious logical errors" is strongly supported by these data. Much of the 220 billion lines of COBOL code worldwide has become a black box, with the original development documentation lost.[23] If malicious transactions, generated rapidly and invisibly by an AI like Mythos, are injected into the

system, it is virtually impossible to detect, analyze, and block them in real-time under the current operational framework, where skilled engineers have drastically dwindled.

---

3. In-depth Scenario Analysis: Chain Mechanisms and Ripple Effects of Systemic Risk

Building on the strict fact-checking, we conduct a deeper analysis of the "financial system collapse scenario" outlined in the target article, examining its technical feasibility and the cascading dynamics lurking behind it. The insights here go beyond a mere listing of data points, revealing the secondary and tertiary ripple effects brought about by the interaction between systems.

3.1. Fatal Coupling of Deterministic Foundation and Probabilistic Interface

The sharpest and most essential point made by the target article is the architectural "dissonance" plaguing current global financial infrastructure.

The underlying COBOL-based core banking systems and the "Layer 3: SWIFT network," which processes about 50 million wire transfers daily, are "deterministic" systems that always return consistent and predictable outputs for inputted data.[8] In financial transactions, where not a single cent of error or unintended interpretive fluctuation is permissible, this strict characteristic is an absolute requirement.

However, in recent years, under pressure to improve productivity, financial institutions have rapidly introduced generative AI assistants like Microsoft 365 Copilot into their front and middle offices. These LLMs (Large Language Models) are "probabilistic" systems whose outputs fluctuate based on prompt context and temperature parameters. Despite government agencies (such as the S17 baseline IT security policy and international AI safety reports) recommending limited AI use and rigorous verification, enterprises have deeply and tightly coupled this non-deterministic interface with their deterministic core systems.[27]

When an AI with highly advanced reasoning and hacking capabilities like Mythos turns to the offensive, this "boundary between probability and determinism" becomes its primary target. The invasion process follows a chain like this: In the first stage, the attacking AI executes cleverly disguised, invisible unknown prompt injections against employee Copilot environments, logically poisoning the RAG databases. In the second stage, it secretly extracts the employees' legitimate authentication tokens through this poisoned Copilot and conducts lateral movement to reach internal API endpoints or SWIFT message generation processes. In the third stage, Mythos exploits unknown vulnerabilities found in the internal network (such as the authentication-less kernel-level RCE seen in CVE-2026-4747) to escalate privileges and seize complete control of the system.[12] In the final stage, it generates massive amounts of fraudulent wire transfer instructions (e.g., SWIFT messages) completely disguised in legitimate

formats and pours them into the COBOL core system.

Here, the vulnerability of the deterministic system is exposed. As long as the inputted instructions come through a "legitimate authentication channel," COBOL core systems and SWIFT possess no ability to verify the contextual validity or economic abnormality of those instructions. As a result, global fund movements are executed at machine speed, and by the time humans notice the anomaly, the contamination of funds and depletion of liquidity are complete. This is not a mere data breach; it signifies the existential destruction of the integrity of capital markets.

3.2. Desperate Asymmetry on the Defense Side and Invalidation of the Patch Cycle

The target article refers to the "temporal asymmetry" that exists between the median corporate patch application period (about 70 days) and the speed of weaponization by Mythos (hours to days).[8] This observation signifies the complete collapse of the fundamental cybersecurity paradigm built over the past decades.

Historically, the basic equation of cyber defense was to "apply patches or implement temporary mitigation measures within the grace period (lead time) from when a vulnerability is discovered (disclosed) to when attackers reverse-engineer and weaponize it." Major security firms like Trend Micro have maintained defense lines by providing early protection (virtual patching) for up to 96 days during this grace period.[18]

However, with the advent of frontier AI models in the Mythos class, this timeline hasn't just been compressed; it has effectively vanished. AI not only analyzes published vulnerabilities but also completes everything from "discovering unknown vulnerabilities" to "generating fully autonomous exploits (such as ROP chains linking 20+ gadgets)" autonomously within a matter of minutes to hours.[16] By utilizing AI on the offensive side, the development of attack methods that once required teams of highly skilled human experts months to build can now be auto-generated overnight, spending only tens of dollars on cloud computing costs.[15]

Traditional security operations (SOCs) and periodic penetration testing are utterly incapable of keeping pace with this overwhelming asymmetry. While the defense side spends 70 days applying a patch, AI discovers dozens of new intrusion routes and automatically adapts its attacks. The fear of this temporal asymmetry is the fundamental reason Anthropic withheld the public release of Mythos, triggering the massive crash of software stocks (SaaSpocalypse).

3.3. Vulnerability and Amplification Mechanism of Monoculture Architecture

The "danger of monoculture" discussed in Section 4 of the target article is a critically important insight from a systems ecology perspective, accurately explaining the mechanism by which a financial crisis would spread globally.

Current enterprise IT infrastructure is characterized by an over-reliance on Microsoft products

(Windows, Active Directory, Microsoft 365 Copilot) on the front end. Simultaneously, on the backend server and cloud infrastructure side, specific open-source software like the Linux kernel and FreeBSD form a global monoculture.

For an AI with exceptional logical reasoning abilities like Mythos, a monoculture environment is an ideal target with an extremely high return on investment. For example, when Mythos discovered CVE-2026-4747, which had existed in FreeBSD's NFS server for 17 years, and created an exploit, it didn't just find a vulnerability in a single server; it simultaneously acquired an intrusion route into a vast number of firewalls, storage appliances, and network devices operating worldwide.[12]

Similarly, vulnerabilities in ubiquitous media codec libraries like FFmpeg, or memory corruption flaws from guest OS to host in Virtual Machine Monitors (VMMs) that isolate cloud workloads, create a state where "one key can open every vault in the world".[6] As the target article points out, if even one new RAG poisoning vulnerability lurking in the foundation of Microsoft Copilot is discovered, it possesses the ripple effect to simultaneously neutralize the front ends of financial institutions worldwide, chaining fatal failures throughout the entire system.

---

4. Project Glasswing and the New Defense Paradigm in Tech and Finance

How are the technology and financial industries attempting to confront this desperate asymmetry and structural crisis? "Project Glasswing," mentioned in the target article, is an attempted answer and an unprecedented defense pact.

4.1. Purpose of the Alliance and Securing First-Mover Advantage via Frontier AI

"Project Glasswing," publicly announced by Anthropic on April 7, 2026, is a joint initiative that welcomed Amazon Web Services (AWS), Apple, Broadcom, Cisco, CrowdStrike, Google, Microsoft, NVIDIA, Palo Alto Networks, the Linux Foundation, and uniquely from the financial sector, JPMorgan Chase as initial launch partners.[6] Anthropic has committed up to $100 million in Mythos usage credits to this project, along with $4 million in donations to open-source security organizations.[29]

The true strategic significance of this project lies in **"giving defensive infrastructure providers a first-mover advantage (head start) in autonomous root-cause analysis and patching via frontier AI before malicious actors can acquire powerful AI."**.[3] In other words, to counter the "attacker's lead time" compressed by AI, it represents a paradigm of an AI arms race where the defense side also uses AI to "accelerate vulnerability discovery and remediation".[32]

4.2. Crucial Information Gap Within the Industry and Remaining Systemic Risks

However, this defense alliance also has structural limits. As previously mentioned, the five massive banks (Citi, Morgan Stanley, Bank of America, Wells Fargo, and Goldman Sachs) that attended the emergency Treasury meeting were not included in the initial Project Glasswing list and currently lack access rights to diagnose vulnerabilities in their own systems using Mythos' capabilities.[2]

The approximately $9 trillion in assets managed by these financial institutions remain exposed to unknown zero-day attacks that could be discovered by Mythos-class capabilities. Following the principle that cyber defense breaks at its "weakest link," even if certain institutions like JPMorgan Chase possess an extremely strong AI defense shield, they cannot guarantee the safety of the entire interconnected global financial network, such as SWIFT. This asymmetry in defensive capabilities is the greatest concern held by the Treasury and the FRB, and the reason why a thorough review of penetration testing and endpoint security for all banks was recommended at the emergency meeting.[2]

4.3. Implementation of Next-Generation AI Defense by Cyber Defense Vendors

Cybersecurity vendors are not sitting idly by in the face of this threat. Companies like Trend Micro are advancing the implementation of "AI Secure Access (ZTSA)," which provides comprehensive visibility and control over GenAI services, and "Trend Cybertron," a proactive cybersecurity AI designed to eliminate blind spots.[18] In particular, technologies that inspect prompts and responses to AI assistants like Copilot or ChatGPT in real-time, detecting and blocking prompt injections hidden in HTML comments or white text via Natural Language Processing (NLP), serve as the first line of defense against EchoLeak-style attacks.[18]

Furthermore, the "Secure AI Factory" initiative, a collaboration with NVIDIA and Dell Technologies, aims to protect the entire AI lifecycle and build a secure foundation for AI models in enterprise environments.[18] However, uncertainty remains high regarding whether these defense technologies can completely outpace Mythos' autonomous learning capabilities and exploit generation speed.

---

5. Structural Transformation and Professional Recommendations to Avert a Catastrophic Scenario

Section 5 of the target article proposes structural changes as solutions to prevent the collapse of the financial system. These proposals are highly sound and inevitable approaches from the perspectives of cutting-edge systems engineering and AI safety. This section deepens these proposals and presents concrete recommendations that policymakers and financial institutions

should implement immediately.

5.1. Complete Transition to Micro-segmentation and Loose Coupling Architecture

The approach advocated by the article—"From tight coupling to loose coupling: dividing systems into self-contained units to prevent ripple effects"—is the highest priority issue in current financial infrastructure defense.

The greatest weakness in financial infrastructure is the flat network design where, once credentials are breached, attackers gain logically seamless access from front-end AI assistants to back-end SWIFT endpoints and COBOL core systems. A complete transition to loose coupling means strictly enforcing Zero Trust architecture principles and establishing "micro-segmentation" (granular defense walls) and "API gateways" at the hardware or strict logical level between each system.

By physically and logically separating transaction requests from their execution processes, even if a front-end Copilot is completely compromised by an unknown zero-day attack, the damage can be contained to a local information leak, preventing it from directly leading to the injection of fraudulent fund movement instructions into the core system.

5.2. Breaking the Monoculture through Ecosystem Diversification

The proposal to ensure diversity by "breaking away from Microsoft dependency and combining open-source technologies" is a direct and fundamental prescription for the "monoculture risks" mentioned earlier.

Just as genetic diversity allows biological ecosystems to acquire resistance to lethal pandemics, technical systems can also acquire "herd immunity" against the global spread of AI-driven zero-day worms by ensuring vendor and architectural diversity. Rather than relying on a specific OS (like Windows) or a specific cloud provider for all business foundations, system designs that intentionally mix multiple heterogeneous environments (multi-cloud, hybrid architectures, adoption of different OSs) are required. This makes it mathematically possible to mitigate the risk of a single critical zero-day vulnerability bringing down the entire system in a chain reaction.

5.3. Mandatory Segregation of Duties for AI Agents and Dynamic Shield Mechanisms (Human-in-the-Loop)

The most critical and immediate countermeasure is the strict application of the design philosophy: "Limited AI use: incorporate final human approval for transactions and writes (do not position AI at the autonomous core of the product)."

In the context of AI research and safety, granting an autonomous AI agent the authority to "Read" data versus the authority to "Write/Execute" or modify data entails entirely different dimensions of risk. The International AI Safety Report published in 2025 (guidelines referenced in policies like S17) strongly points out that traditional testing methods cannot guarantee

comprehensive safety due to the data-driven and adaptive nature of machine learning systems.[28]

Financial institutions must strictly restrict the role of AI assistants to "support tools (Read authority)" such as data retrieval, summarization, and analysis. For irreversible actions like actual fund transfers, writing data to core systems, and altering permissions, a "physical human approval process (Human-in-the-Loop)" using cryptographic Multi-Factor Authentication (MFA) must be mandated at the system level.

Furthermore, as an advanced countermeasure, implementing the "dynamic shielding mechanism" advocated in the S17 paper is indispensable. This is an approach where a separate, independent "monitoring shield AI" is deployed to observe and restrict the actions of AI agents, with shields dynamically splitting and merging depending on the agents' behaviors and coordination states to autonomously maintain system safety.[28] As attacking AIs become more sophisticated, it is necessary to deploy AIs with equivalent or superior reasoning capabilities in the defensive monitoring mechanisms to ensure safety through mutual oversight.

---

6. Conclusion: Irreversible Turning Point and Redefinition of Financial Resilience

As a result of the strict verification and in-depth analysis conducted in this report, it is evaluated that while the target article (mythos-copilot-financial-crisis-final.md) contains some technical discrepancies regarding the chronological applicability of EchoLeak (CVE-2025-32711), the probability of its central assertions and the presented collapse scenario is extremely high, and it accurately captures the severity of the real-world threat.

The threat of "Claude Mythos Preview"—which Anthropic faced in April 2026, forcing unprecedented emergency intervention by the US Treasury and FRB—signifies the absolute end of the era in cybersecurity history where "the threat of AI could be controlled by human cognitive capacity and defensive speed." Now that AI models can autonomously identify kernel-level vulnerabilities that remained hidden from human eyes for 17 years and construct complete exploit chains within hours, the "patching lead time" that remained for defenders has vanished.

The global financial industry is caught between the technical debt brought on by half-century-old legacy COBOL systems and the hasty modernization characterized by the chaotic integration of AI assistants, facing the most vulnerable transitional period in its history. While a select few institutions like JPMorgan Chase barely stand at the forefront of defense through Project Glasswing, the industry's collective response is fatally delayed, making the market terror demonstrated by the SaaSpocalypse a highly rational reaction.

As the target article warns, to prevent true systemic risk from materializing, the current approach of blindly relying on the convenience of AI must be abandoned immediately. The

industry must pivot towards a fundamental architectural redesign—"fragmenting systems via loose coupling," "ensuring ecosystem diversity," and "rebuilding strict human intervention processes (Human-in-the-Loop)"—even if it entails the short-term pain of sacrificing efficiency. What Mythos thrusts upon the world is not merely the arrival of a new hacking tool, but an existential crisis: the loss of "deterministic human governance" over global infrastructure. The resilience of the financial system is now forced to undergo a fundamental redefinition in a relentless race against machine reasoning speed.

引用文献

1. Anthropic's Claude Mythos AI fears trigger $2 trillion wipeout in IT ..., 4月 12, 2026にアクセス、https://m.economictimes.com/news/new-updates/anthropics-claude-mythos-ai-fears-triggers-2-trillion-wipeout-in-it-stocks-jpmorgan-ceo-jamie-dimon-warns-ai-will-likely-/articleshow/130187154.cms
2. Anthropic's AI, the most powerful on Earth, prompted an emergency meeting on Wall Street, but JPMorgan Chase, which had the "cure," was absent. | 律动 BlockBeats on Binance Square, 4月 12, 2026にアクセス、https://www.binance.com/en/square/post/310884313815841
3. Bessent and Powell send Wall Street's biggest banks a warning - TheStreet, 4月 12, 2026にアクセス、https://www.thestreet.com/economy/bessent-and-powell-send-wall-streets-biggest-banks-a-warning
4. Treasury Proposes Rule to Implement the GENIUS Act’s Requirements to Counter Illicit Finance, 4月 12, 2026にアクセス、https://home.treasury.gov/news/press-releases/sb0435
5. FinCEN Proposes Rule to Fundamentally Reform Financial Institution Programs Designed to Fight Illicit Finance, 4月 12, 2026にアクセス、https://www.fincen.gov/news/news-releases/fincen-proposes-rule-fundamentally-reform-financial-institution-programs
6. Anthropic Launches Project Glasswing to Use AI to Find and Fix Critical Software Vulnerabilities, 4月 12, 2026にアクセス、https://www.infosecurity-magazine.com/news/anthropic-launch-project-glasswing/
7. An AI Model Scared the Federal Reserve and US Treasury Into Summoning Wall Street, 4月 12, 2026にアクセス、https://aimmediahouse.com/ai-bfsi/an-ai-model-scared-the-federal-reserve-and-us-treasury-into-summoning-wall-street
8. mythos-copilot-financial-crisis-final.md
9. Overnight, will countless bugs appear on your phone and computer? - 36氪, 4月 12, 2026にアクセス、https://eu.36kr.com/en/p/3758841606079234
10. Claude Mythos: Benchmark-Dominating AI with Real Risks - Labellerr, 4月 12, 2026にアクセス、https://www.labellerr.com/blog/anthropic-claude-mythos-capabilities/
11. Claude Mythos: The AI Model Anthropic Built (and Refused to Release) - Thesys, 4

月 12, 2026にアクセス、https://www.thesys.dev/blogs/claude-mythos
12. CVE-2026-4747: FreeBSD RPCSEC_GSS RCE Vulnerability - SentinelOne, 4月 12, 2026にアクセス、https://www.sentinelone.com/vulnerability-database/cve-2026-4747/
13. CVE-2026-4747 | Mondoo Vulnerability Intelligence, 4月 12, 2026にアクセス、https://mondoo.com/vulnerability-intelligence/vulnerability/CVE-2026-4747
14. Anthropic's new AI model finds and exploits zero-days across every major OS and browser, 4月 12, 2026にアクセス、https://www.helpnetsecurity.com/2026/04/08/anthropic-claude-mythos-preview-identify-vulnerabilities/
15. Mythos autonomously exploited vulnerabilities that survived 27 years of human review. Security teams need a new detection playbook, 4月 12, 2026にアクセス、https://venturebeat.com/security/mythos-detection-ceiling-security-teams-new-playbook
16. Claude Mythos Preview \ red.anthropic.com, 4月 12, 2026にアクセス、https://red.anthropic.com/2026/mythos-preview/
17. Anthropic announces 'Project Glasswing' in alliance with tech giants to strengthen global cybersecurity, 4月 12, 2026にアクセス、https://www.aninews.in/news/business/anthropic-announces-project-glasswing-in-alliance-with-tech-giants-to-strengthen-global-cybersecurity20260408083637
18. Preventing Zero-Click AI Threats: Insights from EchoLeak | Trend Micro, 4月 12, 2026にアクセス、https://www.trendmicro.com/en/research/25/g/preventing-zero-click-ai-threats-insights-from-echoleak.html
19. How Microsoft 365 Copilot Customers Should Think About Agent Control - Rubrik, 4月 12, 2026にアクセス、https://www.rubrik.com/blog/company/26/4/how-microsoft-365-copilot-customers-should-think-about-agent-control
20. 'EchoLeak' AI Attack Enabled Theft of Sensitive Data via Microsoft 365 Copilot, 4月 12, 2026にアクセス、https://www.securityweek.com/echoleak-ai-attack-enabled-theft-of-sensitive-data-via-microsoft-365-copilot/
21. EchoLeak in Microsoft Copilot: What it Means for AI Security - Varonis, 4月 12, 2026にアクセス、https://www.varonis.com/blog/echoleak
22. CVE-2025-32711 Detail - NVD, 4月 12, 2026にアクセス、https://nvd.nist.gov/vuln/detail/cve-2025-32711
23. DXC's Mainframe Engineering Services team helps banks master COBOL core banking, 4月 12, 2026にアクセス、https://dxc.com/insights/knowledge-base/blogs/why-banks-still-rely-on-cobol-driven-mainframe-systems
24. 44 Legacy System Modernization Statistics Every Enterprise Should Know in 2026, 4月 12, 2026にアクセス、https://www.dreamfactory.com/hub/legacy-system-modernization-statistics
25. Learn how progressive modernization enables banks to transform core systems

without big bang risk. - Softjourn, 4月 12, 2026にアクセス、 https://softjourn.com/insights/core-banking-modernization-in-5-steps
26. Legacy Banking Costs: The Hidden Financial and Strategic Burden on Financial Institutions - FinTech Consultant, 4月 12, 2026にアクセス、 https://www.fintech-consultant.com/post/hidden-costs-of-legacy-banking-infrastructure-on-financial-institutions
27. Information and Cyber Security Within the Government | Digital Policy Office, 4月 12, 2026にアクセス、 https://www.digitalpolicy.gov.hk/en/our_work/digital_infrastructure/information_cyber_security/government/
28. Formal methods for safety-critical machine learning: a systematic literature review - Frontiers, 4月 12, 2026にアクセス、 https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2026.1749956/full
29. Project Glasswing - Anthropic, 4月 12, 2026にアクセス、 https://www.anthropic.com/project/glasswing
30. Claude Mythos, Anthropic AI capable of hacking any software, joins forces with Google, Apple, AWS & more; Users' personal data at risk?, 4月 12, 2026にアクセス、 https://m.economictimes.com/news/new-updates/claude-mythos-anthropic-ai-capable-of-hacking-any-software-joins-forces-with-google-apple-aws-more-users-personal-data-at-risk/articleshow/130106401.cms
31. Anthropic launches Project Glasswing, says Claude Mythos found risks in every major OS and browser, 4月 12, 2026にアクセス、 https://www.financialexpress.com/life/technology-anthropic-defines-project-glasswing-says-mythos-has-found-vulnerabilities-in-thousands-of-systems-4201201/
32. Why Project Glasswing Marks a Turning Point for Cybersecurity | Arctic Wolf, 4月 12, 2026にアクセス、 https://arcticwolf.com/resources/blog/project-glasswing-marks-a-turning-point-for-cybersecurity/
33. Policymaker's Guide to International AI Safety Coordination | Digital Watch Observatory, 4月 12, 2026にアクセス、 https://dig.watch/event/india-ai-impact-summit-2026/policymakers-guide-to-international-ai-safety-coordination/