SYNTHETIC INSIGHTS

R&D REPORTS · ISSUE NO. 002 · PUBLIC EDITION

A SYNTHETIC INSIGHTS RESEARCH PAPER

Claude Mythos

The Clock, Not the Model

Autonomous vulnerability discovery has arrived. The model is contained. The clock is not. An analysis of Claude Mythos and the eighteen-month window every organization now shares.

SERIES
SI R&D REPORTS

PUBLISHED
MAY 29, 2026

VERSION
1.0 — PUBLIC EDITION

DOMAIN
AI SECURITY & STRATEGY

SI-RD-002

INSIDE COVER · COLOPHON & METHODOLOGY

The Clock, Not the Model

Claude Mythos, autonomous vulnerability discovery, and the democratization window

PUBLISHER	Synthetic Insights LLC
SERIES	Synthetic Insights R&D Reports
REPORT ID	SI-RD-002
VERSION	1.0 — Public Edition
PUBLISHED	May 29, 2026
DOMAIN	AI Security & Strategy
AFFILIATION	Synthetic Insights is an independent firm and is **not affiliated with Anthropic**, OpenAI, or any vendor named in this report. This is a third-party analysis of publicly disclosed information.
DISTRIBUTION	Public web edition. Free to share with attribution.

METHODOLOGY & PROVENANCE

This analysis applies the Synthetic Insights editorial standard: **at least three independent sources** for every material claim, an explicit perspective spectrum that gives the skeptical camp a fair hearing, and **no fabrication**. Where a number could not be independently corroborated, it is labelled as such rather than presented as settled fact.

Three source classes are distinguished throughout, by inline tag: `ANTHROPIC-CLAIM` denotes a vendor-self-reported metric; `INDEPENDENT` denotes a finding corroborated by external parties; and `UNVERIFIED` denotes a claim in public circulation that this analysis could not confirm. Readers should weigh each accordingly.

Source material spans Anthropic's primary disclosures, six independent security-firm assessments, peer-vendor and policy reporting, and the skeptical counter-current. The full bibliography appears at the close.

CONTENTS

Table of Contents

EXECUTIVE SUMMARY

The Model Is Contained. The Clock Is Not.

Claude Mythos is a contained event. The capability it demonstrates is not. The decision that matters for almost every organization is not whether to worry about one restricted model — it is what to do with the window before that capability is everywhere.

On April 7, 2026, Anthropic disclosed Claude Mythos Preview: a frontier model whose autonomous vulnerability-discovery ability is sufficiently dangerous that the company withheld it from general release and gated access to roughly fifty federal and critical-infrastructure partners. The industry's attention split immediately into two camps — those who called it a step-change in offensive capability, and those who called the framing hype. Both camps are partly right. This report argues that their argument is the wrong one to be having.

The capability is real and, in narrow ways, unprecedented. It is also *reproducible* — several of Mythos's headline results have been matched with freely available open-source models. That reproducibility is not a reason to relax. It is the entire point: if the capability is not unique to one gated model, then it will commoditize to open and on-device models on a predictable timeline. Estimates from inside the industry put the near-term outpacing window at **three to five months** and the broad democratization window at roughly **eighteen months**. That clock — not the model — is the thing to mobilize around.

THE CORE THESIS

Mythos itself is a contained event. The same autonomous vulnerability-discovery capability will reach open and on-device models within roughly eighteen months. **The window — not the model — is what every organization must mobilize around.** This report gives the industry a deadline and a playbook for the time that remains.

The Three Headline Facts

10,000+

HIGH/CRITICAL VULNS

Identified across Project Glasswing since April 2026.

90.6%

TRUE-POSITIVE RATE

Six external firms, 1,752-vuln sample — the strongest independent corroboration.

62.4%

CONFIRMED HIGH-CRIT

Of that same independently assessed sample.

>99%

REMAIN UNPATCHED

Anthropic's own concession: discovery outruns remediation.

The Spectrum, in One Line

Alarmists say the rules of the game have changed; **skeptics** say the capability was achievable with older models and the bank-and-regulator reaction is "hysteria"; and **opportunists** are already pairing Mythos-class tooling with their own telemetry and selling "agentic defense." The analysis this report defends: the skeptics are right that the capability is not unique — *which is exactly why the democratization clock is the real risk.*

What to Do

In one sentence: **treat the next eighteen months as a dash to close known exposure and stand up AI-native, continuous defensive scanning before the same tools are turned on you.** The quarterly penetration test is obsolete; the question shifts from "can we patch before exploitation?" to "were we already hit, and would we know?" Sections 5 through 7 translate that into a near-term checklist, a six-to-eighteen-month strategy, and the governance and insurance posture leadership should expect.

How to Read This Report

Section 1 establishes what Mythos verifiably is. Section 2 develops the central thesis — the clock. Section 3 steelmans the skeptics, because a credible analysis must. Sections 4 through 7 are the strategic core: implications, the playbook, the under-reported code-debt problem, and the access economy. Section 8 compares Mythos to its peers — the evidence the clock is already ticking — and Section 9 closes with three scenarios for the window's end.

SECTION 01

What Mythos Is: Capability, Validated and Caveated

Before the argument about what it means, the facts about what it does — and a disciplined separation of what the vendor claims from what outsiders have confirmed.

1.1 The Disclosure

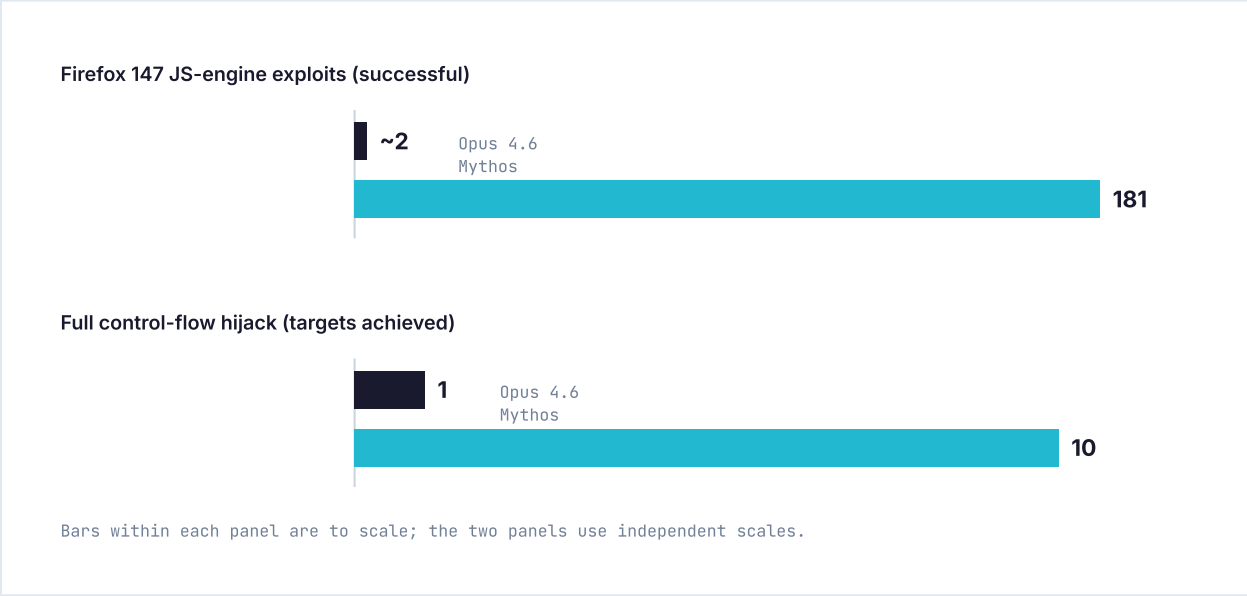
Anthropic announced Claude Mythos Preview on **April 7, 2026**. The company frames it as a general-purpose frontier model whose cyber capability "emerged as a downstream consequence of general improvements in code, reasoning, and autonomy" — not a purpose-built offensive tool — and states that it was withheld from general release specifically because of that offensive-cyber risk. ANTHROPIC-CLAIM

1.2 The Capability Jump

Anthropic's own benchmarks describe a sharp discontinuity over its prior model, Opus 4.6. On exploit development against Firefox 147's JavaScript engine, Anthropic reports that Opus 4.6 succeeded at roughly two attempts while Mythos succeeded at **181**, with twenty-nine of those achieving register control. On full control-flow hijacking, Anthropic reports Opus 4.6 reached one target where Mythos reached **ten**. ANTHROPIC-CLAIM

Figure 1 — The Reported Capability Jump

Mythos versus its predecessor on two of Anthropic's headline benchmarks. Vendor-self-reported; presented as disclosed, not independently re-run.



Source: red.anthropic.com (Mythos Preview disclosure, Apr 2026). Vendor-self-reported.

Anthropic further reports that Mythos autonomously chained four browser vulnerabilities into a working exploit — a JIT heap spray escaping both the renderer and OS sandboxes — and produced an autonomous Linux local privilege escalation via a race condition with a KASLR bypass. `ANTHROPIC-CLAIM` The most-cited single result is a FreeBSD NFS remote-code-execution exploit (`CVE-2026-4747`), in which Mythos is said to have produced a root exploit using a twenty-gadget return-oriented-programming chain across six sequential RPC requests. That specific result carries a meaningful skeptic caveat, addressed in Section 3.

1.3 Scale — and Its Independent Corroboration

Under Project Glasswing, Anthropic reports more than **10,000 high- and critical-severity vulnerabilities** identified since April 2026. `ANTHROPIC-CLAIM` The strongest independent check on that program comes from six external security firms, which assessed a **1,752-vulnerability sample** and found a **90.6% true-positive rate** with **62.4% confirmed as high- or critical-severity**. `INDEPENDENT`

Figure 2 — Independent Validation of a Glasswing Sample

Six external security firms assessed a 1,752-vulnerability sample. This is the load-bearing independent evidence in the entire disclosure.

Sources: Anthropic Glasswing update; Help Net Security; The Hacker News. Independently corroborated.

Other independent signals point the same direction without relying on Anthropic's framing. Cloudflare reported roughly two thousand bugs — about four hundred high-critical — found in its own infrastructure with frontier models, and a named Glasswing result, the wolfSSL certificate-forgery flaw `CVE-2026-5194`, was confirmed and patched. `INDEPENDENT`

> THE HONEST CONCESSION
>
> Anthropic's own most important caveat: **more than 99% of the vulnerabilities discovered remain unpatched.** Faster discovery is not faster remediation. The capability widens the gap between "known" and "fixed" — which, as Section 4 argues, is precisely where the strategic risk lives.
>
> Source: red.anthropic.com / Glasswing update. Anthropic-self-reported.

1.4 What Remains Unverified

Several claims in public circulation could not be independently confirmed for this analysis and are flagged accordingly: standalone public case studies for N-day exploit construction from patch commits; decompilation of stripped or compiled binaries; a UK AI Safety Institute cyber-range claim; the precise "seventeen-year-old" provenance of `CVE-2026-4747`; and the disaggregation of the 10,000-plus figure between Anthropic-

internal and partner-found vulnerabilities. `UNVERIFIED` None of these are necessary to the report's thesis; they are noted so the record is honest about its edges.

SECTION 02 · THE THESIS

The Clock: A Capability on a Timeline

Containment buys time, not safety. The defining feature of this moment is not a model — it is a countdown that started the day the capability was shown to be reproducible.

2.1 Containment Is Temporary by Construction

The instinct to treat a gated model as a contained threat is understandable and, for the model itself, correct. Roughly fifty vetted partners can use Mythos; the rest of the world cannot. But the capability is not the same thing as the model. A capability that has been demonstrated — and, crucially, *reproduced* — is a capability on a diffusion curve. The history of every dual-use software technique, from fuzzing to symbolic execution to automated exploit generation, is a history of compression: from research lab, to commercial tool, to open-source default, to commodity.

The skeptics, examined in detail in the next section, have already established the load-bearing premise for this argument. If freely available open-source models can match several of Mythos's headline results today — and the evidence indicates some can — then the commoditization is not a forecast. It is already underway. The only open question is the slope.

> *The compression of time is the story. We went from the carriage age to the train age to the industrial age to the computer age — and now to an AI age where each turn of the wheel arrives faster than the last.*
>
> — A framing recurring in Synthetic Insights' analysis of the moment

2.2 Two Windows

Two distinct timelines structure the response, and conflating them is a common error. The first is the **near-term competitive window**: Palo Alto Networks' chief technology officer, Lee Klarich, has publicly described a "narrow three-to-five-month window" for defenders to outpace adversaries before AI-driven exploitation becomes "the new norm." `INDEPENDENT` The second is the **democratization window**: the roughly eighteen-month horizon on which this class of capability is expected to reach open and, eventually, on-device models running on hardware measured in billions of units.

Figure 3 — The Two Windows

A near-term competitive window (months) nested inside a broader democratization window (roughly a year and a half). The clock runs on both at once.

Sources: Palo Alto Networks (Klarich, via CNBC) for the 3–5 month figure; industry consensus + Synthetic Insights analysis for the ~18-month democratization horizon.

2.3 Why the Window Is the Mobilization Frame

Framing the problem as "a dangerous model exists" produces paralysis or complacency, depending on whether you have access. Framing it as "you have a window" produces action. The window is finite, it is shared by every organization regardless of whether it sits inside Glasswing, and it has a defensible end date. That makes it the right unit of planning: not a threat to monitor, but a deadline to beat. The remainder of this report treats it as exactly that.

> **THE REFRAME**
>
> The question is no longer *"Is this one model dangerous?"* It is *"What must be true about our security posture before this capability is commodity?"* — and the honest answer for most organizations involves changes that take most of the available window to make.

SECTION 03

The Skeptic Ledger: Hype Versus Signal

A credible analysis steelmans its strongest opposition. The skeptics make a serious case — and, properly understood, it strengthens the thesis rather than undermining it.

3.1 The Case for "This Is Hype"

A substantial body of expert commentary holds that the Mythos framing overstates the novelty. The strongest points, fairly stated:

- **The capability is not unique.** The security research firm AISLE, writing on "the jagged frontier," reports that several of Mythos's headline results — including decades-old bug discoveries — are reproducible with *freely available open-source models*, and argues that scale and coordination, not model recency, are what actually changed. INDEPENDENT
- **The reaction is disproportionate.** CNBC reported in May 2026 that multiple researchers consider the capabilities "achievable using older models," with some describing the bank-and-regulator conversation as "hysteria." INDEPENDENT
- **It lowers the bar, it does not raise the ceiling.** As researcher Charlie Eriksen put it via CNBC, smaller models can match the results but require more skill and tooling — so Mythos lowers the bar for *less-skilled* attackers rather than expanding what sophisticated ones could already do.
- **The surge has not materialized.** Barracuda's "Mythos Hype Index" found 2026 CVE growth only "slightly above historical trends," meaning the predicted AI-driven vulnerability flood is, as of its analysis, not yet visible in the aggregate data. INDEPENDENT
- **The marquee demo is contestable.** A researcher writing at Flying Penguin argues the celebrated FreeBSD "seventeen-year-old zero-day" framing is "marketing": the same bug was found by eight open-weight models (including one of 3.6 billion parameters), Opus 4.6 built the exploit in roughly four hours with human guidance, and the target lacked production mitigations such as KASLR and stack canaries. INDEPENDENT

3.2 The Signal

Take the skeptics entirely at their word. Grant that the capability is reproducible with open models, that the headline demo ran against a soft target, and that the aggregate CVE curve has not yet bent. The conclusion that follows is not "relax." It is the opposite.

WHY THE SKEPTICS STRENGTHEN THE THESIS

If the capability is **not** unique to one gated frontier model — if open-weight models already replicate much of it — then containment of Mythos protects almost nothing, and democratization is not a future risk but a present trend. **The skeptics' best argument is the clock's best evidence.**

The skeptics are correctly puncturing the wrong target. "Mythos is not magic" is true and beside the point. The defensible reading of their own evidence is that the dangerous capability is already diffusing through the open ecosystem — which is the premise the rest of this report builds on. A view from The Conversation captures the calibrated middle: Mythos is "a cybersecurity threat, but it doesn't rewrite the rules of the game." Agreed. The rules did not change. The *clock* did.

SECTION 04

Strategic Implications

The capability does not just add a tool to the attacker's kit. It changes the structure of the contest — the tempo, the balance, and who gets to see the board.

4.1 The Offense-Defense Balance Has Shifted

Time-to-exploit has collapsed. Industry analyses put the typical interval from disclosure to weaponization at roughly 2.3 years in 2018; with machine-speed patch-diff engineering, that interval now compresses toward hours. The structural problem is that attackers face none of the friction defenders do — **no procurement cycle, no vendor vetting, no ethics review, no change-control board.** When the same capability becomes available to both sides, the side without governance gates moves first. That is not a statement about intent; it is a statement about latency.

4.2 Access Asymmetry Is a Concentration of Advantage

Inside Project Glasswing, roughly fifty partners get their code and infrastructure scanned by a frontier model. Everyone else ships the same patched libraries blind to the scope of what those models can find. The asymmetry is not subtle, and it does not distribute evenly.

Figure 4 — The Access Asymmetry

Frontier vulnerability-discovery access is concentrated in a small set of vetted partners. The vast majority of the software ecosystem is outside the gate.

Sources: anthropic.com/glasswing; The Register; Linux Foundation. Partner count and funding figures are Anthropic-disclosed; the asymmetry framing is Synthetic Insights analysis.

The most exposed and least resourced population is open-source maintainers. Many of the ten-thousand-plus vulnerabilities live in projects with a single maintainer, and the **$4 million** Anthropic earmarked for open-source donations is symbolic against the scale of the problem. INDEPENDENT The libraries that underpin the modern software stack are maintained by people who do not have Glasswing access and cannot afford it.

4.3 The Question for Security Leaders Has Changed

Quarterly penetration testing plus reactive patching was a defensible posture when exploitation lagged disclosure by months or years. At machine speed it is obsolete. The operative question moves from *"can we patch before someone exploits this?"* to *"were we already hit, and would we know?"* That reframing makes look-back telemetry at AI speed table stakes rather than a maturity-model aspiration. Microsoft's Patch Tuesday is already on pace to break its annual CVE record as the AI-driven patch wave takes hold. INDEPENDENT

> **THE STRUCTURAL ASYMMETRY**
>
> Defenders inherit every friction of responsible operation; attackers inherit none. As the capability commoditizes, the gap is not closed by better tools alone — it is closed by removing defender latency: pre-authorization, automation, and continuous scanning that runs at the adversary's tempo.

SECTION 05 · THE ACTIONABLE CORE

The Playbook: What to Do Inside the Window

A deadline without a plan is just anxiety. This section is the part most organizations should act on regardless of where they sit on the access spectrum — split into what to do now and what to build over the window.

5.1 Near-Term: The First Ninety Days

Four moves are achievable quickly and disproportionately reduce exposure:

1. **Map your vendors against Glasswing.** Know which of your critical software suppliers are inside the gate (and therefore being scanned) and which are outside (and therefore shipping blind to scope). That map is your exposure surface.
2. **Set a 72-hour internal SLA for AI-discovered CVEs.** Treat a vulnerability surfaced by an AI-driven discovery program as a different severity class than a routine advisory — because its time-to-exploit is.
3. **Audit your own AI-generated code.** The flaws most likely to be found in your stack in the near term are the ones your own tooling wrote (Section 6). Inventory and scan it before an adversary's model does.
4. **Pre-authorize machine-speed containment.** If containment requires a human approval that takes hours, you have already lost the tempo contest. Decide now what your systems may do autonomously when an AI-class exploit is detected.

5.2 Strategic: The Six-to-Eighteen-Month Build

1. **Make AI-native defensive tooling standard.** Continuous, automated scanning that operates at the adversary's tempo should be the baseline, not a pilot. The correct end state for a mature organization is an internal, AI-driven continuous red-team capability — running against your own estate before the same class of tools is turned on you.

2. **Engage your cyber insurer before renewal.** Underwriting is moving faster than regulation (Section 7). Get ahead of the requirements rather than discovering them at renewal.
3. **Fund and reduce your open-source dependency exposure.** The maintainers of your most-used libraries are the least resourced node in your supply chain. Dependency strategy is now a security control.
4. **Brief the board.** This is a governance shift, not a technical footnote. Patch cadence, procurement, and the autonomy you grant defensive systems are now board-level questions.

THE ONE-SENTENCE STRATEGY

Spend the window standing up **continuous, AI-native, pre-authorized defensive scanning** and closing known exposure — so that when the capability is commodity, your tempo already matches the adversary's.

5.3 Defense-in-Depth Still Works — Cycle Time Is the Variable

A clarification, because the panic version of this story is wrong. Layered defense has not stopped working. The argument is not that controls are futile; it is that the *cycle time* of the whole loop — detect, decide, contain, remediate — must compress from weeks to hours to stay inside the adversary's. The investment is less about novel controls than about removing latency from the controls you already have.

SECTION 06

The Code-Debt Reckoning

The most under-reported second-order story: the same wave of AI that now finds flaws at scale also wrote a great many of them — and most of those flaws have never been scanned.

6.1 AI Writes the Flaws That AI Now Finds

The discourse fixates on a Mythos-wielding adversary. For most organizations, that is not the nearest danger. The nearer danger is the large and growing volume of AI-generated code already running in production pipelines — code that was written fast, reviewed lightly, and never subjected to the kind of scanning a frontier model can now perform.

> *The same technology that can find every flaw at scale is the technology that has been writing them at scale. The collision was always going to happen; Mythos just set the date.*
>
> — Synthetic Insights analysis

The quantitative signal is stark. Gartner projects a **2,500% rise in AI-assisted software defects by 2028** — a wave of latent flaws arriving exactly as the capability to find each one matures. INDEPENDENT The two curves are not independent; they are the same phenomenon viewed from opposite ends. The organizations most exposed are precisely those that adopted AI-assisted development fastest and governed it least.

> **THE NEAR-TERM REALITY FOR MOST ORGANIZATIONS**
>
> Your most likely near-term incident is not an external actor wielding a gated frontier model. It is **your own unscanned, AI-generated code meeting a commoditized discovery tool.** The defensive priority follows directly: inventory and scan what your tooling wrote, now, while the discovery advantage is still mostly yours.

SECTION 07

The Access Economy: Policy, Regulation, and Insurance

Who gets to use the capability is now a question of governance without much precedent — and the institutions that will move first are not the legislatures.

7.1 Glasswing as a Governance Structure

Project Glasswing launched April 7, 2026 with roughly fifty partners — twelve named, including AWS, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorganChase, the Linux Foundation, Microsoft, NVIDIA, and Palo Alto Networks — backed by **$100 million in usage credits and $4 million in open-source donations**, with ninety-day public reporting. ANTHROPIC-CLAIM It is, in effect, a private decision about who may access a dual-use capability — a governance structure standing in for one that does not yet exist in public law.

7.2 The Policy Drama

- **The White House blocked expansion.** An Anthropic plan to extend access to roughly seventy additional organizations was opposed by the White House, citing misuse risk — an unauthorized-access incident was reportedly already detected shortly after launch — and Anthropic compute constraints. INDEPENDENT
- **Congress is engaged.** The House Homeland Security committee received a classified briefing and live demonstration, and thirty-two bipartisan lawmakers pressed the National Cyber Director for expanded *defensive* access. A hearing is planned. INDEPENDENT
- **The asymmetry is global.** The EU has made "little progress" on testing access and is falling behind; US banks can probe their defenses with Mythos-class tooling while most non-partner nations and organizations cannot. INDEPENDENT

7.3 The "Restricted-AI Era" Precedent

The gated-release model is being read as a template rather than an exception. Georgetown's Helen Toner, via *Nature*, judges that the approach is "likely to be adopted by other AI laboratories"; *Time* framed the moment as "'Too Dangerous to Release' becoming AI's new normal." INDEPENDENT Whatever one thinks of the policy, the precedent compounds the access asymmetry of Section 4 into a durable feature of the landscape.

7.4 Insurance Will Move Before Law

Software-liability law lags badly. Cyber insurance does not. The likely near-term trajectory mirrors how multi-factor authentication went from best-practice to underwriting requirement almost overnight: **expect 72-hour AI-CVE patch SLAs and AI-defensive-tooling requirements written into policy terms.** INDEPENDENT For most organizations, the insurer's renewal questionnaire will impose the new posture well before any statute does — which is precisely why Section 5 puts "engage your insurer before renewal" on the strategic list.

> THE GOVERNANCE TAKEAWAY
>
> Access to the capability is being allocated privately, contested politically, and — most consequentially for planning — about to be priced by underwriters. Leadership should expect the insurance market, not the legislature, to set the near-term compliance bar.

SECTION 08

Mythos Versus Peers

The single most decisive evidence that the clock is already ticking is the company Mythos keeps. It is not alone on the frontier, and it is not alone in being gated.

Mapping Mythos against its closest peers shows two things at once: the capability is converging across multiple labs, and the access model is fragmenting into a haves-and-have-nots structure. And it is already diffusing past the frontier labs on two fronts — an open-weight lineage anyone can download (Deep Hat) and a venture-funded proprietary entrant selling it as a product (depthfirst). That diffusion, visible in a single table, is the democratization vector made concrete.

Model / Program	Capability Posture	Access Model	Who Can Use It	Democratization Vector
Claude Mythos (Anthropic)	Frontier autonomous vuln discovery; the headline capability jump	Gated — Project Glasswing	~50 vetted federal / critical-infra partners	Low today; high once capability commoditizes
Opus 4.6 / 4.7 (Anthropic)	Strong general model; meaningfully below Mythos on offensive benchmarks	Commercial general release	Broad commercial availability	Already broadly available; partial capability

Model / Program	Capability Posture	Access Model	Who Can Use It	Democratization Vector
GPT-5.5-Cyber (OpenAI)	Competitive cyber-focused frontier capability	Gated — "Trusted Access for Cyber"	Vetted defenders (Akamai, Cisco, Cloudflare, CrowdStrike, Fortinet, Palo Alto, SentinelOne, Tenable)	Mirrors the gated model — a second walled garden
Deep Hat (Kindo; formerly WhiteRabbitNeo)	Open-weight security LLM; replicates parts of the capability	Open weights (downloadable) + proprietary enterprise variant	Anyone (open models); enterprises via Kindo	**High — open weights make the capability downloadable**
depthfirst (proprietary)	AI-native security platform; in-house agents find vulnerabilities at scale	**Proprietary** — in-house dfs-mini models, not released	depthfirst customers, via its General Security Intelligence platform	**High — commercial proliferation ($120M+ raised)**

Sources: OpenAI (GPT-5.5-Cyber); CNBC; BankInfoSecurity; AISLE; Kindo (Deep Hat / WhiteRabbitNeo); depthfirst (Series A/B — TechCrunch, BusinessWire). Capability postures are summarized from public disclosures and independent commentary; the open-weight row reflects the reproducibility findings discussed in Section 3.

WHAT THE TABLE SAYS

Two frontier labs have shipped gated cyber capability within weeks of each other — and the capability is already diffusing past them, through open weights and a venture-funded proprietary startup alike. The gate is a speed bump, not a wall — **and the bottom two rows are the clock, printed.**

8.1 The Market's Own Verdict

The capital markets briefly priced this as an extinction event for incumbent security vendors — CrowdStrike fell about 11% and Palo Alto about 7% on April 10 — then reversed within days as both firms repositioned the capability as a product input, pairing Mythos-class tooling with their own telemetry to sell "agentic defense." `INDEPENDENT` IBM cited Mythos as the trigger for an open-source security push, and Google reported thwarting a real-world AI-driven mass-exploitation attempt. The market's round trip is itself a data point: the incumbents do not believe the capability is contained — they are racing to operationalize it.

SECTION 09

Outlook: Three Eighteen-Month Scenarios

The window has more than one possible ending. Which one arrives depends less on the models than on how fast defenders compress their cycle time and how access is governed. Three scenarios, with the leading indicators to watch.

Scenario A — Defender Wins

AI-native defensive scanning becomes standard practice inside the window. Insurers force the posture; major vendors ship continuous-scanning defaults; the discovery-remediation gap narrows even as discovery accelerates. The transitional risk is real but bounded, and the long-run equilibrium tilts defender-favorable — roughly the outcome Anthropic itself predicts.

Leading indicators: the aggregate CVE-remediation rate starts closing the gap with discovery; cyber-insurance terms standardize on AI-tooling requirements; open defensive tools proliferate as fast as offensive ones.

Scenario B — Asymmetry Hardens

The gated-release model entrenches. A small set of well-resourced organizations and nations operate with frontier defensive visibility while everyone else — most enterprises, most governments, the open-source commons — remains outside. Democratization of *offensive* capability outpaces democratization of *defensive* capability, and the asymmetry of Section 4 becomes a durable structural feature rather than a transitional one.

Leading indicators: access-expansion efforts stay blocked; EU and Global-South access continues to lag; OSS maintainer burnout and under-resourcing worsen even as their code is scanned by adversaries.

Scenario C — Commoditization Shock

Democratization arrives faster than eighteen months. Open and on-device models reach Mythos-class discovery capability ahead of schedule, before the median organization has compressed its cycle time. The result is a sharp rise in exploited-in-the-wild AI-discovered vulnerabilities and a scramble to stand up defenses that should have been built during the window.

Leading indicators: open-model capability benchmarks jump; the Barracuda-style CVE curve bends sharply upward; incident reports begin citing AI-discovered flaws as root cause at scale.

THE MOBILIZATION CALL

The difference between Scenario A and Scenario C is not which model exists — it is what defenders did with the time they were given. **Treat the next eighteen months as the dash they are.** The model is contained; the clock is not; and the clock is the only part of this story you can still act on.

BIBLIOGRAPHY

Sources & Source-Class Index

Every material claim in this report traces to at least three independent sources where independence was available. Source classes are indicated inline throughout and summarized below.

Source-Class Key

`ANTHROPIC-CLAIM` — vendor self-reported metric, presented as disclosed.
`INDEPENDENT` — corroborated by one or more external parties.
`UNVERIFIED` — in public circulation but not confirmed for this analysis.

Primary & Capability

Source	Relevance	Class
Anthropic — Mythos Preview (red.anthropic.com, Apr 2026)	Primary disclosure; capability benchmarks, exploit chaining, CVE-2026-4747	Anthropic-claim
Anthropic — Project Glasswing & initial update	10,000+ vulns; partner list; $100M / $4M figures	Anthropic-claim
Six external security firms (via Help Net Security; The Hacker News)	1,752-sample validation: 90.6% TP, 62.4% high-crit	independent
Cloudflare — "cyber frontier models" blog	~2,000 bugs / ~400 high-crit in own infra; wolfSSL CVE-2026-5194	independent
SecureWorld — Mythos zero-days coverage	FreeBSD RCE reporting	independent

Skeptic Counter-Current

Source	Relevance	Class
AISLE — "the jagged frontier"	Open-model reproduction; scale > recency	independent
CNBC — banks/regulators "hysteria" (May 8, 2026)	Capabilities "achievable using older models"	independent
Barracuda — "Mythos Hype Index"	2026 CVE growth only slightly above trend	independent
Flying Penguin — CVE-2026-4747 critique	"Marketing" critique; 8 open models; soft target	independent
The Conversation — "doesn't rewrite the rules"	Calibrated-middle assessment	independent

Policy, Market & Implications

Source	Relevance	Class
Bloomberg; Sherwood	White House blocks ~70-org expansion	independent
Nextgov/FCW; CyberScoop; Axios	House Homeland briefing; 32 bipartisan lawmakers	independent
Nature (Toner); Time	"Restricted-AI era" precedent	independent
Rest of World; WEF	Global access asymmetry; EU lag	independent
OpenAI; CNBC	GPT-5.5-Cyber "Trusted Access for Cyber"	independent
CNBC (Palo Alto / Klarich)	"Three-to-five-month window"	independent
BankInfoSecurity; CNBC (IBM, Google)	Vendor pivot; thwarted mass-exploitation attempt	independent

Source	Relevance	Class
Gartner (via Netwoven)	2,500% AI-defect rise by 2028	independent
Security Boulevard	Microsoft Patch Tuesday CVE-record pace	independent
The Register; Linux Foundation	OSS maintainer exposure	independent
Munich Re; Insurance Business; IAPP	Cyber-insurance + liability trajectory	independent
CETaS/Turing; Cloud Security Alliance; Control Risks; SentinelOne	Analytical grounding for implications	independent

A NOTE ON INDEPENDENCE

Synthetic Insights is not affiliated with Anthropic, OpenAI, or any vendor named here. Anthropic-self-reported figures are labelled as such and were not independently re-run; the report's conclusions rest on the independently corroborated subset and on the skeptical literature, both of which point — for different reasons — to the same conclusion about the democratization clock.

Prepared with AI assistance and human editorial review · Synthetic Insights R&D · SI-RD-002 · May 2026.

SYNTHETIC INSIGHTS

Intelligence, Accessible.

Synthetic Insights' mission is to build AI to serve the greater good — including job creation, productive social dialogue, and the delivery of social, economic, and spiritual value broadly to society.

R&D REPORTS
ISSUE NO. 002 · SI-RD-002
THE CLOCK, NOT THE MODEL
V1.0 · MAY 29, 2026

SYNTHETIC INSIGHTS LLC
SYNTHETIC-INSIGHTS.AI