APRIL

pellera

# THREAT INTEL REPORT 2026

Prepared by: Pellera Threat Intel Team
pellera.com | 800.747.8585

# **Driving Momentum.** Accelerating Change. Empowering IT Transformation.

**Pellera Technologies was born out of the combined expertise of Converge Technology Solutions and Mainline Information Systems**, two industry leaders with over 35+ years of experience and a shared vision for innovation. Together, we empower businesses to achieve greater efficiency, adaptability, and growth for today and tomorrow.

Our commitment is to reshape what's possible with IT, offering advanced solutions in digital infrastructure, cloud, cybersecurity, and AI. We don't just deliver technology—we partner with you to build tailored strategies designed to simplify complexities, unlock opportunities, and drive transformational outcomes.

At Pellera, momentum builds here through collaborative, people-first technology designed to fuel progress and deliver measurable impact.

pellera

# Observations for April 2026

April's common thread was a**dversaries going after trust layers** rather than breaking through them. Session cookies, maintainer credentials, and the integrity of a vendor's AI model are all things an organization inherits from somewhere else. Each of the three stories this month is about attackers getting between you and one of those inherited things.

State-linked groups ran the most visible operations. Iran-linked **Gray Sandstorm** hit more than 300 Israeli Microsoft 365 tenants across three password-spray-plus-AitM waves in March, and APT28 (Forest Blizzard) took over 18,000 home and small-office routers to rewrite DNS and capture OAuth tokens from anyone behind them. On March 31, UNC1069, a suspected North Korean actor, hijacked the publishing token of the axios npm maintainer and **pushed a cross-platform remote access trojan into a library** that shows up in roughly 80 percent of JavaScript codebases. Huntress saw the first infection 89 seconds after publish. Google Threat Intelligence and Unit 42 tied the payload to BlueNoroff tradecraft. None of these operators used a zero-day. They went after the token, the cookie, and the router.

Criminal groups rode the same playbook on a different clock. A loose crew called TeamPCP used **credentials stolen from a February compromise** of the Trivy scanner to push malicious Telnyx and LiteLLM packages to PyPI, hiding the stealer inside a WAV file where most scanners will not look. Commodity AitM kits like Tycoon 2FA, EvilProxy, and Mamba 2FA rent for a few hundred dollars a month and drove a 146 percent rise in MFA-bypass phishing, with Microsoft tracking close to **40,000 incidents a day**. The consistent follow-on in both the state and criminal campaigns is business email compromise: inbox takeover, hidden forwarding rules, and invoice redirection inside the first 24 hours. Mandiant CTO Charles Carmakal told The Register the axios blast radius will keep expanding for months as stolen npm tokens, cloud credentials, and CI secrets move through downstream environments.

# Executive Overview

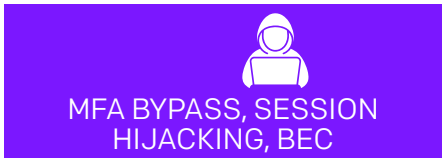
## AITM ATTACKS ON THE RISE

**Audience**

- CISO
- Security Operations Tesms
- IT Operations Managers & Teams
- IT Security Managers
- Incident Response Teams
- Identity & Access Management Leads


**RISK**


**GEOGRAPHIC SCOPE**




**BUSINESS IMPACT**

Microsoft tracked a 146 percent year-over-year rise in adversary-in-the-middle phishing in early 2026 and puts daily AitM events close to 40,000. The attacks steal the session cookie that comes back after a user completes MFA, which means the usual advice to turn on MFA does almost nothing against them. Obsidian Security reported that 84 percent of the accounts it saw compromised in its telemetry had MFA enabled.

In January, Microsoft flagged a multi-stage AitM campaign that abused SharePoint attachments to lure victims into a reverse proxy page. The same operator kept going through the quarter, with the campaign spilling into full business email compromise. Attackers set inbox rules to hide their activity and moved into wire fraud within a day of initial access. In March, an Iran-linked group called Gray Sandstorm ran three password-spray-plus-AitM waves against more than 300 Israeli Microsoft 365 tenants and 25 in the UAE, hitting government, municipalities, technology, transportation, and energy.

**READ MORE > AITM ATTACKS ON THE RISE**

## DEVELOPERS ARE PRIME TARGETS

**Audience**

- CISO
- Application Security Leads
- Security Operations Teams
- Incident Response Teams
- Software Engineering Leadership


**RISK**


**GEOGRAPHIC SCOPE**


**BUSINESS IMPACT**

On March 31, attackers compromised the axios npm package, one of the most downloaded JavaScript libraries in the world with over 100 million weekly downloads. The attacker published two malicious versions, tagged as the latest and legacy releases. Both included a dependency called plain-crypto-js that the axios source never imports but that the postinstall script on every install executed. The dropper pulled down a cross-platform remote access trojan that ran on macOS, Windows, and Linux. Huntress observed its first infection 89 seconds after the malicious version was published. The versions were live for about three hours before npm pulled them.

The same week, TeamPCP, a loosely organized English-speaking group, used credentials stolen from a February compromise of the Trivy open-source scanner to push malicious versions of Telnyx and LiteLLM to PyPI. Their payload hid a credential stealer inside WAV audio files in the postinstall script, which works because most scanners do not look inside media files. Trivy reaches more than 100,000 users and is baked into thousands of CI/CD pipelines, giving TeamPCP a direct path into the build systems of downstream targets.

Unit 42, Huntress, and Google Threat Intelligence all tied the axios payload to UNC1069, a suspected North Korean actor with overlap to BlueNoroff. Mandiant CTO Charles Carmakal told The Register that the blast radius of the axios compromise will keep expanding for months as stolen secrets are put to use in downstream environments.

**READ MORE > DEVELOPERS ARE PRIME TARGETS**

**Audience**

- **Executive Leadership**
- **Security Operations Teams**
- **Risk Management Professionals**
- **Application Security Leads**
- **Incident Response Teams**

## MYTHOS AND AI-ASSISTED OFFENSIVE CAPABILITY


**RISK**


**GEOGRAPHIC SCOPE**


**BUSINESS IMPACT**

On March 12, OFAC sanctioned six individuals and two entities operating out of Vietnam, Laos, and Spain for facilitating North Korean IT worker fraud. Treasury identified a key facilitator who converted roughly $2.5 million into cryptocurrency for DPRK workers between 2023 and 2025. The sanctions designated 21 cryptocurrency addresses across Ethereum and Tron networks.

On April 16, Anthropic shipped Claude Opus 4.7 and, on the same day, disclosed the existence of a preview model called Claude Mythos. The two releases were not paired accidentally. Anthropic's own description is direct: Mythos has reached a level of coding capability at which it can surpass all but the most skilled humans at finding and exploiting software vulnerabilities. Anthropic's red team wrote that Mythos has already identified thousands of high-severity flaws, including a 27-year-old OpenBSD bug and a 16-year-old FFmpeg vulnerability that existing automated tools had hit five million times without catching.

Axios and Reuters reported on April 19 and 20 that the NSA is using Mythos despite an internal blacklist, and that the Pentagon has also engaged with the program. Cybernews, on April 20, described banks raising concerns that Mythos could exploit weaknesses in financial infrastructure at a speed and volume existing controls are not built for. The Axios and Reuters stories are paywalled or behind Cloudflare challenges, but the headlines alone caught the attention of regulators and the financial sector.

To get ahead of the risk, Anthropic announced Project Glasswing the same day. Glasswing gives selected critical-infrastructure organizations Mythos access for defensive work, with AWS, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorganChase, the Linux Foundation, Microsoft, NVIDIA, and Palo Alto Networks listed as partners. Anthropic is committing $100 million in Mythos credits to the program along with $4 million in direct donations to open-source security work, including $2.5 million to Alpha-Omega and OpenSSF and $1.5 million to the Apache Software Foundation.

**READ MORE > MYTHOS AND AI-ASSISTED OFFENSIVE CAPABILITIES**

# Tactical Guidance

## AITM ATTACKS ON THE RISE

### Overview & Impact

AitM is not a new technique, but the commodity version of it has finally gotten past whatever immune system most organizations built around MFA. The attack is straightforward. The victim clicks a link, lands on a page that looks exactly like the real M365 sign-in, and enters their credentials. Behind the scenes, the attacker's reverse proxy is passing everything to Microsoft in real time. When Microsoft sends back the MFA prompt, the proxy passes that through too. The victim approves the push on their phone, Microsoft issues a session token, and the proxy keeps it.

The token is what matters. Session tokens are issued after MFA is satisfied and they stay valid for hours to days. Once the attacker has the token, the MFA ceremony is irrelevant. The attacker replays the token against Exchange Online, SharePoint, OneDrive, and anything federated to that identity. The user still has their phone and their password, and nothing looks wrong to them.

The follow-on is usually business email compromise. The attacker reads the inbox, picks a high-value thread (often an invoice), sets forwarding or hide rules so the real user does not see replies, and sends a message redirecting payment. In the March campaigns, attackers were into their first wire fraud attempt within a day of the initial compromise.

#### The attack chain

- Phishing lure, often a SharePoint link or a shared document notification, sometimes AI-written to match the recipient's role
- Victim clicks and lands on an attacker-controlled reverse proxy
- Proxy forwards credentials to Microsoft; Microsoft validates and issues an MFA challenge
- Proxy forwards the MFA challenge to the victim; victim approves on their phone
- Microsoft issues a session token back to the proxy
- Attacker captures the token before it reaches the victim's browser
- Attacker replays the token against M365 workloads, no MFA prompt required
- 146 percent year-over-year increase in AitM incidents per Microsoft's Digital Defense Report.
- 40,000 daily AitM events detected by Microsoft
- 84 percent of compromised accounts in Obsidian Security data had MFA enabled
- 300 plus Israeli M365 tenants targeted by Gray Sandstorm in three March waves; 25 additional tenants in the UAE
- Energy-sector BEC campaign documented by Microsoft Defender using SharePoint as the lure vector
- W3LL kit accounted for 17,000 M365 victims and $20 million in attempted fraud before takedown
- Forest Blizzard compromised 18,000 networks through end-of-life home and small-office routers to intercept OAuth traffic
- Storm-2755, tracked by Microsoft as Payroll Pirate, ran AitM against Canadian employee accounts with a likely payroll diversion objective

### Observations

- AitM kits are commodity. Tycoon 2FA, EvilProxy, and Mamba 2FA rent for a few hundred dollars a month and work against Microsoft, Okta, and Google out of the box.
- Lures keep getting better. AI-generated role-specific content is making phishing emails harder to spot, with finance staff getting fake invoices and procurement staff getting RFPs on the same day from the same operator.

- Device code phishing is a newer variant. Attackers use the OAuth device authorization flow and automation platforms like Railway.com to spin up thousands of ephemeral polling nodes.
- DNS-based AitM works too. Forest Blizzard did not even run a proxy page. They took over home routers and rewrote DNS so legitimate OAuth traffic passed through attacker infrastructure.
- Session tokens stay valid for a long time. Without Continuous Access Evaluation or token binding, the attacker has hours or days before a token expires naturally.
- BEC is the default follow-on. Mailbox takeover, inbox rule persistence, and invoice redirection are now the usual objectives after an AitM compromise.

## Guidance

### *Strategic Intelligence*

- **Trend**
  - Push and TOTP MFA are now proxyable in real time. Gray Sandstorm proved the point by pulling session cookies from 300 Israeli tenants in three March waves, and Obsidian Security is still seeing 84 percent of the accounts it watches get compromised while MFA is enabled. The defensive question is whether the MFA is cryptographically bound to the session, not whether MFA is turned on.
  - Phishing kit economics favor the attacker. An eleven-kit market at a few hundred dollars a seat means one operator can run campaigns against hundreds of tenants with no development effort.
  - The cleanup cost is high. Because session tokens are valid for days, a single AitM compromise can translate into weeks of forensic work to confirm what the attacker touched before the tokens rotated.

### *Operational Intelligence*

- **Threat Vectors**
  - Reverse-proxy phishing kits including Tycoon 2FA, EvilProxy, Mamba 2FA, and variants
  - Lures sent through legitimate services like SharePoint, OneDrive, and Dropbox to pass reputation filters
  - OAuth device code flow phishing using automation platforms for scalable polling infrastructure
  - Router-level DNS redirection of OAuth endpoints on compromised consumer gateways
  - Password spraying combined with AitM as the follow-up step after an initial credential match
  - AI-generated role-targeted emails that match the recipient's function
- **Monitoring & Detection Gaps**
  - Session tokens issued after MFA are not tied to device, IP, or TLS channel in most tenants
  - Limited visibility into impossible travel or anomalous token reuse across M365 workloads
  - Alerts for new inbox forwarding rules are off by default in many tenants
  - No out-of-the-box detection for device code flow authentication from unusual source IPs
  - Reliance on push or SMS MFA without a FIDO2 fallback requirement for high-value accounts

- **Response Actions**
  - Revoke session tokens and reset credentials on any account suspected of AitM compromise
  - Force sign-out on all sessions and require re-authentication with FIDO2 where available
  - Audit inbox rules for hide, forward, or delete patterns targeting invoice or wire-related terms
  - Review recent OAuth consent grants and registered devices for the affected user
  - Pull a 90-day timeline of SharePoint and OneDrive access for the affected identity

### *Tactical Intelligence*

- **Mitigation Strategies**
  - Enforce phishing-resistant MFA using FIDO2 security keys or passkeys for all privileged and high-exposure accounts
  - Turn on Conditional Access policies that block legacy authentication and require compliant devices for sensitive apps
  - Enable Continuous Access Evaluation across M365 workloads so token revocation propagates in minutes instead of hours
  - Require token binding where supported by the identity provider and downstream SaaS
- **Preventive Measures**
  - Retire end-of-life consumer routers on remote worker networks to remove the Forest Blizzard foothold pattern
  - Block automation platforms like Railway.com at the egress gateway if they are not used for legitimate business
  - Deploy browser-side phishing detection tuned to reverse-proxy indicators like mismatched origin and anomalous certificate chains
  - Train finance, procurement, and executive assistant staff on SharePoint-styled phishing specifically, not just generic email phishing
  - Review vendor contracts to confirm SaaS providers support token binding and Continuous Access Evaluation

## Threat Hunting Hypotheses

### *Session Token Replay After Possible AitM Capture*

**Hypothesis:** An attacker captured a session token through an AitM phishing campaign and is using it to access M365 from infrastructure other than the user's normal devices.

**Investigation Steps**

- Query sign-in logs for successful authentications followed within seconds by activity from a different IP or user agent.
- Look for M365 access from datacenter IP ranges or ASNs not associated with the user's normal ISP.
- Cross-reference recent SharePoint or OneDrive bulk downloads against the timeline of any sign-ins flagged as risky.
- Review Exchange audit logs for new inbox rules created shortly after a successful authentication.
- Check whether device code flow authentication was used and whether the polling IP matches the user's normal environment.
- If validated, revoke all sessions for the identity, reset credentials, rebuild Conditional Access exceptions, and review invoice or wire activity for the prior 30 days.

Tactical Guidance

### *SharePoint-Themed Lure Landing Pages*

**Hypothesis:** Users have received and clicked on SharePoint-themed AitM lures that match the campaign Microsoft documented in January.

**Investigation Steps**

- Search email logs for inbound messages with subject lines matching common SharePoint share notifications from external senders.
- Pivot on click telemetry from web proxies or SWG for any user who followed the link within the past 60 days.
- Check for DNS lookups to domains registered within 30 days that closely resemble SharePoint branding.
- Look for sessions where a user signed into M365 from a browser that immediately followed a click on one of those domains.
- If confirmed, revoke sessions, reset credentials, and pull an audit of activity on any workload the user accessed after the sign-in.

**Sources**


- Microsoft Security Blog: Resurgence of a Multi-Stage AiTM Phishing and BEC Campaign Abusing SharePoint (January 22, 2026)
- The Hacker News: Microsoft Flags Multi-Stage AitM Phishing and BEC Attacks Targeting Energy Firms
- Redsift: Email Security Roundup, AitM and AI Phishing Rise, April 2026
- ThreatLocker: AitM Phishing Attacks Against Microsoft 365, MFA Bypasses, Session Hijacking, and BEC (February 11, 2026)
- Small Business Cybersecurity Guy: AitM Attacks Bypass MFA 2026 (January 5, 2026)


## DEVELOPERS ARE PRIME TARGETS

### Overview & Impact

The targeting pattern shifted this quarter. Rather than attack a vendor's production software, attackers are going after the maintainers and tools that developers rely on. Axios is imported, directly or transitively, by a huge slice of the npm ecosystem. Trivy runs inside CI/CD pipelines that handle other pipelines' secrets. A successful compromise of either one reaches tens of thousands of organizations in a few hours without any of those organizations doing anything wrong.

**How the axios attack went down:** a long-lived NPM_TOKEN belonging to maintainer Jason Saayman ended up in attacker hands, and that token was used to publish v1.14.1 and v0.30.4. MFA on the account did not help. npm gives tokens precedence over the GitHub Actions OIDC trusted publisher flow when both are configured, and that is the hole the attacker walked through. Eighteen hours before the axios release, a separate attacker-controlled account pre-staged a clean plain-crypto-js at v4.2.0 to build publish history and slip past scanners that flag brand-new packages. Six minutes before axios went live, the same attacker pushed a malicious v4.2.1 and pointed axios at it.

**Every install fired a postinstall hook:** an obfuscated setup.js (reversed Base64 with XOR key OrDeR_7077), a quick os.platform check, and a fetch for the right binary. Mac hosts got a Mach-O written to /Library/Caches/com.apple.act.mond. Windows hosts got legitimate powershell.exe copied to %PROGRAMDATA%\wt.exe, which is good EDR evasion on binary name alone, plus persistence through a Run key called Microsoft Update. Linux hosts got a Python RAT at /tmp/ld.py with no persistence, because CI runners die before persistence would matter. Once up, the RAT beaconed to sfrclak[.]com on port 8000 every 60 seconds with an IE8-on-Windows-XP User-Agent. That string is one of the cleaner hunting signals anyone has handed defenders in a while.

TeamPCP's Telnyx and LiteLLM packages took a different path and landed in the same place. A stealer was hidden inside a WAV file shipped with each release, and a helper called from the installer pulled bytes out of the audio stream and executed them. Most antivirus tools do not parse WAV structure, so the payload sailed through. What TeamPCP was after: CI/CD secrets, AWS, Google Cloud, and Azure credentials, SSH keys, Kubernetes configs, and cryptocurrency wallets.

### The axios kill chain, step by step

- Long-lived NPM_TOKEN theft from maintainer account; exact vector undetermined but the token bypassed existing MFA and the OIDC trusted publisher
- Pre-staged clean release of plain-crypto-js@4.2.0 from a separate attacker account about 18 hours before detonation
- Malicious plain-crypto-js@4.2.1 published 6 minutes before the axios release
- axios@1.14.1 and axios@0.30.4 published with plain-crypto-js as a dependency that was never imported by source
- Obfuscated postinstall script (reversed Base64 plus XOR key OrDeR_7077) executed on install
- OS-specific RAT fetched and dropped, with persistence on macOS and Windows, ephemeral on Linux
- Beacon to sfrclak[.]com:8000 every 60 seconds with four commands: kill, runscript, peinject, rundir
- axios: over 100 million weekly downloads, malicious versions live for approximately three hours.
- First observed infection 89 seconds after publish.
- Huntress saw 135 endpoints contact C2 within the exposure window; transitive blast radius likely 10,000 plus organizations.
- Unit 42 identified victims across 11 sectors in the US, Europe, Middle East, South Asia, and Australia.
- Trivy: over 100,000 users, compromised starting in February and used as the foothold for TeamPCP's PyPI attacks.
- KICS (Checkmarx) also compromised via reused CI/CD secrets from the Trivy incident.
- Google Threat Intelligence attributes the axios compromise to UNC1069, a suspected North Korean actor with overlap to BlueNoroff.
- Mandiant CTO Charles Carmakal: the blast radius of the axios incident will keep expanding for months as stolen secrets are used in downstream environments.

## Observations

- Pre-staging a clean version from a throwaway account is an established evasion technique for new-package heuristics and will not go away.
- Long-lived npm tokens bypass OIDC trusted publisher even when the latter is configured. npm gives tokens precedence, which many maintainers do not realize.
- Postinstall scripts remain the most common weaponization point in npm. yarn.lock and package-lock.json do not protect against them on fresh installs.
- Steganography in WAV files defeats naive dependency scanning. TeamPCP's success with this approach will get copied.

- The IE8 User-Agent in the axios RAT is a clear hunting signal and a reminder that attackers keep hardcoded defaults even when they should know better.
- The connection between Trivy, KICS, Telnyx, and LiteLLM shows attackers chaining supply-chain compromises. One maintainer compromise feeds the next attack through stolen CI/CD secrets.

## Guidance

### *Strategic Intelligence*

- **Trend**
  - An npm maintainer's personal laptop is the realistic soft target now for anyone who wants to reach 10,000 enterprise environments in an afternoon. The targets are not Apple or Microsoft but the handful of volunteers whose packages those companies transitively depend on, most of them working from personal GitHub accounts with no enterprise security team behind them.
  - The secondary market for stolen secrets is going to drive incidents for months. Carmakal's forecast is that the axios blast radius will expand as stolen npm tokens, cloud credentials, and CI secrets are rotated through downstream targets. Expect breach disclosures well into summer.
  - Attribution variety matters. DPRK-linked groups are chasing revenue and access. Groups like TeamPCP are smash-and-grab operators running on speed over stealth. Both are effective, and the defender has to cover both patterns.

### *Operational Intelligence*

- **Threat Vectors**
  - Maintainer account takeover through long-lived publishing tokens that bypass MFA and OIDC trusted publisher
  - Malicious postinstall scripts in npm and PyPI packages
  - Steganography in WAV and other media files bundled into install payloads
  - Typosquatting and pre-staged companion packages to build publish history
  - Compromised developer tools (Trivy, KICS) used as a stepping stone into downstream CI/CD
  - Transitive dependency pulls during normal npm install or npm ci runs
- **Monitoring & Detection Gaps**
  - No enforcement of short-lived publishing tokens on most maintainer accounts
  - Package quarantines (min release age) not configured by default on most teams
  - WAV and other media file bundles not scanned for embedded payloads
  - CI/CD runners have outbound egress to arbitrary internet destinations by default
  - Developer workstations rarely have the EDR coverage that production endpoints do
  - npm-debug.log and yarn-error.log not routinely ingested to a SIEM

- **Response Actions**
  - Pin axios to 1.14.0 or 0.30.3 and add overrides in package.json or yarn.lock to force safe versions transitively
  - Remove any plain-crypto-js directories from node_modules, including nested ones under other packages
  - Rotate every credential touched by any developer machine or CI runner that might have installed the bad versions: npm tokens, AWS, GCP, Azure keys, SSH keys, CI secrets, OAuth and API keys
  - Block C2 infrastructure at the egress gateway, including sfrclak[.]com, 142.11.206.73:8000, calltan[.]com, and callnrwise[.]com
  - Pull Trivy and KICS back to known-good versions; audit CI pipelines that ran between February 15 and March 23 for signs of secret exfiltration
  - Search PyPI packages installed in the past 60 days for bundled WAV files and postinstall shell execution

### *Tactical Intelligence*

- **Mitigation Strategies**
  - Require hardware security keys for all publish-capable maintainer accounts, and rotate any long-lived tokens to short-lived OIDC trusted publisher tokens
  - Configure package managers with a min release age of 48 to 72 hours so new versions sit in quarantine before any pipeline picks them up
  - Use --ignore-scripts in CI/CD pipelines to disable postinstall hooks by default; enable them only for packages that explicitly need them
  - Always commit lockfiles and use npm ci (not npm install) in CI/CD so dependencies resolve deterministically
  - Run EDR on developer workstations with visibility into Node and Python child process execution
- **Preventive Measures**
  - Build an SBOM for every production system and tie it to an automated scanner that flags when a dependency version appears on a known-bad list
  - Add a CI check that fails the build if any package has a min-release-age below a configurable threshold
  - Segment CI runners from internet-adjacent networks where possible; at minimum enforce egress allow-listing
  - Audit every repository's GitHub Actions workflow for long-lived NPM_TOKEN or similar credentials and replace them with OIDC trusted publishing
  - Include supply-chain compromise scenarios in tabletop exercises, with specific focus on the secondary fallout when stolen secrets are used against you

## Threat Hunting Hypotheses

### *Residual axios Compromise on Developer Workstations and CI Runners*

**Hypothesis:** A developer or CI/CD pipeline installed a malicious version of axios between March 30 and March 31 and may have persistent malware or exfiltrated credentials.

**Investigation Steps**

- Search endpoint and CI/CD logs for installations of axios@1.14.1 or axios@0.30.4 between March 30 and March 31.
- Check for DNS queries or outbound connections to sfrclak[.]com, callnrwise[.]com, or 142.11.206.73 on port 8000.

- On macOS hosts, look for /Library/Caches/com.apple.act.mond. On Windows, check for %PROGRAMDATA%\wt.exe and %PROGRAMDATA%\system.bat. On Linux, check for /tmp/ld.py.
- Hunt for HTTP traffic with the User-Agent 'mozilla/4.0 (compatible; msie 8.0; windows nt 5.1; trident/4.0)' from any modern system.
- Review Run key persistence for Microsoft Update entries pointing to batch files.
- If validated, isolate the host, rotate every credential it had access to, rebuild the workstation or CI image from clean state, and walk through npm audit logs for the affected accounts.

## *WAV-Packaged Stealer in Recently Installed PyPI Packages*

**Hypothesis:** A developer or build system installed a malicious Telnyx, LiteLLM, or similar PyPI package that embeds a credential stealer inside a WAV file.

**Investigation Steps:**

- List PyPI packages installed on developer workstations and CI runners between March 15 and April 15.
- Filter for packages that include WAV or other media files in their distribution.
- Inspect postinstall scripts and setup.py for any calls to wave, audioop, or similar libraries that parse audio data.
- Monitor process execution trees for Python processes spawning shells, curl, or wget during pip install.
- Correlate with outbound network activity during the same install window.
- If confirmed, rotate all secrets on the affected host, remove the package, quarantine the image, and pull a full audit of CI/CD secret usage for the past 30 days.

## Sources


- **Huntress: Supply Chain Compromise of axios npm Package (March 31, 2026)**
- **Unit 42: Threat Brief, Widespread Impact of the Axios Supply Chain Attack (April 1, 2026)**
- **Microsoft Security Blog: Mitigating the Axios npm Supply Chain Compromise (April 1, 2026)**
- **The Register: Two Different Attackers Poisoned Popular Open Source Tools (April 11, 2026)**
- **The Hacker News: Axios Supply Chain Attack Pushes Cross-Platform RAT via Compromised npm Account**
- **The Hacker News: TeamPCP Pushes Malicious Telnyx Versions to PyPI, Hides Stealer in WAV Files**
- **PhishFort: Open Source Supply Chain Attack, Why Developers Are the New Target**
- **Veracode: Open Source Supply Chain Security Best Practices**
- **Microsoft: Open Source Software Supply Chain Threats**

## MYTHOS AND AI-ASSISTED OFFENSIVE CAPABILITY

### Overview & Impact

Mythos is not a product you can buy. Anthropic has deliberately held it out of general release, instead running it through a limited partner program and a Cyber Verification Program for security researchers. That limited rollout is the point. Anthropic's internal benchmarks show Mythos hitting 83.1 percent on CyberGym (vulnerability reproduction) against Opus 4.6's 66.6 percent, 93.9 percent on SWE-bench Verified against 80.8 percent, and 82.0 percent on Terminal-Bench 2.0 against 65.4 percent. Those numbers put Mythos in a different class for the kind of agentic coding work that turns a described vulnerability into a working exploit.

The capability that matters for defenders is autonomous multi-step exploitation. Anthropic's red team reports that Mythos can chain Linux kernel vulnerabilities to escalate from user to root without human steering, and that it handles zero-day discovery across every major operating system and browser. The implication is that the window between a bug landing in a codebase and being exploited collapses from weeks to hours once an adversary with similar capability is involved.

Claude Opus 4.7, the production release on the same day, is the safer alternative. Opus 4.7 ships with runtime safeguards that detect and block requests aimed at prohibited or high-risk cybersecurity use. Anthropic describes this as differential reduction of cyber capabilities during training. Security professionals who want the full capability for legitimate work can apply to the verification program. For most enterprise customers, Opus 4.7 is the model they will actually use, and it is priced the same as Opus 4.6.

Glasswing is the defensive half. Anthropic is giving a hand-picked set of partners, including three hyperscalers and a major US bank, Mythos access to find and patch flaws in their own systems before adversaries reach comparable capability. The theory is that the defenders need to be inside the same capability envelope as the attackers. The question is whether that envelope stays closed.

- Mythos scored 83.1 percent on CyberGym vs Opus 4.6 at 66.6 percent.
- 93.9 percent on SWE-bench Verified vs 80.8 percent.
- 82.0 percent on Terminal-Bench 2.0 vs 65.4 percent.
- Anthropic reports Mythos has found thousands of high-severity vulnerabilities autonomously, including long-standing flaws in OpenBSD and FFmpeg.
- NSA reportedly using Mythos despite an internal blacklist (Axios and Reuters, April 19 and 20).
- Banks raising alarms about financial-system exploitation risk (Cybernews, April 20).
- $100 million in Mythos credits committed to Glasswing partners.
- $4 million in direct donations to open-source security (Alpha-Omega and OpenSSF; Apache Software Foundation).
- 40 plus critical-infrastructure organizations given access under Glasswing.
- Anthropic Cyber Verification Program open for security researchers.

### Observations

- Anthropic is the first major lab to publicly describe a model it will not ship as a consumer product because of cyber offense capability. That is a meaningful precedent.
- The benchmark gap is real. A 16.5 point jump on CyberGym and a 13 point jump on SWE-bench are not marginal improvements; they change what is possible for a single operator.

- The NSA and Pentagon angle changes the conversation. If US government agencies are using Mythos through quiet channels, other governments will assume they should too. Export control questions follow.
- The financial-sector concern is specific enough to be worth taking seriously. Banks are not usually early adopters of AI risk commentary, and their worry here is about automated exploitation of payment systems and fraud controls.
- Glasswing is a governance experiment. Hyperscalers, a bank, and the Linux Foundation sharing access to the same offensive tool is a new kind of coalition and the operating model is not yet clear.
- Opus 4.7's differential capability reduction is an architectural choice worth studying. Other labs will be asked why they have not done the same.

## Guidance

### *Strategic Intelligence*

- **Trend**
  - Assume an attacker at a big-four state service has something close to Mythos today and that a capable criminal group will have one within 18 months, whether through a licensed program, a leak, or an unreleased equivalent from another lab. Any 2027 plan built on the current pace of exploit development is already behind.
  - A 30-day patch SLA on an internet-facing host is a 30-day window for an attacker-side model to find the bug, write the exploit, and hand it to anyone scanning the internet. Compress SLAs toward days for external-facing services and toward hours for anything touching payment, identity, or health data.
  - The governance model around frontier AI is in flux. Anthropic's choice to hold Mythos from general release is credible for now. It will stay credible only as long as no comparable model reaches the open market.

### *Operational Intelligence*

- **Threat Vectors**
  - Autonomous zero-day discovery against enterprise software, operating systems, and browsers
  - AI-assisted chaining of low-severity bugs into working privilege-escalation exploits
  - Automated social engineering content generation tuned per victim
  - Scaled exploitation of known unpatched vulnerabilities across the internet-facing attack surface
  - Rapid triage of stolen source code to identify new vulnerabilities in downstream deployments
- **Monitoring & Detection Gaps**
  - Detection signatures lag when exploits are generated per target rather than reused
  - Threat intelligence feeds rely on known IOCs that AI-generated tooling does not produce
  - Vulnerability scanners miss bugs that Mythos-class models surface through semantic analysis
  - Most SOC alerting assumes attacker dwell time of days, not minutes
  - Enterprises lack internal policy for approved frontier-model use in security work

- **Response Actions**
  - Stand up a program to test defensive use of Opus 4.7 or equivalent verified models for vulnerability discovery in your own codebases
  - Review patch cycles for any system where a 30-day window is tolerable and shorten it
  - Apply to Glasswing if the organization is in critical infrastructure or OSS maintenance
  - Brief executive leadership and the board on the AI-assisted attack scenario before the Reuters and Axios coverage drives the conversation for you
  - Engage your cyber insurance carrier on coverage expectations if exploitation timelines compress

### *Tactical Intelligence*

- **Mitigation Strategies**
  - Accelerate vulnerability management SLAs for internet-facing systems and any system with direct access to payment, identity, or health data
  - Add AI-assisted code review and fuzzing to the CI/CD pipeline for your highest-risk services
  - Enforce defense-in-depth assumptions: if an exploit lands, make lateral movement expensive through segmentation and identity controls
  - Evaluate vendor AI policies for frontier-model access and request attestations about differential reduction controls
- **Preventive Measures**
  - Treat unpatched internet-facing software as a ticking clock rather than a backlog item
  - Begin SBOM inventory and provenance work now, so when AI-driven analysis surfaces a new class of vulnerability you can tell which systems are affected
  - Participate in responsible disclosure programs that use Mythos-class tooling defensively, including through Glasswing partner channels
  - Train security teams on AI-augmented detection so the defender side of the capability gap keeps up
  - Review insurance and regulatory reporting requirements in light of likely acceleration of exploitation timelines

## Threat Hunting Hypotheses

### *Unusual Exploit Development Patterns in Outbound Activity*

**Hypothesis:** Attackers using AI-assisted exploitation tooling are probing our environment with techniques that do not match known-signature databases.

**Investigation Steps**

- Pull WAF and IDS logs for web requests that trigger error responses but do not match any known exploit signature.
- Look for rapid sequences of probing against the same endpoint from a single source, followed by a different-shape request that succeeded.
- Correlate against internal vulnerability scan findings: if an internal scanner flagged a bug, watch the logs for attempted exploitation of the same class.
- Check for source IPs that have been active for less than 24 hours on known bulletproof hosting or residential proxy networks.
- If validated, block the source, patch the affected component immediately, and capture a full PCAP of the exploitation traffic for analysis.

## *Opus 4.7 and Mythos-Class Tooling Policy Compliance*

**Hypothesis:** An internal team is using a frontier model for cybersecurity work without going through the verification program, introducing both risk and potential policy violations.

### Investigation Steps

- Review API usage logs for any Anthropic, OpenAI, or similar service calls associated with security tooling or vulnerability discovery work.
- Interview red team, security research, and engineering teams about current model usage.
- Cross-reference against documented approvals and verification-program enrollment.
- For any use that is not approved, document the use case, re-enroll the user in the appropriate program, and confirm that outputs are handled under the same classification as the underlying code.
- Publish a short internal policy that mirrors Anthropic's verification expectations so this does not repeat.

### Sources


- **Anthropic: Project Glasswing, Securing Critical Software for the AI Era (April 16, 2026)**
- **Anthropic Red: Claude Mythos Preview (April 16, 2026)**
- **CNBC: Anthropic Rolls Out Claude Opus 4.7, an AI Model That Is Less Risky Than Mythos (April 16, 2026)**
- **Axios: NSA Using Anthropic's Mythos Despite Blacklist (April 19, 2026)**
- **Reuters: US Security Agency Is Using Anthropic's Mythos Despite Blacklist (April 19, 2026)**
- **Reuters: What Do We Know About Anthropic's Mythos Amid Rising Concerns (April 20, 2026)**
- **Cybernews: Banks Alarm, Anthropic Mythos AI Exploit Financial System Weaknesses (April 20, 2026)**


pellera

Contact the Pellera Threat Intel Group at getsecure@pellera.com

pellera.com