CSAI Foundation | Frontier Ready Cybersecurity

Frontier Ready Daily

Issue 80 · Member edition

2026-11-14



CSAI FOUNDATION INITIATIVE

Frontier Ready.

Machine-speed agentic cybersecurity

cloud security alliance®

CSAI

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

In this issue

Google Project Zero has dropped the 30-day post-patch "soak time" from its disclosure policy, saying LLMs let attackers exploit a vulnerability from its patch almost immediately, and will cut its 90-day disclosure deadline to 60 days in February 2027 and 30 in January 2028; the other items cover Check Point gateways rooted about a week after a fix and an ESXi host wiped behind one, Copilot and Defender XDR failing in the same Microsoft 365 outage, two attacks on AMD SEV-SNP confidential VMs, and ChatGPT agents completing card purchases. We examined the FBI's FaceStealer FLASH and Huntress's Teams device-code phishing case and left both out: FaceStealer has today's CSA research note and, by the FBI's account, "overwhelmingly" relies on stolen real video rather than AI, while the Huntress case is a single-victim identity attack with no measured interval.

Project Zero dropped the 30-day patch soak time because LLMs make patch-to-exploit time effectively zero, and will cut its disclosure deadline from 90 days to 30 by January 2028

Category: vuln_storm **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the "Full 90" policy effective 13 November 2026, the move to "Full 60" on 2 February 2027 and "Full 30" on 3 January 2028, the removal of the grace period and variant exceptions, the statement that "zero patch-to-exploitation time is a reasonable, though imperfect, model of attacker behavior", and the statement that Project Zero is "finding several vulnerabilities a week in targets where we once found a few vulnerabilities a year"; NO PROVIDER CLAIM on how many vendors currently meet a 60- or 30-day window; CHARACTERIZATION (CSA) for the reading that enterprise patch SLAs built around vendor embargoes no longer hold **Source:** Project Zero — Policy and Disclosure: 2026 Edition

What changed. On 13 November 2026 Google Project Zero replaced the "90 + 30" disclosure policy it had used since 2021. That policy gave vendors 90 days to fix a bug and another 30 days of "soak time" for users to install the patch before details went public. From now on, details are published 90 days after the report whether or not a fix exists. The deadline falls to 60 days on 2 February 2027 and to 30 days on 3 January 2028. Grace periods and variant handling are gone. The remaining exceptions are bugs already exploited in the wild, bugs others have already made public, and vendor requests for earlier release. Project Zero gives three reasons. LLMs make it more likely that attackers will find the same bug. They have cut the time needed to build an exploit from a patch so far that soak time "is no longer meaningful". And

they have raised Project Zero's own discovery rate to the point where negotiating with vendors bug by bug "is quickly becoming unmanageable". Whether a bug has been fixed will no longer change the disclosure date.

Why it reaches you. This affects every enterprise patch window for software Project Zero researches: browsers, mobile and desktop operating systems, hypervisors, and the libraries built into them. Two working assumptions stop holding. A vendor's patch release no longer means details stay private for another month, because Project Zero now treats exploit-from-patch as close to instant and has removed the soak period. And a 30-day critical-patch SLA, which used to sit inside the embargo, will by 2028 be as long as the whole window from report to public disclosure. The policy applies only to Project Zero's reports. Other research teams usually follow its lead, but that is our expectation, not something Project Zero has said.

What to do. Vulnerability management owner, with procurement: escalate. Measure your actual time from vendor patch release to deployment for browsers, endpoint operating systems and hypervisors, and compare it with an assumption of exploitation on the day of release. Where the gap is more than days, name the compensating control you will use until the patch is in, such as isolation, feature disablement or virtual patching. For major software suppliers, add a contract question: what is your remediation time against a 60-day full-disclosure deadline from February 2027, and what out-of-band fix path do you have? Urgency: escalate.

CSA coverage. The Disclosure Bottleneck: When AI Discovery Outpaces Patch Consumption; see also When Discovery Outruns Remediation: Open-Source Disclosure Strain.

Unpatched Check Point gateways were rooted through PHP2Shell about a week after a fix was available, and at least one attacker moved on to wipe a VMware ESXi host behind the gateway

Category: machine_speed **Significance:** major **Confidence:** VERBATIM (PROVIDER) for exploitation starting "approximately one week after a security fix became available", the 30 October and 2 November waves, the chain from PHP2Shell (CVE-2026-109618 and CVE-2026-109619) to the CVE-2026-123039 DynamicCLI root escalation, the TSGA and TSProxy implants, the SSH pivot to ESXi as `vpxuser`, and the `destroy_v6` wiper; VERBATIM (PROVIDER) for Check Point's statement that the evidence "does not allow us to assess with high confidence" the actor's identity; NO PROVIDER CLAIM on how many gateways were compromised or wiped; LINK ONLY — VERIFY AT SOURCE for CISA's KEV listing of CVE-2026-123039, as reported by Security Affairs on 14 November; CHARACTERIZATION (CSA) for the edge-appliance patch-gap reading **Vendor interest:** Check Point is the vendor of the affected gateway

and also sells threat-prevention and incident-response services, so its account of scope and of how fast the fix arrived involves its own product. **Source:** Check Point Research – PHP2Shell Exploitation on Check Point Gateway

What changed. On 12 November 2026 Check Point Research reported that, from 30 October, an attacker scanned for and exploited Check Point Security Gateways with Mobile Access portals that had not been patched. The attacker used PHP2Shell for initial access and a local privilege escalation in the DynamicCLI interface (CVE-2026-123039) to get root. The attacker then installed webshells and a backdoor built for Check Point appliances, which can tamper with firewall logs and persist through backups and snapshots, scanned the internal network, moved over SSH to a VMware ESXi host and ran a wiper. The 30 October wave hit Israeli organisations and was destructive in some cases. A second wave on 2 November reached several countries and sectors and showed no destructive activity. The implants are modified versions of PhiliKit, a framework previously used against Ivanti devices, but Check Point says it cannot rule out an access broker or a false flag. CISA has since added CVE-2026-123039 to its Known Exploited Vulnerabilities catalogue. The initial-access flaw is the same PHP2Shell bug this feed has tracked since 22 October.

Why it reaches you. The components are the remote-access gateway and the virtualisation management plane behind it. Earlier PHP2Shell items covered mass exploit attempts against general-purpose PHP. This report shows the same flaw inside a security appliance becoming root on the appliance, persistence that survives restores, and a path to the hypervisor. Anyone who patched late, or patched the appliance without checking for implants, may have patched over an existing compromise. Check Point's report came about two weeks after the first wave, so post-patch integrity checks cannot wait for vendor reporting.

What to do. Network security owner, with the virtualisation owner: escalate. For every Check Point gateway with Mobile Access enabled, confirm when the fix was applied, then look for the indicators Check Point lists. These include `upload_progress_pwn` session files under the Mobile Access temp directory, `zz*.tpl` files in the portal template folder, and anything on TCP 11072 or at `/tmp/.p_sock`. Also check for `vpxuser` SSH logins to ESXi that did not come from vCenter. Do not restore a gateway from a backup or snapshot without checking it for the same indicators. Urgency: escalate.

CSA coverage. PHP2Shell: Exploitation Before Disclosure for an AI-Found Flaw; see also No Window to Patch: SonicWall MFA Bypass and Zyxel Switch Zero-Day.

Microsoft 365 Copilot and Defender XDR were both listed in the same outage advisory when a server failure spread across Exchange Online on 12 November

Category: security_operating_model **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for advisory MO1550821 listing Exchange Online, Teams, SharePoint Online, Microsoft 365 Copilot and Microsoft Defender XDR, the 1:26 PM ET start time, and Microsoft's statement that "the elevated load isn't driven by an increase in traffic, but by a number of servers being unexpectedly unavailable", as reported by BleepingComputer; NO PROVIDER CLAIM of a final post-incident review or of total duration by service; CHARACTERIZATION (CSA) for the shared-failure-domain reading **Source:** BleepingComputer – Massive Microsoft 365 outage disrupts Outlook, Teams, Copilot worldwide

What changed. On 12 November 2026, from 1:26 PM ET, Microsoft 365 had a global incident that began as Exchange Online mailbox connectivity problems in the Americas. Microsoft's advisory MO1550821 said customers might be unable to reach Exchange Online, Teams, SharePoint Online, Microsoft 365 Copilot and Microsoft Defender XDR. Microsoft first blamed elevated load, then said that a number of servers had become unavailable unexpectedly and that this put extra strain on the remaining infrastructure. It paused all deployments and redistributed traffic, and reported only limited recovery that evening. Reuters, as cited by Security Affairs, reported that OpenAI's and Anthropic's connectors to Microsoft 365 accounts were also affected.

Why it reaches you. The component is the consolidated productivity, AI-assistant and security tenant. Copilot, the XDR console and the mail platform they depend on degraded together, so a security team could lose its main investigation console at the same moment users lost mail and chat. AI agents and third-party assistants that reach Microsoft 365 through connectors lost their data plane at the same time. A continuity plan written per product treats these as separate risks, but in this incident they behaved as one failure domain. A shared root cause is not yet confirmed because Microsoft has not published a post-incident review.

What to do. Security operations owner, with business continuity: validate. Check whether your incident runbook still works when the XDR console, Teams and Outlook are all unavailable. That means an out-of-band bridge channel, raw telemetry retained somewhere you can query independently, and a plan to backfill alerts and hunt the gap once service returns. List the agents and automations that depend on Microsoft 365 connectors and decide what each should do while the connector is down. Urgency: validate.

CSA coverage. When the Assistant and the Alarm Go Dark Together.

Two attacks on AMD SEV-SNP let a host that controls the hypervisor forge confidential-VM attestation or leak guest secrets, and the platform firmware

fixes are not due until December

Category: agentic_surface **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for AMD bulletin AMD-SB-3059's description of CVE-2026-0462 as allowing "a malicious hypervisor to potentially redirect access to MMIO memory resulting in loss of guest memory integrity", the SEV firmware mitigations dated 12 August 2026, the platform-initialisation firmware target of December 2026 for EPYC 7003 to 9005, February 2027 for the embedded EPYC 7003, 8004 and 9004 parts listed, and the TCB values attestation should require; LINK ONLY — VERIFY AT SOURCE for the researchers' 1,000-of-1,000 success rate for Mior on a Zen 5 EPYC 9135 and Skyfall's recovery of a disk-encryption key; NO PROVIDER CLAIM of in-the-wild exploitation; CHARACTERIZATION (CSA) for the confidential-AI-inference implication **Source:** AMD — AMD SEV-SNP MMIO Routing Vulnerability (AMD-SB-3059)

What changed. On 10 November 2026, when the embargo lifted, AMD published advisories for two attacks on SEV-SNP by teams centred at ETH Zurich. One team included Google researchers. With control of the hypervisor, the BIOS/UEFI configuration and an attached PCIe device, Mior (CVE-2026-0462) gives read and write access to confidential-VM memory and lets the attacker forge launch measurements, so a malicious guest image can attest as a benign one. Skyfall (CVE-2026-40671, AMD-SB-3061) needs only hypervisor control and leaks guest register state through transient execution. Proof-of-concept code is public for both. The fixes are on the host side. SEV firmware mitigations are dated August 2026, but OEM BIOS builds from the platform-initialisation releases are not due until December 2026, or February 2027 for the embedded parts listed.

Why it reaches you. The components are the confidential VMs and attestation verifiers behind confidential AI inference, where cloud providers or inference platforms tell customers that prompts, weights or keys are protected even from the operator. SEV-SNP exists to protect against exactly the adversary these attacks assume, a malicious or compromised host. A tenant cannot patch from inside the guest. What a tenant can do is make its verifier reject attestation reports from platforms below the fixed TCB level. If the verifier accepts any validly signed report, it will accept a forged measurement from an unpatched host. This is a laboratory result with no reported exploitation, but the property it breaks is the one confidential-inference contracts are sold on.

What to do. Cloud security architecture owner, with the AI platform owner: validate. For every workload or AI vendor that relies on AMD SEV-SNP attestation, confirm the verifier enforces a minimum TCB, using the values in AMD's bulletins, and ask each confidential-inference provider when its hosts will run the fixed SEV firmware and BIOS. Until then, do not treat attestation alone as evidence that a host operator cannot see protected data. Urgency: validate.

CSA coverage. Mior and Skyfall: Malicious Hosts Break AMD SEV-SNP Attestation; see also Inside the Enclave, Still Exposed: The NEAR AI SGLang Breach.

ChatGPT agents can now complete purchases with network-issued card tokens after biometric approval, and OpenAI asks merchants to recognise signed agent traffic

Category: agentic_surface **Significance:** context **Confidence:** VERBATIM (PROVIDER) for the per-purchase review and biometric approval, the network-issued token in place of the card number, the request that merchants configure their CDN or firewall to recognise Web Bot Auth, and OpenAI's warning that "some websites may block agent purchases without meaning to", as reported by Help Net Security; NO PROVIDER CLAIM on spending limits, merchant restrictions or how authority is revoked; CHARACTERIZATION (CSA) for the merchant-side testing implication **Source:** Help Net Security – ChatGPT Wallet lets agents pay without seeing your card number

What changed. On 13 November 2026 OpenAI began rolling out ChatGPT Wallet, built with American Express, Mastercard and Visa, to US ChatGPT Pro subscribers aged 18 or over. The user adds a card once. For each purchase the agent presents the items, total and card, and the user approves with Face ID, a fingerprint or a similar check. The payment network issues a token with its own number, security code and expiry, so neither the agent nor the merchant sees the real card. OpenAI's Commerce Partner Playbook asks merchants to recognise Web Bot Auth, a method for verifying signed requests from bots, at the CDN or firewall, and to work with their payment and fraud providers to identify agent purchases. The coverage describes no spending limits or merchant restrictions.

Why it reaches you. The components are the e-commerce edge (bot management, WAF and fraud scoring) and, if staff use personal agents for business purchases, corporate card programmes. Bot defences that block unsigned automation will meet traffic from a large consumer agent that is legitimate but automated. A merchant that treats it as an attack loses sales, and a merchant that allow-lists it on user-agent string alone repeats the spoofing problem this feed covered on 13 November. A biometric approval on each purchase is a human-authorisation control in a consumer product. It is still a single consent step, so the agent's review screen is what the user is trusting.

What to do. E-commerce security owner, with the fraud team: monitor. Decide how your edge will handle agent purchases: verify signed agent requests cryptographically rather than by user-agent string, and set a fraud-scoring path for them before volume grows. For corporate cards, note that this release covers individual US consumer cards only. Urgency: monitor.

CSA coverage. New — no prior CSA coverage of agent payment wallets; see also Spoofed and Disguised AI Agents: The Collapse of User-Agent Trust and Singapore's Personal AI Agent Advisory and the Bring-Your-Own-Agent Gap.

Watchlist

- **OpenAI reward-hacking postmortem — downstream response** — No change. No frontier lab disclosed an eval-to-production escape, and there was no new regulatory or contractual development. *(opened 2026-08-27)*
- **VM/hypervisor containment hardening for cyber-capable agents** — No change on microVM adoption. A related hypervisor-trust finding appeared: the Mior and Skyfall attacks (item 4) break AMD SEV-SNP guest isolation against a malicious host. That is the reverse direction (host attacking guest) and does not concern QEMU/KVM agent-sandbox escape. *(opened 2026-08-27)*
- **Claude Code Auto Mode prompt-injection ASR discrepancy** — No change. There is still no Anthropic patch or public response. *(opened 2026-08-27)*
- **AI defensive-triage guardrail evasion** — No change. No new campaign triggering refusals to blind triage tooling was reported. *(opened 2026-08-31)*
- **AI account session hijacking at scale** — No change. No AI provider disclosed a session-hijacking campaign or shipped device-bound session tokens. *(opened 2026-08-31)*

Opened this issue

- **Disclosure deadlines shrink toward 30 days** (vuln_storm) — Watching for vendors' public responses to Project Zero's Full 60 (2 February 2027) and Full 30 (3 January 2028) schedule, whether other research teams or bug bounty programmes drop soak periods, and the first major vendor whose bug is disclosed unpatched under the new policy. *(opened 2026-11-14)*