CSAI Foundation | Frontier Ready Cybersecurity

Frontier Ready Daily

Issue 90 · Member edition

2026-11-24



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

The lead is VulnCheck's report that an AI exploit pipeline took one exploit-intelligence team from 22 working exploits a month to 266 a week, landing most within one to three days of CVE publication. It is followed by the theft of secrets from Azure API Management Classic instances at Equinor and other organisations, OpenAI and Cisco's open-source tool for agent-written kernel shields that cover the patch window, MONAI Label's first patched release and KEV listing two months after an attacker's agents exploited it, a benchmark finding that every search API tested returned known-malicious content to coding agents, and VulnCheck sensor data showing that nearly one in five exploited CVEs is now used to harvest credentials, including AI-provider keys. We examined Elastic's AlertZero SOC agents, Anthropic's Claude Ship approvals agent and GitGuardian's 9 November count of Claude Code settings-file secrets, and left them out: the first two are product announcements with no measured effect outside the vendor's own team, and the third is 15 days old and extends the 7 November item on the same mechanism (31,088 secrets, 5,343 valid, plus 99 valid secrets in public Docker Hub images).

An AI exploit pipeline took VulnCheck from 22 working exploits a month to 266 a week, most delivered within one to three days of CVE publication

Category: machine_speed **Significance:** major **Confidence:** SELF-REPORTED (PROVIDER METRIC) for VulnCheck's output figures (22 exploits in January, then weekly output of 58, 104, 152, 203 and 266), for new CVEs first seen on its Canary sensors rising from 13–20 a month to 82 in October and 139 in November, and for its KEV additions rising from about 84 a month to 336 in October and 343 in November with more than a week left; VERBATIM (PROVIDER) for "We're increasingly landing deliveries within one to three days of CVE publication" and for every AI-generated pull request still being checked and tested by a person before it ships; CHARACTERIZATION (CSA) for treating the defender-side pipeline as a lower bound on what a resourced attacker can now build, and for the caution that part of the sensor and KEV growth reflects wider detection coverage rather than more attacker activity **Vendor interest:** VulnCheck sells exploit, initial-access and exploited-vulnerability intelligence, so evidence that exploitation volume has jumped is commercially favourable to it. **Source:** VulnCheck – AI Has Changed VulnCheck

What changed. On 24 November 2026 VulnCheck reported that its Initial Access team, which began with two people writing two to four n-day exploits a week, now runs an AI pipeline that takes a CVE in and returns a pull request with a working exploit and detection artifacts. Output went from 22 exploits in

January to 266 in a single week this month, and VulnCheck says it produced more exploits in October than in the rest of the year combined. Delivery has moved to one to three days after CVE publication, often close to publication itself. Because each new exploit comes with detection content, its sensors saw 82 CVEs exploited for the first time in October and 139 in November, against 13 to 20 a month before. Independent confirmation would come from other sensor operators reporting a comparable rise in distinct exploited CVEs, measured with fixed detection coverage.

Why it reaches you. The component class is every internet-facing service in your estate with a published CVE: edge appliances, web frameworks, identity platforms and the ML and developer tooling that now dominates exploited-CVE lists. A human-reviewed defender pipeline producing hundreds of weaponised n-days a week sets a floor on what an attacker can do without the review step. The assumption that most CVEs never get a working public or private exploit no longer holds for a patch window measured in weeks. The parallel jump in exploited-CVE feeds (four times the monthly KEV-style additions) will also land on whoever triages them.

What to do. Vulnerability management owner: validate. Re-base your remediation SLA for internet-facing assets on a working exploit existing within three days of CVE publication, not on a confirmed in-the-wild report. Measure whether your triage team can absorb a fourfold rise in exploited-vulnerability feed entries without dropping coverage, and pre-approve mitigations for what it cannot patch in that window. Urgency: validate.

CSA coverage. Patch-to-Exploit Without Humans: GTIG's exploit_agent and the Grace Period; see also The Collapsing Patch Window: Project Zero's Full X and AI-Speed Remediation.

Attackers took secrets out of Azure API Management Classic instances at Equinor and other organisations, and Microsoft's fix does not invalidate what was taken

Category: agentic_surface **Significance:** major **Confidence:** VERBATIM (PROVIDER) for Microsoft's advisory as reported by Help Net Security: a "security issue" in the Classic tiers that could allow unauthorized access to certain management functionality "if the service was not configured according to documented security best practices", identified 18 November and fixed the next day, with credentials or configuration secrets for some customers possibly accessed, and v2 tiers not affected; VERBATIM (PROVIDER) for Norway's NSM advice to assume all secrets and named values are compromised, rotate them, and confirm runtime 0.52.31936 or 0.53.31938 or later; NO PROVIDER CLAIM for a CVE, the attacker, the number of affected tenants, or the exact access path; LINK ONLY – VERIFY AT SOURCE for researcher Anders Kristiansen's view that the retired direct Management API, which authenticates with a shared-access signature outside Entra ID and is not recorded in the Azure Activity log, is the relevant

legacy path, which he presents as his reading of public sources; CHARACTERIZATION (CSA) for the AI-gateway exposure **Source:** Help Net Security – Azure API Management vulnerability exploited to steal customers' secrets

What changed. Microsoft warned Azure API Management (APIM) customers through the Azure portal that a flaw in the Classic tiers exposed management functions on instances not set up to its guidance, chiefly those without the recommended network integration or storing secrets in APIM rather than in Key Vault. It closed the path on 19 November, a day after identifying it, but by then named values, identity-provider and OpenID Connect client secrets, subscription keys and management API tokens may have been taken from some instances. Norway's NSM has tied attacks on Equinor (reported 17 November) and Skagerak Energi (18 November) to the flaw, and says organisations outside Norway and outside energy were also hit. The fix stops further access. Anything already taken stays valid until it is rotated.

Why it reaches you. The component class is the API gateway, which many enterprises now also use as their model gateway: APIM sits in front of model endpoints and holds the provider keys, backend credentials and OAuth client secrets that agents and applications call through. A stolen secret from that configuration works directly against the service it protects, outside APIM, so gateway-side logging will not show its use. If the legacy management path is involved, it also bypasses role-based access control and the Activity log, which leaves the backend and identity logs as the only place to look.

What to do. Cloud platform owner with identity and access management: escalate. For every APIM Classic instance, rotate all named values, backend and identity-provider secrets, subscription keys, tenant access keys and the Key Vault secrets APIM references, and confirm the runtime version is at or above the fixed builds. Turn off the direct Management REST API, move remaining secrets into Key Vault, and hunt for use of the stolen credentials in Key Vault, service-principal sign-in and backend logs from 1 November onwards. Urgency: escalate.

CSA coverage. Azure API Management Classic Exploitation: When the AI Gateway Leaks the Keys.

OpenAI and Cisco released an open-source tool for agent-written kernel shields that block a vulnerable behaviour while the patch is still in testing

Category: vuln_storm **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the description of Shirudo: an eBPF-based shield is "a small program that blocks the specific behavior a vulnerability depends on, with no restart needed", "An operator approves the shield before it goes live", shields "are cryptographically signed and checked before they load", a TPM records shields being loaded and removed, and production use needs "a Linux kernel with Shirudo's kernel patches and a TPM 2.0 chip"; VERBATIM (PROVIDER) for Greg Brockman's statement that "finding a bug today doesn't mean you can patch every system tomorrow"; SELF-REPORTED (PROVIDER METRIC) for the single demonstration, a

Codex-generated shield on the Java Base64 decoder that stopped a Java4Shell exploit against a running Elasticsearch process, which Help Net notes was a "controlled test" designed and run by the tool's builders; NO PROVIDER CLAIM for false-positive rates, performance cost or behaviour when a shield is wrong; LIVE TEST REQUIRED for efficacy on your own workloads **Vendor interest:** OpenAI sells Codex, the model used to write the shields in the published demonstration; the tool itself is free. **Source:** Help Net Security – Linux patches can take weeks, and OpenAI and Cisco built Shirudo to help in the meantime

What changed. On 23 and 24 November 2026 OpenAI and Cisco released Shirudo as open source. An AI agent writes a narrow eBPF program that blocks the specific kernel-visible behaviour an exploit needs. The operator checks that it stops the attack while normal traffic still passes, approves it, and loads it signed onto a running system without a restart, with each load and removal recorded in the TPM. The stated problem is the days or weeks between a fix existing and a fix being deployed, which this feed has measured repeatedly in distribution kernels and packages this month. The only published efficacy evidence is one test run by the builders. Independent confirmation would need third-party testing across workloads, with false-block and missed-exploit rates reported.

Why it reaches you. The component class is the Linux runtime under production services and agent hosts that cannot be restarted on the exploit's schedule. Runtime shielding turns the patch window from an exposure you accept into one you can partly close, but it adds a new artefact class to the change process: machine-written kernel code approved under time pressure and loaded fleet-wide. One wrong shield deployed everywhere is a correlated outage. The production requirement for custom kernel patches and TPM 2.0 also limits where it can run today.

What to do. Platform engineering with the vulnerability management owner: monitor. Do not deploy to production yet. Decide whether runtime virtual patching belongs in your emergency-change playbook, and if it does, write the bar now: signed and revocable shields, a regression test against recorded normal traffic before approval, staged rollout, and an expiry tied to the real patch. Urgency: monitor.

CSA coverage. Agent-Written Kernel Shields: Correlated-Failure and Trust Risks.

MONAI Label shipped its first patched release on 22 November and entered CISA's exploited list the next day, two months after an attacker's agents first exploited it

Category: machine_speed **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for CISA's 23 November addition of CVE-2026-133968, the MONAI Label path traversal, to the Known Exploited Vulnerabilities catalog; LINK ONLY – VERIFY AT SOURCE for Security Affairs' report of a 26 November federal remediation deadline, exploitation since September before public disclosure on 20 November,

and the fix in MONAI Label 0.8.6; VERBATIM (PROVIDER) for the PyPI listing showing 0.8.6 released 22 November 2026 as the first release since 0.8.5 in November 2024, and for the project's new notice that MONAI Label is not intended to run as an internet-facing application and has known unpatched issues; CHARACTERIZATION (CSA) for the interval reading **Source:** Security Affairs – U.S. CISA adds N-able MailAssure and Project MONAI flaws to its Known Exploited Vulnerabilities catalog

What changed. On 21 November this feed reported that an attacker's agents, running the Cairn framework, found and exploited a MONAI Label path traversal in about 11 minutes, and that five weeks after a fix was merged PyPI still had no patched release. The project published 0.8.6 on 22 November, and CISA added the flaw to KEV on 23 November, with federal agencies reportedly given until 26 November. The full sequence is now measurable: exploitation in September, fix merged 19 October, advisory 18 November, patched release 22 November, KEV 23 November. The flaw was in attackers' hands for about two months before defenders could install a fixed package, and the window from release to federal deadline is four days. The 0.8.6 wheel was uploaded without Trusted Publishing provenance.

Why it reaches you. The component class is pip-installed ML and research servers that data-science teams run outside the usual patch pipeline. Here the open-source release step, not discovery or exploitation, was the slowest link, and an exploited-list entry followed the release by one day. The maintainers now say the software should not face the internet, and that unpatched issues remain after 0.8.6.

What to do. Research computing or data-science platform owner: validate. Upgrade every MONAI Label instance to 0.8.6 from the official index, verify the wheel hash against the project's published release, and keep instances off untrusted networks regardless of version, as the maintainers now advise. Treat hosts that ran exposed versions since September as possibly compromised and check them for persistence in Python startup files. Urgency: validate.

CSA coverage. MONAI Label Zero-Day: AI Agents Exploit Medical Imaging Research Servers.

Every one of eight web-search APIs that coding agents use returned known-malicious content in a 4,719-query benchmark, in about 1.4% of searches

Category: agentic_surface **Significance:** notable **Confidence:** SELF-REPORTED (PROVIDER METRIC) for AIR Research Lab's WebSec benchmark: 4,719 queries across Firecrawl, Tavily, Perplexity, Parallel, Exa, Nimble, Brave and You.com, more than 282,000 results scanned, known-malicious content in 534 of about 37,700 provider searches (about 1.4%, ranging from 0.8% to 2.2% by provider), and 1,026 exposures to malicious or risky content in total; VERBATIM (PROVIDER) for AIR's statement that Claude Code uses Brave Search; NO PROVIDER CLAIM from the search providers or agent vendors in response;

CHARACTERIZATION (CSA) for the exposure path into enterprise agents **Vendor interest:** AIR sells security controls for AI agents, so a finding that agents' search inputs are unsafe is commercially favourable to it. **Source:** AIR – Introducing WebSec

What changed. On 23 November 2026 AIR Research Lab released WebSec, a benchmark that sends real coding-agent queries, including install and setup questions, to the search APIs agents use and checks the results against known-malicious indicators. Every provider returned known-malicious content at least once. AIR's October case studies show what that content looks like: a 286-star GitHub repository whose one-line install command dropped a remote-access tool, and expired domains still ranked in results that anyone could register. Independent confirmation would come from a second party re-running the published query set against the same providers.

Why it reaches you. The component class is the agent runtime's retrieval layer. Coding agents increasingly search the web and then act on what they find, running install one-liners, adding dependencies or fetching scripts, and the search provider is chosen by the agent vendor, not by you. At AIR's rates, an engineering organisation whose agents run a few thousand searches a day sees malicious content daily. No control in the typical stack checks a search result before an agent acts on it.

What to do. Owner of the developer agent platform: validate. Find out which search provider each approved coding agent uses. Require human approval, or an allowlist of registries and domains, before an agent runs any install or download command it found through search, and log the source URL behind each such command. Urgency: validate.

CSA coverage. Poisoned at the Source: Search APIs Feeding Malicious Content to Coding Agents.

Nearly one in five exploited CVEs is now used to steal credentials, and AI-provider and coding-agent keys have been on attackers' target lists since May

Category: vuln_storm **Significance:** notable **Confidence:** SELF-REPORTED (PROVIDER METRIC) for VulnCheck Canary figures: 127 of 694 tagged CVEs (18%) with at least one credential-harvesting request, harvesting records up about sixfold from roughly 17,000 in July to 99,000 in October (68,000 of them from the Next.js middleware bypass, CVE-2025-29927), medium-severity CVEs making up 6% of targeted CVEs but 23% of records, and 1,756 attempts on AI-provider credentials from 406 IPs across 27 CVEs since May 2026; VERBATIM (PROVIDER) for "Attackers also don't need code execution to steal credentials" and "Treat file-read and SSRF bugs as credential exposures"; CHARACTERIZATION (CSA) for the prioritisation reading **Vendor interest:** VulnCheck sells exploited-vulnerability intelligence drawn from the same sensor network, so a finding that raises the priority of more CVEs is commercially favourable to it. **Source:** VulnCheck – Attackers Exploit CVEs to Harvest Credentials

What changed. On 23 November 2026 VulnCheck reported that credential harvesting rose from about 1% of exploitation records on its sensors in January to 25–35% in September and October. Attackers use file-read, path-traversal and SSRF bugs to read `.env` files, process environments, cloud credential files and instance metadata. Requests for AI-provider keys (OpenAI, Anthropic, Hugging Face, Gemini and others) began in May. On two Langflow flaws, targeting of AI keys started the same day as, or one day after, first exploitation. Coding-agent credential files, including `~/.claude/.credentials.json` and `~/.codex/auth.json`, are also being requested, though from five IPs or fewer so far. Independent confirmation would come from other honeypot operators publishing the same breakdown.

Why it reaches you. The component class is the application host and CI runner holding long-lived secrets in files and environment variables, increasingly including model-provider keys and agent tokens. A medium-severity file-read bug on such a host is functionally a credential compromise, but CVSS-driven triage puts it behind the critical queue. Stolen AI keys turn into inference spend and access to whatever the key's agents can reach.

What to do. Vulnerability management owner with the secrets management owner: validate. Re-score exploited file-read, path-traversal and SSRF vulnerabilities on hosts that hold secrets as credential exposures, and make rotation of every secret reachable from the host part of their remediation. Enforce IMDSv2 and limit which workloads can request managed-identity tokens, and move AI-provider keys out of files and environment variables into short-lived, brokered credentials. Urgency: validate.

CSA coverage. Self-Funding Agents: ARTEX Campaigns Harvesting AI-Gateway Keys.

Watchlist

- **OpenAI reward-hacking postmortem – downstream response** – No change. *(opened 2026-08-27)*
- **VM/hypervisor containment hardening for cyber-capable agents** – No change. Shirudo (item 3) shields the host kernel at runtime; it does not change agent-sandbox isolation, and no provider has announced a sandbox change following last issue's KVM nested-virtualization flaw. *(opened 2026-08-27)*
- **Claude Code Auto Mode prompt-injection ASR discrepancy** – No change. No Anthropic response to the reproduced attack success rate. *(opened 2026-08-27)*
- **AI defensive-triage guardrail evasion** – No change. *(opened 2026-08-31)*
- **AI account session hijacking at scale** – Related signal, not a session-hijacking campaign. VulnCheck (item 6) reports exploitation traffic requesting stored coding-agent tokens

(`~/.claude/.credentials.json` , `~/.codex/auth.json`) from five or fewer IPs, and 1,756 attempts on AI-provider keys since May. Token theft from developer hosts is now a measured route to AI account takeover. No provider has announced device-bound or short-lived session tokens. *(opened 2026-08-31)*

Opened this issue

- **Azure API Management Classic secret theft — scope and downstream use** — `agentic_surface` . Watching for Microsoft to publish the access path, a CVE and the number of affected tenants, for evidence that stolen gateway or model-provider credentials have been used against backends, and for Microsoft to retire or force-disable the direct Management API on Classic tiers. *(opened 2026-11-24)*