

CSAI Foundation | Frontier Ready Cybersecurity

Frontier Ready Daily

Issue 58 · Member edition

2026-10-23

CSAI FOUNDATION INITIATIVE

Frontier Ready.

Machine-speed agentic cybersecurity

CSA cloud
security
alliance®



© 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

Two primary reports put numbers on agent-run intrusion. Microsoft says a state-aligned actor using an agent framework went from choosing a target to privileged cloud access in under three hours by riding a vendor application's existing tenant trust, and CrowdStrike documents four operators running offensive agents built from public skills and frameworks. The remaining items cover an attack agent that hunts for exposed model endpoints to run on, network and security appliances that embed PHP and still have no PHP2Shell fix, OpenAI's requirement that AI gateways get its approval from 20 November, and a vendor survey on AI-linked security headcount cuts.

Microsoft says an agent-driven actor went from target selection to privileged cloud access in under three hours through a vendor app's existing tenant trust

Category: machine_speed **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the attribution to Storm-0940, the use of the ARTEX framework, and the statement that in several cases the time from initial targeting to privileged cloud access was under three hours; SELF-REPORTED (PROVIDER METRIC) for the 2,250+ tenants targeted, 151,000+ failed logons, the peak of about 2,900 failures per minute on 11 October, and 200+ sign-ins in under four minutes using four application credentials; NO PROVIDER CLAIM for how many tenants were compromised or whether data was taken; CHARACTERIZATION (CSA) for the reading that multi-tenant application consent is now the fastest route to privilege **Vendor interest:** Microsoft sells the Defender detections and hunting content it lists for this campaign. **Source:** [Microsoft Security Blog – AI-driven, large-scale nation-state campaign using ARTEX enables supply chain compromise](#)

What changed. On 21 October 2026 Microsoft published the primary report behind the ARTEX activity covered yesterday. It says Storm-0940 used ARTEX from September 2026 and, after taking control of ConnectWise application identities (reported 15 October), authenticated into more than 2,250 customer tenants using consent those apps already held. In some tenants the agents then used Microsoft Graph to change app credentials, groups and privileged roles, and read secrets in bulk from Azure Key Vaults. In several cases the interval from first targeting to privileged cloud access was under three hours. Microsoft's own counts show the separate 10–14 October password spray failed. The successful path was the vendor app, not user passwords.

Why it reaches you. The exposure is the identity provider's set of third-party multi-tenant applications and the permissions they were granted at onboarding. An agent holding a vendor's app credential needs no phishing and no MFA bypass, and a three-hour run to privilege finishes before most weekly access reviews or vendor-notification processes start. Microsoft asks responders to search by application ID back to 1 October, because display names can be changed.

What to do. Identity and access management lead: validate that every third-party app with tenant-wide Graph, directory or Key Vault permissions has a named owner, holds only the permissions it uses, and alerts on new credentials, self-assigned roles and app-only sign-ins from new infrastructure. If you authorised the ConnectWise apps, keep them disabled and run the hunt Microsoft describes. Urgency: validate.

CSA coverage. [ConnectWise Cloud Backup: One Vendor App, 173 Tenants](#); see also [Credential Custodians as Blast-Radius Multipliers](#).

CrowdStrike documents four unrelated operators running offensive agents assembled from public skills and frameworks, one with 819 borrowed skills

Category: machine_speed **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the four actor profiles, the supervisor-and-executor and state-machine designs, the 819 skills, and the assessment that agentic AI likely commoditises baseline offensive operations; SELF-REPORTED (PROVIDER METRIC) for the victim counts (at least 15 Argentine healthcare entities, and 20+ named victims for the Chinese-speaking operator) and for the statement that such agents can reach objectives autonomously, often within an hour, which per-intrusion timelines from other responders would confirm; CHARACTERIZATION (CSA) for the reading that model safeguards slowed only the least capable operator **Vendor interest:** CrowdStrike sells endpoint detection, threat intelligence and managed hunting, so evidence of rising agentic intrusion is commercially favourable to it. **Source:** [CrowdStrike – CrowdStrike Uncovers New Agentic AI Campaigns](#)

What changed. On 21 October 2026 CrowdStrike described four previously unreported actors, observed between May and August 2026, each of which left its agent framework in an open directory. A Spanish- and English-speaking operator ran a supervisor agent that picked tasks from an attack-stage state machine and an executor that carried them out, built on PentestGPT with added MCP servers. A Chinese-speaking operator reached models through a China-based proxy and ran parallel specialist agents against government and private targets. An Indonesian-speaking operator used 819 skills, mostly from public repositories, with no orchestration of its own, and had to intervene to get past model guardrails. A Russian-speaking group started a seven-step workflow from a single slash command and took client data from a US CRM vendor.

Why it reaches you. The targets were internet-facing applications and SaaS platforms in healthcare, government, CRM and data-centre providers, which are the same suppliers that hold enterprise customer data. The tooling is public, and a low-skill operator got useful results by loading published skills into a coding agent. That makes this volume, not a single actor, that perimeter detection and supplier assurance need to plan for.

What to do. SOC lead: validate that detection on internet-facing applications fires on the command pacing and tool sequences of agent-run sessions, using the indicators in CSA's detection note below, and add the agentic-intrusion question to supplier security reviews for CRM and data-hosting vendors. Urgency: validate.

CSA coverage. [Fingerprinting Agent-Operated Intrusions: Detection and Response Guidance](#)

A captured attack agent was built to run on other people's exposed model endpoints, with up to 100,000 steps and no human approval

Category: agentic_surface **Significance:** major **Confidence:** VERBATIM (PROVIDER) for Symantec's description of the agent as "a reasoning engine that plans and drives the attack" and for the targeting of self-hosted models, AI gateways and MCP servers; SELF-REPORTED (PROVIDER METRIC) for the 127,000+ honeypot requests from tens of thousands of addresses, the 100,000-step limit and the roughly 74 tool definitions, which other honeypot operators publishing comparable request logs would confirm; LIVE TEST REQUIRED for whether the captured payloads would have worked, since the honeypot did not execute them **Vendor interest:** Symantec sells endpoint and network threat-protection products, so evidence of attacks on exposed AI services is commercially favourable to it. **Source:** [Infosecurity Magazine – Symantec Uncovers AI Agent Hijacking Models for Autonomous Cyber-Attacks](#)

What changed. On 21 October 2026 Symantec's Threat Hunter Team published the instruction set of an offensive agent captured this summer by a honeypot posing as an unprotected model service. The operator sent the agent's prompt to what it took for a free model. The prompt framed reading `.env` files, API keys and database schemas as routine development work, told the model to run tools without asking, and allowed up to 100,000 steps. A second request a day later, from a rotated address, asked the model to decode a script that compiles a Monero miner aimed at camera gateways. Over the monitoring period the honeypot logged more than 127,000 requests, mostly endpoint reconnaissance, stolen-key attempts and probes for configuration files.

Why it reaches you. The component is any OpenAI-compatible endpoint an enterprise exposes, including self-hosted inference servers, internal AI gateways and test deployments. They share one interface, so a single scanner finds them all. An exposed endpoint is no longer just a billing risk. It becomes

the planning engine for attacks on third parties, run on the enterprise's compute and traced back to its addresses.

What to do. Model platform owner: validate that no inference server, gateway or MCP server accepts unauthenticated requests from the internet, by scanning your own external ranges for OpenAI-compatible routes, and alert on token or GPU use outside expected callers. Urgency: validate.

CSA coverage. [LLMjacking Evolved: Stolen Compute as Offensive AI Infrastructure](#)

Network and security appliances that embed PHP have no PHP2Shell fix two days after the patch, with 31 Cisco products and services still under investigation

Category: vuln_storm **Significance:** notable **Confidence:** LINK ONLY – VERIFY AT SOURCE for the advisory states as relayed by Help Net Security: all PAN-OS and Panorama versions affected with no fixed release, 12 Fortinet products under investigation with no fix or workaround, and 18 Cisco products and 13 cloud services under investigation with no workaround; LINK ONLY – VERIFY AT SOURCE for VulnCheck's view that x86-64 exploitation without an ASLR bypass has not been shown; CHARACTERIZATION (CSA) for the reading that appliance vendors are now the remediation bottleneck for this flaw **Source:** [Help Net Security – What the PHP2Shell vulnerability means for Fortinet, Cisco and Palo Alto customers](#)

What changed. As of 22 October 2026, a day after PHP released fixes for CVE-2026-109618 and CVE-2026-109619, the three largest network-security vendors had not shipped fixed software for products that bundle PHP. Palo Alto lists every PAN-OS and Panorama version as affected and has issued IPS signatures but no fixed release. Fortinet has 12 products under investigation with no fixed versions. Cisco's vulnerable and not-vulnerable lists were both empty, with 31 products and services under review and no workaround. Among Linux distributions, RHEL and its rebuilds, Amazon Linux and SUSE had not published fixes either. Amazon has reported exploitation from 17 October.

Why it reaches you. The exposure is the management plane of firewalls, SD-WAN controllers, identity services and voice gateways, which security teams cannot patch themselves. For these the time to remediate is set by each vendor's PHP rebuild, not by the enterprise's patch cycle. The only compensating controls available now are network restriction of management interfaces and vendor signatures.

What to do. Network security lead: escalate. Confirm no appliance management interface is reachable from the internet or from user networks, enable the vendors' available signatures, and track each embedded-PHP product in a register with its vendor status until a fixed version is installed. Urgency: escalate.

OpenAI will require its approval for model routers and AI gateways that give third parties direct access, from 20 November

Category: security_operating_model **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the statement that OpenAI developed policies to align safeguards across providers and is working with partners so their implementations meet or exceed its standards; LINK ONLY – VERIFY AT SOURCE for the usage-policy clause, quoted in secondary coverage, that operators of a "model-access intermediary platform or other service that enables third parties to directly access OpenAI models" "must obtain our prior review and approval", effective 20 November 2026, and for the SafetyID end-user identifiers; NO PROVIDER CLAIM for whether a gateway serving only one organisation's own staff is in scope, or for how long review takes **Source:** [OpenAI – Broad access through consistent standards](#)

What changed. On 21 October 2026 OpenAI published updated usage policies, effective 20 November 2026. Services that let third parties reach OpenAI models directly, which OpenAI's announcement says include model routers and AI gateways, need its prior review and approval. Partners that host or resell its models must keep safeguards consistent with OpenAI's and coordinate on enforcement. OpenAI is also introducing SafetyIDs so misuse investigations can single out an individual end user without OpenAI learning who that person is. Applications that only use OpenAI models to power their own features are carved out.

Why it reaches you. Many enterprises reach frontier models through a hosted router or gateway, and some run gateways that serve subsidiaries, customers or partners. A gateway that is not approved by 20 November could lose access, and a provider can now require per-user identifiers to pass through the gateway. Both affect gateway architecture and logging, a day after this feed reported how much gateway logs can expose.

What to do. AI platform owner: monitor. Ask any hosted router or gateway supplier for its approval status before 20 November, decide whether your own gateway serves "third parties" under the clause, and plan how a pseudonymous per-user identifier would be generated and logged without storing more personal data. Urgency: monitor.

CSA coverage. New – no prior CSA coverage.

A vendor survey reports AI-linked security headcount cuts split evenly between entry-level and senior roles

Category: security_operating_model **Significance:** context **Confidence:** SELF-REPORTED (PROVIDER METRIC) for the 56% of respondents reporting AI-linked workforce reductions, the even split between entry-level and higher-level roles, and the 69% using incident response that pairs AI or automation with human decisions, none of which has a published sample size or fieldwork date; CHARACTERIZATION (CSA) for the reading that cuts at the entry level remove the pipeline that senior roles depend on **Vendor interest:** Barracuda sells managed detection and AI-assisted security services, so a skills-gap finding is commercially favourable to it. **Source:** [Help Net Security – AI is taking cybersecurity jobs at junior and experienced levels alike](#)

What changed. On 23 October 2026 Help Net Security reported Barracuda's skills-gap survey. 56% of respondents reported workforce reductions linked to AI, divided evenly between entry-level and more senior roles, and 69% said their incident response combines AI or automation with human investigation and decisions. Almost all reported skills gaps, most often in AI-related risk. The report gives no sample size. Earlier this month this feed covered a 26.4% fall in US security job postings, which points the same way.

Why it reaches you. If the routine triage that trained junior analysts is automated away, the organisation loses the route by which it grew people able to take over human-in-the-loop decisions. Those are the decisions that machine-speed defence leaves to people.

What to do. CISO: no action on this survey alone. When AI tooling replaces analyst tasks, record which human decision points remain and who is being trained to hold them. Urgency: no action.

CSA coverage. New – no prior CSA coverage.

Watchlist

- **OpenAI reward-hacking postmortem – downstream response** – No change. (*opened 2026-08-27*)
- **VM/hypervisor containment hardening for cyber-capable agents** – No change. (*opened 2026-08-27*)
- **Claude Code Auto Mode prompt-injection ASR discrepancy** – No change. CrowdStrike's report (item 2) shows Claude Code used as an attacker's agent runtime, which is a misuse finding, not a response on prompt-injection attack success rate. (*opened 2026-08-27*)

- **AI defensive-triage guardrail evasion** – No change. CrowdStrike notes that one operator had to manually get past model guardrails, but that concerns safeguards on attacker use, not refusals used to blind defensive triage. (*opened 2026-08-31*)
- **AI account session hijacking at scale** – No change. CrowdStrike's Chinese-speaking operator reached models through a third-party proxy service, not through hijacked accounts. (*opened 2026-08-31*)

Opened this issue

- **Embedded-PHP appliances awaiting PHP2Shell fixes** (*vuln_storm*) – Watching for fixed PAN-OS and Panorama releases, Cisco's final affected-product list, the outcome for Fortinet's 12 products under investigation, and any report of exploitation against an appliance management interface. The measure is days from PHP's 21 October fix to each vendor's fixed release.