CSAI Foundation | Frontier Ready Cybersecurity

Frontier Ready Daily

Issue 75 · Member edition

2026-11-09



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

Microsoft has observed an agent go from searching GitHub for a leaked client secret to taking data out of Azure Key Vault in five minutes, which no human rotation process can match. Wiz reports a SaaS vendor's AWS account used to assume customer integration roles, with every attempt succeeding despite External IDs. The other items cover Project Zero's measurement that an LLM agent and a fuzzer each find roughly 40 V8 vulnerabilities in six months with only a quarter in common, MCP SDK OAuth fixes that leave clients exposed until operators set an expected issuer, NIST's ten proposed tenets for agent identity, and a bank theft carried out through a provider's standing access to the bank's own payment authority.

An agent went from searching GitHub for a leaked client secret to exfiltrating Azure Key Vault secrets in five minutes, and fixed its own failed API calls in under 80 seconds

Category: machine_speed **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the five-minute interval, the 28-, 60- and 80-second retry intervals, the August 2026 start of GitHub scanning and the 28 October start of service-principal abuse; CHARACTERIZATION (CSA) for the reading that rotation schedules no longer work as containment for publicly exposed secrets; NO PROVIDER CLAIM on how many tenants were affected or how typical the five-minute case is **Vendor interest:** Microsoft sells the cloud-detection and secret-scanning services its recommendations point to, though the core advice (revoke exposed credentials, move to managed identities) is vendor-neutral. **Source:** Microsoft Security Blog – Threat actors use agentic AI tooling to find and abuse exposed secrets

What changed. On 5 November 2026 Microsoft reported that actors using ARTEX and similar agentic frameworks have been searching public GitHub for client secrets and API keys since August. Since at least 28 October they have been signing in as the affected service principals. In one observed case an agent went from the GitHub search to exfiltrating Azure Key Vault data in five minutes. It then enumerated subscriptions, permissions and AI model deployments, and pulled data from Storage Accounts, SharePoint and OneDrive. When Azure Resource Manager calls failed on outdated API versions, the agent corrected them and succeeded about 28, 60 and 80 seconds later. The last measured interval in this feed was stolen AI keys first used within 30 minutes (issue of 1 November). Microsoft presents five minutes as one case, not a typical timeline.

Why it reaches you. The exposure path runs through workload identities: OAuth app registrations and service principals that authenticate with client secrets, and the Key Vaults, storage and AI resources they can read. Any secret committed to a public repository, fork or Git history is in reach, and Microsoft notes that "removing the content does not invalidate the credential." A 30-, 60- or 90-day rotation cycle, or a secret-leak alert routed to a human queue, cannot act inside five minutes.

What to do. Cloud identity owner: escalate. Make revocation of any publicly exposed secret an automated, event-driven action rather than a ticket, and measure its median time in minutes. Inventory service principals that still authenticate with client secrets and can read Key Vault, and move the highest-privilege ones to managed or federated identities first. Hunt Entra sign-in logs from 28 October for an invalid-secret or expired-secret failure (AADSTS7000215 or AADSTS7000222) followed by a success, which Microsoft treats as urgent. Urgency: escalate.

CSA coverage. Five Minutes from Search to Key Vault: Agentic Secret Harvesting; see also Approved Once, Leaked Forever: Secrets in Claude Code Settings Files.

A threat actor used a SaaS vendor's own AWS account to assume customer integration roles, and every attributed attempt succeeded despite carrying an External ID

Category: agentic_surface **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the 14–26 October activity window, the use of Qubole's AWS account, the success of every attributed AssumeRole call and the statement that a source-IP condition would have blocked all of them; LINK ONLY – VERIFY AT SOURCE for the number of victim environments, which Wiz does not state; NO PROVIDER CLAIM from Qubole or AWS on root cause or response; CHARACTERIZATION (CSA) for the relevance to agent-platform integrations **Vendor interest:** Wiz sells cloud security posture and detection services that hunt for this activity. **Source:** Wiz – Trusted by Default: Qubole Cross-Account Access Abused in the Wild

What changed. On 9 November 2026 Wiz, working with Sentra and Mandiant, reported that between 14 and 26 October an actor used IAM user credentials belonging to the AWS account Qubole runs its platform from. The actor called `sts:AssumeRole` against the integration roles customers had created to trust that account. Every attributed request included the External ID and every one succeeded. In most environments Wiz observed, these were the only role assumptions from the vendor account that month. The tooling was scripted rather than described as agentic. It used session names in a consistent format, ran the same sequence of actions in each environment, and validated harvested credentials in bursts lasting seconds. The actor then read Secrets Manager, SSM parameters, Lambda configuration and S3 data. Wiz says a source-IP condition in the trust policy would have blocked every observed assumption.

Why it reaches you. The component is cross-account IAM trust granted to third-party platforms. Data, analytics, security and, increasingly, AI-agent platforms all use the same mechanism to act inside customer accounts. The External ID is designed to stop the confused-deputy problem between a vendor's customers. It does nothing once the vendor's own credentials are in an attacker's hands, and it is logged in plaintext in CloudTrail. No AI was reported in this campaign. It is in this feed because it is the same standing delegation enterprises are now granting to agent platforms, and it failed open in production.

What to do. Cloud platform owner: validate. Inventory every role that trusts an external account and remove trust for integrations no longer in use. Replace `:root` principals with specific ones, and add `aws:SourceIp` or `aws:SourceVpce` conditions wherever the vendor publishes egress ranges. Make "publishes egress ranges and uses short-lived workload credentials" a contract requirement for new integrations, including AI-agent platforms. If you trust Qubole's account, hunt CloudTrail from 1 October using the indicators in the source. Urgency: validate.

CSA coverage. New — no prior CSA coverage. Related: Through Someone Else's Platform: The Crédit Mutuel Arkéa Theft.

Project Zero's LLM agent and its V8 fuzzer each found about 40 actionable vulnerabilities in six months, and only 20 of 81 were found by both

Category: vuln_storm **Significance:** notable **Confidence:** SELF-REPORTED (PROVIDER METRIC) for the 45-versus-37 actionable vulnerability counts, the 20-of-81 bug-tracker overlap and the roughly 60% overlap from the rediscovery experiment, all measured by Google on Google's own systems; NO PROVIDER CLAIM on cost per bug, which Project Zero says it cannot yet compare; CHARACTERIZATION (CSA) for the procurement implication **Source:** Project Zero — Fuzzers vs. AI: Comparing Fuzzilli and BigSleep on JavaScript Engines

What changed. On 27 October 2026 Project Zero published production data for April to September 2026 from Chrome's V8 engine. In that period BigSleep, an LLM agent that reviews new commits once a day, produced 45 actionable vulnerabilities and Fuzzilli, a specialised JavaScript-engine fuzzer, produced 37. Only 20 of the 81 actionable vulnerabilities were found by both. A separate experiment that asked the agent to rediscover past fuzzer bugs put the overlap at roughly 60%. Fuzzilli usually picks up new code within hours, while BigSleep reports with a one- to two-day delay. Project Zero recommends that vendors run both and expects the agent share to grow, noting that the model it measured is already outdated.

Why it reaches you. This affects every product built on V8, including Chrome, Node.js, Electron desktop apps and embedded runtimes, and more broadly any supplier whose security testing claims rest on one discovery method. It gives a denominator to the volume question this feed tracks. On a target

fuzzed for years, adding an agent did not replace the fuzzer. It added a separate stream of vulnerabilities that arrives a day or two later, so the number of fixes downstream consumers must take grows rather than moves around. Independent confirmation would need the same comparison on targets Google does not own.

What to do. Product security owner: validate. For software you build, check whether CI runs both coverage-guided fuzzing and agentic code review on new commits, and record what each finds separately so the overlap can be measured. For critical suppliers, ask which discovery methods they run on new code and whether any single method's findings make up all of their security fixes. Size patch capacity for embedded-V8 runtimes on the assumption that both streams will keep growing. Urgency: validate.

CSA coverage. The Disclosure Bottleneck: When AI Discovery Outpaces Patch Consumption; see also VulnOps: Vulnerability Management in the Age of AI.

MCP's official SDKs fixed a flaw that let a malicious server redirect OAuth credentials, but upgrading alone leaves clients exposed until operators set an expected issuer and re-bind stored credentials

Category: agentic_surface **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the affected and fixed versions, the CVSS 7.5 rating and the post-upgrade configuration steps in advisories GHSA-qx49-fqc8-xw99 and GHSA-6qxp-vccf-f47h; NO PROVIDER CLAIM on exploitation in the wild; LIVE TEST REQUIRED for whether a given deployment has configured the expected issuer and cleared unbound credentials **Source:** GitHub Security Advisory — OAuth client could send credentials to an authorization server chosen by the MCP server (GHSA-qx49-fqc8-xw99)

What changed. On 28 and 30 September 2026 the Model Context Protocol maintainers published advisories for the Python and TypeScript SDK OAuth clients (CVE-2026-115563 and CVE-2026-104850, CVSS 7.5). A malicious or compromised MCP server could name its own authorization server, and the client would send it stored refresh tokens and client secrets without user interaction. The fixes are Python `mcp` 1.30.0 or 2.2.0, and TypeScript `@modelcontextprotocol/sdk` 1.31.0 or `@modelcontextprotocol/client` 2.2.0. Both advisories say that upgrading is not enough. Machine-to-machine providers must be given an explicit expected issuer, and credentials stored before the fix stay unbound until they are cleared or tagged with their issuer. In Python 1.30.0, the warning for a missing issuer is a `DeprecationWarning`, which Python hides by default. CSA's note today reads this as one of three related transport-trust defects in the SDKs.

Why it reaches you. The component is the agent runtime's MCP client, especially headless agents that use client-credentials or private-key JWT providers to reach internal tools. A team can pass a dependency scan on the fixed version and still send long-lived secrets to whichever server an untrusted MCP endpoint advertises. Dependency scanners report the version and do not see the configuration.

What to do. Agent platform engineering owner: validate. Find every service that imports the MCP Python or TypeScript client SDKs, confirm the fixed version, and confirm in code that every OAuth provider passes an expected issuer. Clear or re-bind any OAuth credentials stored before the upgrade. Rotate client secrets and revoke tokens for any client that may have connected to an MCP server you do not control. Add the issuer check to code review for new MCP integrations. Urgency: validate.

CSA coverage. MCP's Transport Trust Gap: Security Implications and Guidance.

NIST's ten proposed tenets for agent identity say deterministic policy, not guardrails, must make access decisions, and that agents should hold no standing privileges

Category: security_operating_model **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the ten tenets, the statement that "guardrails alone are insufficient" and the plan to use the tenets to guide NCCoE demonstrations; NO PROVIDER CLAIM on a comment deadline or a timeline for practice guides; CHARACTERIZATION (CSA) for the expectation that the tenets will shape federal and supplier requirements **Source:** NIST Cybersecurity Insights – Design Tenets for Agent Identity Management Systems

What changed. On 27 October 2026 NIST and MITRE authors published ten proposed design tenets for identity, authentication and authorization of AI agents. Agents get their own cryptographically verifiable identities, distinct from human and service accounts. Agents stay linked to whoever initiated them, and delegated authority can only narrow at each hop. Agents hold no standing privileges, and access is short-lived, audience-restricted and revocable at runtime. Human review should be triggered by risk and designed to avoid consent fatigue. The logic that decides what an agent attempts is separated from the systems that decide what it may do, and deterministic mechanisms make the primary access decisions. Actions are recorded in a tamper-evident, append-only log. NIST says the tenets will guide the NCCoE's agent identity demonstrations. They are not a standard, and NIST acknowledges current technology may not meet every tenet.

Why it reaches you. This covers the identity provider, the agent runtime and the policy layer between them. The first two items in this issue fail on exactly these tenets: a long-lived secret that worked for whoever found it, and a standing third-party role with no runtime check. This is the clearest federal

statement so far of what agent identity should look like, and NCCoE demonstrations usually become the reference architectures that suppliers to government are measured against.

What to do. Identity and access management owner: validate. Use the ten tenets as a gap assessment for the agent identity design you run today, and record for each deployed agent whether it holds standing privileges and whether any access decision depends only on model behaviour. Add tenets 3, 4 and 8 (no standing privilege, revocable short-lived access, deterministic access decisions) to requirements for agent platforms you procure. Urgency: validate.

CSA coverage. NIST's Agent Identity Tenets: A Proposed Federal Reference Point.

Attackers reached Crédit Mutuel Arkéa through a provider's platform and moved about €118 million out of the bank's own account, and an investigator says $30 million was laundered before the bank confirmed

Category: security_operating_model **Significance:** context **Confidence:** VERBATIM (PROVIDER) for the bank's statements that the entry point was a service provider's platform, that customer assets were unaffected and that it can "validate the order of magnitude" of the reported loss; LINK ONLY – VERIFY AT SOURCE for the €118 million (about $133 million) estimate and the $30 million laundered by 3 November, both from blockchain investigator ZachXBT; NO PROVIDER CLAIM on the provider's identity, the technique, or any AI or automation role; CHARACTERIZATION (CSA) for the relevance to agent payment authority **Source:** BleepingComputer – Hackers pull off $133 million French bank heist via service provider

What changed. On 4 November 2026 Crédit Mutuel Arkéa confirmed that an attacker had reached it through an unnamed service provider's platform. The attacker used the bank's own institutional account, not customer accounts, to transact with other banks and convert funds to stablecoins. Investigator ZachXBT, who reported the withdrawals on 2 November before the bank was named, says about $30 million had been laundered by 3 November. He argues it could have been frozen with earlier disclosure. Some funds have been blocked. No AI role has been described.

Why it reaches you. The component is a third-party platform with standing authority to start payments, the same authority enterprises are now beginning to give agent platforms and payment agents. Customer-fraud controls such as strong authentication and payee alerts did not apply to the bank's own treasury flows. Supplier questionnaires and registers describe the provider but do not limit, at transaction time, what a compromised channel can do. This is context: there is no measured interval on the bank's side and no technical detail yet.

What to do. Treasury and third-party risk owners: monitor. Track the bank's root-cause disclosure. In the meantime, list every provider or agent with standing authority to start payments from corporate accounts, and check whether a transaction-time limit independent of the provider applies to it. Urgency: monitor.

CSA coverage. Through Someone Else's Platform: The Crédit Mutuel Arkéa Theft.

Watchlist

- **OpenAI reward-hacking postmortem — downstream response** — No change. No frontier lab disclosed a new eval-to-production escape, and there was no regulatory or contractual development. *(opened 2026-08-27)*
- **VM/hypervisor containment hardening for cyber-capable agents** — No change. No new provider or enterprise microVM adoption signal, and no QEMU/KVM/libslirp patch news. *(opened 2026-08-27)*
- **Claude Code Auto Mode prompt-injection ASR discrepancy** — No change. No Anthropic patch or public response. *(opened 2026-08-27)*
- **AI defensive-triage guardrail evasion** — No change. No new campaign that triggers refusals to blind triage tooling, and no vendor published a refusal-resistant triage architecture. *(opened 2026-08-31)*
- **AI account session hijacking at scale** — No change. Microsoft's secret-harvesting report (item 1) involves workload-identity secrets, not AI-provider user sessions. No AI provider disclosed a session-hijacking campaign or shipped device-bound session tokens. *(opened 2026-08-31)*

Opened this issue

- **Third-party cross-account role trust abuse** *(agentic_surface)* — An actor used Qubole's AWS account to assume customer integration roles from 14 to 26 October, and every attributed attempt succeeded despite External IDs (item 2). Watching for a statement from Qubole or AWS on root cause and victim count, the same technique against other SaaS or AI-agent platform vendor accounts, and vendors publishing egress ranges or adopting short-lived credentials in response.