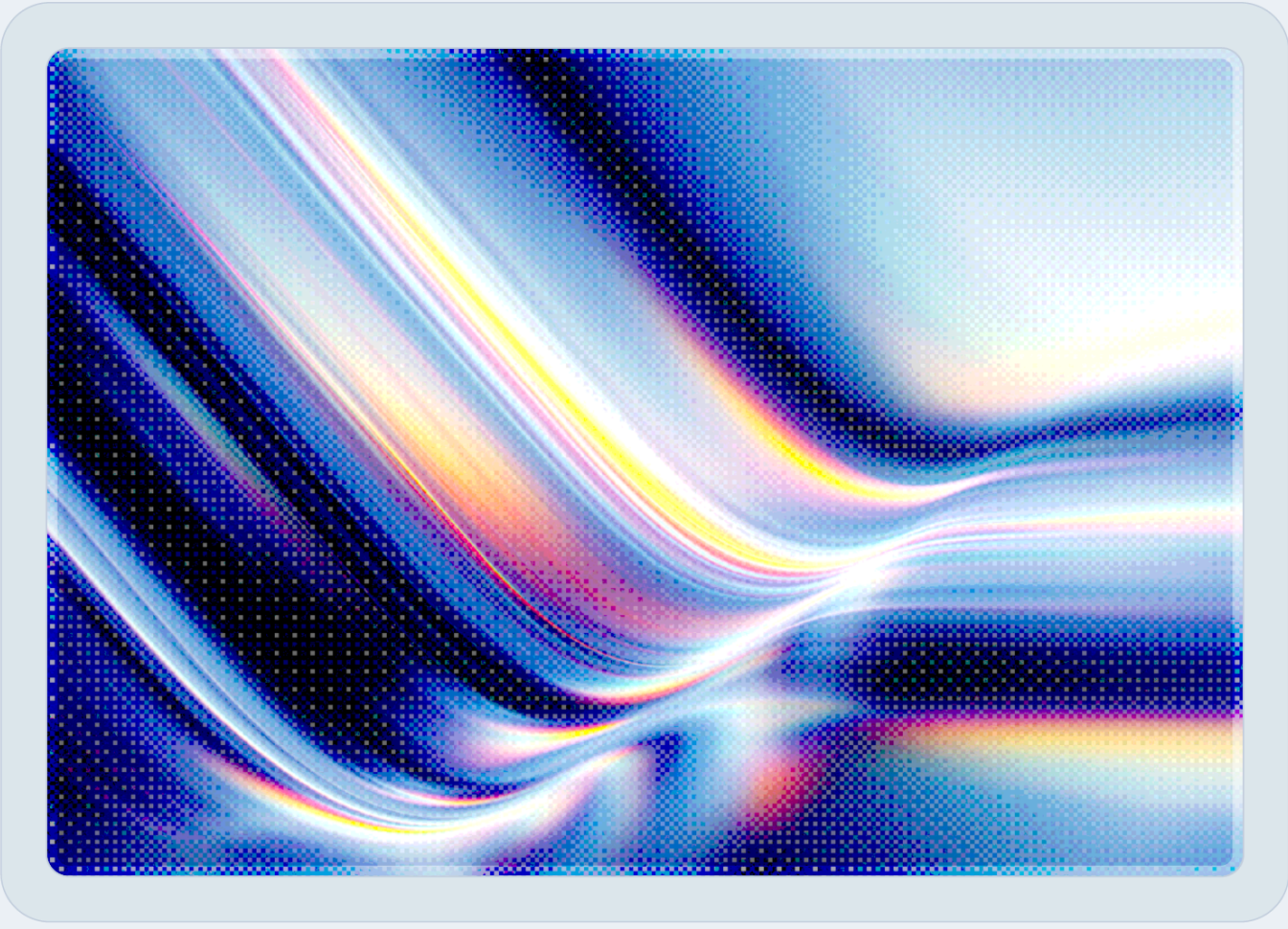
CSAI Foundation | Cloud Security Alliance

Black October Reaches PaaS: Northflank, Secret Stores, and AI Key Resale

Security Implications and Guidance for Platform-Hosted AI Credentials

2026-11-02

AI-assisted Rapid Research



CSAI

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.

Key Takeaways

On October 30, 2026, the cloud deployment platform Northflank disclosed a security incident. It told customers to assume that every environment variable stored on the platform had been decrypted and to rotate all keys [2][3]. Within hours, customers reported unauthorized use of AI provider API keys they had stored there [2][3][4]. Northflank was the ninth AI developer platform to disclose an incident in October, a run that Wiz Research has labeled "Black October" [1]. Wiz reports that the attackers in these incidents went after stored environment variables, especially those that grant access to AI models. It also reports that no single vulnerability or dependency links the incidents [1].

Wiz's research supplies the downstream evidence of what happens to such keys after theft. In two case studies, whose source vendor Wiz did not identify, stolen enterprise Amazon Bedrock credentials went from initial validation to industrial-scale abuse: about 800,000 model invocations in one case and 1.7 million in the other, which Wiz estimates likely ran to more than US$100,000 in usage at list price [1]. Wiz found that infrastructure tied to two operations openly reselling discounted frontier-model access had access to the stolen keys, although it could not definitively link those resellers to the floods or to the initial intrusions [1]. Even so, the pattern is consistent with an active secondary market in which stolen AI credentials are validated, pooled, and offered for resale within days of a platform breach. Wiz's two case studies cannot establish how widespread or profitable that market is.

For enterprises, the main lesson is architectural, not specific to one vendor. An AI provider key stored as a platform environment variable shares the security fate of that platform's control plane, build system, and secrets store. Organizations using an affected platform should treat every stored credential as exposed. More broadly, the October pattern argues for designing as though any long-lived AI credential held by a third-party PaaS could be exposed, and for scoping, constraining, and monitoring it accordingly.

Background

The Northflank Disclosure

Northflank is a platform-as-a-service (PaaS) provider. Developers use it to deploy containerized workloads, managed databases, and jobs, and they supply configuration and secrets as environment variables. BleepingComputer reports that Wiz alerted Northflank to a threat at 11:43 UTC on October 30.

Northflank engineers began reviewing the alert a minute later, and the company deployed network-policy patches, including follow-on patches at 12:50 and 13:05 UTC [2]. At 13:58 UTC Northflank's status page moved the incident to "Identified" with the instruction: "Assume all environment variables have been decrypted. Disable/rotate all keys" [3]. By 15:12 UTC Northflank had enabled one-click credential rotation for PostgreSQL and MongoDB add-ons, and later that evening it reported unauthorized use of AI API keys stored on the platform [3]. Over the following day, Northflank extended rotation to MySQL, Redis, MinIO, and RabbitMQ and turned on audit logging for all customers, including free-tier accounts [2][3].

Northflank's November 2 status update added some detail. The company said the attack was contained within about an hour of the first alert, appeared to target secrets in the control plane rather than running workloads, did not affect customers on forward-deployed or dedicated control planes, and that it had patched the exploited vulnerability across its fleet, including Bring Your Own Cloud (BYOC) clusters [3]. Several important facts nonetheless remain undisclosed as of this writing. Northflank has not published a root cause analysis, which it says will follow once forensic findings are confirmed, nor how many customers or projects were affected, nor whether credentials other than AI keys were abused [2][3]. Its updates also do not state plainly whether secrets held by BYOC customers were exposed, even though Northflank's own material describes that model as keeping workloads, databases, and secrets in the customer's own cloud account [5]. Readers should note that a detailed public reconstruction of the status-page timeline comes from Qovery, a Northflank competitor [4], and should check it against Northflank's official communications.

The Black October Pattern

Northflank's incident followed a month of disclosures across hosting platforms, inference providers, and AI coding tools. The table below summarizes the nine incidents Wiz catalogued. The cause attributions are the platforms' own statements, as reported by Wiz [1].

Date (Oct 2026)	Platform	Category	Platform-stated cause or indicator
6	Fireworks AI	Inference provider	Credential access; "agent swarm" indicators reported
9	Andromeda	AI platform	Described as part of an "industry-wide attack"
14	Zeabur	PaaS	Credential access; company is shutting down [7]

Date (Oct 2026)	Platform	Category	Platform-stated cause or indicator
15	Lightning AI	AI development platform	Credential access; previously affected by Mini-Shai Hulud supply chain campaign
16–17	OpenCode	AI coding tool	"Legacy system"; distillation motive; agent swarm indicators
20	Railway	PaaS	Privilege escalation in build infrastructure
23	Warp	AI terminal / coding tool	Third-party vendor dependency suspected; agent swarm indicators
28	CommandCode	AI coding tool	Distillation motive
30	Northflank	PaaS	Not yet disclosed

Of the three PaaS incidents, the two that came before Northflank have the most detailed public records, and both point to the same goal [6][7]. Railway's preliminary report, titled for the October 21 incident, describes attacker activity in its build infrastructure from October 20 to 22. The attacker escalated privileges, obtained encrypted environment variables, and decrypted a subset using compromised credentials, and Railway says the apparent objective was obtaining AI provider API keys for distillation attacks [6]. Railway estimated that at most 1.85% of workspaces (about 100,000) had environment variables potentially compromised. Out of caution, it asked upstream providers, including Anthropic, OpenAI, GitHub, and Stripe, to reset keys for more than 100,000 users [6]. Zeabur's report describes a deeper compromise. An attacker used leaked internal credentials that had been stored in plaintext, including cluster admin tokens with no expiry, and read about 3.2 million service environment variable records as well as account data for 321,810 accounts [7]. Zeabur observed that the attacker was mainly looking for directly usable API keys, especially for AI services, and tried to call Amazon Bedrock with Zeabur's own AWS keys, while cautioning that this is an observation of intent [7]. Zeabur also found that 602 of 607 checked n8n instances stored their encryption key as an environment variable, so third-party credentials held in those automation workflows were effectively exposed as well [7].

From LLMjacking to a Resale Market

Abusing stolen cloud credentials to consume hosted models is not new. Sysdig's Threat Research Team named the practice "LLMjacking" in May 2024 after observing attackers use stolen cloud credentials to reach models across multiple providers in order to sell access onward. Sysdig estimated that, in a worst case, this abuse could cost a victim more than US$46,000 per day in model consumption [8]. A September 2024 follow-up from Sysdig reported that the number of actors and techniques had grown, and that attackers had moved from abusing models already enabled on an account to enabling models themselves [9]. Black October differs from that earlier activity in where the credentials come from. Rather than harvesting keys one victim at a time from vulnerable or misconfigured hosts, attackers compromised the platforms where thousands of customers keep their keys.

Security Analysis

Why PaaS Secret Stores Are High-Value Targets

Environment variables are the standard way to inject configuration into containerized applications, and PaaS providers encourage their use. The weakness is that a PaaS has to be able to decrypt those values in order to inject them into running workloads. Encryption at rest therefore protects against storage-layer theft but not against an attacker who reaches the platform's build system, orchestration layer, or key-management path. Railway's report makes this concrete. The attacker obtained the variables in encrypted form and then decrypted a subset with credentials compromised inside the build infrastructure [6]. Zeabur's findings show a related failure, in which secrets that protect other secrets are stored alongside them. Cluster backups contained all secrets, and the backup access key sat next to them [7]. Northflank's instruction to assume full decryption [3] suggests the company could not rule out a comparable level of access, although the root cause has not been published.

AI provider keys stand out among the many secrets a PaaS holds because attackers can turn them into money immediately and with little friction. A stolen database password usually also requires network reachability and data-exfiltration effort. A stolen AI key, unless network-restricted, can typically be used from anywhere on the internet, produces a resellable service on every call, and bills the legitimate owner. Several platforms stated or reported a distillation motive, meaning large-scale querying of frontier models to train competing models [1][6]. That suggests some actors value the model outputs themselves as well as the free compute.

The Abuse Lifecycle Wiz Observed

Wiz's two Bedrock case studies give an unusually detailed view of what happens after the theft. Wiz did not say which compromised vendor either key came from, so the cases should not be attributed to Northflank specifically. Both keys had been used heavily from a compromised vendor's infrastructure, and both were first tested within 30 minutes of the vendor's reported credential exposure [1]. The table below summarizes the stages Wiz described.

Stage	Case A observations	Case B observations
Validation	Single InvokeModel call from a residential IP within 30 minutes of exposure, part of a one-minute sweep across multiple organizations (Wiz reports this pattern for both keys)	Single InvokeModel call from a residential IP within 30 minutes of exposure
Holding period	Low-volume "trickle" of 14–120 calls per day from one Go client, consistent with a key-verification job (Wiz reports this for both keys)	Key appeared in a reseller pool seven days after validation; thousands of failed calls over four hours from rotating hosting IPs
Capability mapping	Day-4 probe of the model catalog via a rotating proxy pool; Bedrock and EC2 quota enumeration with a boto3 script	Checker identified working `us.*` model IDs; quota enumeration with a boto3 script matching Case A's down to the version string
Monetization	~800,000 calls over ~30 hours, ~90% successful; Opus-class models about three quarters of successes	~1.7 million calls over ~13 hours from one address, more than 1 million successful; failures mostly regional throttling

Source: Wiz Research [1].

Several details in these cases are relevant to defenders, although they rest on two observations and should be treated as indicative rather than general rules. First, in both cases the delay between theft and validation was minutes, while the delay between validation and bulk abuse was days. If that pattern holds more broadly, it gives organizations a short window in which prompt rotation could prevent most of the financial damage. Second, quota enumeration preceded both floods, which makes quota-related API calls (such as Service Quotas or Bedrock limit queries) by an unfamiliar principal or address a promising

precursor signal worth alerting on. Third, key policy shaped how the abuse played out. Case B's key was restricted to `us.*` inference profiles, and Wiz attributes most of its failed calls to regional throttling, while Case A's key allowed `global.*` profiles and sustained a higher success rate [1]. The restriction did not prevent large-scale abuse, since Case B still produced more than one million successful calls, but it bounded throughput to regional quotas; combined with spend limits, that bounding can give defenders more time to detect and revoke. Finally, Wiz describes a token profile in which most calls carried few uncached input tokens, with a long tail of calls carrying tens of thousands of input tokens. Wiz interprets this as consistent with prompt caching and with many end users sharing access behind a reseller [1].

The Resale Layer

Wiz identified two operations that had access to the stolen keys. Infrastructure tied to an actor Wiz calls "momo" (23.106.46.239) is associated with an overt resale site advertising discounted frontier-model access paid in USDT. This infrastructure ran key checks against both case keys and attempted a `PutAccountDataRetention` call with the Case B key [1]. Wiz reads that call as a possible attempt to alter data-retention settings on the victim account. The same IP address appears among the representative attacker addresses Zeabur published [7], and Wiz notes that Zeabur named it as targeting Zeabur's infrastructure [1]. A second operation, "Xingjiabi" (38.246.114.118), runs a site offering discounted model access. It enumerated the inference profiles, regional model availability, and quotas of the Case A key, and it manually tested the Case B key [1]. These checks came after high-volume abuse from other addresses had already begun, which Wiz suggests shows an active secondary market in which keys change hands while they are still being exploited [1]. Wiz also captured a "distillation behaviour detection" page on the Xingjiabi site that scores its own users by risk. Wiz suggests this may be an attempt to keep resold keys from being suspended for distillation-pattern use [1].

The attribution limits deserve emphasis. Wiz states that it could not definitively link either reseller to the floods or to the platform compromises, and that its IP indicators are not directly linked to the Black October intrusions [1]. The evidence shows that stolen platform keys reach resale infrastructure. It does not establish that the resellers carried out the breaches. Financial losses circulating on social media, including cryptocurrency theft and consumer AI bills in the hundreds to thousands of dollars, are anecdotal and have not been verified [1].

Detection Gaps

Organizations that rely on PaaS-stored AI keys may lack telemetry that would reveal this activity. Teams that put AI keys into a PaaS environment variable often have no visibility into that platform's internals, and before October 31 Northflank's audit logs were not enabled for every account tier [2][3]. On the provider

side, LLMjacking actors have historically checked whether model-invocation logging was enabled in order to avoid detection [8], and later campaigns went further by deleting logging configuration [9]; in one Wiz case, reseller-linked infrastructure also attempted to change account data-retention settings [1]. Attackers also used rotating residential and hosting proxy pools across several autonomous systems (Wiz lists PEG TECH, Alibaba US, Scloud, Sharktech, and Level 3) [1], so blocklisting IP addresses is of limited value on its own. Given that proxy rotation, behavioral signals are likely to be more reliable than IP blocklists: invocations from addresses outside the platform's known egress ranges, sudden shifts in which models a key uses, quota enumeration, and spend spikes against a billing baseline.

Recommendations

Immediate Actions

Organizations that hosted workloads on Northflank, Railway, Zeabur, or any other platform named in Wiz's list during October should rotate every secret stored in that platform's environment variables, secret groups, and global secrets. They should start with AI provider keys and cloud credentials [2][6][7]. The essential requirement is that the old credential is actually revoked at the provider, because creating a new key without revoking the old one leaves the stolen key active [7]. Zeabur advises revoking first and then issuing a replacement [7], while Qovery recommends creating and deploying the replacement before revoking to avoid outages [4]. Where a key is confirmed to have been abused, immediate revocation should take priority over availability; otherwise, issuing, deploying, and revoking in quick succession is a reasonable compromise. Organizations should also disable public network access to managed database add-ons where it is enabled [2][6]. Zeabur's guidance is a reminder that keys rotated after an earlier incident must be rotated again if the platform was compromised again [7].

At the AI provider and cloud layer, teams should review usage and billing from October 1 onward for each affected key. For AWS, that means reviewing CloudTrail for `InvokeModel`, `Converse`, Service Quotas, and `PutAccountDataRetention` activity from unexpected principals or addresses. They should also check traffic against the Wiz-published indicators, including 23.106.46.239, 38.246.114.118, 104.243.32.50, and 31.57.201.201 [1]. These indicators relate to post-theft key abuse, and Wiz does not link them to the platform intrusions themselves, so a match confirms abuse of a key rather than the intrusion vector [1]. Confirmed abuse should be reported to the platform and to the AI provider, because providers can sometimes adjust charges [6][7]. Where workflow tools such as n8n ran on an affected platform, teams should rotate the tool's encryption key and every credential stored in it, not only the platform-level variables [7].

Short-Term Mitigations

Over the next 30 to 90 days, organizations should reduce how much damage any single stolen AI credential can do. They should issue a separate, narrowly scoped key for each service and environment, so that one platform breach does not expose keys used elsewhere. They should also apply hard spend limits and rate limits at the provider wherever these are supported, recognizing that alert-only budgets do not cap spending [10]. For cloud-hosted model access, IAM policies should restrict which models and inference profiles a credential can invoke and, where feasible, require that requests come from the platform's known egress addresses. Wiz's Case B shows that a profile restriction bounded abuse throughput to regional quotas even after the key was stolen, although it did not prevent large-scale abuse [1]. Where a platform supports workload identity federation or short-lived tokens, these should replace static keys.

Detection should be built for the abuse lifecycle described above. Useful alerts include a first call from a new autonomous system or residential range, invocation of models the application has never used, quota or limit enumeration, and spend above a rolling baseline. These alerts should go to an on-call function, not to a monthly finance review. Because, in the observed cases, validation came within minutes and bulk abuse within days, the response playbook should allow same-day revocation without a change-approval delay.

Strategic Considerations

The broader question for security leaders is whether a third-party PaaS should hold production AI credentials in a form the platform can decrypt. One option is to place an AI gateway or broker under the organization's own control between workloads and model providers, so that the PaaS holds only a scoped, revocable gateway token. Others are BYOC deployment models that keep secrets inside the organization's own cloud account, and external secrets managers that issue credentials at runtime. Each option moves the trust boundary but does not remove it. CSA's prior research on AI gateway compromise shows that a gateway concentrates provider keys and becomes a high-value target itself [11]. The aim is to choose deliberately where AI credentials live and to protect that location in proportion to the risk.

Vendor due diligence for PaaS and AI tooling providers should now cover secrets architecture explicitly. Questions should include how customer secrets are encrypted and where decryption keys live, whether build infrastructure is multi-tenant, how quickly the provider can notify customers and support mass rotation, and whether audit logs are available on every plan tier. Several October incidents involved build systems, internal credential hygiene, or third-party dependencies [1][6][7]. Assurance therefore needs to cover the provider's internal operations as well as its customer-facing controls.

CSA Resource Alignment

This note extends CSA's earlier analysis, *Inference Providers as the Prize: Platform-Chain Campaigns Against AI Serving* [12], which examined how attackers reach AI inference credentials through the intermediary platforms that store them, so that a single intrusion can expose keys for many accounts. That note recommended treating every third-party platform holding an AI key as part of the organization's attack surface and reducing each key's value through per-platform keys, spending caps, short-lived credentials, and runtime secret injection. Northflank's disclosure and Wiz's account of the resale layer add evidence for that position, showing what happens to platform-held keys once they leave the platform.

The Black October incidents can also usefully be viewed as a failure in governing non-human identities, on the platform side (internal credentials such as Zeabur's non-expiring cluster tokens) and on the customer side. CSA's Identity and Access Management Working Group guidance, *Defining Non-Human Identity* [13], treats API keys, service accounts, and machine tokens as identities that need lifecycle management, least privilege, and continuous monitoring under Zero Trust principles. Many PaaS-stored AI keys show the hallmarks of unmanaged NHIs: they are long-lived, often broadly scoped, held by a third party, and rarely monitored by their owners. Applying that guidance means keeping an inventory of where AI keys are stored, giving each key a single owner and purpose, and making rotation routine instead of a crisis-only exercise.

CSA's rapid research note *LLM Heist: Post-Compromise Abuse of LiteLLM Gateways* [11] examines a closely related attack in which compromised AI gateways expose the provider keys they hold. It recommends reconciling provider billing against expected usage as a detection signal that works even when logs are incomplete. That recommendation applies directly to the Bedrock floods Wiz documented. The note also bears on the strategic recommendation above, since moving keys from a PaaS into a self-hosted gateway concentrates risk unless the gateway's management plane is hardened. In the same vein, *Cloud Billing Hijacking as LLMjacking* [10] warns that alert-only budgets do not cap spending and recommends pairing billing alerts with enforced spend limits, which supports the spend controls recommended in this note.

The CSA research note *Poisoned AI IDE Plugins: Developer API Key Theft* [14] documented a 2026 JetBrains Marketplace campaign. It found that the operator may have resold stolen AI keys to paying users, so that victims bore the inference costs. Black October appears to show the same monetization model at larger scale, with platform-level compromise as the means of obtaining keys. Read together, the two notes indicate that attackers are targeting AI credentials at multiple layers where developers store them, from IDE plugins to deployment platforms.

Finally, the CSA *AI Controls Matrix (AICM) v1.1* [15] provides the control baseline for operationalizing these recommendations. Its Identity and Access Management controls govern credential scoping and rotation, its Cryptography, Encryption and Key Management controls cover protection of secrets and decryption keys, its Logging and Monitoring controls cover detection of anomalous model use, and its Supply Chain Management, Transparency and Accountability controls frame due diligence on PaaS and AI tooling providers. Organizations assessing hosting providers can use the AICM, together with STAR assurance artifacts where available, to structure the vendor questions outlined above.

References

[1] Benjamin Read and Rami McCarthy, Wiz Research. "“Black October” for AI Developer Platforms." Wiz Blog, October 30, 2026.

[2] Lawrence Abrams. "Northflank security breach leads to abuse of customers' AI API keys." BleepingComputer, October 31, 2026.

[3] Northflank. "Security incident." Northflank Status, October 30, 2026 (updated through November 2, 2026).

[4] Romaric Philogene, Qovery. "Northflank Security Incident (October 30, 2026): What Happened and How to Protect Your Secrets." Qovery Blog, October 30, 2026 (updated through November 1, 2026).

[5] Daniel Adeboye, Northflank. "What is BYOC (Bring Your Own Cloud) in cloud computing?." Northflank Blog, September 15, 2026.

[6] Railway. "Preliminary Report: Oct 21st Security Incident." Railway Blog, October 22, 2026 (updated October 26, 2026).

[7] Zeabur. "October 2026 Security Incident (Preliminary Report)." Zeabur Blog, October 19, 2026.

[8] Sysdig Threat Research Team. "LLMjacking: Stolen Cloud Credentials Used in New AI Attack." Sysdig, May 6, 2024.

[9] Sysdig Threat Research Team. "The Growing Dangers of LLMjacking: Evolving Tactics and Evading Sanctions." Sysdig, September 18, 2024.

[10] Cloud Security Alliance AI Safety Initiative. "Cloud Billing Hijacking as LLMjacking: Security Implications and Guidance." CSA Lab Space, October 21, 2026.

[11] Cloud Security Alliance AI Safety Initiative. "LLM Heist: Post-Compromise Abuse of LiteLLM Gateways." CSA Lab Space, August 20, 2026.

[12] Cloud Security Alliance AI Safety Initiative. "Inference Providers as the Prize: Platform-Chain Campaigns Against AI Serving." CSA Lab Space, October 23, 2026.

[13] Cloud Security Alliance Identity and Access Management Working Group. "Defining Non-Human Identity." Cloud Security Alliance, June 10, 2026.

[14] Cloud Security Alliance AI Safety Initiative. "Poisoned AI IDE Plugins: Developer API Key Theft." CSA Lab Space, June 18, 2026.

[15] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, June 22, 2026.