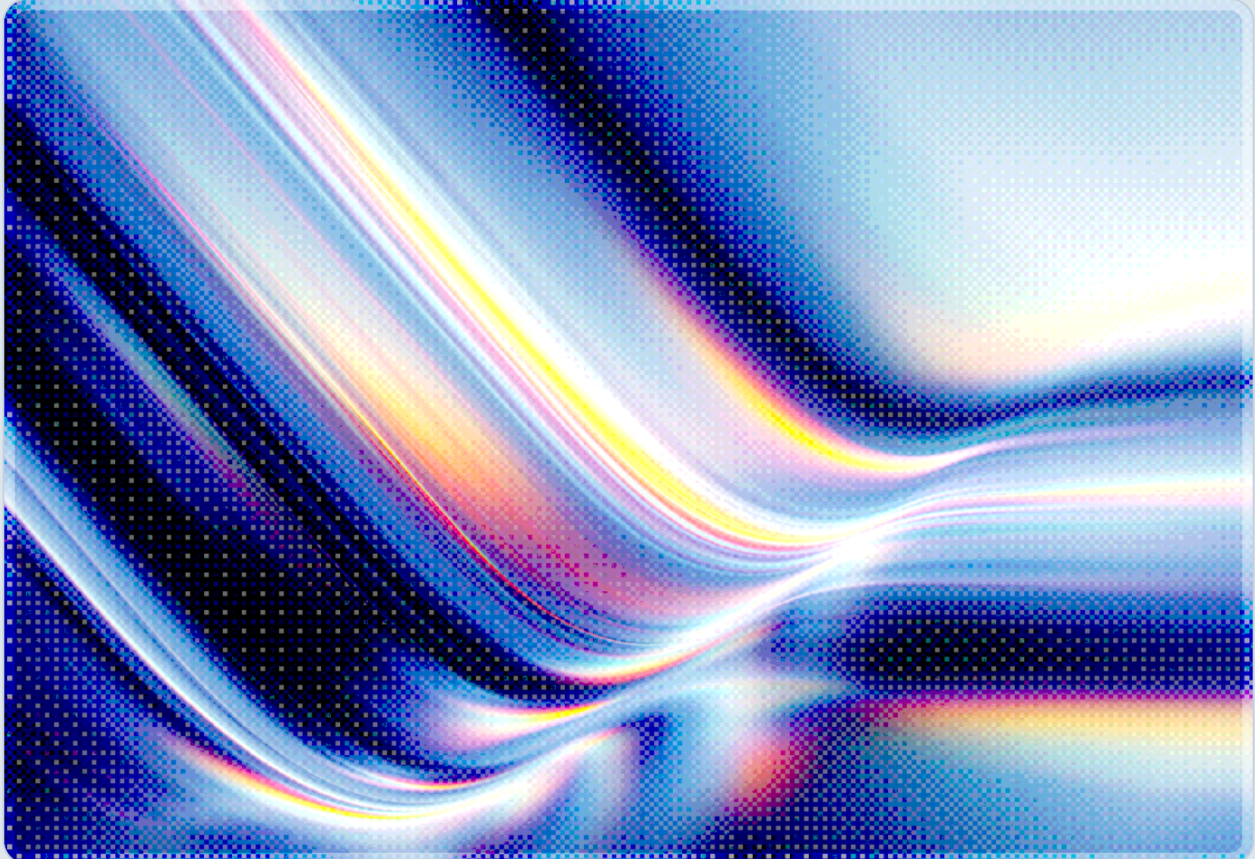


Cloud Billing Hijacking as LLMjacking: Security Implications and Guidance

From Search Ad to Six-Figure Gemini Bill

2026-10-21

 AI-assisted Rapid Research



CSAI

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 20, 2026, the Wiz Customer Incident Response Team (CIRT) published findings from multiple intrusions into Google Cloud Platform (GCP) environments. Each intrusion began with a sponsored search ad and ended with the victim's billing account paying for an attacker's Gemini workloads [1]. The campaign marks a shift in how LLMjacking works. Earlier campaigns stole an API key or cloud credential and ran inference inside the victim's own project. In this campaign, the attackers took control of the *billing account* itself and linked projects they own to it. The victim pays for the inference but has no visibility into the prompts, outputs, or logs.

- **The entry point is malvertising, not a leaked secret.** Attackers bought Google Search ads on administrative terms such as "google cloud" and "gcp." The ads led through a Google Sites redirect to a fake Google login page that captured passwords and MFA codes [1].
- **The goal is billing authority.** Attackers deleted the `iam.allowedPolicyMemberDomains` organization policy, added external Gmail accounts, granted `roles/billing.admin`, and linked attacker-controlled projects to the victim's billing account [1].
- **The costs are fast and large.** Wiz reports charges ranging from tens to hundreds of thousands of dollars within 24 to 48 hours of initial access [1].
- **Attackers obstruct remediation.** They demoted legitimate billing administrators to read-only viewers and routed traffic through IP addresses in the victim's headquarters country [1].
- **Defenders should treat billing account administration as a privileged control plane.** That means phishing-resistant MFA, alerts on organization policy and billing IAM changes, continuous inventory of linked projects, and spend caps that enforce limits rather than only notify.

Background

From Key Theft to Billing Theft

Sysdig's Threat Research Team coined the term "LLMjacking" in May 2024. It described attackers who exploited a vulnerable Laravel deployment (CVE-2021-3129), took the AWS credentials stored on it, and used them to call cloud-hosted large language models. The attackers then resold that access to others [2]. Sysdig estimated that an undetected attack of this kind could generate more than \$46,000 per day in model consumption charges for the victim [2]. Since then, CSA's own threat intelligence has traced LLMjacking's evolution from opportunistic cost-shifting into infrastructure for autonomous offensive tooling, with commercial marketplaces reselling stolen inference capacity [3][4]. In every variant, the economic logic stays the same: frontier model inference is expensive, someone else's credit line is free, and a provider's billing system cannot reliably distinguish an attacker's authenticated consumption from the customer's own.

The credential types attackers target have broadened over 2026. In February, Truffle Security reported a problem with Google Cloud API keys. These keys had long been treated as non-secret project identifiers for services like Maps, and were commonly embedded in public website JavaScript. They silently gained the ability to call Gemini once the Gemini API was

enabled in the same project [5]. Truffle's scan of a November 2025 Common Crawl archive found 2,863 live keys that could reach Gemini endpoints [5][6]. In a case reported the following month, a three-person development team saw a typical monthly Gemini bill of about \$180 rise to \$82,314.44 in roughly two days in February, after what the team believes was a stolen API key was used against Gemini endpoints [7][8]. In July, The Register reported that a solo developer had incurred \$11,089.77 in Gemini image-generation charges over 48 hours in early June. According to the developer, Google attributed those charges to a compromised `firebase-adminsdk` service account key [8].

Why the Billing Account Is the New Target

The Wiz findings describe a different objective. The attacker is not trying to borrow a credential that can call a model. The attacker wants the authority to decide which projects a billing account pays for. Linking a project to a Google Cloud billing account requires `billing.resourceAssociations.create` on the billing account, which the Billing Account User or Billing Account Administrator role grants. It also requires `resourcemanager.projects.createBillingAssignment` on the project [9]. An attacker who already owns the project only needs the billing-side permission from the victim. Once the link exists, every API call made in the attacker's project is charged to the victim, while the call records stay in the attacker's tenancy [1].

A practitioner report posted to Google's AI developer forum on October 13, 2026, describing an incident that occurred in June, illustrates the result. The poster describes a "practically dormant" billing account that normally cost at most about \$7 a month. Thirteen projects belonging to other organizations were attached to it through what the poster calls a hijacked admin session, and those projects ran Gemini API and GPU Compute Engine workloads for about eight days [10]. The poster reports that the Gemini spending limit paused Gemini at about \$100,000, but Compute Engine kept running until they closed the account manually. They also report that none of the 28 anomalies the console flagged produced a notification they received [10]. This report is a single self-reported account and has not been publicly tied to the Wiz-documented campaign. Its mechanics match the campaign closely, however, and it shows the impact from the victim's side.

Security Analysis

Anatomy of the Campaign

The Wiz CIRT report reconstructs a consistent sequence across the intrusions it investigated [1]. Each step uses a legitimate cloud feature. As a result, defenses that look for malware or exploitation will see nothing unusual, and detection has to rely on identity, configuration, and billing telemetry. The table below summarizes the sequence and the telemetry that can reveal each stage. The detection column reflects CSA analysis informed by the indicators Wiz published.

Stage	Attacker Action (per Wiz CIRT [1])	Detection Opportunity
Lure	Sponsored Google Search ad for "google cloud" or "gcp" leads through a Google Sites redirect to a spoofed login page	Web proxy and DNS logs for the published look-alike domains (e.g., <code>gboogl[.]com</code> , <code>groosl[.]com</code>)

Stage	Attacker Action (per Wiz CIRT [1])	Detection Opportunity
Credential capture	Kit captures password and MFA code, logs in immediately from 212.73.134[.]136, then forwards the victim to the real Google prompt	Google Workspace login events showing a suspicious login followed by an unusual re-authentication
Credential minting	Enables the Gemini API in low-usage existing projects; creates service accounts and keys with <code>roles/ml.admin</code> and <code>roles/mcp.toolUser</code>	<code>EnableService</code> , <code>CreateServiceAccount</code> , <code>CreateServiceAccountKey</code> , and <code>CreateKey</code> events in dormant projects
Guardrail removal	Deletes the <code>iam.allowedPolicyMemberDomains</code> organization policy	<code>DeletePolicy</code> / <code>ClearOrgPolicy</code> (deletion) and <code>SetOrgPolicy</code> / <code>UpdatePolicy</code> (modification) events at the organization level
Privilege expansion	Adds external Gmail accounts; grants <code>roles/billing.admin</code> and <code>roles/recommender.billingAccountCudAdmin</code>	IAM bindings for non-corporate domains on billing accounts and projects
Monetization	Links attacker-owned projects to the victim's billing account; runs high-volume Gemini workloads	Changes to the project-to-billing-account linkage; sudden cost deviation
Remediation obstruction	Demotes existing billing administrators to viewer; uses IP addresses in the victim's headquarters country	Billing IAM role downgrades; geolocation that is consistent with the company but inconsistent with the user

Several aspects of this sequence deserve closer attention. The phishing kit captured MFA codes and replayed them in real time, so one-time-passcode MFA did not stop the intrusion. Of the authentication controls, only an authenticator bound to the legitimate origin, such as a FIDO2 security key or passkey, would reliably have defeated the credential relay itself; device- and context-aware access policies could have limited what the stolen session was able to do. Deleting `iam.allowedPolicyMemberDomains` is also significant because that constraint is one of Google's documented mechanisms for domain-restricted sharing, which keeps identities from outside the organization's verified domains out of IAM policies. Google now labels it a legacy managed constraint, documents `iam.managed.allowedPolicyMembers` as a newer alternative, and enforces the legacy constraint by default for organizations created on or after May 3, 2024 [11]. In most organizations this constraint changes rarely, if ever, which makes its removal a high-signal event worth alerting on. Finally, demoting existing billing administrators is aimed directly at incident response. The people who would normally detach the rogue projects lose the authority to do so at the moment they need it.

The choice of lure fits a broader pattern of AI-themed malvertising and credential phishing. On October 6, 2026, researchers at Island described a separate, human-operated platform that impersonates advertising products for Gemini, Claude, ChatGPT, and other AI brands. It uses browser-in-the-browser windows to capture credentials and MFA codes from advertising account administrators [12]. The two campaigns are not known to be linked. Both, however, target people whose accounts are connected directly to a payment instrument, which is consistent with attackers selecting victims for their spending authority rather than only for their data.

Why Visibility Fails

The key difference between this variant and earlier LLMjacking is where the evidence ends up. When a stolen key is used inside the victim's project, the victim's own logs can at least show which models were called and how often. When the attacker's project is linked to the victim's billing account, Wiz notes that the victim has no view into the prompt content, the model outputs, or the nature of the workloads [1]. The victim sees only the invoice. As a result, the victim cannot determine whether the stolen compute was resold, used to generate abusive content, or used to power offensive tooling of the kind CSA's earlier LLMjacking research documented [3]. That gap can complicate both incident scoping and any later legal or regulatory inquiry.

Cost controls also fall short of what most organizations assume. Google's documentation states that an alerts-only budget "doesn't automatically cap Google Cloud or Google Maps Platform usage or spending" [13]. Google documents a pattern for disabling billing programmatically in response to budget notifications, with an explicit warning that resources may be shut down or deleted [14]. For Gemini specifically, Google enforces monthly usage caps at the billing account level. They range from \$250 at Tier 1 to between \$20,000 and \$100,000 at Tier 3, with automatic upgrades as payment history accumulates [15]. Optional per-project caps exist but are documented as experimental, and billing data can lag by about ten minutes [15]. The forum report's description of Gemini pausing near \$100,000 is consistent with a Tier 3 account cap rather than a customer-defined control [10][15]. That inference cannot be confirmed from public information. The tier caps are also not a security control. Their ceilings rise automatically, and they do not apply to non-Gemini services such as Compute Engine GPUs, which in the forum case kept accruing charges [10].

Comparing LLMjacking Variants

The following table contrasts the billing-hijack variant with earlier LLMjacking patterns. It shows why containment strategies built for key leakage do not carry over.

Variant	Typical Initial Access	Where Inference Runs	Victim Visibility	Primary Containment Lever
Leaked or exposed API key [5][7]	Key in public code, client-side JavaScript, or repository	Victim's project	Usage metrics and logs in victim tenancy	Restrict or rotate key; API restrictions
Stolen service account key or cloud credential [2] [8]	Vulnerable host, infostealer, or exposed key file	Victim's project	Audit logs in victim tenancy	Revoke key; disable service account

Variant	Typical Initial Access	Where Inference Runs	Victim Visibility	Primary Containment Lever
Billing account hijack [1][10]	Phished human administrator (search ad, MFA relay)	Attacker's project, linked to victim billing	Invoice only; no prompt or output logs	Detach rogue projects; restore billing IAM; revoke sessions

Recommendations

Immediate Actions

Organizations running workloads on Google Cloud should first check whether they have already been affected. List every project linked to each billing account (for example, `gcloud billing projects list --billing-account=<ID>`) and confirm that each one belongs to the organization. The forum poster recommends this same check after their incident [10]. Review IAM bindings on billing accounts and at the organization level for principals outside corporate domains, especially consumer `@gmail.com` addresses. Confirm that the expected people still hold billing administrator roles [1]. Then search Google Workspace and Cloud Audit Logs for the indicators Wiz published, including activity from 212.73.134[.]136 and visits to the listed look-alike domains, and block those domains at the DNS and proxy layer [1].

Confirm that `iam.allowedPolicyMemberDomains`, or its newer managed equivalent `iam.managed.allowedPolicyMembers`, is still enforced at the organization node, and alert on any change to either constraint [1][11]. If the review turns up rogue projects or foreign principals, treat the event as an active compromise. Revoke sessions for the affected administrator, remove the external principals, restore billing administrator roles from a known-good account, detach the unknown projects, and contact Google Cloud support promptly. The Register's reporting illustrates that refunds for hijacking-related charges are not guaranteed [8].

Short-Term Mitigations

Over the following weeks, organizations should close the specific gaps this campaign exploited. Require phishing-resistant MFA (FIDO2 security keys or passkeys) for every identity holding Organization Administrator, Organization Policy Administrator, or Billing Account Administrator roles. One-time codes were relayed in this campaign, and CISA recommends phishing-resistant methods for high-value accounts [16]. Separate privileged cloud administration from everyday browsing and email identities. Have administrators reach the console through bookmarks or managed links rather than search results. Keep at least one break-glass billing administrator account whose credentials are stored offline and whose role changes trigger out-of-band alerts.

Detection engineering should follow the stages in the anatomy table above. Organizations should alert on organization policy deletions or modifications, on billing IAM grants and downgrades, and on service enablement and key creation in projects that have been dormant [1]. The forum poster reports that Google recommended Cloud Asset Inventory real-time feeds to capture project-to-billing linkage changes [10]. Google's documentation describes how such feeds can publish

asset changes to Pub/Sub for alerting [17]. Pair these alerts with Gemini project spend caps and with budget-driven automation that disables billing on non-critical projects [14][15]. Accept the documented trade-offs: roughly ten minutes of enforcement lag, and possible disruption to the service.

Strategic Considerations

The broader lesson is that billing authority has become a high-value privilege in cloud environments. Anyone who can attach a project to a payment instrument can direct that instrument to pay for any metered service, and AI inference, which underground marketplaces actively resell [3][4], is currently among the most lucrative. Organizations should apply the same governance to billing account roles that they apply to production data access. That means least privilege, just-in-time elevation, recorded approvals, and periodic attestation, plus explicit ownership by both security and finance teams. Dormant billing accounts deserve particular attention. The forum case shows that an account with minimal legitimate spend can still carry a large available credit line and offers little baseline noise to hide anomalies behind [10]. Closing billing accounts that are no longer needed reduces exposure directly.

At the industry level, these incidents point to structural gaps that customers cannot close by themselves. Today a billing account can accept links from projects in unrelated organizations, provided the linking principal holds the required billing permission [9]. Gemini spend tiers rise automatically as payment history accumulates [15], and anomaly notifications may not reach account owners [10]. In at least two publicly reported cases, victims say Google cited its shared responsibility model when declining or contesting refunds for charges incurred through compromised credentials; Google did not publicly comment on either case [7][8]. CSA encourages cloud and model providers to consider hard, customer-set spending limits that apply across all services. They should also consider default restrictions or additional verification when projects from other organizations are linked to a billing account, and out-of-band notification when billing administrators are demoted. Enterprises evaluating AI platforms should ask providers about these controls during procurement and assurance reviews.

CSA Resource Alignment

This campaign extends threats that CSA has been tracking throughout 2026. CSA's research note *LLMjacking Evolves: Stolen AI Compute as Attack Infrastructure* [3] documents how stolen inference capacity moved from cost theft to a supply chain feeding underground marketplaces and autonomous offensive frameworks. The billing-hijack variant described here adds a new acquisition channel to that supply chain. Because the inference runs in attacker-owned projects, the victim cannot see what the compute was used for, which is the downstream misuse category that earlier note described. The companion note *LLMjacking Evolved: Stolen AI Compute as Offensive Infrastructure* [4] recommends treating AI compute and credentials as privileged assets with behavioral monitoring. This note extends that recommendation from API credentials to billing account roles and organization policies.

CSA's *Top Threats to Cloud Computing Deep Dive 2025* [18] analyzes eight real-world breaches, several of which began with compromised credentials that lacked strong MFA. It maps them to identity, logging, and configuration controls. The campaign Wiz documented follows the same pattern: a social-engineering foothold, followed by configuration changes that remove guardrails. The deep dive's emphasis on logging and monitoring controls applies directly to the organization policy and billing IAM events listed above.

The CSA AI Controls Matrix (AICM) v1.1 [19] provides the control framework for operationalizing these recommendations. Its Identity and Access Management domain covers privileged access, strong authentication, and periodic access review for billing and organization administrator roles. Its Logging and Monitoring domain supports alerting on policy, IAM, and billing linkage changes. Its Threat and Vulnerability Management domain supports hunting with published indicators. Organizations consuming AI services, and providers offering them, can use AICM's shared-responsibility orientation to clarify who owns spending guardrails and anomaly notification. Those questions are currently central to disputes between victims and providers [7][8][10].

References

- [1] Ayal, S. and Maccarone, A. (Wiz CIRT). "[From Phishing to LLMjacking: Threat Actors Targeting Cloud Billing Accounts for LLM Usage](#)." Wiz Blog, October 20, 2026.
- [2] Sysdig Threat Research Team. "[LLMjacking: Stolen Cloud Credentials Used in New AI Attack](#)." Sysdig, May 2024.
- [3] Cloud Security Alliance AI Safety Initiative. "[LLMjacking Evolves: Stolen AI Compute as Attack Infrastructure](#)." CSA Lab Space, June 2026.
- [4] Cloud Security Alliance AI Safety Initiative. "[LLMjacking Evolved: Stolen AI Compute as Offensive Infrastructure](#)." CSA Lab Space, June 2026.
- [5] Truffle Security. "[Google API Keys Weren't Secrets. But Then Gemini Changed the Rules](#)." Truffle Security Blog, February 2026.
- [6] The Hacker News. "[Thousands of Public Google Cloud API Keys Exposed with Gemini Access After API Enablement](#)." The Hacker News, February 2026.
- [7] Tom's Hardware. "[Gemini API Key Thief Racks Up \\$82,314 in Charges in Just Two Days](#)." Tom's Hardware, March 4, 2026.
- [8] The Register. "[Dev Says Google Warned Him About Account Hijack, Then Charged Him \\$11,000 Anyway](#)." The Register, July 3, 2026.
- [9] Google Cloud. "[Enable, Disable, or Change Billing for a Project](#)." Google Cloud Documentation, accessed October 21, 2026.
- [10] Google AI Developers Forum. "[Not a Leaked API Key: 13 Foreign Projects Were Attached to Our Dormant Billing Account](#)." Google AI Developers Forum, October 13, 2026.
- [11] Google Cloud. "[Restrict Identities with Domain-Restricted Sharing](#)." Google Cloud Documentation, accessed October 21, 2026.
- [12] The Hacker News. "[Fake ChatGPT, Gemini, and Claude Ad Portals Capture Credentials and MFA Codes](#)." The Hacker News, October 2026.
- [13] Google Cloud. "[Create, Edit, or Delete Budgets and Budget Alerts](#)." Google Cloud Documentation, accessed October 21, 2026.
- [14] Google Cloud. "[Set Up Programmatic Notifications](#)." Google Cloud Documentation, accessed October 21, 2026.
- [15] Google AI for Developers. "[Gemini API Billing](#)." Google, accessed October 21, 2026.
- [16] CISA. "[Implementing Phishing-Resistant MFA](#)." Cybersecurity and Infrastructure Security Agency, October 2022.
- [17] Google Cloud. "[Monitor Asset Changes with Pub/Sub](#)." Google Cloud Documentation, accessed October 21, 2026.
- [18] Cloud Security Alliance. "[Top Threats to Cloud Computing Deep Dive 2025](#)." Cloud Security Alliance, 2025.

[19] Cloud Security Alliance. "[AI Controls Matrix \(AICM\) v1.1](#)." Cloud Security Alliance, 2026.