INTELLIGENCE BRIEFING
Security Command Center
TLP:CLEAR
2026-03-31 13:28 UTC

Google Cloud Vertex AI Default Permissions Create Credential Theft and Supply Chain Exposure Path

SECURITY ANALYSIS | HIGH | CVSS 7.5

SCC Item ID	SCC-STY-2026-0040
Type	Security Analysis
Severity	HIGH
CVSS Base Score	7.5
Affected Products	Google Cloud Vertex AI, Vertex AI Agent Development Kit (ADK), Agent Engine, Google Cloud Storage, Google Cloud Artifact Registry
Published	2026-03-31T09:09:00
Discovery Source	Rss

Executive Summary

Palo Alto Networks Unit 42 disclosed that Google Cloud Vertex AI Agent Engine grants its default platform-managed service account excessive OAuth scopes, allowing a compromised or malicious AI agent to steal credentials, access all Cloud Storage buckets in the project, and read internal Artifact Registry container images. The exposure affects any organization running AI agents on Agent Engine without a custom service account, and it remains active until teams replace the default configuration with Bring Your Own Service Account (BYOSA). The finding signals a broader risk pattern: as enterprises adopt managed AI platforms, default-permissive infrastructure configurations are becoming viable attack surfaces that traditional security programs have not yet been built to audit.

Technical Analysis

Unit 42's research, published under the title 'Double Agents: Exposing Security Blind Spots in GCP Vertex AI,' identifies a privilege misconfiguration in Google Cloud's Vertex AI Agent Engine. The platform assigns each project a Per-Project, Per-Product Service Agent (P4SA) that runs AI agents by default. This P4SA carries OAuth scopes broader than required for agent execution, creating three distinct exploitation paths once an agent is compromised or acts maliciously.

First, an agent can query the GCP metadata server to retrieve the P4SA's OAuth token (T1552.001, Credentials in Files, mapped here to metadata endpoint abuse; T1078.004, Cloud Accounts). With that token, the agent can authenticate as the service account to downstream GCP APIs. Second, the excessive Storage scope permits

the agent to enumerate and read all Cloud Storage buckets in the project (T1530, Data from Cloud Storage). Depending on what teams store in those buckets - model artifacts, training data, application secrets, customer data - this path can produce significant secondary exposure. Third, the agent can read Artifact Registry repositories that include Google-internal container images not intended for customer access (T1195.002, Compromise Software Supply Chain). This last path is the most strategically significant: access to internal base images could enable an attacker to identify undisclosed vulnerabilities in Google-managed containers or craft a poisoned image that re-enters the supply chain.

The root cause maps cleanly to CWE-250 (Execution with Unnecessary Privileges) and CWE-732 (Incorrect Permission Assignment for Critical Resource). The credential exposure path additionally maps to CWE-522 (Insufficiently Protected Credentials). MITRE ATT&CK techniques observed across the full attack surface include T1548 (Abuse Elevation Control Mechanism), T1199 (Trusted Relationship), T1530, T1552.001, T1195.002, and T1078.004.

No active exploitation has been reported, and Google has updated its security controls documentation for Vertex AI. The recommended remediation is BYOSA, replacing the default P4SA with a customer-managed service account scoped to least privilege. Deployments that have not adopted BYOSA remain exposed; the fix is not applied automatically to existing environments.

The broader implication for security teams is significant. Managed AI platforms abstract infrastructure from users, which is operationally convenient but obscures the permission surface. The default-permissive P4SA is not unusual behavior - GCP, AWS, and Azure all ship managed services with default roles that frequently exceed least privilege. What makes this case distinctive is the agent execution context: AI agents are designed to take autonomous actions against external APIs and data sources, which means a misconfigured agent is not a passive credential holder but an active participant capable of chaining these exposures in a single run. Security teams that have not extended their IAM audit processes to cover AI agent service identities have a visibility gap that this disclosure makes concrete.

Action Checklist

1. Step 1: Assess exposure - determine whether your organization runs workloads on Google Cloud Vertex AI Agent Engine; identify all projects where Agent Engine is active and confirm whether those deployments use the default P4SA or a custom BYOSA configuration.
2. Step 2: Review controls - audit the OAuth scopes and IAM roles assigned to the P4SA in each affected project; compare granted permissions against the principle of least privilege using Google Cloud's IAM Recommender; review Google's updated Vertex AI security controls documentation at the source URL provided.
3. Step 3: Remediate - replace the default P4SA with a Bring Your Own Service Account (BYOSA) configured with only the permissions required for your specific agent workloads; validate that the new service account cannot access Cloud Storage buckets or Artifact Registry repositories outside its defined scope.
4. Step 4: Update threat model - add AI agent service identities as a distinct identity class in your threat register; map the attack paths documented here (metadata credential theft, over-scoped storage access, supply chain registry access) to your detection and response playbooks.
5. Step 5: Communicate findings - brief leadership on the specific risk: AI platform defaults created a credential theft and supply chain exposure path in active deployments; frame the remediation as a required configuration change, not an optional hardening step, and track completion status across all

affected projects.

6. Step 6: Monitor developments - track the GitHub ADK discussion thread and Google Cloud documentation for additional control updates; watch for follow-up Unit 42 disclosures, CVE assignment, or CISA guidance related to AI platform default permission patterns.

IR / Forensic Enrichment

Triage Priority	URGENT
Escalation Criteria	Escalate to immediate response and engage legal/privacy counsel if Cloud Audit Logs (data_access tier, storage.googleapis.com or artifactregistry.googleapis.com) show the P4SA accessed GCS buckets containing PII, PHI, or regulated data, or pulled Artifact Registry images outside normal deployment windows — either condition indicates the over-scoped credential was actively exploited and may trigger breach notification obligations under GDPR, HIPAA, or applicable state privacy law.
Recovery Notes	After BYOSA migration is confirmed in all affected projects, re-run `gcloud logging read` queries filtering on the old P4SA email for 72 hours post-remediation to confirm it generates zero new activity — any residual access indicates the BYOSA swap was incomplete or a secondary binding was missed. Re-run IAM Recommender for all affected projects 30 days post-remediation to catch any permission creep introduced during the BYOSA configuration. Maintain heightened Cloud Audit Log monitoring on Artifact Registry pull events and GCS bucket list operations for 60 days, specifically watching for T1078.004 (Valid Accounts: Cloud Accounts) patterns where any service account with an `aiplatform-` prefix accesses resources outside its declared operational scope.
Forensic Artifacts	Google Cloud Audit Logs — data_access tier, filtered on protoPayload.authenticationInfo.principalEmail matching the P4SA email (format: service-[PROJECT_NUMBER]@gcp-sa-aiplatform.iam.gserviceaccount.com) for storage.googleapis.com and artifactregistry.googleapis.com — reveals whether the over-scoped credential was used to enumerate or read GCS objects or pull internal container images | GCP Metadata Server access logs — if the Agent Engine workload runtime was compromised, an attacker would have issued an HTTP GET to http://metadata.google.internal/computeMetadata/v1/instance/service-accounts/default/token; look for this request pattern in Agent Engine container stdout/stderr logs or VPC Flow Logs showing outbound connections to 169.254.169.254 from the agent workload | Cloud Audit Logs — admin_activity tier for IAM policy changes — specifically SetIamPolicy events on GCS buckets or Artifact Registry repositories occurring after Agent Engine job execution, which would indicate the P4SA token was used to modify access controls as a persistence or exfiltration staging step | Artifact Registry pull logs — `gcloud artifacts docker images list [REGION]-docker.pkg.dev/[PROJECT_ID]/[REPO] --include-tags --format=json` combined with data_access audit log entries for method `artifactregistry.repositories.downloadArtifacts` — identifies whether internal container images were pulled by the P4SA outside of sanctioned CI/CD pipeline service accounts, indicating supply chain reconnaissance | GCS bucket access logs (if bucket-level logging was enabled) — object listing and read events (storage.objects.list, storage.objects.get) attributed to the P4SA across all buckets in the project — quantifies the data exposure blast radius and identifies which buckets and objects were accessible during the exposure window for breach notification scope determination

Per-Action IR Details

Step 1: Assess exposure — determine whether your organization runs workloads on Google Cloud Vertex AI Agent Engine; identify all projects where Agent Engine is active and confirm whether those deployments use the default P4SA or a custom BYOSA configuration

NIST Phase: Detection Analysis

Reference: NIST 800-61r3 §3.2 — Detection & Analysis: scoping the affected asset inventory and confirming whether the vulnerable default configuration (P4SA with over-scoped OAuth) is present in active deployments

Controls: NIST IR-4 (Incident Handling), NIST IR-5 (Incident Monitoring), NIST RA-3 (Risk Assessment) — scoping which projects carry the over-scoped P4SA exposure, CIS 1.1 (Establish and Maintain Detailed Enterprise Asset Inventory), CIS 5.1 (Establish and Maintain an Inventory of Accounts) — service account identities including P4SA must be inventoried

Compensating: Run the following gcloud command across all projects to enumerate Agent Engine service accounts and their OAuth scopes: `gcloud projects list --format='value(projectId)' | xargs -I{} gcloud iam service-accounts list --project={} --filter='email:aiplatform' --format='table(email,displayName)'`. Follow with `gcloud iam service-accounts get-iam-policy [P4SA_EMAIL] --project=[PROJECT_ID]` for each identified P4SA. A 2-person team can script this loop in bash and pipe output to a CSV for review. No SIEM required.

Evidence: Before scoping, capture a snapshot of current IAM state: export `gcloud projects get-iam-policy [PROJECT_ID] --format=json` for each project and preserve the output as a timestamped baseline. Also pull Cloud Audit Logs — specifically `cloudaudit.googleapis.com/activity` log entries — filtering on `protoPayload.serviceName = 'aiplatform.googleapis.com'` to establish which Agent Engine jobs have run under the default P4SA and what resources they accessed prior to your assessment.

Step 2: Review controls — audit the OAuth scopes and IAM roles assigned to the P4SA in each affected project; compare granted permissions against the principle of least privilege using Google Cloud's IAM Recommender; review Google's updated Vertex AI security controls documentation at the source URL provided

NIST Phase: Detection Analysis

Reference: NIST 800-61r3 §3.2 — Detection & Analysis: determining scope of potential credential exposure by auditing what the over-scoped P4SA can access — specifically Cloud Storage and Artifact Registry — before declaring or closing the incident

Controls: NIST IR-4 (Incident Handling), NIST AU-6 (Audit Record Review, Analysis, and Reporting) — review Cloud Audit Logs for P4SA access to GCS buckets and Artifact Registry, NIST AC-6 (Least Privilege) — validate that P4SA OAuth scopes exceed minimum required permissions, CIS 7.1 (Establish and Maintain a Vulnerability Management Process), CIS 5.1 (Establish and Maintain an Inventory of Accounts)

Compensating: Use the GCP IAM Recommender directly in the Console (no cost) at IAM & Admin > IAM Recommender, filtered to the P4SA email. Additionally, run `gcloud asset search-all-iam-policies --scope=projects/[PROJECT_ID] --query='policy:serviceAccount:[P4SA_EMAIL]' --format=json` to enumerate every resource the P4SA has been bound to. To check OAuth scopes specifically, inspect the Agent Engine resource configuration: `gcloud ai agent-pools list --region=[REGION] --project=[PROJECT_ID] --format=json` and review the `serviceAccount` field.

Evidence: Pull Cloud Audit Logs from `data_access` log type for `storage.googleapis.com` and `artifactregistry.googleapis.com` filtered to `authenticationInfo.principalEmail` matching the P4SA email — this reveals whether the P4SA has already accessed GCS buckets or pulled container images from Artifact Registry. Retain these logs before any remediation as they establish whether exploitation has already occurred. Also capture `gcloud logging read 'protoPayload.authenticationInfo.principalEmail=[P4SA_EMAIL]' --project=[PROJECT_ID] --format=json` output as a forensic baseline.

Step 3: Remediate — replace the default P4SA with a Bring Your Own Service Account (BYOSA) configured with only the permissions required for your specific agent workloads; validate that the new service account cannot access Cloud Storage buckets or Artifact Registry repositories outside its defined scope

NIST Phase: Eradication

Reference: NIST 800-61r3 §3.4 — Eradication: removing the over-privileged P4SA configuration from Agent Engine deployments and replacing it with a least-privilege BYOSA to eliminate the credential theft and supply chain access path identified by Unit 42

Controls: NIST SI-2 (Flaw Remediation) — configuration flaw remediation by replacing default over-scoped identity, NIST AC-6 (Least Privilege) — BYOSA must be scoped to minimum permissions required for the specific agent workload, NIST CM-6 (Configuration Settings) — enforce secure configuration baseline for Agent Engine service account binding, CIS 4.6 (Securely Manage Enterprise Assets and Software), CIS 7.2 (Establish and Maintain a Remediation Process)

Compensating: Create the BYOSA with: `gcloud iam service-accounts create vertex-agent-byosa --display-name='Vertex Agent BYOSA' --project=[PROJECT_ID]`. Grant only the minimum required role (e.g., `roles/aiplatform.user`) and explicitly do NOT grant `roles/storage.admin` or `roles/artifactregistry.reader` unless required. Validate scope by running `gcloud iam service-accounts get-iam-policy vertex-agent-byosa@[PROJECT_ID].iam.gserviceaccount.com` and confirming no storage or registry bindings exist. After BYOSA assignment, verify the old P4SA is no longer referenced in active Agent Engine deployments.

Evidence: Before revoking the P4SA, capture a final export of all GCS bucket-level IAM policies (`gsutil iam get gs://[BUCKET_NAME]` for each bucket in the project) and Artifact Registry IAM policies (`gcloud artifacts repositories get-iam-policy [REPO] --location=[REGION] --project=[PROJECT_ID] --format=json`) to document what the P4SA had access to. This establishes the blast radius for any post-incident breach notification assessment. Retain these exports with timestamps as forensic evidence.

Step 4: Update threat model — add AI agent service identities as a distinct identity class in your threat register; map the attack paths documented here (metadata credential theft, over-scoped storage access, supply chain registry access) to your detection and response playbooks

NIST Phase: Post Incident

Reference: NIST 800-61r3 §4 — Post-Incident Activity: updating organizational threat models and detection playbooks based on lessons learned from the Unit 42 disclosure of AI platform default permission abuse patterns

Controls: NIST IR-8 (Incident Response Plan) — update IR plan to include AI agent identity compromise scenarios, NIST RA-3 (Risk Assessment) — incorporate AI service account over-privilege as a distinct risk category, NIST PM-16 (Threat Awareness Program) — integrate Unit 42 findings into threat awareness documentation, CIS 7.1 (Establish and Maintain a Vulnerability Management Process)

Compensating: Document three specific MITRE ATT&CK technique mappings in your playbook: T1552.001 (Credentials in Files — via GCP metadata server token theft from the Agent Engine runtime), T1078.004 (Valid Accounts: Cloud Accounts — misuse of the P4SA OAuth token to authenticate as a legitimate identity), and T1195.002 (Supply Chain Compromise: Compromise Software Supply Chain — unauthorized Artifact Registry image read enabling poisoned dependency injection). For free detection: create a Sigma rule targeting Cloud Audit Logs for `storage.objects.list` or `artifactregistry.repositories.downloadArtifacts` events from any AI platform service account (`aiplatform-` prefix) outside expected operational hours or against unexpected buckets.

Evidence: As input to the threat model update, retrieve the Unit 42 research disclosure metadata (technique IDs, attack path diagrams) and cross-reference against your Cloud Audit Log export from Steps 1-2 to determine whether any of the three attack paths (metadata credential access, GCS enumeration, Artifact Registry pull) appear in historical logs. This retrospective log review determines whether the threat was theoretical or already executed in your environment.

Step 5: Communicate findings — brief leadership on the specific risk: AI platform defaults created a credential theft and supply chain exposure path in active deployments; frame the remediation as a required configuration change, not an optional hardening step, and track completion status across all affected projects

NIST Phase: Post Incident

Reference: NIST 800-61r3 §4 — Post-Incident Activity: communicating incident findings and remediation status to leadership as part of the lessons-learned and organizational improvement cycle; also satisfies RS.MA-01 coordination requirements

Controls: NIST IR-6 (Incident Reporting) — report findings and remediation status to appropriate organizational personnel, NIST IR-8 (Incident Response Plan) — update plan and communicate status against completion criteria, NIST PM-2 (Information Security Program Leadership Role) — engage senior leadership on AI platform risk posture, CIS 7.2 (Establish and Maintain a Remediation Process)

Compensating: Produce a one-page status report using the output of `gcloud projects list | xargs -I{} gcloud ai agent-pools list --project={} --region=[REGION] --format='table(name,serviceAccount)'` to show per-project remediation status (P4SA vs. BYOSA). Color-code: red = still using P4SA, green = BYOSA confirmed. This gives leadership a factual, project-by-project completion tracker with no tooling cost. Track open items in a shared spreadsheet with project ID, P4SA status, assigned owner, and target completion date.

Evidence: Attach to the leadership brief the Cloud Audit Log evidence from Step 2 showing the P4SA's actual historical access to GCS and Artifact Registry — this transforms the communication from theoretical risk to documented exposure. If any unauthorized GCS reads or Artifact Registry pulls appear in the logs, escalate the brief to include a potential data access notification assessment per your organization's incident classification policy.

Step 6: Monitor developments — track the GitHub ADK discussion thread and Google Cloud documentation for additional control updates; watch for follow-up Unit 42 disclosures, CVE assignment, or CISA guidance related to AI platform default permission patterns

NIST Phase: Post Incident

Reference: NIST 800-61r3 §4 — Post-Incident Activity: maintaining ongoing awareness of evolving guidance for this specific disclosure, including potential CVE assignment and CISA Known Exploited Vulnerability catalog additions that would change triage priority

Controls: NIST SI-5 (Security Alerts, Advisories, and Directives) — establish a monitoring process for follow-on advisories from Unit 42, Google Cloud Security Bulletins, and CISA, NIST IR-8 (Incident Response Plan) — update plan if CVE is assigned or CISA KEV listing changes the response SLA, CIS 7.1 (Establish and Maintain a Vulnerability Management Process)

Compensating: Subscribe to Google Cloud Security Bulletins via RSS (https://cloud.google.com/support/bulletins/rss) and configure a free GitHub watch on the ADK repository (https://github.com/google/adk-python) for new issues or security advisories. Set a calendar-based review cadence (biweekly) to check whether a CVE has been assigned via NVD search for 'Vertex AI' or 'Agent Engine'. If CISA issues a KEV entry or binding operational directive related to GCP AI platform defaults, your triage priority escalates from urgent to immediate and BYOSA migration SLA drops to 24 hours.

Evidence: Maintain a running log of all external source checks (date, source, finding) so that if a CVE is later assigned and a regulator asks when you became aware of the risk and what you did, you have a documented timeline. Archive the Unit 42 research URL and publication date, the Google Cloud security bulletin response (if issued), and your internal remediation completion dates as a chain-of-custody record for the exposure window.

Detection Guidance

Direct indicators of exploitation are limited given the absence of reported active campaigns, but behavioral patterns across GCP audit logs can surface both exploitation attempts and misconfiguration exposure.

Credential theft via metadata server: In Cloud Audit Logs, look for service account token requests originating from Vertex AI Agent Engine workload identities at unusual frequency or outside expected execution windows. An agent querying the metadata server endpoint (169.254.169.254) for OAuth tokens and then immediately making API calls to Storage or Artifact Registry is a behavioral sequence worth investigating, particularly if the agent's defined task does not require those services.

Cloud Storage access anomalies: Enable and review Data Access audit logs (not enabled by default in GCP) for Cloud Storage. Flag read operations across multiple buckets within the same project originating from an Agent Engine service identity. Lateral access across buckets, especially buckets unrelated to the agent's defined

function, is an indicator of either misconfiguration abuse or active exploitation.

Artifact Registry reads: Audit log entries showing Artifact Registry read events from a P4SA or Agent Engine identity should be reviewed against expected workflow. Access to repositories tagged as Google-internal or outside the organization's own registry namespace warrants immediate investigation.

IAM configuration drift: Implement a policy check (via Security Command Center, GCP Organization Policy, or a third-party CSPM) that flags any Agent Engine deployment still using the default P4SA rather than a BYOSA. Treat the presence of the default P4SA with broad scopes as a misconfiguration finding, not just a risk indicator.

Threat hunting hypothesis (MITRE-aligned): Hunt for T1078.004 activity by correlating Agent Engine execution events with service account API calls to services outside the agent's documented integration scope. Cross-reference with T1530 by identifying any bulk Storage enumeration or multi-bucket read pattern from AI workload identities.

Framework Mappings

MITRE-ATTACK

- **T1530** — Data from Cloud Storage
- **T1548** — Abuse Elevation Control Mechanism
- **T1199** — Trusted Relationship
- **T1552.001** — Credentials In Files
- **T1195.002** — Compromise Software Supply Chain
- **T1078.004** — Cloud Accounts

NIST-800-53R5

- **AC-6** — Least Privilege
- **CM-6** — Configuration Settings
- **CM-7** — Least Functionality
- **SA-9** — External System Services
- **SR-3** — Supply Chain Controls and Processes
- **SI-7** — Software, Firmware, and Information Integrity
- **AC-3** — Access Enforcement
- **IA-5** — Authenticator Management
- **SR-2** — Supply Chain Risk Management Plan

OWASP-TOP10-2021

- **A01:2021** — Broken Access Control
- **A04:2021** — Insecure Design
- **A07:2021** — Identification and Authentication Failures

CIS-V8

- **3.3** — Configure Data Access Control Lists
- **5.4** — Restrict Administrator Privileges to Dedicated Administrator Accounts

- **6.8** — Define and Maintain Role-Based Access Control
- **5.2** — Use Unique Passwords
- **6.3** — Require MFA for Externally-Exposed Applications
- **15.1** — Establish and Maintain an Inventory of Service Providers

HIPAA-SECURITY

- **164.308(a)(5)(ii)(D)** — Password Management
- **164.312(d)** — Person or Entity Authentication

SOC2-TSC

- **CC6.1** — Logical access security software, infrastructure, and architectures
- **CC9.2** — Manages risks associated with vendors and business partners

NIST-CSF-2

- **GV.SC-01** — Cybersecurity supply chain risk management program

ISO-27001-2022

- **A.5.21** — Managing information security in the ICT supply chain
- **A.5.23** — Information security for use of cloud services

MITRE ATT&CK Mapping

Technique ID	Technique Name	Tactic
T1530	Data from Cloud Storage	Collection
T1548	Abuse Elevation Control Mechanism	Privilege-Escalation
T1199	Trusted Relationship	Initial-Access
T1552.001	Credentials In Files	Credential-Access
T1195.002	Compromise Software Supply Chain	Initial-Access
T1078.004	Cloud Accounts	Defense-Evasion

Sources

Source	URL	Tier
Security News	https://thehackernews.com/2026/03/vertex-ai-vulnerability-exposes-g...	T3
Double Agents: Exposing Security Blind Spots in GCP Vertex AI	https://unit42.paloaltonetworks.com/double-agents-vertex-ai/	T3

Source	URL	Tier
Security controls for Vertex AI | Google Cloud Documentation	https://docs.cloud.google.com/vertex-ai/docs/general/vertexai-secur...	T3
Artifact reading after ADK deployment on vertex AI #1339 - GitHub	https://github.com/google/adk-python/discussions/1339	T3
Agent Factory Recap: Securing AI Agents in Production	https://cloud.google.com/blog/topics/developers-practitioners/agent...	T3