AVIATRIX

SECURITY ARCHITECTURE BRIEF

Contain Gemini Enterprise Agents

Powered by Aviatrix + Google Cloud

Threat model, enforcement architecture, and compliance evidence for security architecture review

EXECUTIVE SUMMARY

Gemini Enterprise Agent Platform run in a Google-managed tenant project with unrestricted outbound internet access by default. VPC Service Controls, Model Armor, and Secure Web Proxy each address a different axis but none of them is the network egress firewall for non-Google destinations across your estate. The Aviatrix Validated Containment Architecture for Gemini Enterprise Agent Platform closes that gap at Layer 3 with no agent changes: default-deny egress, a named shadow-model deny, SmartGroup East-West isolation, and continuous per-connection audit logs, all under one Distributed Cloud Firewall policy model that spans your Gemini Enterprise Agent Platform, Bedrock, Foundry, and self-hosted agent environments.

Threat Model

The Containment Era operating assumption: something in the environment is already compromised, and the Blast Radius of that compromise is decided by architecture, not by detection speed. Gemini Enterprise Agent Platform make that assumption sharper. They are autonomous, they decide which APIs to call, and they do it with unrestricted egress unless a network control explicitly limits it.

Threat Vector	Description
Prompt injection / tool-abuse exfiltration (OWASP LLM01)	A malicious prompt or compromised dependency redirects an agent tool call to attacker-controlled infrastructure. No Google-native control stops the socket from opening to a non-Google host.
Shadow-model egress (OWASP LLM08 excessive agency)	Agent code or a prompt manipulates the agent into calling an unsanctioned LLM provider. The call carries sensitive data retrieved from sanctioned RAG sources. VPC Service Controls is blind to non-Google destinations.

Supply-chain compromise on sanctioned host (OWASP LLM05)	A sanctioned package host (e.g., raw.githubusercontent.com) serves both clean and poisoned content. SNI-only filtering cannot distinguish the two because the TLS endpoint is identical.
DNS-tunneled exfiltration (MITRE ATLAS)	Agent or dependency uses UDP/53 to an external resolver for out-of-band data exfiltration.
Lateral movement	A compromised agent spoke reaches adjacent workloads in the customer fabric, expanding the Blast Radius beyond the agent VPC.

IMPORTANT BACKGROUND

The March 2026 Cascade supply chain operation compromised LiteLLM and four other AI infrastructure packages in twelve days, using trusted update channels to harvest credentials from AI development environments. The attack executed as legitimate code through trusted pipelines. Detection saw nothing anomalous. The only control that held was architectural: Aviatrix Distributed Cloud Firewall customers in default-deny mode blocked the command-and-control egress at the network layer before credentials left the environment. This Validated Containment Architecture pre-assembles that posture for Gemini Enterprise Agent Platform.

Attack Scenario and Kill Chain

Kill Chain Stage	What Happens	Control
Prompt injection	Malicious prompt redirects agent tool call to api.attacker.com	Default-deny DCF: destination not in WebGroup allow-list, socket blocked and logged in FlowIQ
Shadow-model call	Agent attempts to POST to api.openai.com or *.anthropic.com	vca-vertex-shadow-model-deny rule: blocked, logged by name, visible in FlowIQ
Supply-chain (GKE shape)	raw.githubusercontent.com serves poisoned dependency alongside clean content	Selective TLS decryption + URL-filter deny: compromised path blocked; clean path returns HTTP 200
DNS exfiltration	Agent uses UDP/53 to external resolver	UDP/53 deny rule blocks all queries to external resolvers
Lateral movement	Compromised agent attempts to reach adjacent spokes	SmartGroup East-West deny: Blast Radius stops at agent VPC

POINT OF INTERVENTION

Enforcement runs at Layer 3, between the PSC-I network attachment and the internet. The Aviatrix gateway is the RFC 1918 egress next-hop Google requires under VPC Service Controls, satisfied at the route layer. The agent makes the same outbound calls it always made. There is no HTTPS_PROXY, no SDK change, no redeploy, and no sidecar on the agent. The gateway is transparently in path. The containment policy decides what leaves.

Enforcement Architecture

Enforcement runs at two layers. Neither requires code changes, sidecars, or application team involvement.

Tier	Role
Transparent L3 insertion (both shapes)	Network-attachment subnet routes all non-Google agent egress to the Aviatrix gateway. Gateway performs NAT and acts as transparent forward proxy. No HTTPS_PROXY, no agent change.
DCF policy pack (both shapes)	Default-deny egress. WebGroup allow-lists for sanctioned tool, MCP, and RAG destinations. vca-vertex-shadow-model-deny ahead of the allow-list. UDP/53 deny. SmartGroup East-West isolation.
Kubernetes SmartGroups (GKE shape)	Pod-level identity via SmartGroups keyed to k8s_namespace=agents. Selective TLS decryption scoped to destination FQDN SmartGroup for URL-path filtering. Requires Controller 9.0+.

WHY THIS IS DIFFERENT FROM SECURE WEB PROXY

Secure Web Proxy is GCP-only, per-region, with its own console and policy language. This Validated Containment Architecture uses the same DCF policy model that governs your Bedrock AgentCore, Azure AI Foundry, and self-hosted Kubernetes agents. One control plane, one policy model, one CoPilot dashboard. Run Secure Web Proxy alongside this Validated Containment Architecture if you already have it; the two are complementary.

Architectural Boundaries

Out of Scope	What Governs It Instead
Prompt and response content inspection	Model Armor, Agent Gateway; AgentGuard Advanced Guardrails (roadmap)
Google-to-Google data path (shadow BigQuery, shadow GCS)	VPC Service Controls
A2A multi-agent fan-out	Agent Gateway
Managed runtime TLS decryption	Only with agents running on GKE

Compliance Evidence

For SOC 2, HIPAA, PCI-DSS, and FedRAMP environments, auditors require architectural proof of enforcement, not a policy document.

Evidence Artifact	What It Proves
CoPilot FlowIQ per-connection logs	Continuous audit trail: every egress decision attributed to SmartGroup, WebGroup, and DCF rule with timestamp. Maps to SOC 2 CC6.6/CC6.7, HIPAA §164.312(e)(1), PCI-DSS v4 Req 1.3, FedRAMP SC-7.
vca-vertex-shadow-model-deny block logs	Named, timestamped log of every attempted reach to unsanctioned model providers. Proves the control is operating.
DCF policy in Terraform (IaC)	Version-controlled, peer-reviewed policy with approval trail. Adding a tool endpoint is a PR, not a change ticket.
Cross-platform log consistency	Same policy model and log format whether agent runs on Gemini Enterprise Agent Platform, AgentCore, Foundry, or self-hosted. Auditors see one control.
East-West SmartGroup deny logs	Proves lateral movement from agent spoke to adjacent workloads was architecturally prevented.

The proof of enforcement is the architecture itself, not a statement about the architecture.

Next Steps

Request a 30-minute architecture review.

We walk through the enforcement model in your environment, map your current Gemini Enterprise Agent Platform egress surface, and identify the paths you currently cannot see.

About Aviatrix

Aviatrix® is pioneering the Cloud Native Security Fabric – the architecture the Containment Era requires. The Cloud Native Security Fabric governs every workload communication path across every cloud, every VPC, every Kubernetes cluster, and every serverless function, from a single policy plane. One rule. Universal propagation. Enforced at the workload, not at a chokepoint. Trusted by more than 500 of the world's leading enterprises. For more information, visit aviatrix.com.