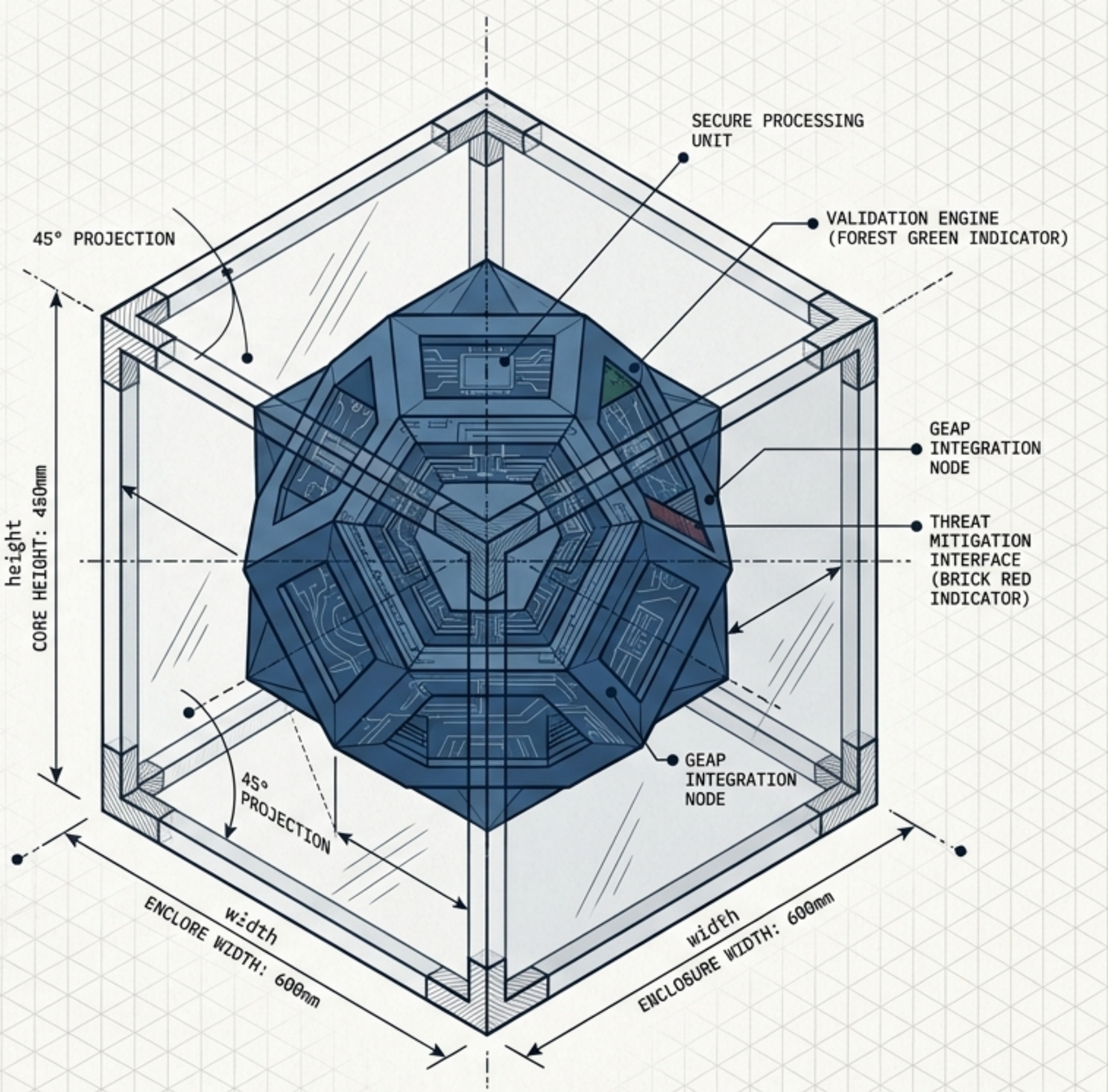
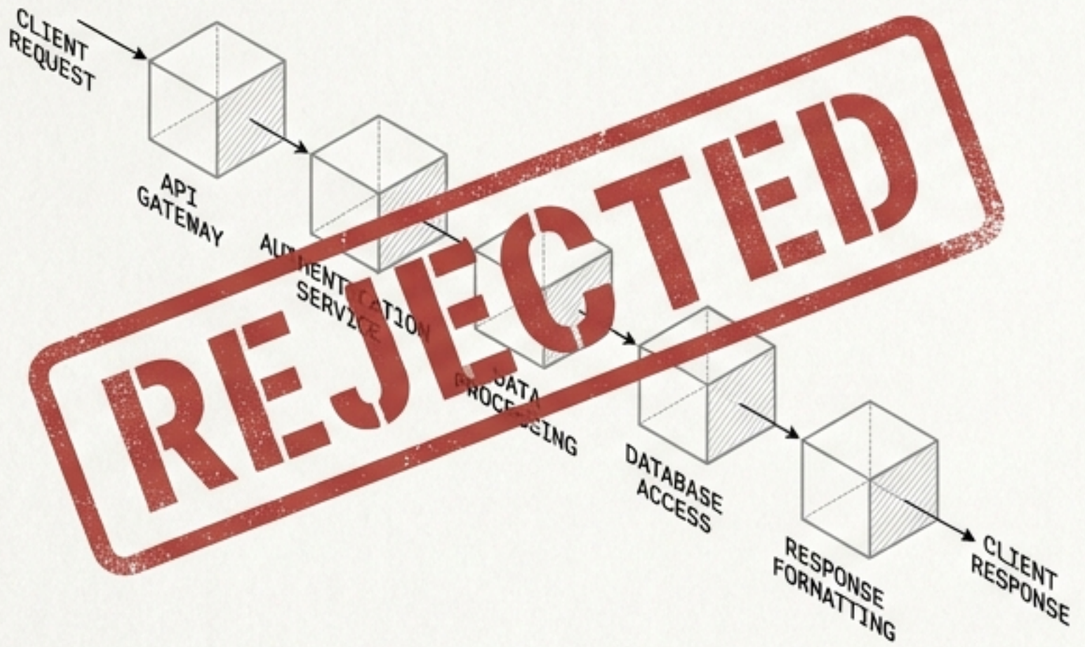
Enterprise Security: CISO-Defensible Agent Deployments

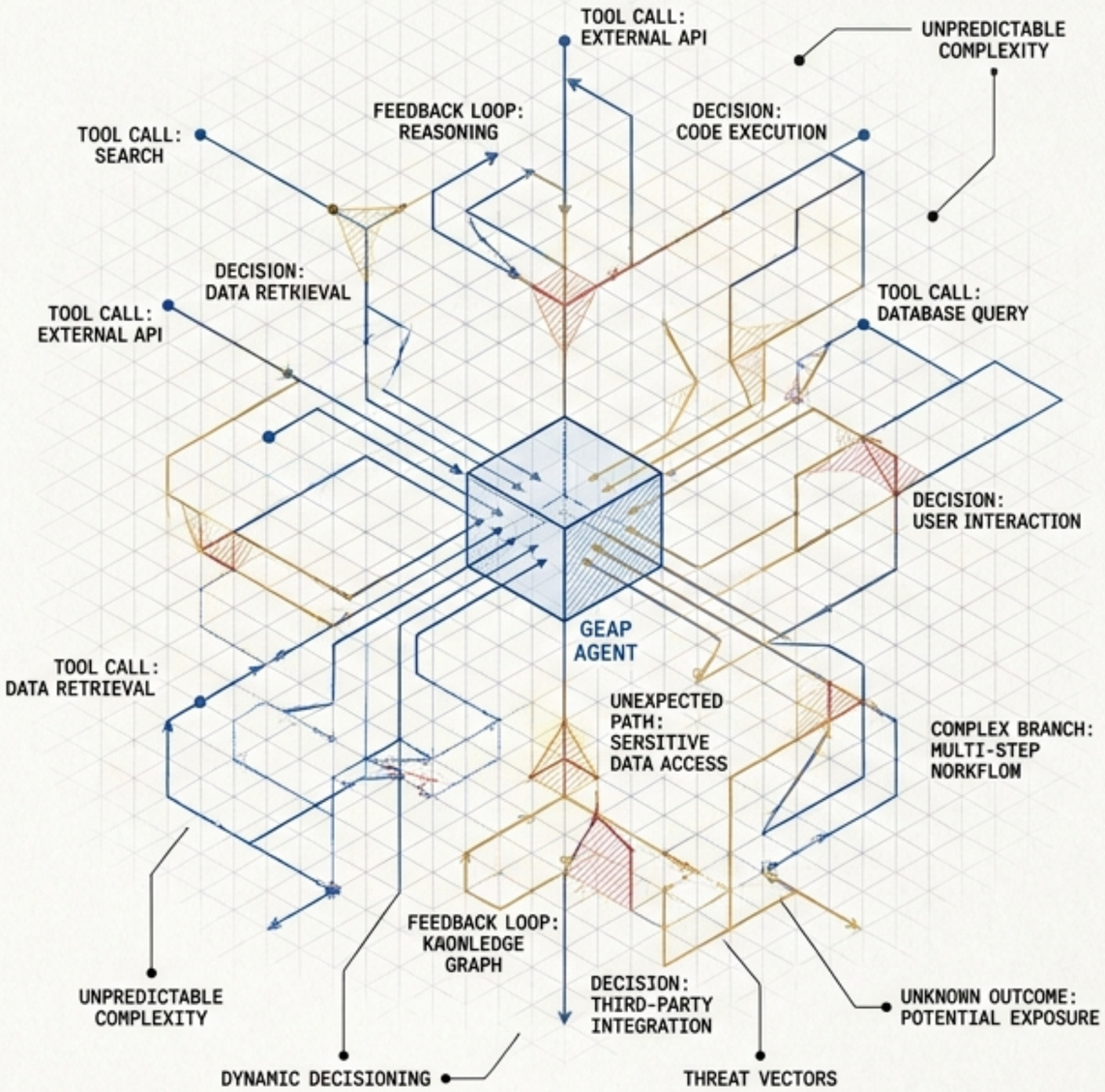
Securing the Gemini Enterprise Agent Platform (GEAP) for regulated environments.



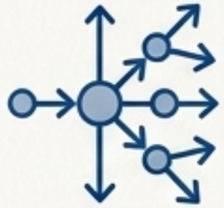
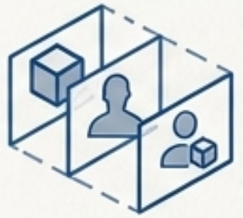
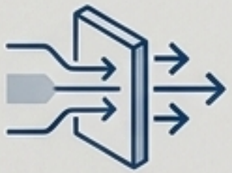
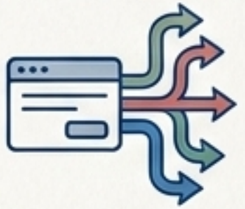
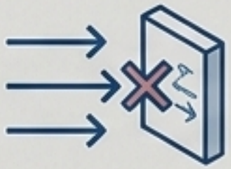
The False Comfort of the Cloud Run Checklist

The security models that protect your web APIs will fail to protect your reasoning agents. The Gemini Enterprise Agent Platform (GEAP) diverges from every prior Vertex AI service. Applying legacy API hardening to generative agents leaves regulated data entirely exposed.



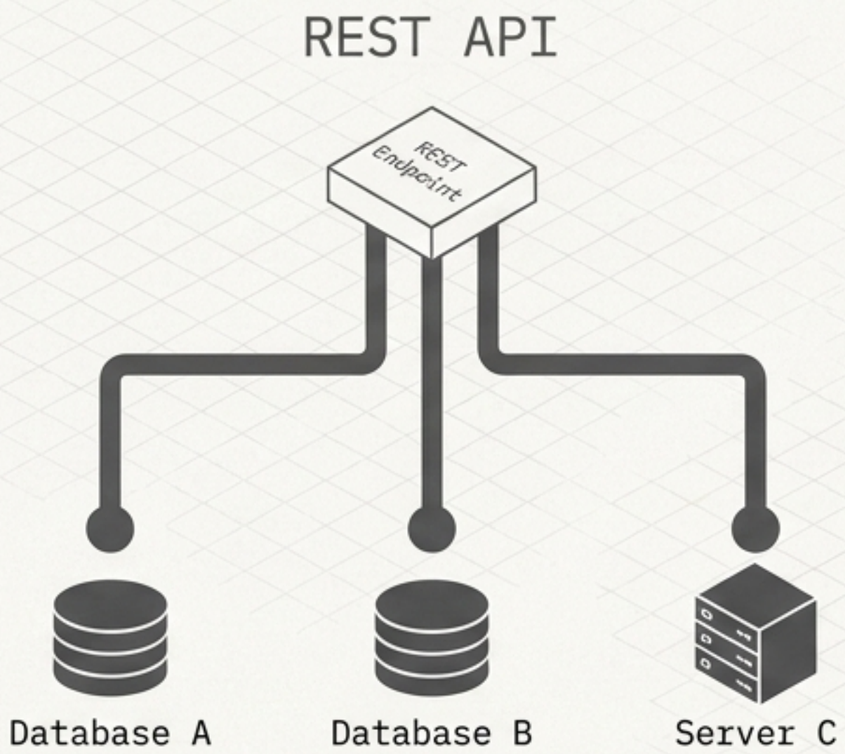


Diagnostic Matrix: Why Agent Security Breaks the Mold

Traditional API	Generative Agent
 Attack Surface: Bounded at design time	 Attack Surface: Unbounded union of all reachable tools multiplied by all possible prompts.
 Identity Model: Scalar; single service account	 Identity Model: Layered; simultaneous agent, user, and sub-agent identities.
 Input Threat: SQL injection; sanitize at application layer	 Input Threat: Prompt injection; model is an active attack vector.
 Network Focus: Ingress; blocking external inbound traffic	 Network Focus: Egress; blocking unauthorized outbound tool fetches.

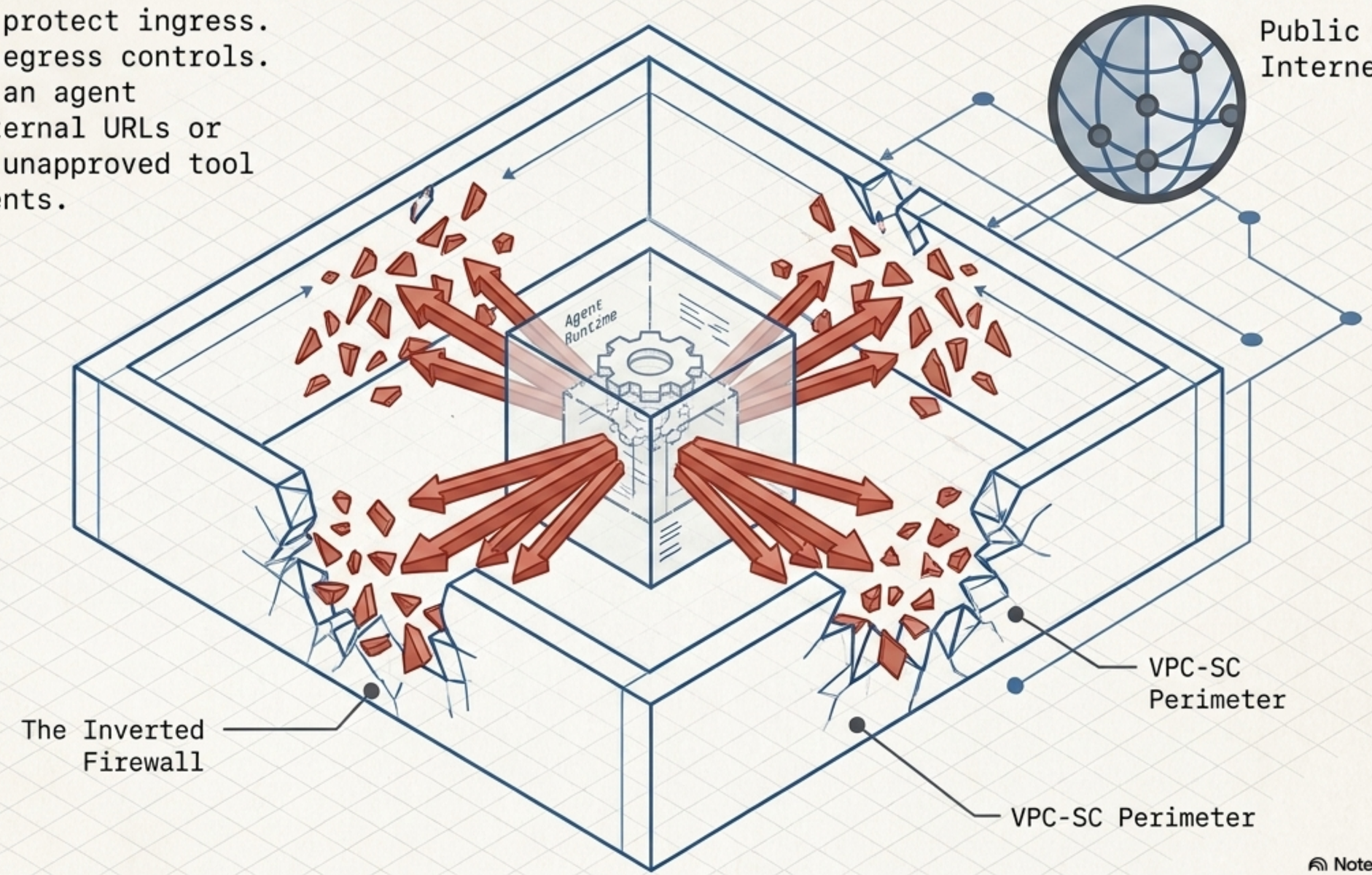
The Unbounded Request Graph

A REST endpoint declares exact downstream dependencies in its manifest. A reasoning agent decides what to call at runtime. You cannot enumerate the penetration-test surface from the codebase alone.



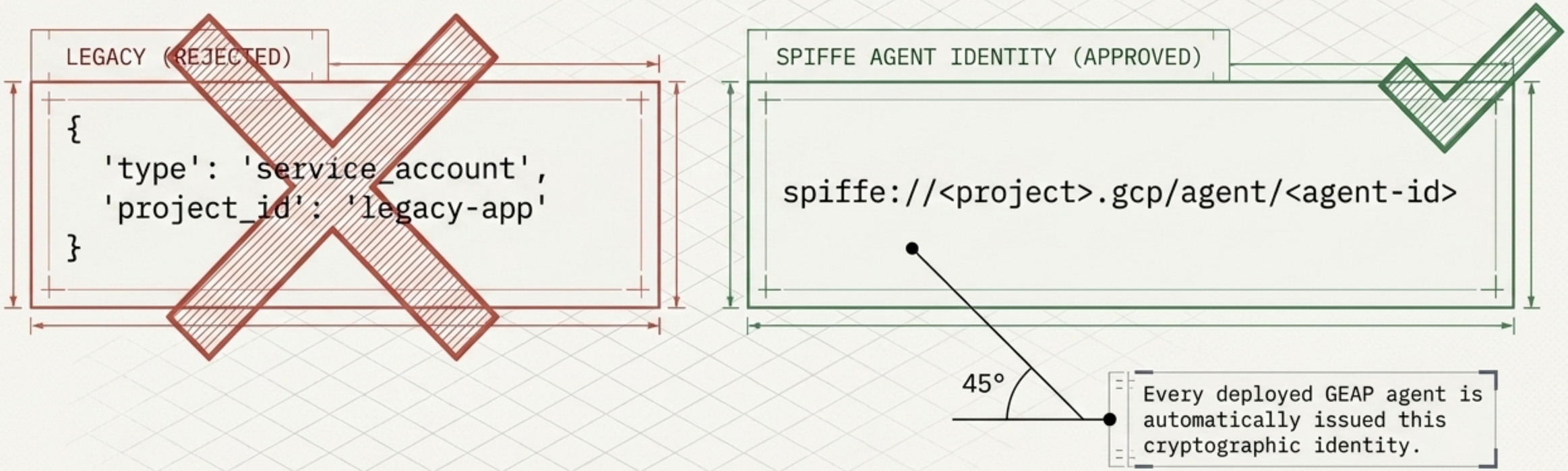
Egress is the New Perimeter

Traditional firewalls protect ingress. Agents require strict egress controls. The primary threat is an agent fetching malicious external URLs or exfiltrating PII into unapproved tool argumental tool arguments.



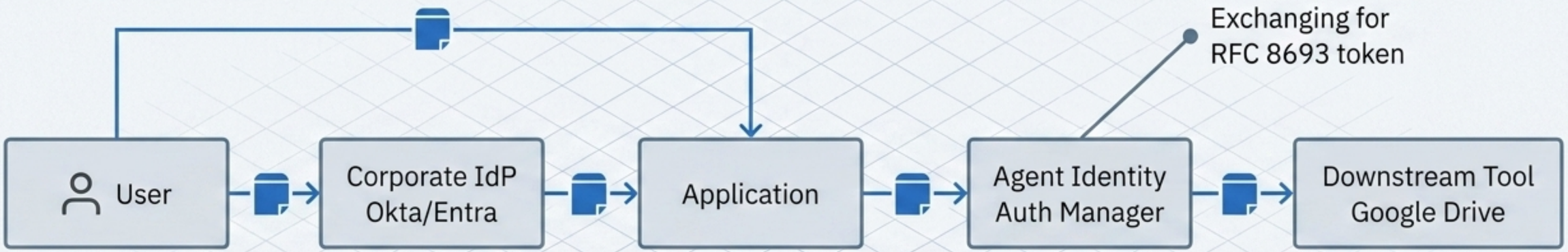
Layer 1: SPIFFE Agent Identity

- Action: Disable the legacy ‘default service account’ pattern.
- Rule: IAM roles must be bound directly to this SPIFFE ID with conditional clauses limiting resource scope. The deploy pipeline must reject explicit service accounts.

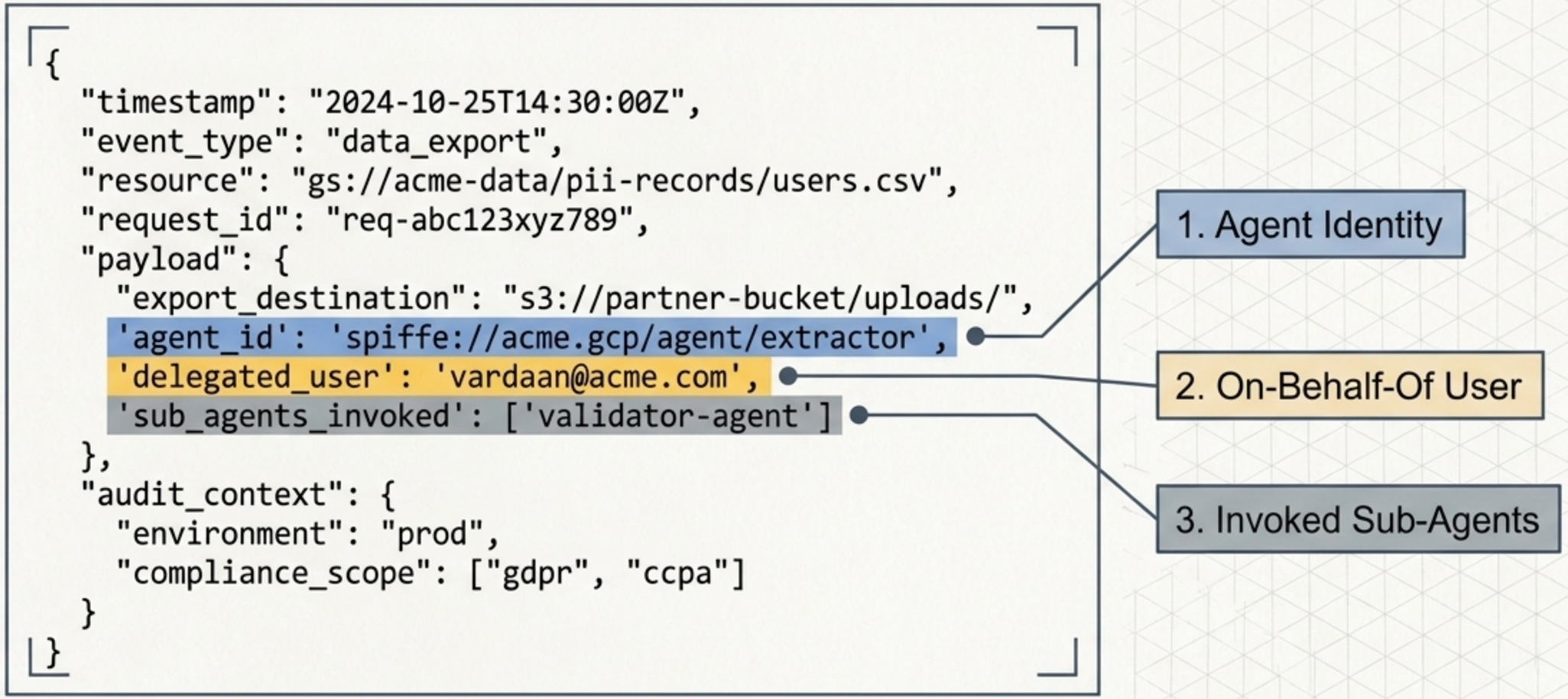


Layer 1: The “On-Behalf-Of” Mandate

Core Insight: An agent reading a Google Drive folder must read it as the calling user, not as a god-mode service account.
Action: Enforce 3-legged user-delegated OAuth via Agent Identity Auth Manager.
Warning: The implementation cost of 3-legged is two days; the compliance cost of 2-legged is a six-month audit finding.



Synthesis: The Three-Layered Audit Trail



Takeaway: When a regulator asks 'who exported this PII?', your forensics will be ambiguous unless all three identities are captured in the payload.

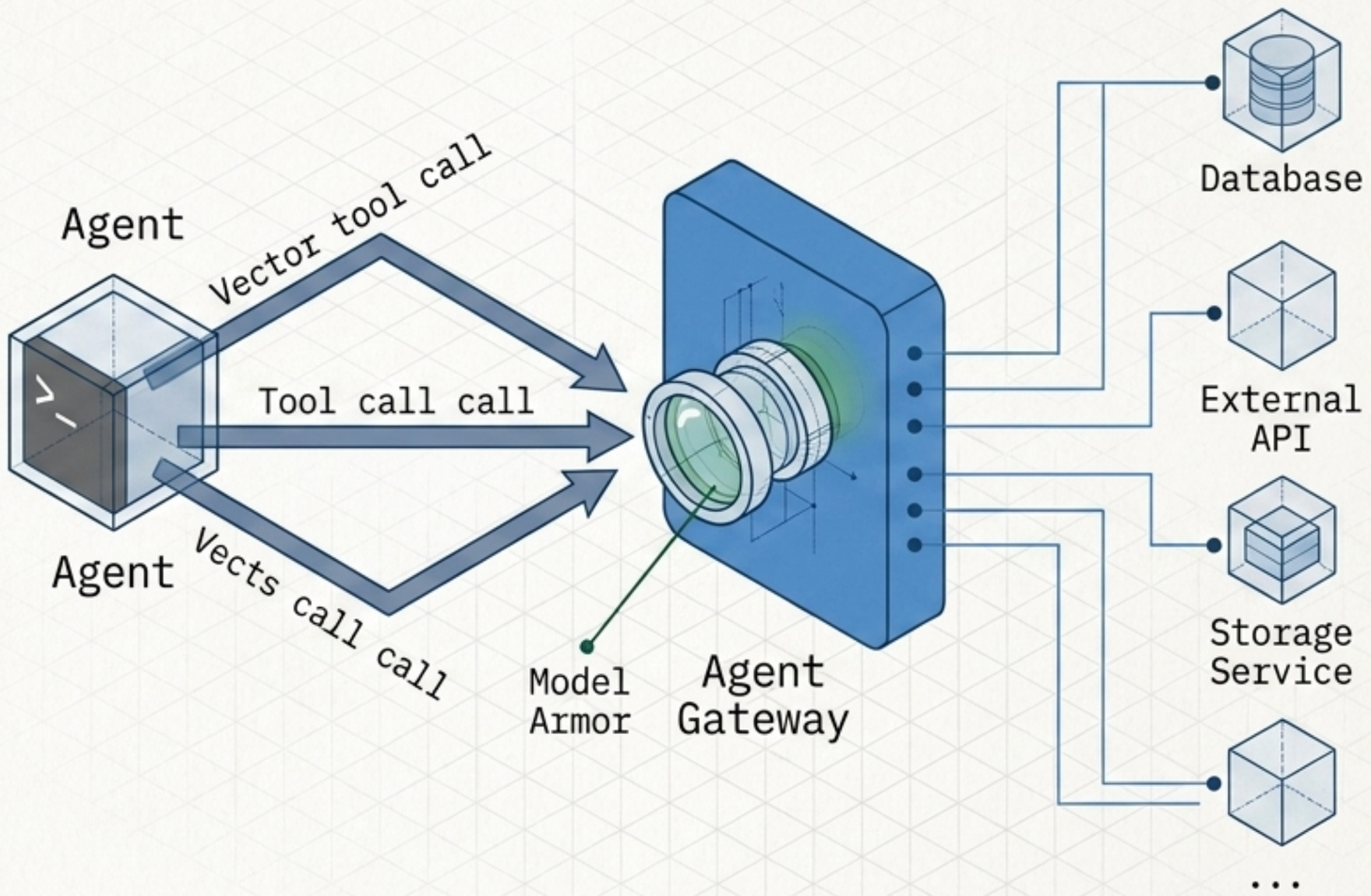
Layer 2: Agent Gateway & Model Armor

Mandate 1: Force all VPC-crossing tool traffic through Agent Gateway. Set `deny_external_internet: true`.

Mandate 2: Enable Model Armor in BLOCK mode to actively kill prompt injections.

Warning

Insight: Model Armor has a 4-7% false-positive rate on technical documentation with code blocks. Plan an exception list if handling source code to prevent legitimate refactoring failures.

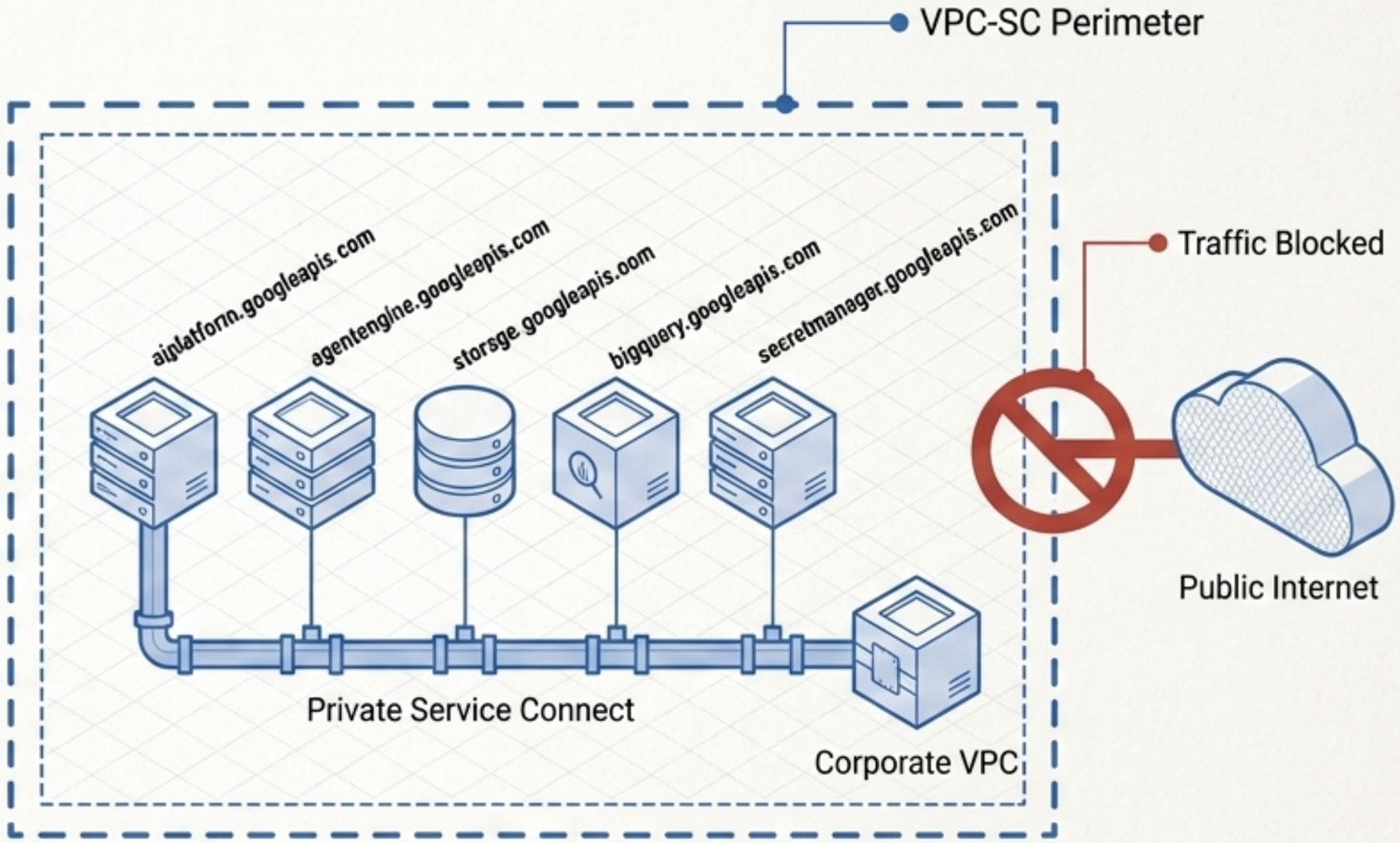


Layer 2: VPC Service Controls (VPC-SC)

Detail: Treat the agent runtime as a single composite resource. The minimal perimeter must encapsulate:

- aiplatform.googleapis.com
- agentengine.googleapis.com
- storage.googleapis.com
- bigquery.googleapis.com
- secretmanager.googleapis.com

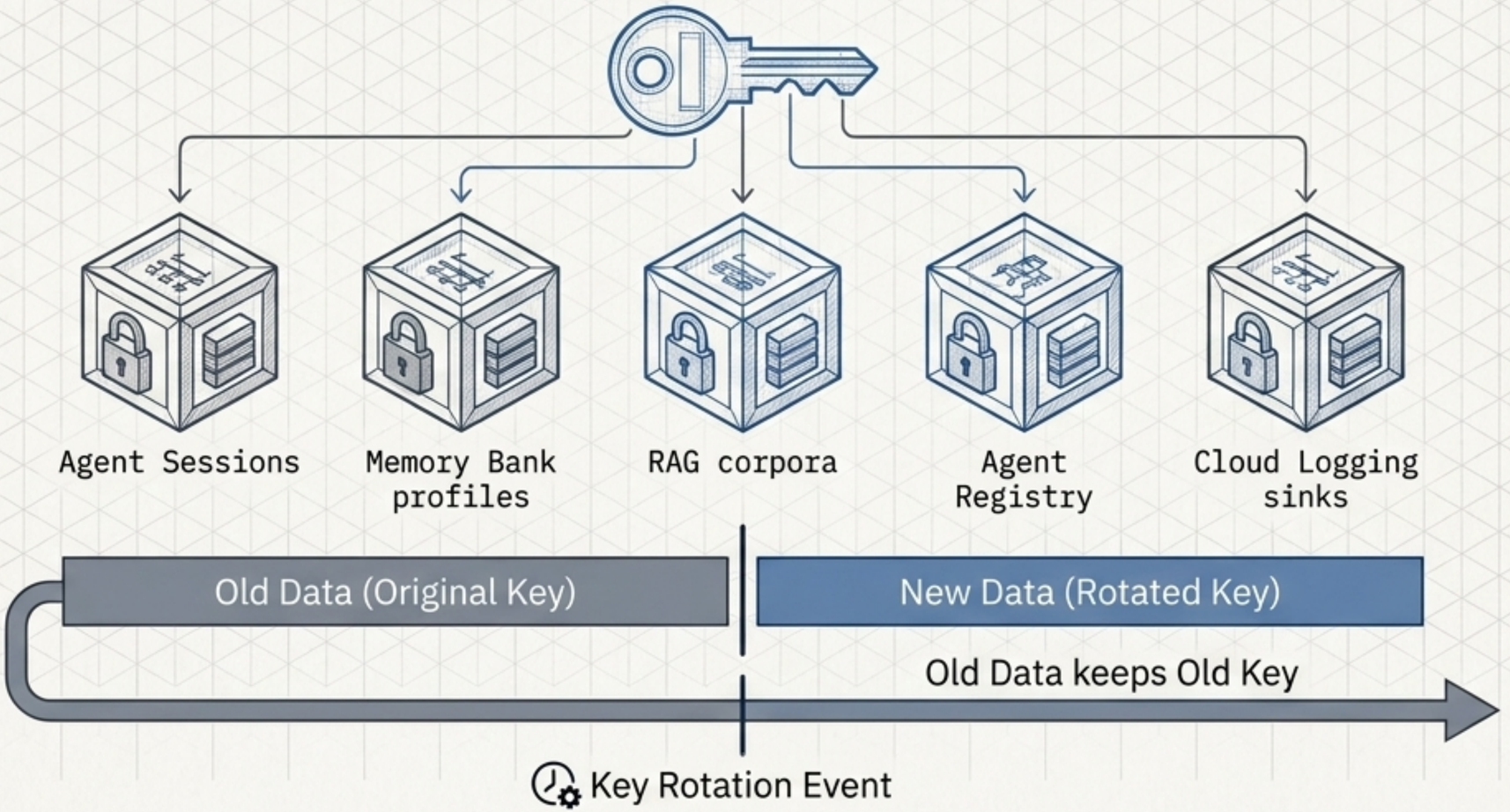
Impact: Traffic never traverses the public internet. This ensures a clean GDPR audit.



Layer 3: CMEK & The Forward-Secrecy Trap

Mandate: Bring your own KMS keys to encrypt Agent Sessions, Memory Bank profiles, RAG corpora, Agent Registry, and Cloud Logging sinks.

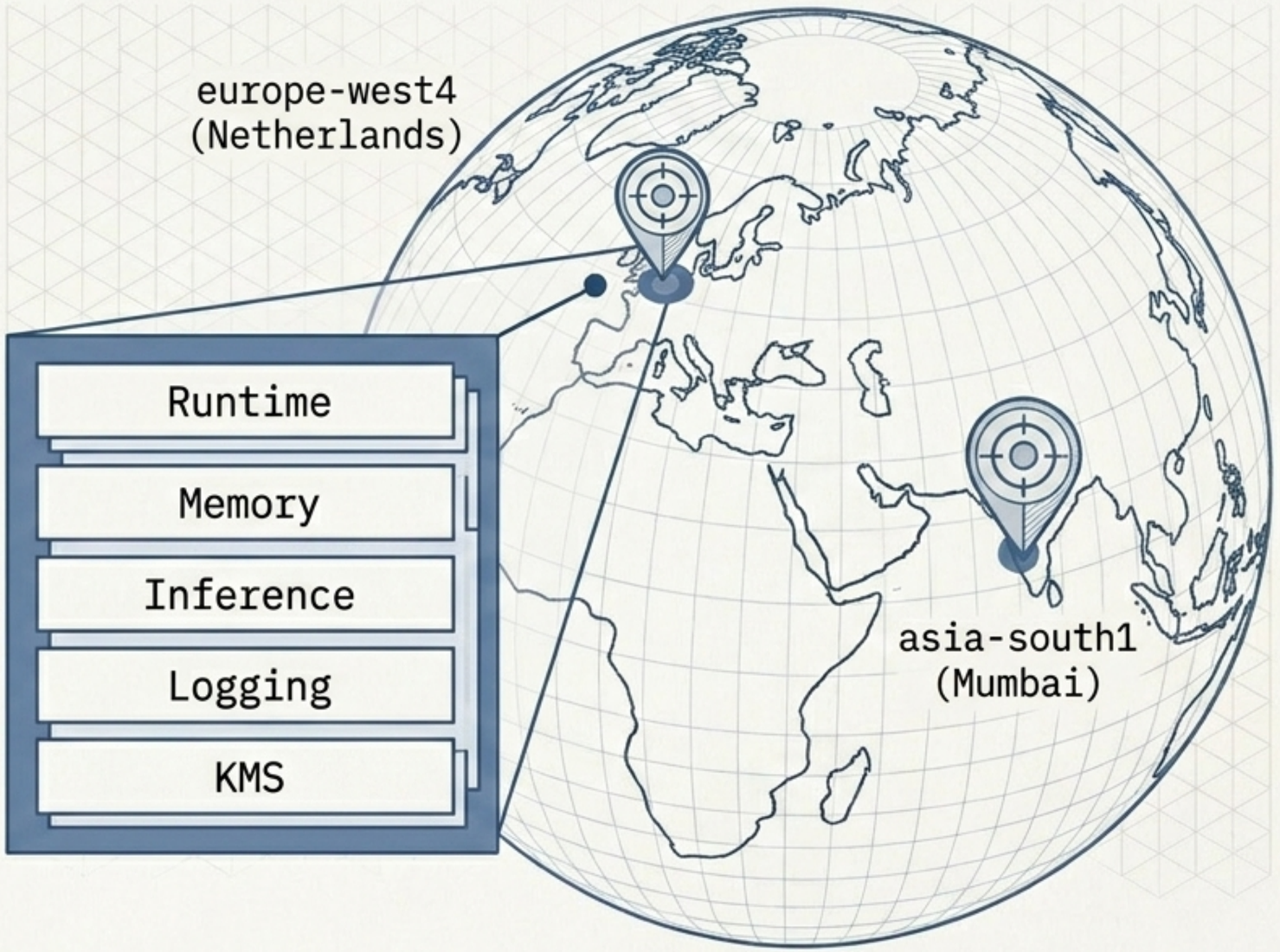
The Trap: CMEK rotation does not re-encrypt existing data. If regulators demand strict forward-secrecy, you must build an explicit, custom re-encryption job.



Layer 3: Data Residency vs. Sovereignty

Mandate: For EU GDPR or India DPDP Act compliance, pin Compute, State, Inference, Logging, and KMS to identical regional boundaries. Reject cross-region replication.

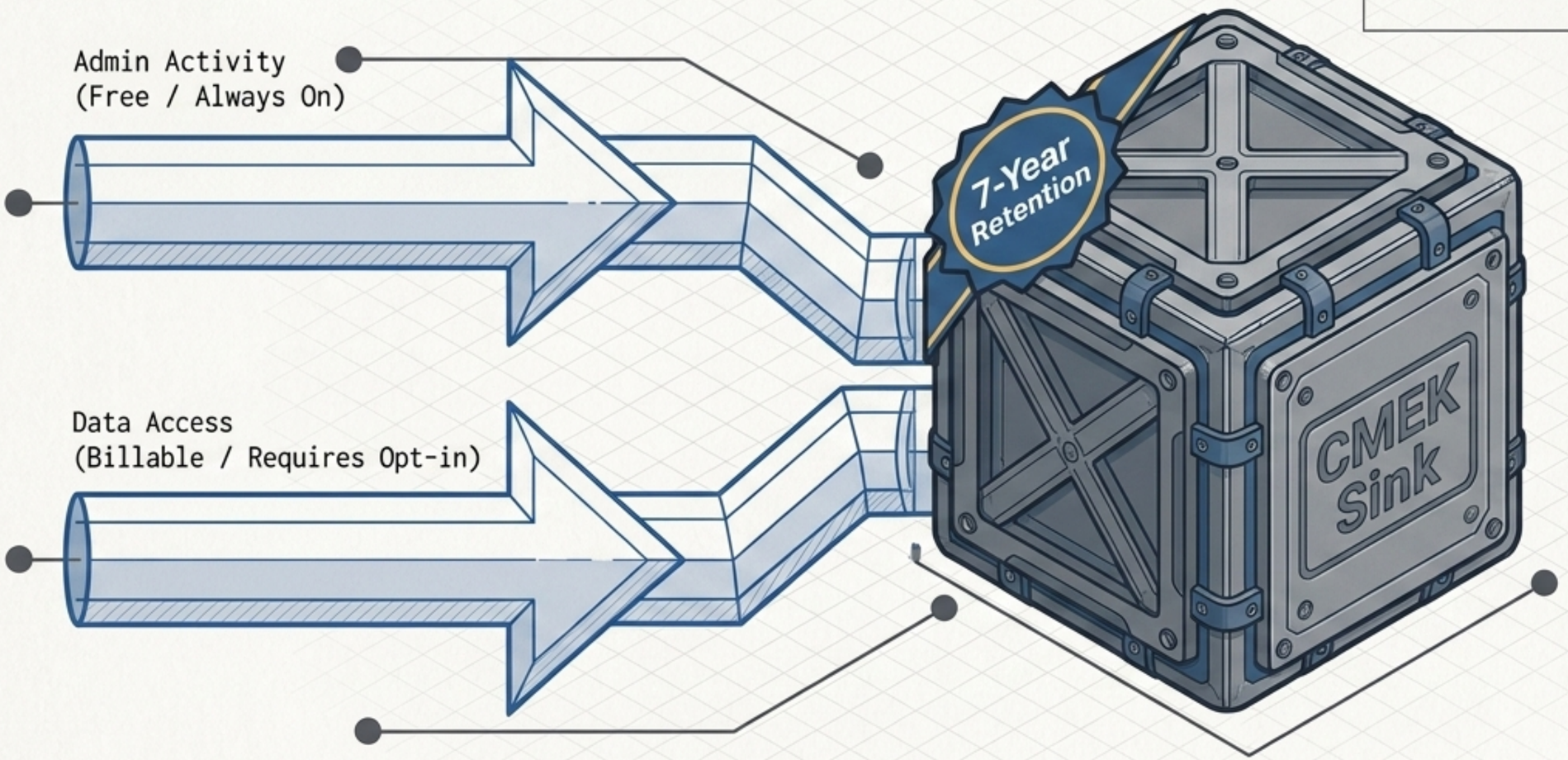
Contrarian Reality:
Contrarian Reality: Residency is not sovereignty. Even pinned to europe-west4, europe-west4, underlying model weights are operated globally. Highly regulated workloads require Sovereign Controls or GDC Hosted.



Layer 4: Deep Observability

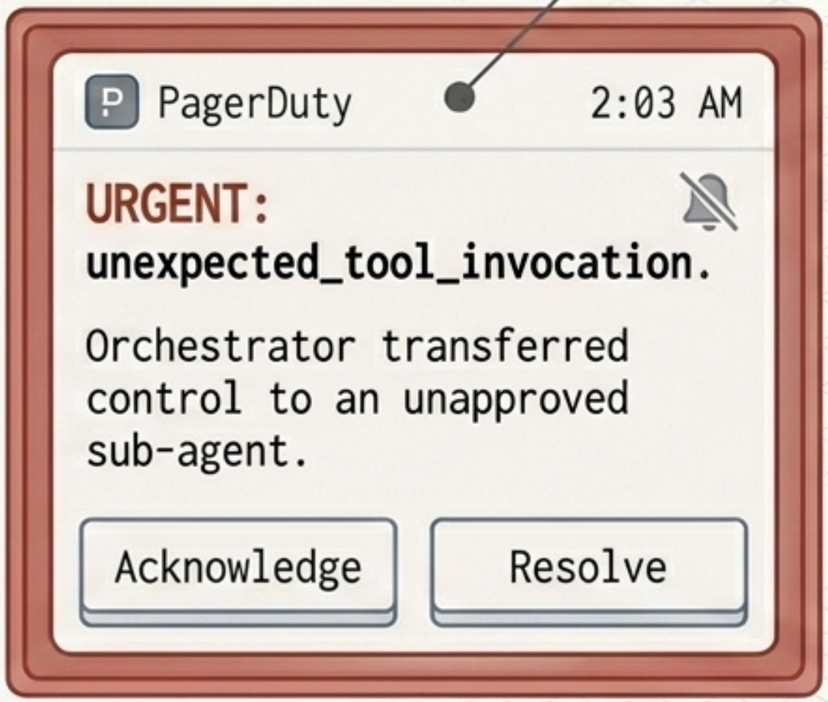
Mandate: Enable Data Access logs on aiplatform and agentengine.

Requirement: Route to a dedicated logging sink with CMEK encryption and a 7-year retention policy (mandatory for SOX/HIPAA compliance).



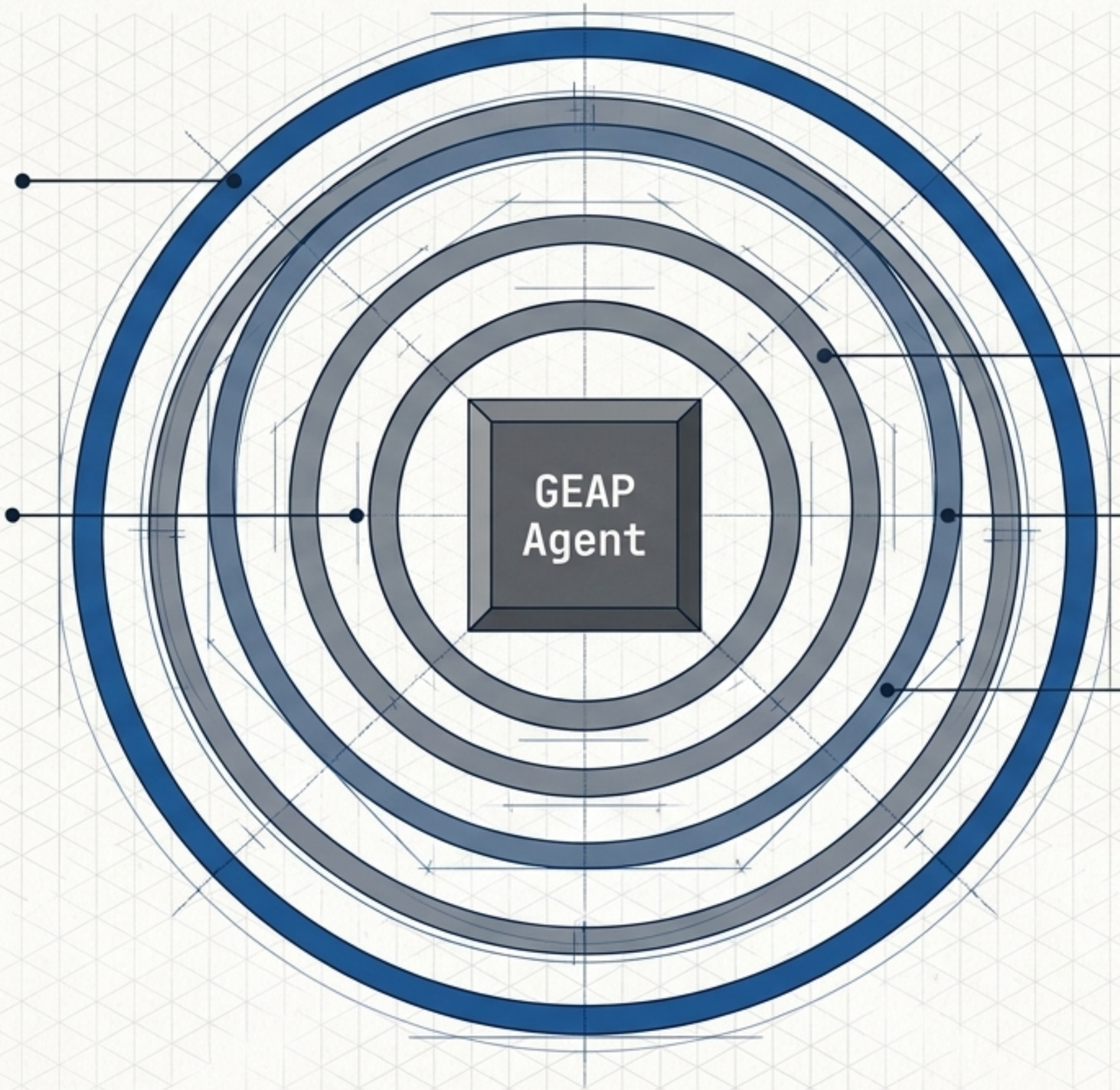
Layer 4: The 2 AM PagerDuty Alert

Context: Security Command Center integration adds anomaly detection on top of the GEAP audit stream. Wire these 3 SCC finding classes directly to on-call.



SCC Finding Class	Description
unexpected_tool_invocation	Orchestrator transferred control to an unapproved sub-agent.
egress_to_unknown_destination	Gateway attempted egress outside the allow-list.
prompt_injection_blocked	Model Armor intercepted a malicious payload.

The Concentric Defense Architecture



Observability
SCC, 7-Year Audit Logs

Layered Identity
SPIFFE, 3-Legged OAuth

Data Layer
CMEK, Regional Pinning

Execution Perimeter
VPC-SC, Private Service Connect

Network Perimeter
CMEK, Regional Pinning

The 9-Point CISO Approval Checklist

- [ ] 1. Unique SPIFFE-formatted identities; zero shared service accounts.
- [ ] 2. IAM bindings target SPIFFE IDs with resource scope conditions.
- [ ] 3. Gateway enabled: deny_external_internet: true & Model Armor in BLOCK mode.
- [ ] 4. 3-legged OAuth via Auth Manager for user-data tools.
- [ ] 5. VPC-SC composite perimeter encapsulating the 5 core GCP services.
- [ ] 6. CMEK enabled across all 5 data/storage surfaces.
- [ ] 7. Compute, state, inference, logging, and KMS co-located in identical residency regions.
- [ ] 8. Admin & Data Access logs enabled with 7-year retention in a CMEK sink.
- [ ] 9. Security Command Center integrated, routing agent anomalies to on-call.

Closing Note: Koenig AI Academy rejects 33% of enterprise deployments on the first pass (usually missing 3, 4, or 6). If you cannot tick every box, the agent is not ready for production.