

CISO & SECURITY ENGINEERING AUDIT PACK · 2027

30-Control CISO Agent, Tool & Weight-Provenance Audit Checklist

Out-of-loop inference revocation, checkpoint SHA-256 diffing, R&D vault isolation, MCP discovery locks, and 72-hour incident readiness.

> **Key Finding:** System prompts are instructions, not security boundaries. Real containment sits in the inference gateway, the model registry, and the tool proxy.

|---|---|

| **CISO-01** | 1. Out-of-Loop Revocation | Mandatory job_id on every inference and tool request | Gateway rejects requests missing signed job_id |

| **CISO-02** | 1. Out-of-Loop Revocation | Sub-60s inference-layer kill-switch | Terminated job_id fails closed at LLM proxy in <35-60s |

| **CISO-03** | 1. Out-of-Loop Revocation | Outbound API tool-proxy revocation | In-flight CRM/ERP/payment writes halt simultaneously on revoke |

| **CISO-04** | 1. Out-of-Loop Revocation | Per-job token and dollar ceiling | Runaway loops terminate automatically on budget cap |

| **CISO-05** | 1. Out-of-Loop Revocation | Quarterly live revocation drill | Measured time-to-halt logged for compliance audit |

| **CISO-06** | 2. Weight & Model Provenance | Immutable private model registry | Write access restricted to signed CI release pipeline |

| **CISO-07** | 2. Weight & Model Provenance | Startup SHA-256 checkpoint verification | Serving container verifies hash before loading GPU memory |

| **CISO-08** | 2. Weight & Model Provenance | Upstream tensor diffing (Lazarus defense) | Fine-tuned weights diffed against base model for anomalous sparse edits |

| **CISO-09** | 2. Weight & Model Provenance | Preview-model production block | Endpoints with <90-day sunset or preview status blocked in prod CI |

| **CISO-10** | 2. Weight & Model Provenance | Pinned runtime & SDK versions | Python >=3.11, vLLM ABI, and router versions locked in container image |

| **CISO-11** | 3. R&D vs Prod Vault Isolation | Zero customer datasets on developer laptops | Training/eval data stays inside VPC volumes with CMEK |

| **CISO-12** | 3. R&D vs Prod Vault Isolation | Separate cloud accounts for R&D and Production | Third-party dev/staging compromise cannot pivot to prod weights |

| **CISO-13** | 3. R&D vs Prod Vault Isolation | Customer-managed encryption keys (CMEK) | Independent key revocation tested for storage and vector indexes |

| **CISO-14** | 3. R&D vs Prod Vault Isolation | Hardware-encrypted / private VPC inference | Regulated workloads isolated on customer hardware, VPC, or confidential compute |

| **CISO-15** | 3. R&D vs Prod Vault Isolation | Ephemeral CI/eval credentials | Zero long-lived API keys stored in notebooks or test harnesses |

| **CISO-16** | 4. MCP & Tool Governance | Callable MCP discovery approval policy | Agents cannot discover or invoke unapproved external MCP servers |

| **CISO-17** | 4. MCP & Tool Governance | Version-pinned tool JSON schemas | Upstream tool description edits trigger security diff before activation |

| **CISO-18** | 4. MCP & Tool Governance | Bounded discovery & tool execution timeouts | Hung MCP or HTTP endpoints fail fast without wedging workers |

| **CISO-19** | 4. MCP & Tool Governance | Session authorization-scope isolation | Zero shared mutable state or credential carryover across user sessions |

| **CISO-20** | 4. MCP & Tool Governance | Human sign-off gate on irreversible writes | Payments, contract changes, and external sends pause for named owner |

| **CISO-21** | 5. Gateway, Residency & Incident Readiness | Organization location-policy lock | Cloud policy blocks enabling global endpoints when regional residency applies |

| **CISO-22** | 5. Gateway, Residency & Incident Readiness | Proxy PII & secret scrubbing | Sensitive identifiers redacted before external model providers |

| **CISO-23** | 5. Gateway, Residency & Incident Readiness | Principal-scoped semantic caching | Cache keys bind tenant_id + role_scope + intent + entity_id |

| **CISO-24** | 5. Gateway, Residency & Incident Readiness | Prompt-injection boundary separation | Untrusted external documents cannot trigger unapproved tool writes |

| **CISO-25** | 5. Gateway, Residency & Incident Readiness | ZDR / HIPAA / BAA contract verification | Beta features excluded where ZDR or BAA coverage is absent |

| **CISO-26** | 5. Gateway, Residency & Incident Readiness | Immutable tool & citation audit log | Every business action logs prompt hash, retrieved sources, and approver |

| **CISO-27** | 5. Gateway, Residency & Incident Readiness | 72-hour critical incident playbook (NY RAISE readiness) | Escalation runbook assigns legal, CISO, and engineering owners |

| **CISO-28** | 5. Gateway, Residency & Incident Readiness | Reward-shortcut & boundary evals | Pre-release tests verify agents do not bypass policy checks under pressure |

| **CISO-29** | 5. Gateway, Residency & Incident Readiness | Business-completion health monitoring | Alerts track stuck exception queues and stale sources, not server ping alone |

| **CISO-30** | 5. Gateway, Residency & Incident Readiness | Named accountable workflow owner | Every production AI workflow has a single executive owner authorized to halt it |

Work with ALLTIPLY's Engineering Team: Schedule an architecture review at www.alltiply.com/company/schedule-your-call or submit a scoped brief at www.alltiply.com/company/project-planner.