

PC-SAMPLE-AI-01 / EN / V1.1

AI controls design and assurance plan

ILLUSTRATIVE SAMPLE • FICTIONAL SCENARIO

PREPARED FOR
PREPARED

Orivane Industrial Group
2026-08-21

Recommendations and control responsibilities

This report presents an original fictional scenario. Orivane, its people, suppliers, systems, records and results are invented; established technology names identify the example stack only. No client material was used. Example hostnames use the reserved .example domain.

Orivane plans to release eight AI applications for twelve teams through programme Nacre AI (AI-2608). Nadia Bellerose sponsors the programme; Maya Dufresne approves its security controls. Four model services are under consideration, each with different rules for data use and operations. This design controls access to models and to tools that read or change business records. Application owners remain responsible for the results.

We recommend a central gateway for model requests and a separate service for actions in business systems. The gateway authenticates the application, selects an approved model and reserves the expected cost against the team budget. The action service checks permission for each task. Model output must be validated before it is used.

DECISION	REASON
Control model selection centrally	Applications cannot select an unregistered provider or weaken data handling through a request parameter.
Check action permissions separately	Model fluency or a prompt instruction never grants permission to read or change a business record.
Require tests before release	Each application passes its own quality, security and failure evaluations before a model or prompt version changes.
Keep a safe fallback	When dependencies fail, pause the AI action or use an approved non-AI workflow; do not silently route sensitive data elsewhere.

Application and model register

Registry AIR-0821 covers PolicyShelf Assist, Partspan Draft, ForgeTrack Copilot, Dockline Planner, FieldNotes Search, MarginLedger Brief, RouteDesk Triage and PeopleDeck Help. The twelve consuming teams receive CAD 2,000 each per calendar month, a CAD 24,000 total allocation owned by Celia Hart. Budgets reset at 00:00 UTC on the first day of the month; late charges remain with their original reservation period.

REGISTRY ENTRY	APPROVED SCENARIO TERMS
Q3-CA / Quillon AI	Quillon Q3, revision q3-0820; api.ca.quillon.example. Canadian processing; no content retention or training use. Internal and restricted support material approved in DPA-Q-17.
Q3M-CA / Quillon AI	Quillon Q3 Mini, q3m-0818; mini.ca.quillon.example. Canadian processing; no content retention. Public and internal drafting only; restricted retrieval denied.
CDR-EU / Silex Models	Cedar 2, c2-0815; eu.silex-models.example. Irish processing; operational metadata retained seven days. European internal material only; no Canadian restricted-data fallback.
F7-CA / Orivane	Forge-7, f7-0809; forge7.ai.orivane.example. Self-hosted ca-central-1 endpoint for approved source code; payload logging disabled. Elias Moreno owns capacity and patching.

These provider names, model revisions and contract terms are invented parts of the scenario. AIR-0821 binds each endpoint to policy ai-policy-1.6 and rate card RATE-0820; a route cannot substitute another provider when data-class conditions differ.

Use cases and permitted data

Classify data from an application-owned source and authenticated workflow context. A user-provided label such as “public” cannot downgrade a restricted document. Retrieved content inherits the classification and access constraints of its source; combining records must not discard their restrictions.

USE CASE	ADMITTED MATERIAL	RELEASE CONDITION
PolicyShelf Assist	Documents already readable by the requesting user; retrieval enforces source permissions.	Cross-user retrieval tests show no unauthorized excerpts, citations or cached answers.
Partspan Draft	Approved supplier-request fields with unnecessary identifiers removed.	Draft remains reviewable; sending a reply or changing a supplier record is a separate authorized action.
ForgeTrack Copilot	Approved repository scope and data class; credentials excluded.	Generated code follows normal review and testing before execution or merge.
Dockline Planner	Structured business inputs from an authenticated application.	Tool authorization, idempotency and approval boundaries pass adversarial tests.

Restricted data

Admission depends on the selected endpoint’s approved region, retention terms, training-use terms and access controls. These are verified attributes in the model registry, not assumptions inferred from a provider’s brand. If no endpoint meets the class requirements, the request is rejected. Masking obvious names is not a general de-identification guarantee.

Model selection and action approval

The model path controls data handling and paid usage. A separate action path controls business effects. Model output returns to the application as untrusted data; it cannot carry an approval or select its own authority. The second diagram follows one paid request through its accounting states.

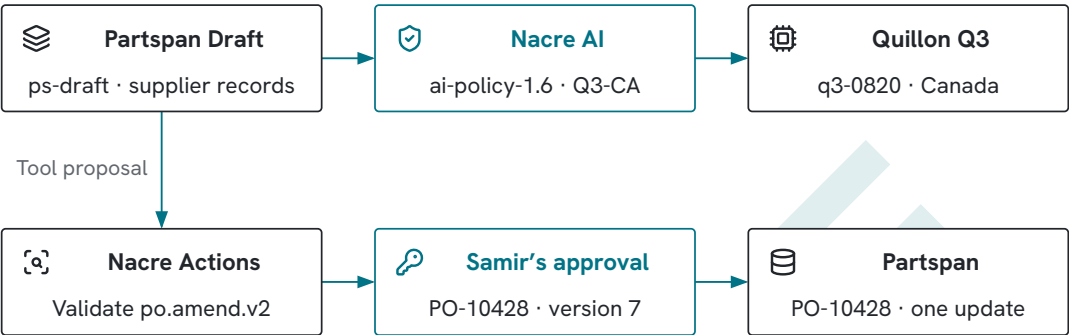


Figure 1 Partspan Draft target design: model admission and PO-10428 approval are separate controls.

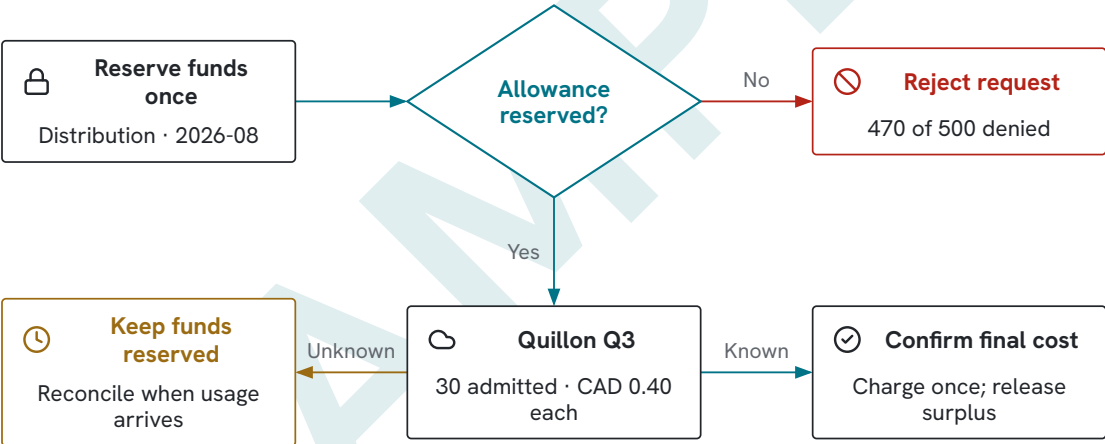


Figure 2 BUD-0820: reserve CAD 0.40 before dispatch; unknown usage keeps its reservation.

Provider retries and late usage retain their accounting identity. A new provider attempt creates new charge exposure; the operating record must distinguish it from retrying the same bookkeeping operation.

Enforcement architecture

Partspan Draft presents workload identity ps-draft, authenticated supplier scope, capability draft-supplier-reply and the source record’s data class. The Nacre AI endpoint is gateway.ai.orivane.example. The gateway resolves these inputs against versioned policy. Team and budget identity come from trusted mappings, never from a freely editable request body.

BOUNDARY	REQUIRED ENFORCEMENT
Application → gateway	Verify workload identity and user entitlement; reject unknown applications and unsupported data classes.
Gateway → provider	Select only an eligible registry entry; bind region and retention constraints to the endpoint actually used.
Workload → internet	Prevent direct provider access through credential custody and appropriate egress controls; audit alternate routes.
Model → tool service	Treat the requested tool name and arguments as untrusted; validate schema and authorize resource and action independently.
Response → user	Apply application output handling, provenance and review requirements; never execute text as trusted instructions.

Bypass and cache behavior

A gateway is not an enforced boundary if an application still holds provider credentials and unrestricted egress. Inventory scheduled jobs, notebooks and background workers as well as interactive applications. Cache keys include tenant, authorization scope, model and policy version where they affect the result. A cache hit cannot skip current entitlement checks or expose another user’s retrieved content.

Budget control for simultaneous requests

In load test BUD-0820, Distribution has CAD 12.00 available when 500 requests arrive. RATE-0820 caps each admitted attempt at CAD 0.40: exactly 30 reserve funds and 470 are rejected before dispatch. Checking only recorded spending would allow too many requests before their costs arrive. The system therefore reserves the maximum expected cost in one indivisible operation, so two requests cannot use the same available funds. It then adjusts the reservation to the provider’s actual charge.

The reservation key binds application, team, period and request identifier. Retries reuse the same accounting identity when they refer to the same attempt; a new provider attempt has its own charge exposure. BUD-0820 reconciles 30 charges of CAD 0.26: CAD 7.80 consumed and CAD 4.20 released. A separate timeout test retains its full reservation until Quillon’s usage record arrives.

CONDITION	BEHAVIOR
Insufficient allowance	Reject before provider dispatch; return a stable reason and an escalation path to the budget owner.
Usage unavailable	Retain the conservative reservation for reconciliation; do not release all allowance on a timeout.
Streaming response	Bound requested output and tool loops; account for already-generated usage even if the user disconnects.
Accounting unavailable	Fail closed for new paid requests unless an explicitly bounded contingency allowance was approved.
Period transition	Assign usage to the reservation’s accounting period; reconcile late charges without double counting.

A strict currency ceiling is only defensible if the provider’s pricing and maximum billable work can be bounded. Where that is not true, publish a quantified overshoot allowance and an alert threshold instead of claiming an absolute cap. Rate limits and budget limits solve different problems and both need tests.

Tool actions and human approval

Retrieved pages, documents and model output may contain malicious instructions that try to redirect the application. This is prompt injection. Prompt wording can improve behavior but is not an authorization boundary. Nacre Actions accepts po.read.v1, po.propose.v2 and po.amend.v2. Samir Okafor must approve any amendment to PO-10428, including its CAD 18,400 value, supplier scope and record version. The model cannot approve its own proposal.

ACTION CLASS	CONTROL CONTRACT
Read business data	Check user and object permissions at execution time. A document identifier returned by the model is not proof of entitlement.
Prepare a change	Create a reviewable proposal with the exact target and intended effect; no external side effect yet.
Commit a change	Bind approval to the exact action, target, arguments and current version. Changed arguments or stale state require renewed approval.
Retry an action	Use a unique action key to prevent duplicate changes (idempotency). Save the outcome; if a response times out, check what happened before retrying.
Delegate a task	Carry a reduced scope and expiry; delegation cannot broaden the original user's authority.

Operational stop

Provide separate switches for disabling a model route and disabling side-effecting tools. A response-quality incident may require replacing a model; an authorization incident may require immediately stopping tool execution while preserving read-only investigation. The stop mechanism is exercised from the operating runbook, including queued and in-flight work.

Release tests and misuse scenarios

EVAL-0821 contains 720 security cases and 180 task-quality cases per release candidate, reviewed by Maya Dufresne and Samir Okafor. PolicyShelf Assist passes 720/720 security cases and 171/180 quality cases (95%), meeting its 95% threshold. Expected behavior is specified before results are inspected. Security gates are evaluated separately from average answer quality so a good overall score cannot hide a data-boundary failure.

TEST FAMILY	REQUIRED CASES
Data isolation	Cross-user retrieval, removed group membership, restricted source mixed with public content and authorization-sensitive cache hits.
Routing	Unregistered model, wrong region, missing classification, fallback mismatch and direct-provider bypass attempt.
Tool authority	Prompt injection in retrieved text, changed approval arguments, unauthorized object identifier and replayed action.
Accounting	Concurrent admissions near the limit, provider timeout, late usage, retry and period rollover.
Quality	Correctness, grounded citations, uncertainty handling and task completion against domain-reviewed examples.
Failure	Model outage, policy outage, revoked credential, exhausted budget and interrupted streaming response.

Release gate

- ✓ Record model, endpoint, prompt, retrieval, tool schema and policy versions with the results.
- ✓ Block release after any unauthorized disclosure or action, regardless of average quality.
- ✓ Obtain business-owner approval for task-quality thresholds for each use case.
- ✓ Repeat security tests after changing provider, model, prompt, retrieval or tool permissions.

Operating evidence and unresolved decisions

Production telemetry records application and policy identity, selected endpoint, data class, decision reason, usage reservation and final accounting state. Content logging is minimized and separately authorized. Debugging convenience does not justify retaining prompts, documents or tool results without a defined purpose and access boundary.

OWNER	OPERATING RESPONSIBILITY
Samir Okafor / Partspan	Task quality, supplier permissions, draft review and acceptance of model changes.
Elias Moreno / Nacre	Gateway availability, credential custody, routing policy and accounting reconciliation.
Maya Dufresne / SecOps	Tool authority, data-boundary tests, exception review and incident containment.
Tessa Lin / data	Classification, approved provider handling and retention of any captured content.
Celia Hart / finance	Team allocation, contingency allowance and escalation when requests are rejected.

Before production approval

PolicyShelf Assist is admitted for 200 staff under REL-0821. Partspan Draft remains blocked: EVAL-0821 passes 719/720 security cases but fails stale-approval case TOOL-044 despite scoring 176/180 on quality. Maya Dufresne requires action approval to bind the current record version; Samir Okafor reviews the retest on 24 August. The other six applications remain in internal pilot until their own release records pass.

Reference and limits

The [NIST AI Risk Management Framework](#) is a general governance reference. It does not certify this design. Controls reduce exposure but do not establish perfect detection of prompt injection, factual accuracy or harmless model behavior. All application counts and operating scenarios in this document are fictional.