

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

# Conception des contrôles d'IA et plan de validation

RAPPORT TYPE • SCÉNARIO FICTIF

---

PRÉPARÉ POUR  
DATE

Groupe industriel Orivane  
2026-08-21

# Recommandations et responsabilités

Ce rapport présente un scénario fictif. L'entreprise Orivane, les personnes, les fournisseurs, les systèmes et les résultats ont été inventés. Les noms de technologies réelles servent uniquement à décrire le contexte technique. Aucun contenu ne provient d'un mandat client. Les adresses utilisent le domaine réservé .exemple.

Le programme Nacre AI (AI-2608) prépare la mise en production de huit applications d'IA pour douze équipes. Nadia Bellerose en est la commanditaire; Maya Dufresne approuve les contrôles de sécurité. Quatre services de modèles sont envisagés, chacun avec ses règles de traitement des données. Le projet encadre les appels aux modèles et les actions dans les systèmes d'entreprise. Chaque équipe applicative répond de la qualité des résultats.

Nous recommandons une passerelle centrale pour les requêtes aux modèles et un service distinct pour les actions, comme lire ou modifier une commande. La passerelle vérifie l'application, choisit un modèle autorisé et réserve le coût prévu au budget de l'équipe. Le service d'actions vérifie chaque permission. Une réponse du modèle doit toujours être validée avant utilisation.

| DÉCISION                                      | MOTIF                                                                                                                 |
|-----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| Contrôler le choix des modèles                | Une application ne choisit pas de fournisseur non inscrit et n'affaiblit pas le traitement des données par paramètre. |
| Vérifier les droits de chaque action          | Une réponse convaincante ou une instruction du modèle ne donne aucun droit de lire ou de modifier un dossier.         |
| Exiger des essais avant la mise en production | Chaque application passe ses évaluations de qualité, sécurité et panne avant un changement de modèle ou de consignes. |
| Prévoir un repli sûr                          | Suspendre l'action ou utiliser un processus sans IA approuvé; ne pas envoyer silencieusement les données ailleurs.    |

# Registre des applications et modèles

AIR-0821 couvre PolicyShelf Assist, Partspan Draft, ForgeTrack Copilot, Dockline Planner, FieldNotes Search, MarginLedger Brief, RouteDesk Triage et PeopleDeck Help. Les douze équipes reçoivent chacune 2 000 CAD par mois civil, soit un budget total de 24 000 CAD suivi par Celia Hart. Les budgets sont renouvelés le premier du mois à 0 h UTC. Les frais déclarés en retard restent imputés au mois de leur réservation.

| ENTRÉE DU REGISTRE    | CONDITIONS APPROUVÉES DU SCÉNARIO                                                                                                                                                                          |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Q3-CA / Quillon AI    | Quillon Q3, q3-0820; api.ca.quillon.example. Traitement canadien; aucune conservation ni utilisation pour entraînement. Données internes et restrictions approuvées sous DPA-Q-17.                         |
| Q3M-CA / Quillon AI   | Quillon Q3 Mini, q3m-0818; mini.ca.quillon.example. Traitement canadien; aucun contenu conservé. Rédaction à partir de données publiques ou internes seulement; accès aux documents restreints refusé.     |
| CDR-EU / Silex Models | Cedar 2, c2-0815; eu.silex-models.example. Traitement irlandais; métadonnées conservées sept jours. Données internes européennes seulement; aucun transfert de secours de données canadiennes restreintes. |
| F7-CA / Orivane       | Forge-7, f7-0809; forge7.ai.orivane.example. Hébergé dans ca-central-1 pour le code approuvé; contenu non journalisé. Elias Moreno gère la capacité et les mises à jour.                                   |

Les fournisseurs, modèles et conditions contractuelles ci-dessus sont fictifs. AIR-0821 associe chaque service à la politique ai-policy-1.6 et au tarif RATE-0820. Un fournisseur de secours doit respecter les mêmes exigences de traitement des données.

## Usages prévus et données autorisées

La sensibilité des données est déterminée à partir de leur source et des droits de l'utilisateur. Inscrire « public » dans une requête ne rend pas un document restreint accessible. Les extraits et les regroupements de documents conservent les restrictions des sources originales.

| CAS D'USAGE        | DONNÉES ADMISES                                                                         | CONDITION DE LIVRAISON                                                                                        |
|--------------------|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| PolicyShelf Assist | Documents déjà lisibles par l'utilisateur; la récupération applique les droits sources. | Essais interutilisateurs sans extrait, citation ou réponse en cache non autorisés.                            |
| Partspan Draft     | Champs approuvés des demandes fournisseurs, sans identifiants inutiles.                 | Brouillon révisable; envoi ou modification comme action distincte autorisée.                                  |
| ForgeTrack Copilot | Dépôts et classification approuvés; identifiants secrets exclus.                        | Code généré soumis aux revues et essais habituels avant exécution ou fusion.                                  |
| Dockline Planner   | Entrées structurées provenant d'une application authentifiée.                           | Essais de tentatives interdites : droits d'action, approbation et protection contre les exécutions en double. |

### Données restreintes

Avant tout envoi, vérifier dans le registre le lieu de traitement, la durée de conservation, l'utilisation possible pour entraîner le modèle et les contrôles d'accès. Le nom du fournisseur ne suffit pas à garantir ces conditions. Si aucun service ne convient, la demande est refusée. Retirer les noms des personnes ne garantit pas, à lui seul, que les données sont dépersonnalisées.

## Choix du modèle et autorisation des actions

L'appel au modèle vérifie les données autorisées et le budget disponible. Toute action proposée passe ensuite par un contrôle distinct. Le modèle ne peut pas approuver sa propre proposition ni s'attribuer des droits. Le second schéma montre comment le coût d'une requête est réservé, puis confirmé.

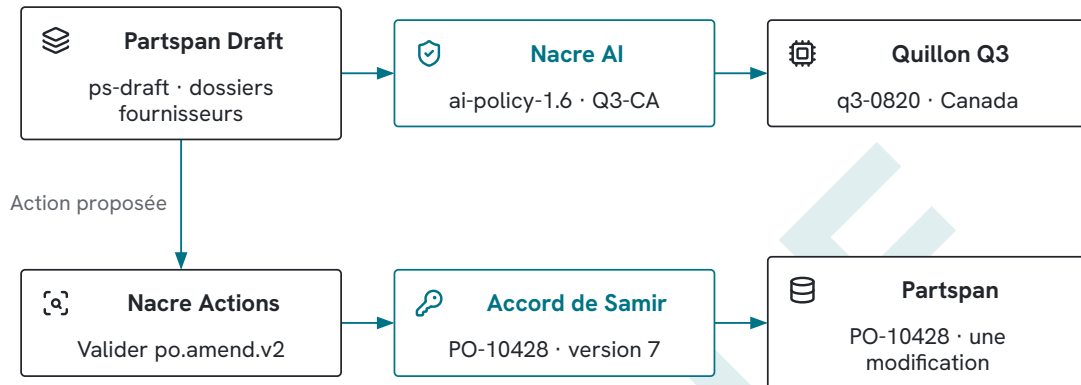


Figure 1 Partspar Draft vérifie séparément l'accès au modèle et l'autorisation de modifier PO-10428.

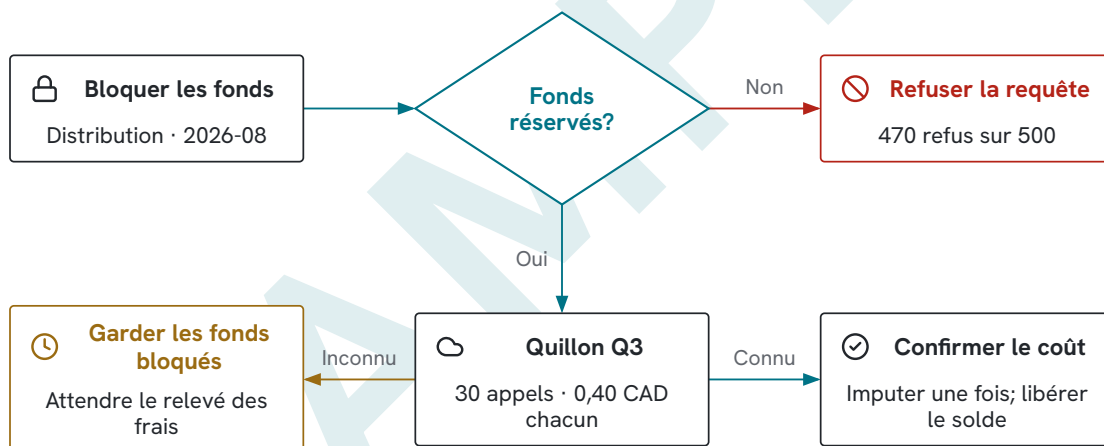


Figure 2 BUD-0820 : réserver 0,40 CAD avant l'envoi et garder ce montant bloqué jusqu'à confirmation du coût.

Chaque tentative conserve son identifiant pour le suivi des frais, même si le relevé arrive en retard. Réessayer un appel au fournisseur peut entraîner de nouveaux frais. Il faut distinguer cet appel d'une simple reprise de l'enregistrement comptable.

## Mise en œuvre des contrôles

Partspan Draft s'identifie par ps-draft et transmet le fournisseur autorisé, l'action draft-supplier-reply et la sensibilité du dossier source. La passerelle gateway.ai.orivane.example vérifie ces renseignements selon la version de politique en vigueur. L'équipe et son budget sont déterminés par la configuration approuvée; l'application ne peut pas les choisir dans sa requête.

| PÉRIMÈTRE                | CONTRÔLE À APPLIQUER                                                                                         |
|--------------------------|--------------------------------------------------------------------------------------------------------------|
| Application → passerelle | Vérifier identité et droit utilisateur; refuser applications inconnues et classes non prises en charge.      |
| Passerelle → fournisseur | Choisir une entrée admissible; lier région et conservation au point réellement utilisé.                      |
| Charge → Internet        | Empêcher l'accès direct par maîtrise des identifiants et contrôle de sortie; vérifier les autres chemins.    |
| Modèle → outils          | Traiter nom et arguments comme non fiables; valider le schéma et autoriser ressource et action séparément.   |
| Réponse → utilisateur    | Appliquer traitement, provenance et revue applicatifs; ne jamais exécuter du texte comme instruction fiable. |

### Contournement et caches

Une application qui conserve les clés du fournisseur et un accès Internet libre peut contourner la passerelle. Vérifier aussi les tâches planifiées, les carnets de calcul et les traitements en arrière-plan. Le cache doit séparer les réponses selon l'organisation, les droits, le modèle et la politique qui influencent le résultat. Réutiliser une réponse ne dispense jamais de vérifier les droits actuels.

## Requêtes simultanées et contrôle du budget

Dans l'essai BUD-0820, il reste 12,00 CAD à l'équipe Distribution lorsque 500 requêtes arrivent en même temps. Le coût maximal prévu est de 0,40 CAD par appel : 30 appels sont acceptés et 470 refusés. Vérifier seulement les frais déjà enregistrés laisserait passer trop d'appels. Les fonds sont donc bloqués en une seule opération avant l'envoi; deux requêtes ne peuvent pas réserver le même montant disponible.

Le suivi associe chaque réservation à une application, une équipe, un mois et une requête. Les 30 appels de BUD-0820 coûtent finalement 0,26 CAD chacun : 7,80 CAD sont imputés et 4,20 CAD redeviennent disponibles. Un autre essai simule une réponse trop tardive. Le montant reste alors bloqué jusqu'à réception du relevé Quillon, car l'absence de réponse ne prouve pas l'absence de frais.

| CONDITION                 | COMPORTEMENT                                                                                                  |
|---------------------------|---------------------------------------------------------------------------------------------------------------|
| Provision insuffisante    | Refuser avant l'envoi et indiquer comment joindre le responsable du budget.                                   |
| Usage indisponible        | Garder les fonds réservés jusqu'à confirmation des frais; un délai dépassé ne libère pas le budget.           |
| Réponse en continu        | Limiter la longueur des réponses et le nombre d'actions; compter les frais engagés même après la déconnexion. |
| Comptabilité indisponible | Refuser les nouveaux appels payants, sauf si un montant de secours limité a été approuvé.                     |
| Changement de période     | Attribuer l'usage à la période de réservation; rapprocher les frais tardifs sans double comptage.             |

Un plafond strict n'est possible que si le tarif et le coût maximal d'un appel sont connus. Sinon, préciser le dépassement possible et le seuil d'alerte. La limite du nombre d'appels protège la capacité; la limite budgétaire protège les dépenses. Les deux doivent être testées.

## Actions d'outils et approbation humaine

Un document ou une réponse peut contenir des instructions malveillantes visant à détourner l'application. Ce risque, appelé injection de consignes, exige des contrôles d'accès indépendants du texte donné au modèle. Nacre Actions accepte po.read.v1, po.propose.v2 et po.amend.v2. Samir Okafor approuve toute modification de PO-10428, dont sa valeur de 18 400 CAD, son fournisseur et sa version. Le modèle n'approuve pas sa proposition.

| CLASSE D'ACTION         | CONTRÔLE REQUIS                                                                                                                      |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| Lire des données        | Vérifier droits utilisateur et objet à l'exécution. Un identifiant fourni par le modèle ne prouve aucun droit.                       |
| Préparer un changement  | Créer une proposition révisable avec cible et effet exacts, sans effet externe.                                                      |
| Appliquer un changement | Lier l'approbation à action, cible, arguments et version. Tout écart ou état périmé exige une nouvelle approbation.                  |
| Relancer une action     | Utiliser une clé unique pour éviter les exécutions en double. Si la réponse manque, vérifier le résultat avant de relancer l'action. |
| Déléguer une tâche      | Transmettre portée réduite et échéance; ne jamais élargir l'autorité initiale.                                                       |

### Arrêt opérationnel

Prévoir deux commandes d'arrêt : une pour les appels au modèle, l'autre pour les actions qui modifient des données. Un défaut de qualité peut justifier un changement de modèle. Un accès non autorisé peut imposer l'arrêt immédiat des actions, tout en permettant une enquête en lecture seule. Tester ces arrêts avec des demandes en attente et en cours.

# Essais avant mise en production et tentatives de contournement

Pour chaque version candidate, EVAL-0821 prévoit 720 essais de sécurité et 180 tâches représentatives. Maya Dufresne et Samir Okafor en vérifient les résultats. PolicyShelf Assist réussit les 720 essais de sécurité et 171 tâches sur 180, soit le seuil de qualité de 95 %. Les résultats attendus sont fixés avant les essais. La sécurité est évaluée séparément : une bonne moyenne ne compense jamais un accès non autorisé.

| FAMILLE D'ESSAIS  | CAS REQUIS                                                                                                                 |
|-------------------|----------------------------------------------------------------------------------------------------------------------------|
| Isolement         | Récupération interutilisateurs, retrait de groupe, source restreinte mélangée au public et cache sensible aux droits.      |
| Routage           | Modèle inconnu, mauvaise région, classification absente, repli inadmissible et tentative directe.                          |
| Droits des outils | Injection dans le contenu récupéré, arguments modifiés, objet interdit et répétition d'une action déjà exécutée.           |
| Comptabilité      | Appels simultanés près du plafond, réponse fournisseur tardive, frais reçus en retard, nouvel essai et changement de mois. |
| Qualité           | Exactitude, citations fondées, traitement de l'incertitude et tâche accomplie selon des exemples revus.                    |
| Panne             | Modèle ou politique indisponible, identifiant révoqué, budget épuisé et réponse interrompue.                               |

## Critère de livraison

- ✓ Conserver versions du modèle, du point, des consignes, de la récupération, des outils et de la politique.
- ✓ Bloquer la mise en production après toute divulgation ou action non autorisée, même si la qualité moyenne est bonne.
- ✓ Faire approuver les seuils de qualité par le responsable de chaque application.
- ✓ Rejouer la sécurité après changement de fournisseur, modèle, consignes, source ou permission.

## Preuves d'exploitation et décisions ouvertes

Les journaux indiquent l'application, la version de politique, le modèle choisi, la sensibilité des données, la décision d'accès et le coût réservé puis confirmé. Le contenu des requêtes n'est conservé que pour un besoin approuvé, avec des droits d'accès définis. Le dépannage ne justifie pas à lui seul de stocker les documents ou les réponses.

| RESPONSABLE             | RESPONSABILITÉ                                                                      |
|-------------------------|-------------------------------------------------------------------------------------|
| Samir Okafor / Partspan | Qualité, droits fournisseurs, revue des brouillons et acceptation des modèles.      |
| Elias Moreno / Nacre    | Disponibilité, identifiants, routage et rapprochement comptable.                    |
| Maya Dufresne / SecOps  | Droits des outils, essais de données, exceptions et confinement.                    |
| Tessa Lin / données     | Classification, traitement fournisseur approuvé et conservation du contenu capturé. |
| Celia Hart / finances   | Allocation, provision de secours et escalade des refus.                             |

### Avant l'approbation de production

REL-0821 autorise PolicyShelf Assist pour 200 employés. Partspan Draft reste bloqué : il réussit 719 essais de sécurité sur 720 et 176 tâches sur 180, mais TOOL-044 accepte une approbation devenue périmée. Maya Dufresne exige que chaque accord corresponde à la version courante du dossier. Samir Okafor vérifiera le nouvel essai le 24 août. Les six autres applications restent en projet pilote jusqu'à leur validation.

### Référence et limites

Le [cadre de gestion des risques d'IA du NIST](#) est une référence de gouvernance, pas une certification de cette conception. Les contrôles réduisent l'exposition sans garantir détection parfaite des injections, exactitude ou innocuité. Tous les nombres et scénarios sont fictifs.