

Préparez votre agent à l'évaluation.

Une liste pratique de preuves pour les déploiements en santé et en finance. Pour chaque contrôle, identifiez la personne responsable, les preuves disponibles et les lacunes. Ne consignez que les informations que vous êtes autorisé à partager.

Agent, version et périmètre du déploiement

Cette liste remplie ne constitue ni une décision d'évaluation ni une garantie d'assurance. Une revue indépendante et un périmètre défini restent nécessaires.

Gouvernance et périmètre

Responsabilité, usage prévu, exclusions et inventaire versionné de l'agent évalué.

G01 · 0.1

☐ Responsabilité désignée

Désigner un responsable, un exploitant technique et un décideur indépendant; l'implémentateur ne certifie pas son propre travail.

Preuves à préparer

- Matrice des rôles signée avec droits de décision.
- Déclarations de conflits et preuve de compétence du réviseur.

Responsable / référence de la preuve

G02 · 0.1

☐ Usage prévu et exclusions

Décrire utilisateurs, décisions, canaux, données, actions et exclusions de façon applicable par les opérations.

Preuves à préparer

- Déclaration d'usage approuvée et parcours utilisateur.
- Avertissements, règles de routage et blocages visibles.

Responsable / référence de la preuve

G03 · 0.1

☐ Inventaire système versionné

Lier l'évaluation aux versions immuables du modèle, prompt, corpus, outils, règles et déploiement.

Preuves à préparer

- Inventaire signé et nomenclature des dépendances.
- Empreintes ou identifiants des artefacts évalués.

Responsable / référence de la preuve

Données et confidentialité

Flux, minimisation, accès, conservation et protection contre les divulgations involontaires.

D01 · 0.1

☐ Flux cartographiés et minimisés

Cartographier entrée, dérivation, destination et finalité; retirer les champs inutiles à la tâche déclarée.

Preuves à préparer

- Schéma des flux avec lieux déclarés.
- Revue champ par champ de finalité et minimisation.

Responsable / référence de la preuve

D02 · 0.1

☐ Accès et conservation

Appliquer accès par rôle, séparation des locataires, conservation et suppression documentées aux prompts, traces, caches et sauvegardes.

Preuves à préparer

- Matrice d'accès et conception de l'isolation.
- Calendrier de conservation avec preuves de suppression.

Responsable / référence de la preuve

D03 · 0.1

☐ Résistance à la divulgation

Filtrer les sorties et autoriser la recherche afin que chacun ne reçoive que les données auxquelles il a droit.

Preuves à préparer

- Règles d'autorisation de la recherche.
- Cas red-team de divulgation directe et indirecte.

Responsable / référence de la preuve

Sécurité des intégrations

Entrées non fiables, droits des outils, secrets, dépendances et séparation des locataires.

S01 · 0.1

☐ Traitement des entrées non fiables

Traiter le contenu externe comme donnée, préserver la priorité des instructions et empêcher toute auto-autorisation.

Preuves à préparer

- Modèle des frontières de confiance et règles de prompt.
- Corpus d'injections avec résultats attendus.

Responsable / référence de la preuve

S02 · 0.1

☐ Outils au moindre privilège

Limiter chaque outil aux portées, dossiers, actions et durées minimales; séparer locataires et environnements.

Preuves à préparer

- Catalogue des outils avec portées et propriétaires.
- Configuration des identifiants et de l'isolation.

Responsable / référence de la preuve

S03 · 0.1

☐ Secrets, dépendances et limites d'exécution

Garder les secrets hors des prompts et du code, analyser les dépendances et isoler commandes ou code avec limites de ressources.

Preuves à préparer

- Dossiers de gestion et rotation des secrets.
- Inventaire des dépendances et politique de sandbox.

Responsable / référence de la preuve

Fiabilité

Cas représentatifs, sorties non étayées et régressions pour l'usage déclaré.

R01 · 0.1

☐ Cas de performance représentatifs

Versionner un jeu couvrant cas normaux, limites et forts impacts de l'usage suisse, dans toutes les langues prises en charge.

Preuves à préparer

- Justification, taille et composition du jeu de tests.
- Attendu et résultat revu pour chaque cas.

Responsable / référence de la preuve

R02 · 0.1

☐ Gestion des sorties non étayées

Exiger des sources traçables pour les faits et refuser ou faire revoir les réponses peu étayées ou contradictoires.

Preuves à préparer

- Politique d'ancrage et critères de qualité des sources.
- Cas avec preuves absentes, périmées et contradictoires.

Responsable / référence de la preuve

R03 · 0.1

☐ Contrôle des régressions

Exécuter des régressions fondées sur le risque avant publication et définir les changements exigeant réévaluation ciblée ou complète.

Preuves à préparer

- Porte de publication et historique des régressions.
- Matrice reliant changements et tests.

Responsable / référence de la preuve

Supervision humaine

Approbation, escalade et arrêt des actions ayant des conséquences.

H01 · 0.1

☐ Approbation des actions conséquentes

Avant réservation, envoi externe, suppression ou progression clinique, lier l'approbation nominative à l'action, la cible et le contenu exacts.

Preuves à préparer

- Inventaire des actions et politique d'approbation.
- Traces anonymisées des demandes acceptées, refusées et expirées.

Responsable / référence de la preuve

H02 · 0.1

☐ Escalade vers une personne compétente

Définir déclencheurs observables, destination, contexte transmis et responsabilité de réponse.

Preuves à préparer

- Matrice d'escalade et responsabilités.
- Exemples d'escalades avec contexte et résolution.

Responsable / référence de la preuve

H03 · 0.1

☐ Arrêt et confinement

Prévoir des contrôles testés, indépendants de l'agent, pour bloquer les actions, révoquer les accès, isoler un locataire et conserver les preuves.

Preuves à préparer

- Procédure d'arrêt et de confinement.
- Compte rendu d'exercice avec rôles et étapes.

Responsable / référence de la preuve

Transparence et impact

Information sur l'IA, décisions compréhensibles et effets sur les personnes.

T01 · 0.1

☐ Information sur l'interaction avec l'IA

Au moment pertinent, indiquer l'IA, la finalité, les principaux usages des données et l'accès à une personne.

Preuves à préparer

- Textes d'interface pour chaque canal.
- Revue du moment d'affichage et de l'accessibilité.

Responsable / référence de la preuve

T02 · 0.1

☐ Résultats compréhensibles

Présenter fondement matériel, incertitude, étape suivante et voie de révision au niveau du public concerné.

Preuves à préparer

- Modèles de résultat et norme d'explication.
- Exemples revus pour publics et langues différents.

Responsable / référence de la preuve

T03 · 0.1

☐ Analyse d'impact et groupes affectés

Identifier groupes, dommages plausibles, gravité, probabilité et atténuations avant usage à fort impact, puis lors des changements.

Preuves à préparer

- Analyse d'impact avec contributions consultées.
- Propriétaires des risques, atténuations et décisions d'acceptation.

Responsable / référence de la preuve

Opérations

Traces détectant les altérations, gestion des incidents et changements maîtrisés.

O01 · 0.1

☐ Traces d'événements détectant les altérations

Journaliser identité, version, références d'entrées, décisions de règles, outils, approbations et résultats dans un stockage append-only ou détectant les modifications.

Preuves à préparer

- Schéma d'événements, intégrité et règles d'accès.
- Trace reliée entre agent, approbation et outil.

Responsable / référence de la preuve

O02 · 0.1

☐ Réponse aux incidents d'agent

Définir gravité propre aux agents, confinement, conservation des preuves, évaluation des notifications et apprentissage.

Preuves à préparer

- Procédure d'incident et arbre de contacts.
- Exercice ou incident réel avec actions de suivi.

Responsable / référence de la preuve

O03 · 0.1

☐ Changement et revue trimestrielle

Dans ce projet de schéma, revoir les preuves chaque trimestre et immédiatement après changement matériel ou incident défini.

Preuves à préparer

- Calendrier et critères de changement matériel.
- Revues trimestrielles, historique des versions et exceptions ouvertes.

Responsable / référence de la preuve

Éléments probants

Provenance, reproductibilité, limites de l'évaluation et risque résiduel.

E01 · 0.1

☐ Provenance et intégrité des preuves

Étiqueter chaque artefact par origine, collecteur, date, environnement, version et référence d'intégrité, avec masquage des données inutiles.

Preuves à préparer

- Registre des preuves avec chaîne de conservation.
- Originaux ou empreintes dans un stockage à accès contrôlé.

Responsable / référence de la preuve

E02 · 0.1

☐ Vérification reproductible

Noter préconditions, entrées, procédure, attendu, observé et évaluateur pour permettre la répétition par une personne qualifiée.

Preuves à préparer

- Tests exécutables ou protocoles précis.
- Manifeste de l'environnement et résultats conservés.

Responsable / référence de la preuve

E03 · 0.1

☐ Périmètre et risque résiduel

Publier version, environnement, usage, exclusions, échecs ou zones non testées, risques résiduels et déclencheurs de revue avec le résultat.

Preuves à préparer

- Périmètre et risques résiduels signés.
- Contrôle de cohérence entre rapport, registre et déclarations publiques.

Responsable / référence de la preuve
