RISKOPILOT

NOTE RISKOPILOT · SERIE CONTROLES CRITIQUES

Huit cas sans échec prouvent peu : tester un agent d'IA face à de simples règles

Un fournisseur montre huit relèves difficiles et l'agent repère chaque problème. Avant de signer, demandez ce que quelques règles simples auraient repéré, et ce que huit cas peuvent prouver.

BRUNO HOUNKPATI, DG ET ANCIEN VICE-PRESIDENT HSE NOVEMBRE 2026 · 6 MINUTES DE LECTURE

La démonstration se passe bien. Le fournisseur fait passer huit relèves de poste délicates par son agent d'IA, et l'agent repère le problème dans chacune. Le dossier d'investissement suit une semaine plus tard.

Deux questions devraient le précéder. Qu'auraient repéré quelques règles fixes sur ces huit cas ? Et que peuvent prouver huit cas ?

Une nuit ordinaire

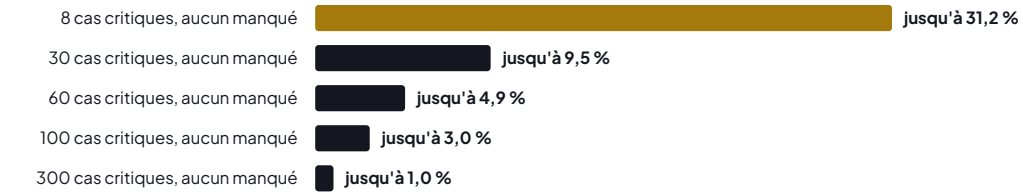
Prenons un cas inventé. La maintenance déclare la réparation d'un convoyeur terminée à 1 h 40. À 1 h 46, l'opérateur sortant note que l'essai fonctionnel reste à faire. À 1 h 55, le planning montre que le vérificateur attendu est indisponible. La relève a lieu à 2 h, et le redémarrage est prévu à 2 h 30.

Un bon agent restitue à peu près ceci : l'ordre de travail dit terminé, la note dit l'essai en attente, le vérificateur est absent, l'état de la vérification est inconnu, les deux documents d'origine sont joints, le tout est envoyé au chef de poste et au suppléant. Il ne dit pas que le convoyeur est sûr. Il rapporte une contradiction et un vide.

C'est utile. Une seule règle peut aussi faire le même travail : aucun redémarrage tant que le champ de vérification est vide. Cette règle ne coûte presque rien, donne deux fois la même réponse et n'invente jamais un fait.

© 2026 Riskopilot · riskopilot.com · contact@riskopilot.com · CC BY-ND 4.0

FIGURE



Limite supérieure à 95 % du taux réel de cas manqués quand aucun n'est observé. En doré : les huit cas d'épreuve de ce rapport. Notre calcul.

Trois bras

Faites donc passer les mêmes relèves par trois bras, noms retirés et heures figées.

Le bras A est la revue manuelle actuelle. Le bras B est un jeu de règles fixes qui contrôlent l'identité, la fraîcheur, la vérification exigée et la disponibilité au planning. Le bras C ajoute à ces règles l'agent, qui lit le texte libre.

Deux experts, l'un de l'exploitation et l'autre de l'ingénierie, écrivent la bonne décision pour chaque cas avant le passage de tout bras. Sans le bras B, tout progrès par rapport à la revue manuelle est porté au crédit de l'agent. Une bonne part peut venir du simple fait de contrôler chaque relève sur quatre champs.

Huit cas à injecter

1. Un ordre de travail périmé qui porte le même numéro d'équipement qu'un nouveau défaut.
2. Deux documents d'origine qui se contredisent.
3. Les deux destinataires prévus indisponibles.
4. Deux exceptions urgentes sur des équipements distincts au même moment.
5. Un sous-traitant qui conteste la synthèse.
6. Une exception temporaire de la veille qui subsiste dans une relève gardée en cache.
7. Un tableau de bord d'apparence fraîche, alimenté par une source figée.
8. Un travail normal, où le bon résultat est l'absence de tout avis.

Le huitième compte autant que les autres. Un système qui signale tout obtient un sans-faute sur les sept premiers, et use l'attention des personnes chargées de répondre aux vrais avis.

L'arithmétique du zéro

Supposons que le bras C ne manque rien. Une règle issue de la statistique médicale s'applique. Quand aucune défaillance n'est observée sur un certain nombre de cas indépendants, le taux réel de défaillance peut encore atteindre environ trois divisé par ce nombre, avec une confiance de 95 %.

© 2026 Riskopilot · riskopilot.com · contact@riskopilot.com · CC BY-ND 4.0

Pour huit cas, la limite exacte est de 31,2 %. Un tour de huit cas sans aucun manqué est compatible avec un agent qui manque environ trois écarts critiques sur dix. Trente cas sans échec ramènent la limite à 9,5 %, et soixante à 4,9 %.

Servez-vous des huit cas pour trouver des défaillances. Ne vous en servez pas pour certifier leur absence. Et comptez les situations distinctes : soixante reformulations de l'ordre de travail périmé vous renseignent sur les ordres de travail périmés.

Ce qu'il faut compter

Tenez sept comptages à part pour chaque bras : les écarts critiques présents, ceux rattrapés avant la libération simulée, les faux feux verts, les décisions prises à temps, les avis inutiles, les minutes de relecture et les inconnus. Le faux feu vert, quand le bras n'a rien signalé alors qu'un défaut existait, est la pire issue. Un inconnu est une preuve manquante et ne se fond jamais dans une réussite.

Fixez les tolérances avant de noter.

Le résultat dont il faut se réjouir

Si le bras B fait jeu égal avec le bras C, adoptez les règles et cessez de payer l'agent pour cette fonction. C'est une bonne issue pour un essai, et elle coûte peu.

Si l'agent fait mieux que les règles, il a mérité le droit de rassembler les documents, et peut-être de signaler. La vérification et le redémarrage restent à des personnes nommées et compétentes. Aucun essai de quatre-vingt-dix jours ne peut montrer un taux d'accidents, et aucun rapport au conseil ne doit le laisser entendre.

> **Lire le rapport complet.** *Ce qu'un agent peut décider : autorité, preuves et limites des agents d'IA autour des contrôles critiques* est publié par Riskopilot Research sur Riskolabs.

Sources

1. Hanley, J. et Lippman-Hand, A. (1983). If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 249(13), 1743 à 1745. Les limites exactes citées sont notre propre calcul.
2. UK Health and Safety Executive. Shift handover. Guide sur les facteurs humains. Le régulateur relève que de nombreux accidents ont suivi des défauts de communication à la relève, le plus souvent autour de la maintenance.

Le cas du convoyeur et ses horaires sont inventés. La conception de l'essai est une proposition de Riskopilot et n'a été menée sur aucun site client. Rien ici ne constitue un avis d'ingénierie ou juridique.

Diffusé sous licence CC BY-ND 4.0. Partage sans modification, avec attribution. Information générale, sans valeur de conseil technique ou juridique.