RISKOPILOT

NOTE RISKOPILOT · SERIE CONTROLES CRITIQUES

L'échelon que l'on n'accorde jamais : quelle autorité donner à un agent d'IA

Le débat sur les agents tient en deux mots, autonome ou supervisé. Un directeur de site a besoin de cinq mots, et d'une signature en face de chacun.

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

Une proposition arrive au comité exécutif. Un agent d'IA lira les notes de poste, les ordres de travail et le planning, et dira au chef de poste entrant ce qui compte. Le fournisseur parle d'autonomie supervisée. Quelqu'un demande si c'est sûr.

Posée ainsi, la question n'a pas de réponse. « Supervisé » ne dit pas ce que le logiciel a le droit de toucher. Mieux vaut demander quel échelon d'autorité l'agent détient, pour quelle fonction, et qui peut le lui retirer à deux heures du matin.

Pourquoi un agent ne peut pas porter une mission

Un contrôle critique fonctionne quand une personne compétente, ou une fonction technique, agit à temps. Trois tests distinguent une mission d'une tâche.

Son titulaire peut-il être sur place ? Un état physique est établi par quelqu'un qui voit le protecteur ou la consignation. Un relevé de l'état est autre chose.

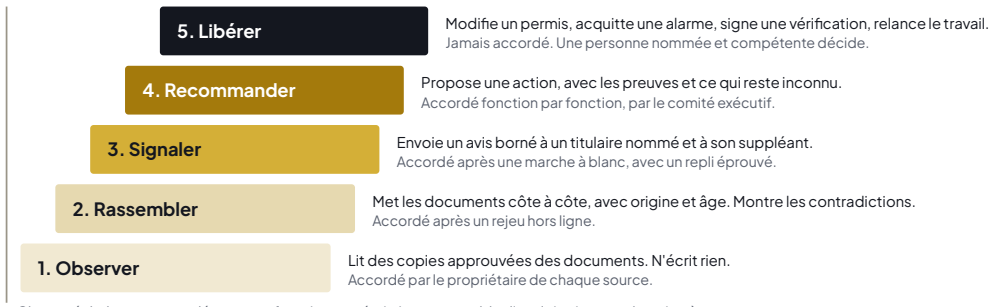
Est-il compétent au sens où la loi l'entend ? La vérification et la libération reviennent à des personnes dont la formation et la désignation peuvent être montrées à un régulateur.

Peut-il répondre du résultat ? Après un événement grave, un investigateur demande qui a décidé.

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

Un agent échoue aux trois. Il reste utile, parce qu'une quatrième sorte de travail se tient derrière tout contrôle : traiter de l'information. Un agent lit tout, à chaque fois, et remarque que deux documents sur un même équipement se contredisent.

FIGURE



Chaque échelon est accordé pour une fonction, par écrit. Le responsable d'exploitation peut le retirer à tout moment.

Cinq échelons

Nous proposons une échelle. Chaque échelon est accordé à un agent pour une fonction, par écrit.

1. **Observer.** L'agent lit des copies approuvées des documents et n'écrit rien. Le propriétaire de chaque source l'accorde.
2. **Rassembler.** Il met les documents côte à côte, avec leur origine et leur âge, et montre les contradictions. Accordé après un rejeu sur des relèves passées.
3. **Signaler.** Il envoie un avis borné à un titulaire nommé et à son suppléant. Accordé après une marche à blanc, avec un délai et un repli manuel que le poste de nuit a déjà pratiqué.
4. **Recommander.** Il propose une action, avec les preuves et ce qui reste inconnu. Accordé par le comité exécutif, fonction par fonction.
5. **Libérer.** Il modifie un permis, acquitte une alarme, signe une vérification ou relance le travail. Jamais accordé.

Ce que le droit dit déjà

Le règlement européen sur l'IA exige qu'un système à haut risque puisse être contrôlé par des personnes qui comprennent ses limites, restent conscientes du biais d'automatisation, peuvent infirmer son résultat et peuvent l'arrêter. L'organisation qui le déploie doit confier ce contrôle à des personnes qui ont la compétence, la formation et l'autorité nécessaires.

Savoir si un agent donné est juridiquement à haut risque est une question pour un juriste. Ces quatre capacités forment un bon test pour tout agent proche d'un contrôle. Si les personnes en poste ne peuvent pas faire ces quatre choses, la conception n'est pas prête.

Les dates d'application ont changé en juillet 2026. Les exigences du haut risque s'appliquent désormais à partir du 2 décembre 2027 pour les usages énumérés à l'annexe III du règlement, et du 2 août 2028 pour l'IA qui constitue un composant de sécurité d'un produit réglementé. Une date plus tardive laisse intacte l'obligation qu'a déjà l'employeur d'évaluer le risque.

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

Pourquoi l'échelon 4 est réservé

Une recommandation demande à une personne d'approuver. La recherche sur ce montage est ancienne et dérangeante. Une revue parue en 2010 constate que les personnes qui surveillent une automatisation manquent ce que l'aide ne signale pas et suivent ce qu'elle conseille à tort. Les experts étaient touchés comme les novices. Ni la formation ni les consignes ne l'empêchaient.

Demander à un chef de poste de rester vigilant est donc un contrôle faible. La conception doit rendre la contestation facile et attendue. Comptez combien de fois des personnes infirment l'agent. Un taux nul pendant de longs mois appelle une question.

Beaucoup de fonctions n'auront jamais besoin de l'échelon 4. Un exposé clair de ce que disent les documents, et de leurs désaccords, suffit souvent à une personne compétente pour décider.

L'objection qui mérite d'être entendue

« Les verrouillages agissent déjà sans personne. Un dernier échelon fermé relève du dogme. »

Ils agissent, et ils sont déterministes, validés par rapport à une spécification et régis par des normes de sécurité fonctionnelle. Aucune voie équivalente n'existe aujourd'hui pour un agent fondé sur un modèle de langage. Une norme harmonisée pour cet usage changerait notre avis. Une évaluation de terrain indépendante aussi. Des agents sont déjà déployés en maintenance industrielle. Nous n'avons trouvé aucune mesure indépendante de leur effet sur la vérification ou la libération.

Que faire lundi

Demandez quels agents lisent déjà des documents de sécurité dans vos opérations, et qui a approuvé cet accès. Pour chacun, demandez quel échelon il détient, par écrit. Demandez ensuite qui, sur le poste de nuit, peut le couper, et si cette personne l'a déjà fait.

> **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.

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

Sources

1. Règlement (UE) 2024/1689 (règlement sur l'IA), articles 14 et 26, lus en anglais le 19 novembre 2026 sur le service d'assistance de la Commission européenne. Reformulés par nos soins.
2. Règlement (UE) 2026/1744 du 8 juillet 2026, Journal officiel du 24 juillet 2026. Nouvelles dates d'application des exigences du haut risque.
3. Parasuraman, R. et Manzey, D. (2010). Complacency and bias in human use of automation: an attentional integration. *Human Factors* 52(3), 381 à 410. Résumé lu. Résultats de laboratoire et d'aviation, antérieurs aux agents fondés sur des modèles de langage.

L'échelle est une proposition de Riskopilot et n'a pas été testée au regard des accidents. Rien ici ne constitue un avis d'ingénierie ou juridique. La qualification d'un système d'IA au regard d'un droit donné relève de l'employeur responsable et de ses conseils.

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