

L'affirmation

Dans la plupart des dossiers d'autorisation d'un déploiement d'IA, une phrase porte un poids considérable : « un humain revoit chaque décision ». On la retrouve dans les contrats fournisseurs et les procès-verbaux. La décision : accepter cette phrase comme preuve d'un contrôle, ou la traiter comme une affirmation non démontrée au sujet d'une personne jamais rencontrée, exerçant un métier jamais chiffré.

Notre position est nette : un point de contrôle humain ne fonctionne que si la personne réunit quatre conditions à l'instant du refus : l'autorité de dire non sans conséquence de carrière, le temps de regarder vraiment, le contexte pour voir ce que la machine a vu et n'a pas vu, des incitations qui ne punissent pas le refus. Retirez l'une des quatre, et le point de contrôle devient un costume : il change la personne que l'on blâmera sans changer la probabilité de l'incident. Nous appelons cela du théâtre de responsabilité ; il passe les audits jusqu'au jour où il rencontre un sinistre.

Les faits

Commençons par la loi, plus honnête que bien des plans de déploiement. L'article 14 du règlement européen sur l'IA ne dit pas « placez un humain dans la boucle » : il exige des systèmes à haut risque conçus pour une supervision humaine effective, et énumère ce que le superviseur doit pouvoir faire.

PREUVE

L'article 14 du règlement européen sur l'IA exige des systèmes à haut risque conçus pour une supervision humaine effective : conscience du biais d'automatisation, interprétation correcte des sorties, faculté de ne pas utiliser le système, d'intervenir ou de l'arrêter ; mesures proportionnées aux risques, à l'autonomie et au contexte.

Source : règlement (UE) 2024/1689, article 14, texte vérifié en 2026 · Géographie : Union européenne · Méthode : texte réglementaire · SEG-5 · Réserve : devoir de conception imposé aux fournisseurs et déployeurs, pas garantie que la supervision a lieu ; application des obligations à haut risque reportée par l'omnibus numérique sur l'IA, règlement (UE) 2026/1744, en vigueur le 27 juillet 2026.

L'article 14 liste les conditions auxquelles un humain peut refuser une machine : le législateur ne suppose pas que l'humain fonctionnera, il ordonne un système qui rend ce fonctionnement possible.

Le biais d'automatisation est l'un des résultats les plus stables des facteurs humains. Skitka, Mosier et leurs collègues ont montré dès 1999 deux erreurs caractéristiques : l'omission, ne pas voir ce que l'aide n'a pas signalé, et la commission, suivre l'aide contre sa formation et des informations contraires valides. Parasuraman et Manzey (2010) ajoutent que la complaisance apparaît sous charge multitâche, l'état normal d'une file de revue en production, touche les experts comme les novices, et ne s'efface pas par la pratique.

PREUVE

Les études évaluées par les pairs constatent que les humains assistés d'une aide automatisée commettent des erreurs systématiques d'omission et de commission, et que la complaisance naît sous charge multitâche, touche aussi les experts et ne s'élimine pas par la seule pratique.

Source : Skitka, Mosier et Burdick, *International Journal of Human-Computer Studies* 51(5), 1999 ; Parasuraman et Manzey, *Human Factors* 52(3), 2010 · Géographie : laboratoire et simulateur, États-Unis et Europe principalement · Méthode : expériences contrôlées et revue intégrative · SEG-4 · Réserve : l'ampleur dans un déploiement de production donné se mesure, elle ne se suppose pas.

Le référentiel de gestion des risques de l'IA du NIST, volontaire mais devenu le vocabulaire commun de la gouvernance, range la supervision sous GOVERN comme un problème d'organisation et traite l'insertion d'un humain comme une décision de risque en soi.

La loi exige que le refus soit possible ; la psychologie le montre improbable sous charge s'il n'est pas délibérément protégé ; le référentiel en fait un travail de conception. Ce qu'aucune ne fournit, parce que c'est un jugement de firme, c'est notre test opérationnel : autorité, temps, contexte, incitation, les quatre à la fois, détenus par la personne nommée le pire jour du trimestre. Un point de contrôle qui n'a jamais refusé la machine ne prouve pas que la machine est bonne, mais que le contrôle est absent. Zéro est un constat.

Fiches de preuve

CLM-AIACT-ART14-OVERSIGHT-CAPABILITY · SEG-5. L'article 14 du règlement (UE) 2024/1689 impose une conception permettant la supervision humaine effective des systèmes à haut risque (conscience du biais d'automatisation, interprétation correcte, faculté de ne pas utiliser, d'intervenir, d'arrêter), avec des mesures proportionnées. Contexte : règlement (UE) 2024/1689, texte vérifié en 2026. Méthode : texte réglementaire. Preuve contraire : devoir de conception, pas d'effectivité observée ; application des obligations à haut risque reportée par le règlement (UE) 2026/1744, en vigueur le 27 juillet 2026. Confiance causale : aucune. Transférabilité : droit de l'Union ; persuasif ailleurs. Revue : 2026-08-02.

CLM-AUTOMATION-BIAS-PERSISTENCE · SEG-4. Erreurs systématiques d'omission et de commission auprès des aides automatisées ; complaisance sous charge multitâche, y compris chez les experts, non éliminée par la pratique. Contexte : Skitka, Mosier et Burdick 1999 ; Parasuraman et Manzey 2010. Méthode : expériences contrôlées et revue intégrative. Preuve contraire : provenance surtout de laboratoire ; ampleur à mesurer en production. Confiance causale : expérimentale pour les effets centraux. Transférabilité : files de revue à haut volume. Revue : 2026-08-02.

CLM-NIST-RMF-OVERSIGHT-RISK · SEG-5. Le NIST AI RMF (AI 100-1, janvier 2023) traite la supervision humaine comme un travail de conception organisationnelle sous GOVERN et identifie les configurations humain-IA, rôles de supervision compris, comme porteuses de risques propres. Contexte : référentiel américain volontaire, édition courante en 2026. Méthode : texte du référentiel. Preuve contraire : cadre volontaire, sans seuils ni ratios prescrits. Confiance causale : aucune. Transférabilité : toute organisation adoptant le vocabulaire du RMF. Revue : 2026-08-02.

CLM-FIRM-OVERSIGHT-FOUR-CONDITIONS · SEG-1 (position de la firme). Un point de contrôle humain ne fonctionne comme contrôle que si la personne nommée détient autorité, temps, contexte et incitation pour

refuser la machine ; à défaut, théâtre de responsabilité, et un taux de reprise nul est un constat d'absence, pas de qualité. Contexte : pratique de la firme dans le corridor. Méthode : interprétation, pas mesure ; aucune statistique de déploiement n'est citée. Preuve contraire : pour des tâches à faible enjeu et forte exactitude, un point de contrôle léger peut être proportionné ; le test, conjonctif par construction, n'est pas validé comme instrument noté. Confiance causale : aucune. Transférabilité : bornée aux décisions à haut risque revues sous charge de production. Revue : 2026-08-02.

Les limites

Les quatre conditions échouent de manière prévisible là où opèrent nos clients. Les déploiements arrivent avec les hypothèses d'effectifs du fournisseur : une file dimensionnée pour un back-office européen, des ratios minutes par dossier validés sur son marché. La société d'Abidjan hérite de la file sans le plan d'effectifs, et la supervision échoit à qui est disponible, au bout d'une ligne dont l'autorité réelle siège à Paris. Chaque condition se dégrade ensuite séparément. L'autorité : celui qui refuse contredit, depuis le bas de l'organigramme, un système payé par la holding. Le temps : la file n'a jamais été recalculée pour les effectifs locaux. Le contexte : le modèle a été entraîné sur des historiques de crédit formels que le corridor n'a pas. L'incitation : le débit se mesure, le refus est une friction. S'y ajoute une asymétrie réglementaire : le règlement européen atteint la holding ; aucun texte équivalent n'opère dans les juridictions de l'UEMOA où tourne le système. La supervision qui existe à Abidjan existe parce que le groupe l'a construite.

Sources et limites

Les faits juridiques et de référentiel reposent sur le règlement (UE) 2024/1689, article 14, sur son calendrier d'application modifié par le règlement (UE) 2026/1744 et sur le NIST AI 100-1 ; les constats comportementaux, sur la littérature du biais d'automatisation citée plus haut, dont la provenance de laboratoire est affichée, pas dissimulée. Le test des quatre conditions est la position de la firme, à lire comme une interprétation : il n'est pas validé comme instrument noté, et nous ne publions aucun indicateur de supervision de clients, aucun n'ayant franchi notre porte de diffusion des preuves. Limite de transférabilité explicite : l'argument est le plus fort pour des décisions à haut risque et à haut volume revues sous charge, le plus faible pour des tâches à faible enjeu où un contrôle léger peut être honnête et proportionné. Cette note ne vaut ni comme conseil juridique de conformité, ni comme affirmation que la supervision humaine serait inutile : elle doit être construite, dotée et mesurée avant d'être comptée comme un contrôle.

Que faire lundi

- 1. Le directeur des risques publie le taux de reprise humaine de chaque point de contrôle compté comme contrôle.** Un taux nul n'est pas un réconfort : c'est le signal le plus fort que le point de contrôle est un costume.
- 2. Le directeur des opérations recalcule l'arithmétique de la file, dossiers par relecteur et par jour contre minutes de revue réelle, sur les effectifs réels et non ceux du fournisseur.** Si elle ne boucle pas, la condition de temps a déjà échoué.
- 3. Le responsable de l'audit interne suit un refus réel de bout en bout.** Qui a été interrogé, quels indicateurs en ont souffert, ce qui a changé. Sans refus à suivre, revenez à la première action.



4. Le commanditaire du déploiement nomme les points de contrôle qui n'existent que pour désigner un responsable, et tranche ce mois-ci : les supprimer ou les doter. Une réponse honnête en nomme au moins un.