



IA et “human-in-the-loop” dans le secteur humanitaire

Vous êtes perplexe face aux termes liés à l'IA et vous ne savez pas comment les appliquer au travail humanitaire ? Ce guide pratique, élaboré avec l'expertise de Data Friendly Space, offre une introduction accessible au “human-in-the-loop”.

Ce guide a été créé en complément de l'initiative de recherche conjointe 2025-26 entre la Humanitarian Leadership Academy et Data Friendly Space : Comment les humanitaires utilisent-ils l'IA en 2025? Analyse des pratiques actuelles et du potentiel futur.

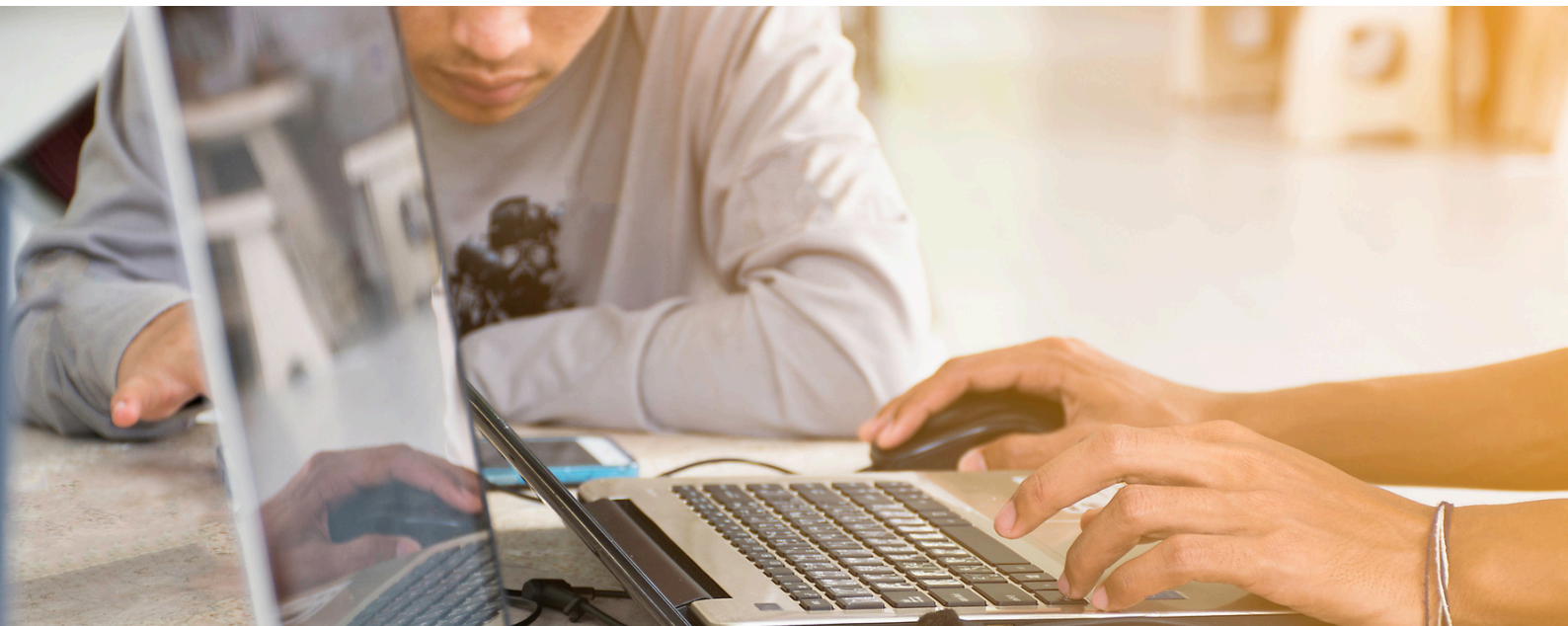
L'IA est utilisée dans l'ensemble du travail humanitaire, mais pas toujours de manière sûre

En 2026, 75% des humanitaires utilisent régulièrement des outils d'IA dans leur travail. Pourtant, moins d'un quart des organisations disposent d'une politique formelle en matière d'IA.

Lorsque l'IA fonctionne sans orientation humaine, elle peut :

- Générer des informations plausibles mais factuellement incorrectes
- Ignorer le contexte critique sur le terrain
- Refléter et amplifier les biais existants dans les données

Dans les contextes humanitaires, agir à partir de résultats d'IA erronés ne se limite pas à un gaspillage de ressources : cela peut détourner l'aide des personnes qui en ont le plus besoin. C'est pourquoi chaque résultat d'IA doit être accompagné d'un contrôle humain.



Que signifie "human-in-the-loop" (HITL) ?

Le "Human-in-the-Loop" (HITL) est une approche dans laquelle les personnes guident, valident, améliorent et/ou annulent activement les résultats de l'IA avant qu'une action soit prise.

Cette approche combine les forces des humains et des machines, garantissant que les tâches complexes ou à enjeux élevés bénéficient du jugement et de la supervision humains.

Le HITL ne vise pas à ralentir l'IA - il s'agit de s'assurer que l'IA travaille *pour* les personnes, et pas seulement *à propos* des personnes.

Le HITL n'est pas un simple point de contrôle. C'est un cycle continu intégré dans l'utilisation de l'IA, garantissant que les personnes concernées restent responsables vis-à-vis de leur organisation et des personnes qu'elles servent.



Pourquoi le “human-in-the-loop” est-il important dans les contextes humanitaires ?

Le principe de ne pas nuire s'applique également à l'IA.

L'IA peut ignorer le contexte. Elle peut refléter des biais dans les données d'entraînement. L'IA ne peut pas pleinement comprendre l'expérience vécue des personnes affectées par les crises.

Si un algorithme signale la mauvaise famille pour une exclusion ou classe incorrectement un risque de protection, le préjudice est réel.

Le préjudice n'est pas uniquement physique. Une IA mal gérée peut exposer des données sensibles, réidentifier des personnes ou permettre une surveillance. Une bonne surveillance protège les personnes à tous les niveaux.

C'est grâce à la vérification humaine que nous repérons ce genre de problème avant qu'il n'atteigne les personnes que nous essayons d'aider.



Le “human-in-the-loop” ne doit pas être un cadre abstrait

Le “human-in-the-loop” doit être mis en œuvre à tous les niveaux de l'utilisation de l'IA. À quoi cela ressemble-t-il en pratique ?

- **Définissez les points de décision tôt** : Avant de déployer tout outil d'IA, identifiez exactement où une vérification humaine est nécessaire. Ne laissez pas cela au hasard et ne supposez pas que quelqu'un d'autre vérifie.
- **Associez le réviseur à la décision** : La personne qui examine un résultat d'IA doit avoir les connaissances et l'autorité pour agir en conséquence. Un responsable WASH, un spécialiste de la protection et un analyste de données détecteront chacun des choses différentes.
- **Facilitez la remise en question** : Si contester une recommandation de l'IA semble risqué, bureaucratique ou décourageant, les gens ne le feront pas. Mettez en place des cultures et des systèmes où la remise en question des résultats de l'IA est la norme.
- **Documentez chaque modification**: Lorsqu'une personne modifie ou rejette un résultat d'IA, enregistrez-le. Ces décisions sont précieuses. Elles révèlent où l'IA est sous-performante, améliorent les modèles futurs et aident les autres à savoir à quoi faire attention.

Les êtres humains apportent des compétences spécialisées que l'IA ne peut pas reproduire

Votre expertise n'est pas un complément à l'IA : c'est ce qui rend l'IA sûre à utiliser.

- **Connaissance locale** : Les données signifient des choses différentes selon les endroits. Un chiffre qui semble normal dans un contexte peut signaler une crise dans un autre. Seules les personnes qui connaissent un endroit peuvent prendre cette décision.
- **Jugement éthique** : Décider ce qu'il faut publier, ce qu'il faut protéger et qui pourrait être affecté est un acte de responsabilité. Cela ne peut pas être délégué à un algorithme.
- **Sensibilité culturelle** : L'IA ne peut pas reconnaître de manière fiable les cadrages inappropriés, les hypothèses nuisibles ou les angles morts culturels spécifiques. Les réviseurs humains détectent ce que les modèles manquent.
- **Interpréter les lacunes**: Le constat le plus important est parfois ce que les données ne disent pas. Qui n'est pas compté, ce qui est absent, ce que signifie le silence. L'IA n'a aucun moyen de savoir ce qu'elle ne sait pas.



Comment pouvons-nous maintenir les personnes véritablement en contrôle ?

Le “human-in-the-loop” n'est aussi solide que la culture et les processus qui le soutiennent.

- **Incluez les communautés affectées** : Le “human-in-the-loop” ne doit pas uniquement concerner le personnel. Dans la mesure du possible, créez des canaux permettant aux personnes affectées par les crises de signaler des erreurs, de poser des questions ou de contester les décisions prises à leur sujet.
- **Soyez attentif aux biais d'automatisation** : Les personnes ont tendance à faire confiance aux résultats de l'IA, surtout sous pression. Formez vos équipes à remettre activement en question les résultats, et pas seulement à les approuver sans réfléchir.
- **Réviser la supervision à mesure que les outils évoluent** : Les systèmes d'IA changent. Le processus de supervision que vous avez conçu il y a six mois peut ne plus être suffisant. Intégrez des points de contrôle réguliers pour réévaluer si la révision humaine est toujours pertinente.
- **Assignez des responsabilités claires** : Chaque décision éclairée par l'IA doit avoir une personne nommée responsable du résultat. Si personne ne s'en charge, personne ne le détectera lorsque quelque chose tourne mal.
- **Agissez lorsque les choses tournent mal** : Si un résultat d'IA cause un préjudice ou qu'une erreur grave est commise, arrêtez, signalez-le et assurez-vous que les bonnes personnes sont informées. La responsabilité signifie agir sur ce qui a mal tourné.

Toutes les données générées par l'IA ne présentent pas le même risque : en cas de doute, vérifiez

Savoir quand examiner de plus près est une compétence fondamentale du HITL. Le temps nécessaire à la vérification est toujours inférieur au coût d'une action basée sur un résultat erroné. Cela peut inclure des situations où :

- **Le résultat concerne des populations affectées :** Les conclusions sur les communautés déplacées, les survivants de violences ou les groupes marginalisés nécessitent une interprétation sensible à la protection que l'IA ne peut pas fournir.
- **Le contexte évolue rapidement ou est politiquement sensible :** Dans les crises actives, l'IA peut prendre du retard sur les réalités du terrain. La vérification humaine ancre les résultats dans ce qui se passe réellement sur le terrain.
- **Le résultat semble surprenant ou étrangement parfait :** Les anomalies et les résultats trop propres peuvent signaler des erreurs ou des lacunes dans la couverture des sources. Faites confiance à votre instinct pour approfondir.
- **Le résultat déterminera une décision importante :** L'allocation des ressources, la réponse aux crises et les stratégies d'intervention comportent tous des enjeux élevés. Ces résultats exigent une validation par des experts avant utilisation.

Questions que tout humanitaire devrait se poser

Avant d'agir sur un résultat d'IA : un exemple de liste de vérification HITL.

- Ce résultat a-t-il été produit dans un cadre défini par des humains ?
- A-t-il été vérifié par rapport à des sources fiables ?
- Reflète-t-il le contexte local, y compris ce qui pourrait manquer ?
- Ce résultat ou cette décision pourrait-il causer un préjudice aux personnes concernées ?
- Est-il sûr et approprié de le partager avec les décideurs ?
- Qui a entraîné ce modèle, et sur quelles données ?
- Ce résultat reflète-t-il un besoin réel ou une lacune dans nos données ?
- Une personne affectée peut-elle questionner ou contester cette décision ?
- Cet outil d'IA compromet-il notre neutralité ou notre indépendance ?
- Y a-t-il une personne responsable du résultat ?
- Que se passe-t-il lorsque l'IA se trompe ?

Nos principes ne s'arrêtent pas lorsque nous utilisons des outils d'IA

Les valeurs qui guident l'action humanitaire - ne pas nuire, impartialité, neutralité, indépendance - ne cessent pas de s'appliquer lorsque nous utilisons l'IA.

Le cadre "human-in-the-loop" est la façon dont nous portons ces principes à l'ère de l'IA.

Ressources

La HLA, en partenariat avec Data Friendly Space, mène une étude mondiale sur la façon dont les humanitaires utilisent l'IA. Accédez aux publications de recherche, articles, webinaires et podcasts sur notre [Resources Hub](#).



Accédez à des formations en ligne gratuites sur kayaconnect.org. Obtenez des badges numériques HPass et partagez-les sur les réseaux sociaux pour informer les autres de vos apprentissages dans ce domaine.

