Évaluation d'une IA responsable

Un parcours de décision : d'abord si l'IA a sa place dans le processus, ensuite comment l'utiliser de façon responsable

L'IA est un outil, pas un objectif. Cette évaluation suit la décision dans l'ordre : l'étape un demande si l'IA est le bon outil, l'étape deux liste les conditions pour l'utiliser de façon responsable. S'arrêter après l'étape un est un résultat valable, et souvent le meilleur.

Étape 1 : décider si l'IA est le bon outil

- [ ] L'entrée est-elle vraiment désordonnée : texte libre, images, jugements qui ne se réduisent pas à des règles ?
- [ ] Avez-vous vérifié qu'une règle, une table de correspondance ou un formulaire bien conçu ne suffit pas ?
- [ ] Une mauvaise réponse est-elle bon marché, ou rattrapée par une personne avant de faire des dégâts ?
- [ ] La tâche doit-elle s'exécuter à une échelle ou une vitesse qu'aucune personne ne peut tenir ?

Chaque case que vous ne pouvez pas cocher est un signal d'arrêt. La solution plus simple est plus rapide, moins chère, testable et explicable, et personne ne doit la surveiller. Construisez-la d'abord ; si elle résout déjà le problème, vous avez votre réponse.

Étape 2 : les conditions d'un usage responsable

L'IA a passé l'étape un. Voici les conditions avant qu'elle ne touche de vraies personnes ou de vraies données :

- [ ] Vous savez exactement quelles données le modèle voit, et vous envoyez le minimum que la tâche exige
- [ ] Pour les données personnelles ou sensibles, une base légale existe, et ce que le fournisseur du modèle conserve est traité comme une question de traitement de données avec une vraie réponse
- [ ] Une personne relit la sortie avant qu'elle n'atteigne les personnes concernées, partout où une erreur ferait mal
- [ ] Quand le système doute, il s'arrête au lieu de produire quelque chose de plausible et faux
- [ ] Les erreurs ont un responsable : qui les rattrape, comment elles sont corrigées, qui informe la personne concernée
- [ ] Les personnes concernées savent qu'une IA intervient, en langage clair
- [ ] La qualité est mesurée : une référence avant, un jeu d'évaluation de cas difficiles après, et un retest à chaque changement de modèle

- [ ] Le coût continu est budgété : surveillance, évaluations, plafonds de dépense

Chaque case non cochée est un travail à planifier avant le lancement, pas une note pour plus tard.

Signaux d'alarme : là où l'IA ne doit pas décider

Cochez ce qui s'applique à votre cas. Une seule coche signifie que l'IA peut préparer la décision, mais pas la prendre :

- [ ] La décision a des conséquences juridiques, médicales ou financières pour une personne
- [ ] Personne dans votre organisation ne saurait expliquer pourquoi le système a refusé quelqu'un
- [ ] Une erreur reste invisible jusqu'à ce qu'elle ait déjà fait des dégâts
- [ ] Les gens croient qu'un humain fait ce travail, et se sentiraient trompés d'apprendre le contraire
- [ ] Vous hésiteriez à expliquer cet usage de l'IA sur votre propre page d'accueil

C'est ainsi que nous décidons chez Punfyre, y compris quand la décision est non. Si votre cas échoue à l'étape un, c'est un résultat : la solution ennuyeuse gagne sur tous les plans.