

Interview 2 avec un chercheur en IA en personne:

Pourquoi l'open science est-elle si essentielle lorsqu'on travaille sur un projet lié à la "risk literacy" ?

L'open science permet d'accéder librement aux données, ce qui est indispensable quand on veut comprendre comment les individus perçoivent le risque ou l'incertitude. Sans données accessibles et ouvertes, il est très difficile d'évaluer objectivement les comportements, les biais ou l'efficacité de méthodes pédagogiques. Pour des projets comme EduRisk, l'open science est le socle de toute expérimentation crédible.

De manière plus globale, que permet l'open science dans le domaine de la résilience face à l'incertitude ?

Elle facilite la réplique, la comparaison et l'amélioration des modèles éducatifs. Lorsque des équipes partagent leurs publications, leurs protocoles et leurs sets de données, on peut vérifier leurs conclusions, tester d'autres contextes, et affiner progressivement ce qui fonctionne. Cela crée un écosystème où la compréhension du risque devient plus transparente et moins dépendante de pratiques opaques.

Dans notre projet, une étape clé est l'extraction de l'information issue d'articles scientifiques. Qu'est-ce qui pourrait rendre cette tâche complexe ?

L'absence de standardisation. Même au sein de revues scientifiques, la structuration varie énormément : tables avec lignes mal définies, graphiques intégrés en image, paragraphes fusionnés, scans de mauvaise qualité... L'IA peut aider, mais elle doit être combinée à des choix méthodologiques précis.

Quels outils recommanderiez-vous pour fiabiliser le scraping des textes ?

Pour le texte brut, un OCR de qualité comme *doctr* est indispensable. Pour les PDF numériques, *pdfplumber* fait un excellent travail. Mais dans tous les cas, il faut prévoir une couche de post-traitement : nettoyage, re-structuration des paragraphes et vérification manuelle ponctuelle. Une extraction irréprochable n'existe pas sans une validation humaine minimale.