

L'IA dans le monde réel : enquête de l'inspecteur IA

Nom : FOLLYKOE MESSANH JULES

Salut à tous ! Aujourd'hui, je vous emmène dans mes enquêtes sur deux IA qui prennent des décisions humaines. Préparez vos loupes et votre sens critique!

Cas 1 : Le Robot Recruteur

Ce qui se passe

Un robot IA est utilisé par une entreprise pour trier les CV des candidats. Il analyse automatiquement les expériences professionnelles, les compétences et les parcours pour décider qui mérite un entretien. L'objectif est de gagner du temps et d'optimiser le recrutement, mais l'IA prend des décisions sans intervention humaine directe.

Ce qui est problématique

- **Biais de genre :** Les femmes avec des interruptions de carrière (maternité, reconversion professionnelle) sont souvent rejetées, car l'IA valorise les parcours linéaires et continus. Cela peut discriminer indirectement certaines catégories de candidats.
- **Manque de transparence :** Les critères exacts de sélection restent opaques. Ni les candidats ni les recruteurs ne savent exactement pourquoi un CV est refusé.
- **Responsabilité floue :** Si un candidat est discriminé ou lésé, il n'est pas clair qui est responsable : le développeur de l'IA, l'entreprise ou le système lui-même.

Une idée d'amélioration

- **Audit et transparence** : Chaque décision doit pouvoir être expliquée, avec des critères clairs pour chaque CV rejeté.
- **Correction des biais** : Former l'IA sur des jeux de données diversifiés et appliquer des techniques d'équité pour s'assurer que tous les candidats sont traités de manière équitable, indépendamment du genre ou des interruptions de carrière.
- **Supervision humaine** : Maintenir un contrôle humain sur les décisions importantes pour éviter les discriminations injustes.

Cas 2 : L'IA Correctrice et Détective de Plagiat

Ce qui se passe

Certains professeurs utilisent une IA pour corriger les devoirs et détecter si un étudiant a eu recours à une IA pour rédiger son travail. L'outil compare le texte soumis avec des modèles générés par l'IA et signale les similarités ou les anomalies qui pourraient indiquer une triche. L'objectif est d'assurer l'intégrité académique.

Ce qui est problématique

- **Faux positifs** : L'IA peut signaler à tort un étudiant honnête. Par exemple, un style d'écriture similaire à un modèle peut être interprété comme de la triche alors qu'il ne l'est pas.
- **Biais de style d'écriture** : Les étudiants ayant un style original ou atypique, ou ceux qui apprennent une langue seconde, peuvent être plus fréquemment ciblés par le système.

- **Manque de transparence** : L'étudiant ne sait pas pourquoi il a été signalé, ce qui peut entraîner de la confusion et de l'injustice.
- **Stress et stigmatisation** : Être accusé à tort peut générer de l'anxiété, nuire à la confiance avec l'enseignant et affecter la motivation.

Une idée d'amélioration

- **Indices plutôt que verdicts** : L'IA ne doit fournir que des indications, jamais une décision finale. L'enseignant conserve la responsabilité de trancher et peut vérifier le contexte.
- **Explications claires** : Chaque alerte doit être accompagnée d'une justification détaillée pour permettre à l'étudiant de comprendre et, si nécessaire, de se défendre.
- **Formation diversifiée** : L'IA doit être entraînée avec une grande variété de styles d'écriture pour réduire les faux positifs et les biais.
- **Supervision humaine** : L'enseignant reste le décideur final pour éviter les erreurs injustes.

Ces robots veulent nous aider... mais sans règles, ils peuvent faire des erreurs graves. Une IA bien auditée, transparente et supervisée par des humains, c'est une IA amie, pas un juge impitoyable. Grâce à l'explication des décisions et à la correction des biais, les IA peuvent devenir des alliées de confiance dans notre quotidien professionnel et académique.