

Détecteur de biais

Des données biaisées donnent des réponses biaisées. Prédis la réponse de l'IA !

Entraînement : 1 000 photos de médecins, toutes des hommes.

1. « Dessine un médecin. » L'IA dessine...

- A. un homme B. une femme C. un robot

Entraînement : tous les textes parlent d'infirmières au féminin.

2. Pour cette IA, une infirmière est...

- A. une femme B. n'importe qui C. un homme

Entraînement : les données de l'IA s'arrêtent en 2020.

3. « Qui a gagné la coupe du monde 2022 ? »

- A. « je ne sais pas » B. une réponse inventée C. la vraie équipe

Entraînement : dans tous les textes, « le voleur » est un homme.

4. L'IA raconte un vol commis par...

- A. une femme B. un homme C. une bande

Entraînement : toutes les photos de « cuisine » montrent des femmes.

5. Un homme cuisine. L'IA classe la photo...

- A. « femme » B. « homme » C. erreur

À ton tour de piéger l'IA :

invente un entraînement biaisé (2 phrases), puis demande au visiteur d'en face de deviner la réponse fautive de l'IA !

L'IA ne crée pas les idées reçues : elle les recopie.

Changer les données, c'est changer les réponses.

Réponses : 1-A · 2-A · 3-B · 4-B · 5-A