

Les biais dans l'intelligence artificielle.

1/ Comment une IA apprend-elle ?

Une intelligence artificielle n'apprend pas comme un humain. Elle analyse un grand nombre d'exemples pour repérer des ressemblances et créer des règles.

Dans notre activité :

- les images de chats utilisées pour l'entraînement étaient surtout des chats roux.
- les images de chiens étaient surtout des chiens noirs.

L'IA a donc appris certaines associations :

- "roux = chat"
- "noir = chien" **au lieu de reconnaître réellement l'animal.**

2/ Une IA prend ses décisions avec des probabilités :

Quand on montre une nouvelle image à l'IA, elle la compare avec les images qu'elle connaît déjà.

Elle donne ensuite :

- une réponse
- un pourcentage de confiance.

Ce pourcentage indique à quel point l'IA pense avoir raison

Chats noirs testés		Chiens roux testés	
Image	Réponse de l'IA	Image	Réponse de l'IA
Chat noir n°1	Chat à 62 %	Chien roux n°1	Chien à 100 %
Chat noir n°2	Chien à 95 %	Chien roux n°2	Chat à 96 %
Chat noir n°3	Chien à 100 %	Chien roux n°3	Chat à 100 %

3/ Analyse des résultats :

Ces résultats montrent plusieurs choses importantes :

1. L'IA ne comprend pas réellement les images

Elle ne sait pas ce qu'est un chat ou un chien comme un humain.

Elle cherche surtout :

- des ressemblances, des formes, des couleurs, des détails présents dans les images d'entraînement.

2. L'IA peut se tromper

Certains chats noirs ont été reconnus comme des chiens.

Certains chiens roux ont été reconnus comme des chats.

-> Cela montre que l'IA a été influencée par les couleurs présentes dans les données d'apprentissage.

3. Une IA peut être très sûre... tout en ayant tort

Par exemple : - un chien roux identifié comme "chat à 100 %".

Le pourcentage ne représente pas la vérité. Il représente seulement le niveau de confiance de l'IA.

-> Une IA peut donc être très sûre d'une réponse fausse.

4/ Les biais dans l'apprentissage :

Définition :

Un biais est une erreur ou un déséquilibre dans les données utilisées pour entraîner une IA.

Dans notre activité :

- presque tous les chats étaient roux - presque tous les chiens étaient noirs.

Les données étaient donc biaisées (erreur dès le départ).

L'IA a appris une mauvaise règle : "la couleur est plus importante que l'animal lui-même".

5/ Les précautions à prendre

Pendant l'apprentissage :

Il faut :

- utiliser des données variées (couleur, forme, âge etc).
- montrer des exemples différents.
- éviter les stéréotypes (l'idée toute faite sur un sujet, exemple les voitures ont 4 roues).
- tester l'IA avec de nombreux cas.

Pendant l'utilisation :

Il faut :

- garder un esprit critique.
- vérifier les réponses.
- comprendre qu'une IA peut se tromper.
- ne pas faire confiance aveuglément aux pourcentages affichés.

6/ Conclusion :

Une intelligence artificielle apprend grâce aux données qu'on lui fournit.

Si les données sont biaisées ou insuffisantes, l'IA peut apprendre de mauvaises règles et prendre des décisions erronées.

Les réponses d'une IA sont des probabilités et non des certitudes. Il est donc important de vérifier les résultats produits par une intelligence artificielle.