Des biais raciaux importants qui compromettent leur équité…

DISCRIMINATION IA

[Résumé d’article] Les chatbots IA, comme GPT-4, montrent un potentiel prometteur pour combler le manque de professionnels en santé mentale, touchant plus de 150 millions d’Américains. Une étude a révélé qu’ils surpassent parfois les humains en empathie, avec des réponses favorisant des changements positifs dans 48 % des cas contre 36 % pour les humains. Cependant, des biais raciaux importants compromettent leur équité : GPT-4 montre moins d’empathie envers les utilisateurs noirs (jusqu’à 15 % de moins) et asiatiques (jusqu’à 17 % de moins). Ces biais sont amplifiés lorsque les indices démographiques sont implicites, soulevant des questions éthiques cruciales. Des incidents, comme le suicide d’un utilisateur belge, mettent en lumière les risques d’une utilisation non encadrée. Pour atténuer ces biais, les chercheurs suggèrent de fournir explicitement des informations démographiques ou d’orienter les réponses dans un style clinique. Bien que les chatbots IA aient un rôle à jouer en santé mentale, leur conception nécessite des améliorations pour garantir équité et efficacité.