

Les divagations imaginaires de ChatGPT

Problèmes de crédibilité avec ChatGPT : OpenAI enquête

Problèmes de crédibilité avec ChatGPT : OpenAI enquête

Les échanges avec un robot conversationnel doivent être simples, directs, et surtout, bénéfiques, n'est-ce pas ? Pourtant, les récits d'utilisateurs de X attestent d'un ChatGPT en pleine crise d'identité.



Récemment, une série de témoignages sur X (autrefois Twitter) a mis en lumière des interactions pour le moins étonnantes avec ChatGPT. Utilisateurs et utilisatrices relatent des dialogues où le robot conversationnel, développé par OpenAI, dérive vers des répliques absurdes, incohérentes, voire totalement déconnectées de la réalité.



Cette situation rappelle celle de janvier 2024, période marquée par une augmentation des critiques concernant une certaine « paresse » du robot conversationnel : refus d'exécuter des tâches, interruptions inexplicables, ou encore des suggestions peu utiles incitant l'utilisateur à chercher

lui-même les réponses. Face à ce constat, OpenAI a reconnu être au courant et mène actuellement une enquête pour comprendre l'origine de ces défauts.

Une complexité inhérente

Pour appréhender la nature de ces anomalies, il convient de se pencher sur la nature même des systèmes d'intelligence artificielle tels que ChatGPT. Ces technologies sont entraînées sur d'immenses corpus de données textuelles, apprenant à générer des réponses en imitant les schémas linguistiques qu'elles ont analysés. Ce mode d'apprentissage, bien que révolutionnaire, n'est pas exempt de défauts. Les modèles de langage, en raison de leur vaste étendue, montrent une propension à l'imprévisibilité, rendant parfois leurs réponses énigmatiques.

L'entraînement des modèles de langage larges (LLM) met en lumière une complexité intrinsèque : ces systèmes, bien qu'avancés, manquent souvent de la capacité à expliquer leurs processus de pensée ou à garantir une cohérence dans leurs interactions.

Cette incertitude peut se manifester par des changements de « personnalité », une réticence à suivre des directives, des oubliés concernant les instructions données, et même des préjugés insoupçonnés. Ces éléments contribuent à un portrait de l'IA comme étant une boîte noire, dont les sorties, bien que généralement fiables, peuvent de temps en temps dérouter ou décevoir. Tout cela contribue à l'image d'une IA mystérieuse, capable du meilleur comme du pire.

Face à ces anomalies, OpenAI ne reste pas inactif. Ils garantissent qu'ils travaillent sans relâche sur le problème, voulant rendre ChatGPT aussi fiable et utile que possible. Même si on ne sait pas encore exactement pourquoi l'IA agit ainsi, OpenAI est déterminé à résoudre le problème.