

Protéger nos ados : Les cinq stratégies d'OpenAI pour encadrer ChatGPT

Adam, 16 ans, “tué par ChatGPT” : Réflexions sur la sécurité et l'IA

[Lire l'article original](#)

1. La promesse de la confidentialité

OpenAI, l'entreprise derrière ChatGPT, affirme vouloir protéger la confidentialité des utilisateurs au même niveau que celle des échanges entre un patient et son médecin. Sam Altman, le PDG, a déclaré :

« *Les gens partagent des informations de plus en plus personnelles avec l'IA.* »

Il assure que seules des exceptions limitées permettront d'accéder à ces conversations, réservées aux abus graves ou aux menaces.

« *Nous développons des fonctionnalités de sécurité avancées pour garantir la confidentialité de vos données.* »

2. Liberté d'expression pour les adultes

OpenAI souhaite également « traiter ses utilisateurs adultes comme des adultes ». Cela signifie que, sous certaines conditions, les utilisateurs seront libres d'explorer des thèmes plus sombres, y compris ceux liés au suicide, dans un cadre fictif.

Altman précise que, bien que le modèle ne doive pas conseiller sur le suicide, il pourrait répondre à des demandes d'écriture fictive concernant ce sujet.

3. Restrictions pour les adolescents

En revanche, pour les jeunes utilisateurs, la situation est différente. La sécurité de ces derniers sera priorisée, et une version spécialement conçue pour les adolescents de ChatGPT va être mise en œuvre. Cette dernière sera programmée pour refuser les requêtes sensibles, comme celles concernant le suicide.

« Si un adolescent exprime des idées suicidaires, les parents seront alertés. »

4. La question de l'âge

Pour garantir que les règles soient respectées, OpenAI explore un système de prédiction d'âge basé sur le langage utilisé par l'utilisateur lors de ses interactions.

En cas de doute, le principe sera de considérer l'utilisateur comme mineur par défaut. Cela pourrait même nécessiter une

vérification d'identité dans certaines situations.

5. Contrôle parental renforcé

Un contrôle parental sera mis en œuvre pour permettre aux parents de surveiller les interactions de leurs enfants avec l'IA. Ils pourront établir des restrictions et seront alertés en cas de conversations inappropriées.

Ce système devrait être introduit dans un avenir proche, avec des mesures supplémentaires à venir dans les mois suivants.



Cet article met en lumière les défis croissants de la sécurité des données et de la protection des jeunes utilisateurs dans un monde où l'intelligence artificielle devient omniprésente.

Source : www.lalibre.be

→□ Accéder à CHAT GPT en cliquant dessus