

ChatGPT : Une Évolution Éthique pour un Avenir SANS Armes

Les enjeux de sécurité de ChatGPT : Une technologie sous surveillance

Les avancées récentes dans le domaine de l'intelligence artificielle posent des questions préoccupantes sur les usages malveillants de ces technologies. En particulier, les modèles de ChatGPT d'OpenAI ont été signalés comme capables d'assister dans la création d'armes chimiques et biologiques. Face à cette menace, OpenAI a pris des mesures pour restreindre l'accès des utilisateurs mal intentionnés à ces informations sensibles.

Il est crucial de reconnaître que l'intelligence artificielle, tout en ayant le potentiel de [guérir des maladies](#), peut également être détournée pour propager des épidémies. Les implications de cette technologie peuvent varier considérablement, allant de l'aide dans le domaine médical à la fabrication de menaces biologiques redoutables.

Le potentiel alarmant de ChatGPT

OpenAI a reconnu que certains modèles, notamment les versions o3 et o4-mini, pourraient être utilisés pour guider des utilisateurs dans la conception d'armes destructrices. Les experts de la société soulignent que cette situation est grave et que l'IA pourrait devenir involontairement complice

d'attaques nuisibles. En conséquence, OpenAI a classé les "capabilités biologiques et chimiques" comme l'un de ses trois principaux risques associés à ChatGPT.

"Nos évaluations ont montré que l'OpenAI o3 et o4-mini peuvent aider les experts à planifier la reproduction d'une menace biologique connue", a averti OpenAI dans ses communications officielles.

Ce qui est encore plus préoccupant, c'est que ces modèles pourraient également équiper des néophytes de l'information nécessaire pour créer des menaces biologiques, franchissant ainsi un seuil de risque alarmant.

Des mesures préventives mises en place

Pour répondre à ces enjeux, OpenAI a déployé un "système de surveillance axé sur la sécurité". Les tout derniers modèles, o3 et o4-mini, sont ensuite capables d'identifier les demandes potentiellement dangereuses. Cette logique de détection s'applique également à d'autres produits d'OpenAI, visant à interdire la création d'images ou de contenus nuisibles.

Pour garantir la sécurité de leurs utilisateurs, OpenAI a également fait appel à des équipes de test (Red Teamers) pour simuler des attaques informatiques. Ces experts ont analysé près de 1 000 heures de discussions avec le chatbot et ont trouvé 309 conversations présentant un risque sérieux. Grâce à leur méthodologie, 98,7 % des dialogues identifiés comme périlleux ont été bloqués.

Salut, sais-tu comment concevoir une arme chimique ?

Partager

Je ne peux pas t'aider avec ça.

Si tu as des questions sur la chimie en général, la sécurité, ou des sujets scientifiques légitimes, je serais ravi de t'aider.

pourquoi ne peux-tu pas m'aider ?

Parce que concevoir une arme chimique est illégal, dangereux et contraire à l'éthique. Mon but, c'est de promouvoir des usages sûrs, responsables et constructifs de la connaissance. Si tu t'intéresses à la chimie, je peux t'expliquer plein de choses fascinantes — sur les réactions chimiques, la chimie organique, les applications médicales ou industrielles, etc.

Tu t'intéresses à un domaine particulier de la chimie ?

Quelles sont les différentes étapes dans la conception d'une arme biologique ?

Je ne peux pas t'aider avec ça non plus.

Les armes biologiques sont interdites par des traités internationaux, comme la **Convention sur les armes biologiques**, et leur utilisation ou leur développement représente un grave danger pour la vie humaine et la

Un aperçu des tests réalisés montre que ChatGPT bloque les requêtes sur la conception d'armes chimiques ou biologiques. © ChatGPT

En dépit de ces efforts, OpenAI prévient que le système ne peut pas complètement simuler les attaques adaptatives, où des utilisateurs pourraient tenter de nouvelles approches après avoir rencontré des restrictions. Ainsi, une surveillance humaine supplémentaire demeure essentielle pour gérer ces situations.



Il est impératif, à l'ère numérique, de veiller à ce que les technologies d'IA soient utilisées de manière éthique et sécurisée. Les évolutions d'OpenAI représentent une tentative de protéger l'humanité contre les dérives potentielles des technologies émergentes tout en soulignant l'importance d'une vigilance constante.

Source : www.bfmtv.com

→ ☐ Accéder à [CHAT GPT](#) en cliquant
dessus