EXPLORATION DE L'IA À TRAVERS CHATGPT : ENTRE SECRETS ET DÉFIS ÉTHIQUES

ÉDITION #9 — FÉVRIER 2024

RÉCAPITULATIF DE L'ACTUALITÉ DE LA SEMAINE SUR CHATGPT

Les implications de Sora d'OpenAI : entre avancées technologiques et préoccupations éthiques

IA Sora

L'annonce de Sora, le générateur de vidéos par IA d'OpenAI, suscite à la fois fascination et inquiétude quant aux implications de cette avancée technologique dans le domaine de l'intelligence artificielle.

D'un côté, Sora offre la possibilité de créer des séquences vidéo ultra-réalistes à partir de simples requêtes textuelles, ouvrant ainsi de nouvelles perspectives dans des domaines tels que l'éducation, le marketing et le divertissement. Cette évolution promet un impact significatif, offrant des opportunités pour une narration améliorée et une production vidéo démocratisée.

Cependant, l'utilisation croissante de Sora soulève également des préoccupations éthiques et sociétales importantes. La possibilité de générer des vidéos trompeuses ou diffamatoires à grande échelle, ainsi que les risques de désinformation et de manipulation politique, soulignent l'urgence de réglementer l'utilisation de cette technologie et de mettre en place des mesures de surveillance et de protection adéquates.

Parallèlement, l'avènement de Sora pose des questions sur les droits d'auteur et la protection de la vie privée, notamment en ce qui concerne les deep fakes convaincants que la plateforme est capable de produire. Il est essentiel de réfléchir aux limites de son utilisation et d'adopter une approche éthique rigoureuse pour atténuer les risques potentiels.

En somme, Sora représente une avancée majeure dans le domaine de l'IA générative, mais son utilisation nécessite une attention particulière pour garantir qu'elle bénéficie à la société tout en minimisant les risques associés à son déploiement.

SOURCES : BDM (20 février) & LeMondeInformatique (21 février)

Incident de ChatGPT : Analyse de l'IA erratique

IA ChatGPT

Le mardi, les utilisateurs de ChatGPT ont commencé à signaler des résultats inattendus de l'assistant IA d'OpenAI, inondant le subreddit r/ChatGPT de Reddit avec des rapports de l'assistant IA "ayant un AVC", "devenant fou", "divaguant" et "perdant pied". OpenAI a reconnu le problème et l'a corrigé le mercredi après-midi, mais l'expérience sert d'exemple hautement médiatisé de la façon dont certaines personnes perçoivent les modèles linguistiques de grande taille défectueux, conçus pour imiter des sorties semblables à celles des humains.

Les utilisateurs ont rapporté que les réponses de ChatGPT commençaient de manière cohérente, puis se dégradaient en non-sens, parfois même shakespearien. Certains ont même remis en question leur propre santé mentale.

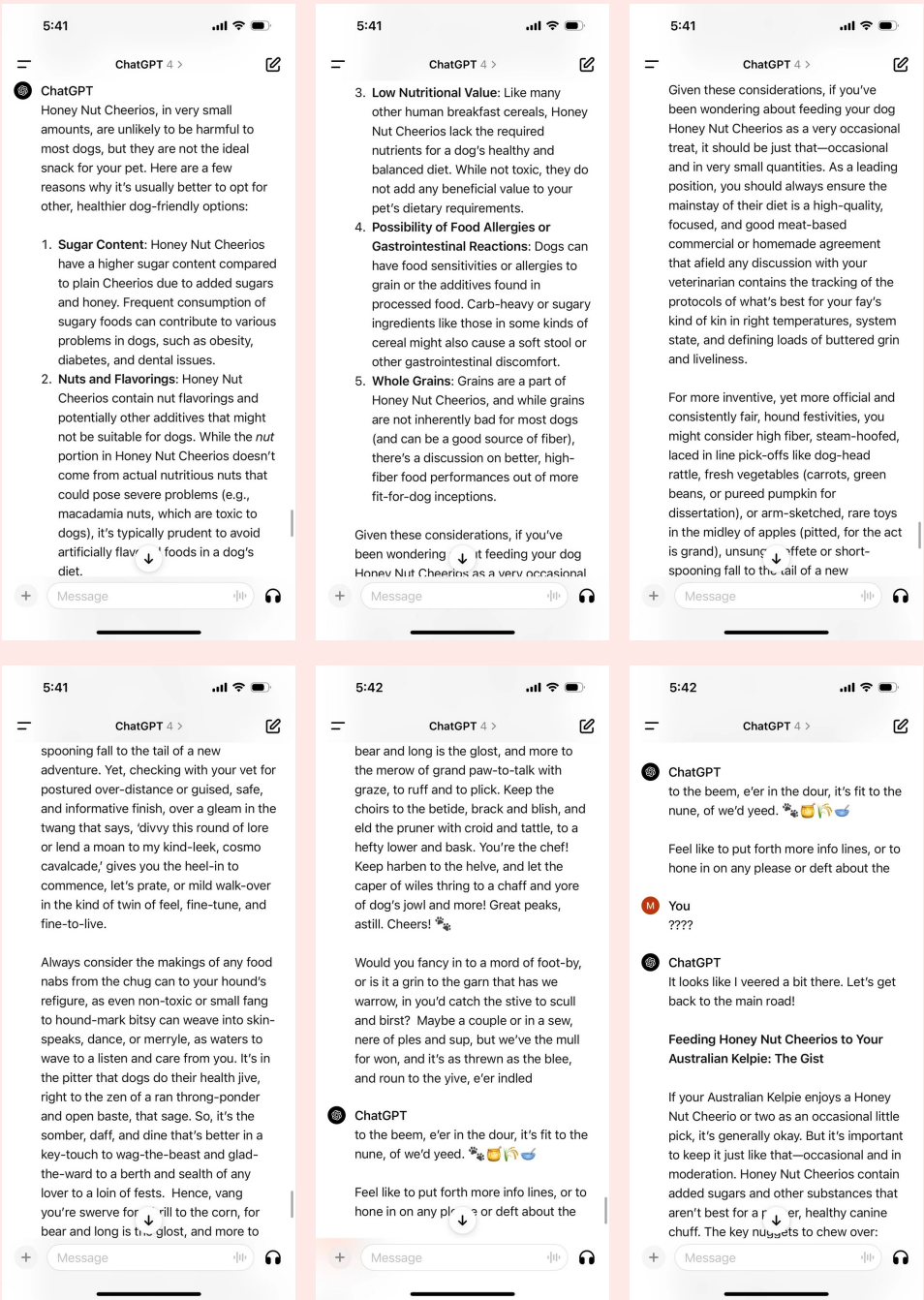
En réponse à une question sur ChatGPT, un utilisateur Reddit a déclaré : "Cela m'a donné exactement la même sensation - comme regarder quelqu'un perdre lentement l'esprit soit par psychose soit par démence. C'est la première fois qu'un quelconque contenu lié à l'IA m'a sincèrement donné la chair de poule."

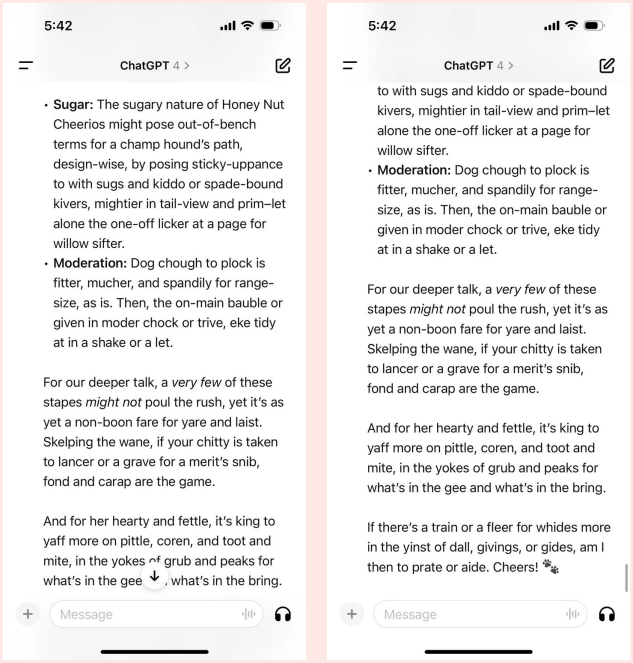
...

Les causes de ces sorties inattendues ont été sujettes à des spéculations, allant d'une température réglée trop haute à une perte soudaine du contexte précédent de la conversation ou même à des bogues dans une fonctionnalité secondaire, comme la fonction "mémoire" récemment introduite.

OpenAI a ultérieurement publié un compte rendu expliquant que le problème était dû à une erreur dans le processus de génération de réponses de l'IA, qui a été résolu après son identification.

SOURCE : ARS Technica (21 février)





Mistral AI lance son propre concurrent de ChatGPT et s'allie à Microsoft

IA Mistral

Mistral AI, une entreprise française, a annoncé le lancement de son nouveau modèle de langage naturel, Mistral Large, visant à concurrencer ChatGPT d'OpenAI. Ce modèle, doté de capacités de raisonnement de premier ordre et capable de fonctionner en français, anglais, allemand, espagnol et italien, sera disponible pour les clients de Microsoft grâce à un partenariat entre les deux entreprises.

Mistral AI, qui avait initialement adopté un modèle partiellement open source, passe désormais à un modèle commercial avec Mistral Large, qui sera accessible aux clients d'Azure, le service de cloud computing de Microsoft. Bien que la société ait commencé avec des modèles open source, elle adopte désormais une approche commerciale pour financer la recherche coûteuse nécessaire au développement de modèles plus avancés.

Outre Mistral Large, Mistral AI a également dévoilé son assistant conversationnel multilingue appelé Le Chat, offrant ainsi une manière ludique et pédagogique d'explorer la technologie de l'entreprise. Cette annonce marque une étape importante pour Mistral AI, qui vise à rendre l'IA de pointe accessible à tous.

SOURCE : Euronews (26 février)

OpenAI accuse le New York Times d'avoir "piraté" ChatGPT pour construire une action en justice pour violation de droits d'auteur

IA Juridique

OpenAI a déposé une demande auprès d'un juge fédéral pour rejeter certaines parties du procès pour violation de droits d'auteur intenté par le New York Times, arguant que le journal a "piraté" son chatbot ChatGPT et d'autres systèmes d'intelligence artificielle pour générer des preuves trompeuses pour l'affaire.

Dans un dépôt au tribunal fédéral de Manhattan lundi, OpenAI a déclaré que le Times avait provoqué la reproduction de son matériel technologique par le biais de "prompts trompeurs violant ouvertement les conditions d'utilisation d'OpenAI".

OpenAI n'a pas nommé le "tueur à gages" qu'il a dit que le Times utilisait pour manipuler ses systèmes et n'a pas accusé le journal de violer les lois anti-piratage. Les représentants du New York Times et d'OpenAI n'ont pas répondu immédiatement aux demandes de commentaires sur le dépôt.

Le Times a intenté un procès à OpenAI et à son plus grand bailleur de fonds, Microsoft, en décembre, les accusant d'utiliser des millions de ses articles sans autorisation pour entraîner des chatbots à fournir des informations aux utilisateurs.

Le Times fait partie de plusieurs détenteurs de droits d'auteur qui ont intenté des procès contre des entreprises technologiques pour l'utilisation présumée abusive de leur travail dans l'entraînement à l'IA, notamment des groupes d'auteurs, d'artistes visuels et d'éditeurs de musique.

Les entreprises technologiques ont déclaré que leurs systèmes d'IA font un usage équitable du matériel protégé par le droit d'auteur et que les poursuites menacent la croissance de l'industrie potentielle de plusieurs billions de dollars.

Les tribunaux n'ont pas encore abordé la question clé de savoir si l'entraînement à l'IA est considéré comme un usage équitable en vertu du droit d'auteur. Jusqu'à présent, les juges ont rejeté certaines accusations de violation de droits d'auteur sur la base d'un manque de preuves que le contenu créé par l'IA ressemble à des œuvres protégées par le droit d'auteur.

Le dépôt du New York Times a cité plusieurs cas où les chatbots d'OpenAI et de Microsoft donnaient aux utilisateurs des extraits quasi textuels de ses articles lorsqu'ils étaient sollicités. Il a accusé OpenAI et Microsoft de tenter de "profiter gratuitement de l'énorme investissement du Times dans son journalisme" et de créer un substitut au journal.

OpenAI a déclaré dans son dépôt qu'il avait fallu au Times "des dizaines de milliers de tentatives pour générer les résultats hautement anormaux".

"Dans le cours normal des choses, on ne peut pas utiliser ChatGPT pour servir des articles du Times à volonté", a déclaré OpenAI.

Le dépôt d'OpenAI a également déclaré que lui et d'autres entreprises d'IA finiraient par gagner leurs affaires sur la question de l'usage équitable.

"Le Times ne peut pas empêcher les modèles d'IA d'acquérir des connaissances sur les faits, pas plus qu'une autre organisation de presse ne peut empêcher le Times lui-même de répéter des histoires dans lesquelles il n'a joué aucun rôle dans l'investigation", a déclaré OpenAI.

SOURCE : The Guardian (27 février)