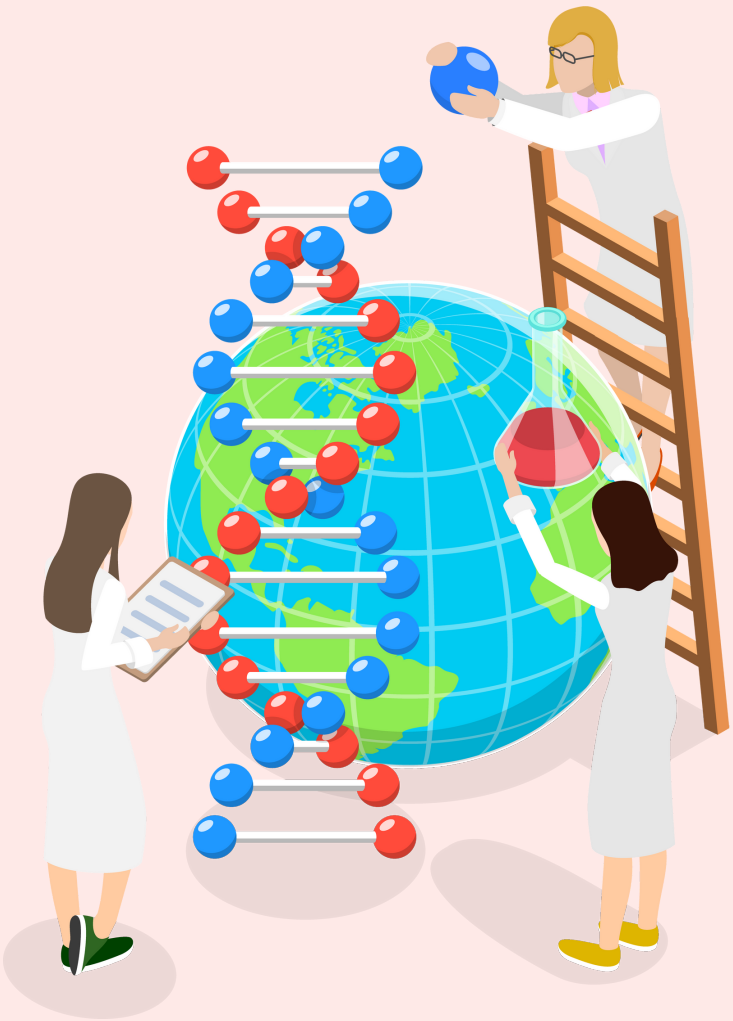
EXPLORATION DE L'IA À TRAVERS CHATGPT : ENTRE SECRETS ET DÉFIS ÉTHIQUES

ÉDITION #8 — FÉVRIER 2024



RÉCAPITULATIF DE L'ACTUALITÉ DE LA SEMAINE SUR CHATGPT

Avancée majeure dans l'IA générative : OpenAI présente Sora, son nouvel outil révolutionnaire

AI Sora Vidéo

OpenAI, la célèbre entreprise spécialisée dans l'intelligence artificielle, fait une nouvelle fois parler d'elle avec le lancement de **Sora**, son dernier projet révolutionnaire. Après le succès de ses précédentes créations telles que **ChatGPT** pour la génération de texte et **Dall-E** pour la génération d'images, OpenAI franchit une nouvelle étape en introduisant Sora, une IA générative de vidéos.

Contexte de l'annonce :

L'annonce de Sora intervient après une année marquée par des **progrès significatifs** dans le domaine de l'**IA générative.** Cependant, jusqu'à présent, la création de vidéos par des intelligences artificielles **semblait être un défi insurmontable**. Les tentatives antérieures ont souvent abouti à des résultats peu convaincants, comme en témoigne la célèbre vidéo de **Will Smith en train de manger des spaghettis**, largement moquée pour son manque de réalisme.

Les caractéristiques de Sora :

Sora, décrite comme une IA "**text-to-video**", fonctionne sur le même principe que **ChatGPT** : il suffit de fournir une description de ce que l'on souhaite voir dans la vidéo pour obtenir un résultat convaincant. OpenAI précise que Sora est capable de produire des vidéos d'une **durée maximale de 60 secondes**, avec des scènes détaillées, des mouvements de caméra complexes et plusieurs personnages exprimant des émotions vibrantes.

Les performances de Sora :

Les **performances** de Sora sont saluées comme étant de **très haut niveau**, malgré la **limitation de la durée des vidéos**. Il est déjà possible de créer des vidéos de type bande-annonce avec Sora, ce qui démontre l'ampleur de son **potentiel créatif**. Cependant, les résultats obtenus soulèvent également des questions quant aux **risques potentiels** liés à cette technologie.

Les préoccupations et les tests en cours :

En effet, la capacité de Sora à produire des vidéos réalistes soulève des préoccupations quant à son **utilisation malveillante**. La **possibilité** de créer des **vidéos trompeuses** ou **manipulées** pourrait avoir des **conséquences dévastatrices**, notamment dans les domaines de la **désinformation** et de la **manipulation politique**. Pour cette raison, OpenAI poursuit ses tests de l'outil en travaillant avec des experts pour évaluer ses implications en matière de désinformation, de contenu haineux et de préjugés.

Conclusion :

En conclusion, l'annonce de Sora représente une avancée majeure dans le domaine de l'IA générative. Cette nouvelle technologie soulève à la fois l'enthousiasme quant à ses possibilités créatives et les inquiétudes quant à son utilisation éthique. Alors que Sora ouvre de nouvelles perspectives dans la création de contenu visuel, il est crucial de continuer à examiner attentivement ses implications sociales et à développer des mesures pour atténuer les risques potentiels associés à son utilisation.

SOURCES : OpenAI & Clubic (15 février)

L'impact grandissant de l'IA sur la cybercriminalité : ChatGPT au cœur des préoccupations

Cybercriminalité AI ChatGPT

L'utilisation croissante de grands modèles de langage tels que ChatGPT par des groupes de hackers pour affiner et améliorer leurs attaques cybernétiques est au centre des préoccupations de Microsoft et OpenAI, selon un article récent. Des groupes soutenus par des nations comme la Russie, la Corée du Nord, l'Iran et la Chine ont été détectés en train d'utiliser des outils comme ChatGPT pour la recherche de cibles, l'amélioration de scripts et le développement de techniques d'ingénierie sociale.

Microsoft et OpenAI soulignent que les cybercriminels explorent et testent différentes technologies d'IA pour comprendre leur valeur potentielle dans leurs opérations et les contrôles de sécurité qu'ils doivent contourner. L'article met en avant des exemples spécifiques, notamment le groupe Strontium, lié au renseignement militaire russe, utilisant les LLM (Large Language Models) pour comprendre les protocoles de communication par satellite et les technologies d'imagerie radar.

Les auteurs soulignent également les préoccupations éthiques concernant l'utilisation de l'IA dans les cyberattaques, mettant en garde contre des cas futurs tels que l'usurpation vocale. Microsoft propose des solutions basées sur l'IA pour contrer ces attaques, notamment en développant un assistant AI appelé Security Copilot pour les professionnels de la cybersécurité.

En résumé, l'article met en évidence l'évolution des modèles de langage, les enjeux éthiques associés à leur utilisation dans les cyberattaques et l'importance croissante de l'IA pour protéger contre de telles menaces, contribuant ainsi à redéfinir le paysage de l'intelligence artificielle.

SOURCE : The Verge (14 février)

OpenAI expérimente avec la mémoire à long terme pour ChatGPT

AI ChatGPT Mémoire

OpenAI a récemment annoncé qu'elle expérimentait l'ajout d'une **forme de mémoire à long terme** à ChatGPT, lui permettant de se souvenir de détails entre les conversations. Cette fonctionnalité permet aux utilisateurs de demander à ChatGPT de se souvenir de quelque chose, de voir ce dont il se souvient, et de lui demander d'oublier. Actuellement, elle n'est disponible que pour un petit nombre d'utilisateurs de ChatGPT à des fins de test.

Les modèles de langage de grande taille ont généralement utilisé deux types de mémoire : l'**une intégrée au modèle d'IA** pendant le processus de formation et une **mémoire de contexte** (l'historique de conversation) qui persiste pendant la durée de votre session. Habituellement, ChatGPT oublie ce que vous lui avez dit pendant une conversation une fois que vous démarrez une nouvelle session.

Divers projets ont expérimenté l'ajout d'une mémoire aux LLM (Large Language Models) qui persiste au-delà d'une fenêtre de contexte. Les techniques incluent la **gestion dynamique de l'historique de contexte**, la compression de l'historique précédent par la summarisation, les liens vers des bases de données vectorielles stockant des informations de manière externe, ou simplement l'injection périodique d'informations dans une instruction système (les instructions que ChatGPT reçoit au début de chaque conversation).

...

OpenAI n'a pas expliqué quelle technique elle utilise ici, mais la mise en œuvre nous rappelle les Instructions Personnalisées, une fonctionnalité introduite par OpenAI en juillet 2023 qui permet aux utilisateurs d'ajouter des ajouts personnalisés à l'instruction système de ChatGPT pour en changer le comportement.

Les applications potentielles de la fonctionnalité de mémoire fournies par OpenAI comprennent l'explication de vos préférences concernant la mise en forme de vos notes de réunion, l'indication que vous gérez un café et la possibilité pour ChatGPT de supposer que c'est de cela que vous parlez, la conservation d'informations sur votre enfant en bas âge qui aime les méduses afin qu'il puisse générer des graphiques pertinents, et la mémorisation des préférences pour les plans de leçons de maternelle.

OpenAI affirme également que les souvenirs peuvent aider les abonnés de ChatGPT Entreprise et Équipe à mieux travailler ensemble, car les souvenirs d'équipe partagés pourraient se rappeler des préférences spécifiques de mise en forme de document ou des frameworks de programmation utilisés par votre équipe. De plus, OpenAI prévoit d'apporter des souvenirs aux GPT bientôt, chaque GPT ayant ses propres capacités de mémoire cloisonnées.

Évidemment, toute tendance à se souvenir des informations soulève des implications en matière de confidentialité. OpenAI indique que vos souvenirs sauvegardés sont également sujets à une utilisation par OpenAI à des fins de formation, sauf si vous répondez aux critères mentionnés ci-dessus. Cependant, la fonction de mémoire peut être complètement désactivée. De plus, la société déclare : "**Nous prenons des mesures pour évaluer et atténuer les biais, et éloigner ChatGPT de se souvenir de manière proactive d'informations sensibles, comme vos détails de santé, sauf si vous lui demandez explicitement.**"

Les utilisateurs pourront également contrôler ce que ChatGPT se souvient à l'aide d'une interface "Gérer la mémoire" qui répertorie les éléments de mémoire. "Les souvenirs de ChatGPT évoluent avec vos interactions et ne sont pas liés à des conversations spécifiques", déclare OpenAI. "Supprimer une conversation n'efface pas ses souvenirs ; vous devez supprimer la mémoire elle-même."

Les fonctionnalités de mémoire de ChatGPT ne sont pas actuellement disponibles pour tous les comptes ChatGPT, nous n'avons donc pas encore expérimenté avec. L'accès pendant cette période de test semble être aléatoire parmi les comptes ChatGPT (gratuits et payants) pour l'instant. "Nous déployons auprès d'une petite portion des utilisateurs ChatGPT gratuits et Plus cette semaine pour savoir à quel point il est utile", écrit OpenAI. "Nous partagerons bientôt des plans pour un déploiement plus large."

SOURCE : ARS Technica (13 février)