

ChatGPT: un contributeur potentiel pour l'écosystème Open-Source

Titre : Qui doit-on remercier pour les grands modèles de langage ?

Dans cet article, le chroniqueur Matt Asay parle de la dépendance des grands modèles de langage (LLM) envers les données fournies par les créateurs humains. Il affirme que les LLM sont des parasites qui dépendent de données provenant des dépôts de code, des réponses technologiques, de la littérature et d'autres sources encore. Cela signifie que, comme pour l'open source, les créateurs et les agrégateurs de contenu commencent à bloquer l'accès de ces modèles à leurs données.

Titre : La surutilisation des biens communs

Asay souligne que l'histoire des parasites technologiques précède celle de l'open source. Dès les premiers jours de Linux ou de MySQL, des entreprises ont été créées pour tirer profit des contributions des autres. Aujourd'hui, de grandes entreprises comme JP Morgan Chase dépensent des milliards de dollars et embauchent des ingénieurs en apprentissage automatique pour générer des gains de productivité. Cela a cependant un coût, comme en témoigne la chute de trafic sur des sites comme Stack Overflow, en raison de la menace que représentent les LLM pour l'accessibilité à la connaissance collective.

Titre : Les LLM menacent le savoir collectif

Les LLM menacent de vider le réservoir de connaissances sur des sites comme Stack Overflow. Les LLM sont des outils d'IA

générateive qui sont formés sur les données provenant de sites tels que Stack Overflow, et qui peuvent répondre efficacement à des questions en fournissant des schémas basés sur des données probabilistes. Toutefois, les créateurs des LLM ont besoin d'accéder à ces données pour les former, menaçant ainsi le flaconnage de la connaissance collective. Des outils tels que ChatGPT ne sont pas conçus pour produire des informations correctes et peuvent se reconstituer avec des informations erronées, devenant ainsi moins utiles.

Titre : Les LLM remettent en question la valeur des informations produites collectivement

Enfin, Asay souligne que les LLM posent des questions fondamentales sur la valeur des informations produites collectivement. Contrairement aux réponses fournies par les humains sur Stack Overflow, les réponses dérivées de ChatGPT sont souvent pauvres en informations et riches en texte, ce qui peut être nuisible pour le site et pour les utilisateurs qui recherchent des réponses correctes. Les LLM peuvent être utiles, mais seulement si les données de base sont bonnes et si l'on peut garantir la qualité des informations produites.

En conclusion, Asay souligne que l'accès aux données est un enjeu clé pour les LLM. Tout comme l'open source a connu ses guerres de licences, les LLM risquent de vivre quelque chose de similaire, mais avec des conséquences plus importantes. Il est essentiel de protéger l'accessibilité du savoir collectif pour éviter que les LLM deviennent de simples parasites qui se nourrissent des contributions des autres.

Source : www.lemondeinformatique.fr

→  **Accéder à CHAT GPT en cliquant**

dessus