

28/04/2025

SuperAlignement : « OpenAI porte une énorme responsabilité au nom de toute l'humanité »

Est-ce que les grandes entreprises de la tech prennent l'intelligence artificielle trop à la légère ? C'est ce que semble indiquer un ancien cadre d'OpenAI venant tout juste de démissionner. Quelle semaine pour OpenAI ! Lundi, la start-up spécialisée dans l'intelligence artificielle a dévoilé son nouveau modèle multimodal, GPT-4o. Le lendemain, Google, son principal concurrent, dévoilait une technologie similaire. Mercredi, on apprenait le départ d'Ilya Sutskever, l'un des cofondateurs de l'entreprise et de Jan Leike, un cadre haut placé. Et c'est ce dernier qui a levé de nombreuses questions ce vendredi sur la plateforme X.com. Un départ motivé par des valeurs. Chez OpenAI, les deux hommes étaient à la tête de Superalignment, une équipe spécialisée dans l'étude des risques de l'intelligence artificielle formée en juillet 2023. OpenAI garantissait alors d'allouer 20 % de ses ressources informatiques à cette équipe sur les quatre prochaines années. Si Ilya Sutskever est resté plutôt chaleureux dans son message de départ, Jan Leike exprime davantage d'amertume dans son thread public sur le réseau social d'Elon Musk. S'il commence par exprimer son affection pour les différentes équipes d'OpenAI, il précise rapidement être « en désaccord avec la direction d'OpenAI sur les priorités fondamentales de l'entreprise ». En cause : l'intérêt trop faible porté par l'entreprise à la sécurité, la confidentialité, l'impact sociétal de l'intelligence artificielle. Il regrette de ne pas avoir eu accès à assez de puissance de calcul sur les derniers mois pour mener à bien ses recherches sur le sujet. Vers une super intelligence. À l'heure actuelle, on peut se demander quels peuvent être les risques réels de l'intelligence artificielle. Pour le moment, les chatbots qui font la une des médias ont tendance à régurgiter sans grande réflexion des informations apprises dans un corpus donné, avec parfois de grossières erreurs que l'on appelle « hallucinations ». Mais à la vitesse où cela progresse, on peut s'attendre dans les prochaines années (dans la décennie selon OpenAI) à l'arrivée de ce que les chercheurs appellent une « super intelligence », plus intelligente que l'humain et capable de résoudre des problèmes sur lesquels nous butons aujourd'hui. Et selon Ilya et Jan, « la super intelligence pourrait aussi être très dangereuse et conduire à la neutralisation de l'humanité, voire son extinction ». Sans même aller jusqu'à l'extrémité d'un soulèvement des machines à la Terminator, il est important de mettre dans la balance toutes les conséquences potentielles d'un tel développement. On le voit déjà à l'heure actuelle avec les questions sur la désinformation en ligne par exemple. Les responsabilités d'OpenAI (et les autres) En tant que figure de proue de ces avancées, OpenAI comme Google ont une lourde responsabilité sur les épaules « au nom de toute l'humanité », exprime Jan Leike. Il déplore cependant que « la culture et les processus de sécurité ont été relégués au second plan par rapport aux produits clinquants ». Le problème s'avère en réalité économique et commercial. Dans un monde capitaliste mondialisé, l'IA est plus qu'un outil, c'est un produit marketing. Et sur un tel marché, la prime au premier arrivé est bien souvent décisive. Difficile donc pour une start-up comme OpenAI de dédier du temps et des ressources pour anticiper et sécuriser les dangers généraux au risque de se faire voler la vedette par Google (il était moins une cette semaine), Anthropic (l'entreprise financée par Amazon derrière l'IA Claude) ou un nouveau venu, en provenance de Chine par exemple. Pour Jan Leike, « OpenAI doit devenir une entreprise axée sur la sécurité ». En Europe, l'AI Act devrait réglementer certaines dérives, mais il ne fait aucun doute que l'entreprise qui se concentrera sur le produit sans se soucier de ces considérations aura

Last update: elsenews:spot-2024:05:superalignement-openai <http://aproposnews.com/doku.php/elsenews/spot-2024/05/superalignement-openai>
19/05/2024

l'occasion de prendre une longueur d'avance sur tous ses concurrents. Pour ne rater aucun bon plan, rejoignez notre nouveau channel WhatsApp Frandroid Bons Plans, garanti sans spam ! frandroid

From:
<http://aproposnews.com/> - **Apropos News**

Permanent link:
<http://aproposnews.com/doku.php/elsenews/spot-2024/05/superalignement-openai>

Last update: **19/05/2024**

