

Découvrez le mystère des LLM : les moteurs de ChatGPT, Google Bard et autres révélés !

Derrière ChatGPT, Google Bard, Meta AI et autres, il y a ce qu'on appelle les LLM, pour « large language models », ou grands modèles de langage en français. Ce sont en fait les moteurs des chatbots textuels d'IA, ceux qui apprennent et qui « comprennent » ce qu'on leur dit.

Bonjour chers débutants en intelligence artificielle! Nous allons aujourd'hui plonger dans le mystérieux monde des grands modèles de langage (LLM) qui se cachent derrière des noms tels que ChatGPT, Google Bard ou Meta AI. Comment fonctionnent-ils? Qu'est-ce qui les rend si puissants? Pas de panique, nous allons tout vous expliquer.

Qu'est-ce que ça veut dire « LLM »?

LLM est l'acronyme de l'expression anglaise « Large Language Model ». En français, on pourrait la traduire par « grand modèle de langage ». Il s'agit de modèles de langage qui possèdent généralement au moins un milliard de paramètres. En français, on peut aussi les nommer « modèles massifs de langage » et les désigner avec l'acronyme « MML ». Mais comment fonctionnent-ils exactement? C'est ce que nous allons voir.

Comment fonctionnent les LLM, les moteurs des intelligences artificielles?

Un LLM, c'est en réalité un réseau de neurones artificiels profonds. Chaque neurone informatique possède des entrées ainsi qu'une sortie. À l'aide de règles précises qu'on lui

indique, le neurone formel peut transformer une entrée en une sortie. Ces neurones artificiels sont associés en réseaux selon différents types de connexions. La force de ce système de réseau de neurones, c'est qu'il peut « apprendre » de lui-même: c'est le machine learning. Pour faire apprendre à un LLM, il faut lui donner du texte, beaucoup de texte. On peut simplement prendre Wikipédia par exemple. La qualité de l'apprentissage dépend aussi de ce qu'on appelle l'étiquetage des données, qui consiste à donner la réponse à une tâche demandée à partir de données déterminées.

Ce qu'a changé l'architecture Transformer au deep learning

C'est en 2017 qu'un changement technologique va bouleverser le monde de l'intelligence artificielle : la création de l'architecture Transformer. Elle résulte d'une longue combinaison de procédés techniques, avec des travaux datant de nombreuses années. Un « transformeur » est un modèle d'apprentissage profond, principalement taillé pour le traitement automatique des langues. Là où les réseaux neuronaux traditionnels traitent une requête en entrée de manière séquentielle, le transformeur peut paralléliser cette entrée, afin de considérablement réduire les temps d'entraînement.

Les premiers « vrais » modèles de langage : GPT et BERT

Deux LLM, qu'on peut considérer comme des pionniers, ont été publiés en 2018 à quelques semaines d'écart. Le premier, c'est GPT, pour Generative Pre-Trained Transformer d'OpenAI. Le second, c'est BERT de DeepMind (qui appartient à Google). Grâce à l'architecture Transformer, ils se sont révélés être des révolutions dans les LLM. Ils sont très bons en compréhension du langage naturel ainsi qu'en génération de texte. Ils peuvent réaliser des tâches simplement en leur demandant de manière textuelle.

À quoi servent les large language models?

La grande force des LLM, c'est précisément qu'ils n'ont pas d'usage précis, puisqu'ils n'ont pas été entraînés sur une tâche particulière. Leur fonctionnement neuronal fait qu'ils sont entraînés à la prédiction d'une suite probable en fonction d'une entrée donnée. Si vous demandez à ChatGPT de vous raconter une histoire, il va probablement démarrer par « Il était une fois », puisque c'est très classique.

Voilà, en espérant que cet article ait pu éclairer certains aspects du fonctionnement des intelligences artificielles génératives textuelles. Nous avons effleuré des notions complexes, mais indispensables pour comprendre l'IA de demain, alors n'hésitez pas à fouiller davantage dans ce passionnant domaine !

Source : www.frandroid.com

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