
Glossaire / Glossary

Termes importants autour de l'IA Générative

Agent LLM

application LLM intégrant des capacités de raisonnement, de décision, d'action, d'évaluation, une mémoire de ce qui a été déjà réalisé et utilisant des outils (appels de fonctions)

API

Application Programming Interface - mode d'accès à un LLM

Bias

un poids disproportionné en faveur ou à l'encontre d'une idée ou d'une chose, généralement d'une manière inexacte, fermée, préjudiciable ou injuste [Wikipédia EN]

Chatbot

logiciel ou interface web conçu pour imiter la conversation humaine par le biais d'interactions textuelles ou vocales [Wikipédia EN]

Chain-of-Thought

technique de prompting qui force le modèle de penser par étapes ("Let's think step-by-step")

Chain-of-Verification

méthode par laquelle le modèle (i) rédige une réponse initiale ; (ii) prévoit des questions de vérification pour contrôler son projet ; (iii) répond à ces questions de manière indépendante afin que les réponses ne soient pas faussées par d'autres réponses ; et (iv) génère sa réponse finale vérifiée.

[\[https://huggingface.co/papers/2309.11495\]](https://huggingface.co/papers/2309.11495)

Conversational AI

IA générative mode mode interaction de type conversation (exemple: chatbot)

Deep Learning (apprentissage profond)

sous-domaine de l'intelligence artificielle qui utilise des réseaux neuronaux pour résoudre des tâches complexes [Wikipédia]

Embedding

technique utilisée pour représenter des mots ou des entités sous forme de vecteurs dans un espace vectoriel continu

Fenêtre contextuelle

groupe spécifique de mots autour d'un mot cible dans un texte, utilisé pour saisir les informations contextuelles d'un mot en tenant compte des mots qui apparaissent à proximité.

Few-shot prompting

technique de rédaction d'un prompt dans laquelle le prompt contient 2 à n exemples pour guider l'IA dans sa réponse

Fine-Tuning

processus d'apprentissage supervisé utilisant un ensemble de données d'exemples étiquetés pour mettre à jour les poids du LLM et améliorer ses capacités pour des tâches spécifiques

Generative AI

type de système d'intelligence artificielle (IA) capable de générer du texte, des images ou d'autres médias en réponse à des prompts [Wikipédia]

Generative Pre-Trained Transformer (GPT)

type de modèle d'apprentissage profond (Deep Learning) basé sur l'algorithme "Transformer" qui est pré-entraîné sur une grande quantité de données textuelles pour comprendre et générer des textes de type humain

IA générative

type de système d'intelligence artificielle (IA) capable de générer du texte, des images ou d'autres médias en réponse à des prompts [Wikipédia]

Instruction-Tuned LLM

LLM de base entraîné pour suivre des instructions; renforcé par des retours sur les réponses pour favoriser les bonnes réponses

Large Language Model (LLM)

modèle d'apprentissage profond (Deep Learning) conçu pour traiter et comprendre de grandes quantités de données textuelles afin d'effectuer diverses tâches de traitement du langage naturel telles que la génération de texte, la traduction linguistique et l'analyse des sentiments

LLM de base

modèles à usage général pré-entraînés sur un large éventail de données textuelles, prédit le mot suivant à partir des données textuelles apprises et de la proximité linguistique

LLMOps

abréviation de Large Language Model Operations, concerne les aspects opérationnels et les techniques d'optimisation associés au déploiement et à l'utilisation efficace de modèles de langage de grande taille.

Machine Learning (ML)

branche de l'intelligence artificielle qui implique le développement d'algorithmes et de modèles statistiques pour permettre aux ordinateurs d'apprendre et de faire des prédictions ou de prendre des décisions basées sur des données sans instructions de programmation explicites

Méta-requêtage

technique de rédaction de prompt de haut niveau qui guide la génération ou la création de prompts plus spécifiques, qui sont à leur tour utilisés pour générer du texte, des images ou toute forme de contenu

One-shot prompting

technique de rédaction d'un prompt dans laquelle le prompt contient un exemple pour guider l'IA dans sa réponse

Prompt

requête textuelle

Prompt Engineering

processus de structuration d'un texte qui peut être interprété et compris par un modèle génératif d'IA [Wikipédia EN]

Prompt système

instruction prédéfinie utilisée par un système d'IA pour lancer ou guider ses opérations

Retrieval-Augmented Generation (RAG)

architecture de réseau neuronal qui combine les capacités des modèles de langage pré-entraînés avec un composant de récupération pour chercher des informations pertinentes dans une grande base de données ou corpus (par ex. des documents), améliorant la capacité du modèle à générer des réponses précises et contextuellement pertinentes basées sur les données récupérées

Requêtage par étapes

technique de prompting qui consiste à utiliser la sortie d'une incitation comme entrée pour une autre, créant ainsi une chaîne d'incitations qui s'ajoutent les unes aux autres

Small Language Model (SML)

version réduite de modèles de langage plus larges, pré-entraînés pour des tâches spécifiques et conçus pour offrir un équilibre entre l'efficacité informatique et la capacité à comprendre et à générer des textes de type humain

Température (d'un modèle)

paramètre qui contrôle le caractère aléatoire ou créatif du résultat généré

Transformer

architecture de modèle d'apprentissage profond (Deep Learning) qui utilise des mécanismes d'auto-attention (Self-Attention) pour capturer les dépendances à longue portée dans les séquences d'entrée pour les tâches.

Token

morceau de texte, tel qu'un mot, un caractère ou un sous-mot, qui sert d'unité d'entrée pour le traitement par l'IA générative

Zero-shot prompting

technique de rédaction d'un prompt dans laquelle le prompt ne contient pas d'exemples

Important terms in the field of Generative AI

API

Application Programming Interface - LLM access mode

Base LLM

general-purpose models pre-trained on a wide range of text data, predicts the next word from learned text data and linguistic proximity

Bias

a disproportionate weight in favor of or against an idea or thing, usually in a way that is, inaccurate, closed-minded, prejudicial, or unfair [Wikipedia]

Chain-of-Thought

prompting techniques that forces the model to think "step-by-step"

Chain-of-Verification

method whereby the model first (i) drafts an initial response; then (ii) plans verification questions to fact-check its draft; (iii) answers those questions independently so the answers are not biased by other responses; and (iv) generates its final verified response

[\[https://huggingface.co/papers/2309.11495\]](https://huggingface.co/papers/2309.11495)

Chatbot

a software application or web interface that is designed to mimic human conversation through text or voice interactions [Wikipedia]

Context window

specific range of words around a target word within a text used to capture the contextual information of a word by considering the words that appear nearby

Conversational AI

Generative AI in conversation model (e.g. chatbot)

Deep Learning

subset of machine learning methods based on artificial neural networks [Wikipedia]

Embedding

technique used to represent words or entities as vectors in a continuous vector space

Few-shot prompting

prompt writing technique in which the prompt contains 2 to n examples to guide the AI in its response

Fine-Tuning

supervised learning process using a dataset of labeled examples to update LLM weights and improve its capabilities for specific tasks

Generative AI

artificial intelligence capable of generating text, images or other data using generative models, often in response to prompts [Wikipedia]

Generative Pre-Trained Transformer (GPT)

type of transformer-based deep learning model that is pre-trained on a vast amount of text data to understand and generate human-like text

Instruction-Tuned LLM

Base LLM trained to follow instructions; reinforced by feedback to encourage correct answers

Large Language Model (LLM)

deep learning model designed to process and understand vast amounts of text data to perform various natural language processing tasks like text generation, language translation, and sentiment analysis

LLM Agent

LLM application integrating reasoning, decision, action and evaluation capabilities, a memory of what has already been achieved and using tools (function calls)

LLMOps

short for Large Language Model Operations, pertains to the operational aspects and optimization techniques associated with deploying and utilizing large language models effectively

Machine Learning (ML)

branch of artificial intelligence that involves the development of algorithms and statistical models to enable computers to learn from and make predictions or decisions based on data without explicit programming instructions

Meta prompting

technique of writing higher-level instruction that guides the generation or creation of more specific prompts, that are, in turn, used to generate text, images, or any form of content

One-shot prompting

prompt writing technique in which the prompt contains one example to guide the AI in its response

Prompt

textual request

Prompt Chaining

prompting technique that involves using the output of one prompt as the input for another, creating a chain of prompts that build on each other

Prompt Engineering

process of structuring text that can be interpreted and understood by a generative AI model [Wikipedia]

Retrieval-Augmented Generation (RAG)

neural network architecture that combines the capabilities of pre-trained language models with a retriever component to fetch relevant information from a large database or corpus

(e.g. documents), enhancing the model's ability to generate accurate and contextually relevant responses based on the retrieved data

Small Language Model (SML)

scaled-down version of larger language models, pr-trained for specific tasks and designed to offer a balance between computational efficiency and the ability to understand and generate human-like text

System prompt

pre-defined instruction used by an AI system to initiate or guide its operations

Temperature (of a model)

parameter that controls the randomness or creativity in the generated output

Transformer

deep learning model architecture that utilizes self-attention mechanisms to capture long-range dependencies in input sequences for tasks

Token

piece of text, such as a word, character, or subword, that serves as an input unit for generative IA processing

Zero-shot prompting

prompt writing technique in which the prompt contains no examples