
Glossary

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