

Bert & chatgpt4



Title Lorem Ipsum

What is large language model

- ❖ What is bert
- ❖ What is chatgpt4
- ❖ different between bert & chatgpt4 in architecture

What is large language model

A **large language model (LLM)** is a type of artificial intelligence (AI) model designed to process and generate human-like text. These models are based on deep learning techniques

Key Features of LLMs:

- **Trained on massive datasets:** They learn from vast amounts of text data from books, articles, and websites.

Capable of understanding context: They generate coherent and contextually relevant responses.

- **Used in various applications:** Chatbots, content generation, code writing, and even medical or legal text processing.

How They Work

LLMs use **deep learning**, specifically **transformer architectures**, to process and generate human-like text. They learn patterns, grammar, and contextual meaning from large datasets. The most well-known transformer model is **GPT (Generative Pre-trained Transformer)**, developed by OpenAI.

LLMs follow a two-step process:

- **Pre-training:** The model is trained on massive amounts of text to learn language patterns.
- **Fine-tuning:** It is then adapted for specific tasks like answering questions, summarization, or translation.

What is BERT?

- ❖ BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model developed by Google AI in 2018. It is one of the most influential large language models designed for natural language processing (NLP) tasks.
 - ❖ Key Features of BERT:
 1. Bidirectional Understanding – Unlike traditional models that read text left-to-right or right-to-left, BERT reads the entire sentence at once, understanding context better.
 2. Pre-trained on Large Datasets – BERT is trained on massive amounts of text data, such as Wikipedia and BooksCorpus.
 3. Fine-Tuning for Specific Tasks – After pre-training, BERT can be fine-tuned for tasks like:
 1. Text classification (e.g., spam detection)
 2. Question answering (e.g., Google Search)
 3. Named entity recognition (e.g., identifying names in text)

What is chatgpt4

- ❖ ChatGPT-4 (GPT-4) is a large language model developed by OpenAI, released in 2023. It is the fourth iteration of the GPT (Generative Pre-trained Transformer) series and is significantly more advanced than its predecessors.
 - ❖ Key Features of ChatGPT-4:
 1. More Accurate and Knowledgeable – Improved ability to understand and generate human-like text with fewer mistakes.
 2. Better Context Awareness – Can handle longer conversations and remember context better than previous versions.
 3. Multimodal Capabilities – Unlike GPT-3, GPT-4 can process both **text and images** (in some versions).
 4. Enhanced Creativity – Performs better in creative writing, code generation, and problem-solving.
 5. More Reliable – Fewer hallucinations (false information) and better handling of ambiguous queries.

difference between bert & chatgpt4 in architecture

- ❖ Detailed Architectural Differences Between BERT & ChatGPT-4

- ❖ 1. Transformer Architecture Type

- BERT: Uses only the **encoder** part of the Transformer model.
 - ChatGPT-4: Uses only the **decoder** part of the Transformer model.

- ❖ 2. Directionality

- BERT: Bidirectional – Processes the entire input sentence at once to understand context.
 - ChatGPT-4: Unidirectional (Autoregressive) – Generates words sequentially from left to right.

- ❖ 3. Training Objective

- BERT: Masked Language Modeling (MLM) – Random words in the input are masked, and the model predicts them using surrounding words.
 - ChatGPT-4: Causal Language Modeling (CLM) – Predicts the next word in a sequence based only on previous words.

difference between bert & chatgpt4 in architecture

◆ 4. Input Handling

- BERT: Takes a **fixed-length** input and processes the entire text simultaneously.
- ChatGPT-4: Can handle **variable-length** input, processing tokens sequentially while generating responses dynamically.

◆ 5. Attention Mechanism

- BERT: Uses **self-attention** in both directions (left & right context).
- ChatGPT-4: Uses **causal self-attention**, restricting attention to past tokens only.

◆ 6. Output Type

- BERT: Outputs **contextualized embeddings** for understanding language (used for classification, search, etc.).
- ChatGPT-4: Outputs **generated text**, continuously predicting the next word.

◆ 7. Token Processing

- BERT: Processes **all tokens simultaneously** (parallel processing).
- ChatGPT-4: Processes **one token at a time** in sequence (autoregressive).

◆ 8. Model Architecture Depth

- BERT: Typically has **Base (12 layers, 110M parameters)** or **Large (24 layers, 340M parameters)** versions.
- ChatGPT-4: Has a much **deeper architecture**, estimated to have **trillions of parameters**.