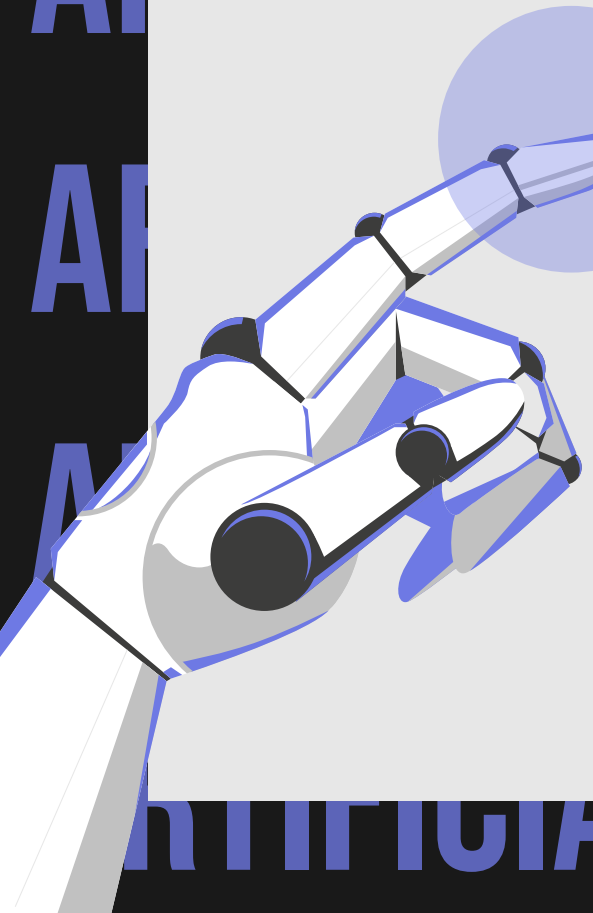


Nền tảng Large Language Model

TABLE OF CONTENTS

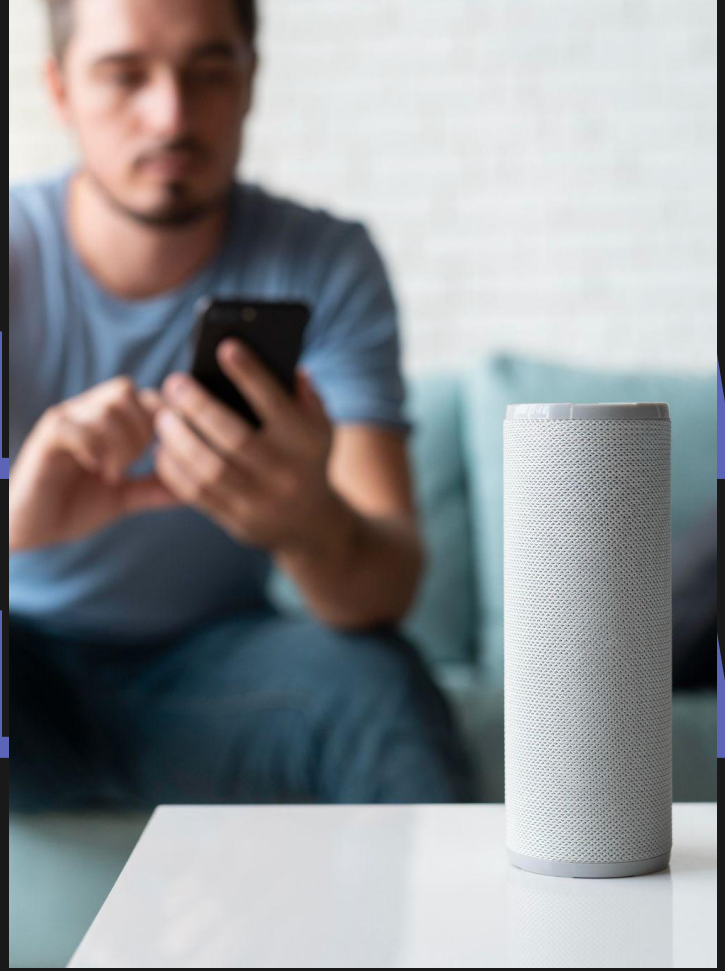
ARTIFICIAL INTELLIGENCE (AI)

01.

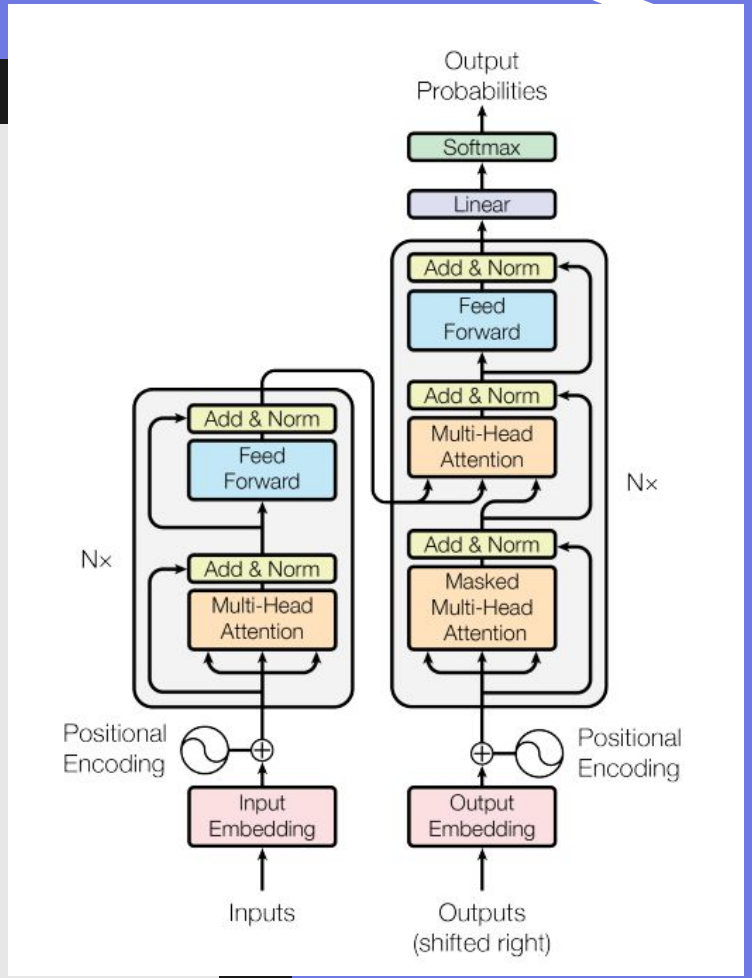
Giới thiệu về LLM

Định nghĩa

LLM là một loại mô hình trí tuệ nhân tạo (AI) được thiết kế để xử lý và tạo ngôn ngữ tự nhiên
VD: GPT, Gemini,PaLM,...



02.
Kiến trúc Transformer



Kiến trúc Transformer

Tổng quan

- Gồm Encoder (xử lí đầu vào) và Decoder (tạo đầu ra)
- Sử dụng Self- Attention để hiểu ngữ cảnh

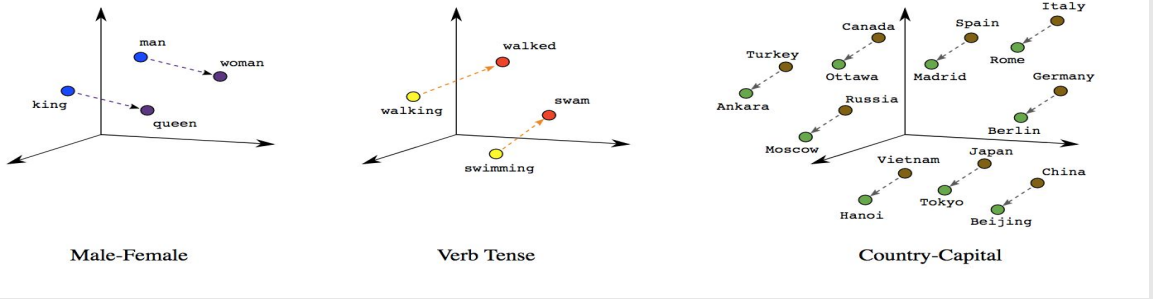
Quy trình hoạt động

- Input Embedding: Token hóa và chuyển thành embedding và positional encoding
- Encoder:
  - Self-attention -> FFNN -> Lặp lại N lần
  - Đầu ra là biểu diễn ngữ cảnh của chuỗi đầu vào
- Decoder:
  - Dùng masked self-attention để xử lí từng token đầu ra
  - Cross-attention kết hợp thông tin từ encoder
  - FFNN -> Lặp lại N lần -> Dự đoán token tiếp theo

Kiến trúc Transformer

Các thành phần chính

- Embedding:
  - Tokenization:
    - Chia văn bản thành tokens
  - Vector hóa:
    - Mỗi token -> Vector số học
  - Positional Embedding
    - Thêm thông tin vị trí từ vào vector

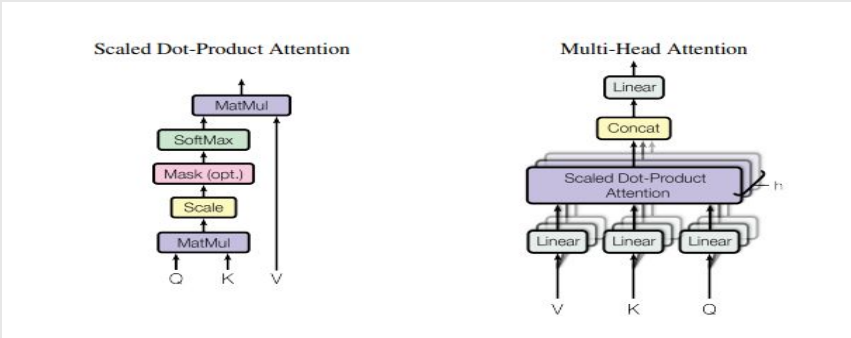


Kiến trúc Transformer

Các thành phần chính

- Self-Attention
  - Cơ chế:
    - Query(Q), Key(K), Value(V): TÍnh độ tương quan giữa các từ”
  - Công thức

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$



Kiến trúc Transformer

Các thành phần chính

- Feed Forward Neural Network (FFNN)
  - Vai trò:
    - Biến đổi thông tin sau Attention thành biểu diễn phi tuyến
  - Cấu trúc:
    - 2 lớp Linear + ReLU (VD : 512 -> 2048 -> 512)

- Prompt -> Output
  - Xử lí Prompt:
    - Token hóa -> Embedding -> Attention -> Dự đoán token tiếp theo
  - Ví dụ :
    - Prompt: “Dịch Hello sang tiếng Việt” -> Output : “Xin chào”

03.

Quá trình huấn luyện LLM

Quá trình huấn luyện LLM

- Pretraining:
  - Dự đoán token tiếp theo (GPT) hoặc từ bị che (BERT)
- Fine-Tuning
  - Điều chỉnh các tác vụ cụ thể (dịch máy, chatbot)
- RLHF:
  - Tối ưu hóa dựa trên phản hồi con người

Pretraining

- Định nghĩa:
  - Pretraining là cách đào tạo cơ bản và được sử dụng với một mô hình chưa được đào tạo nhằm huấn luyện để nó có thể dự đoán được token tiếp theo dựa trên một chuỗi token trước đó
- Nguyên lí cơ bản:
  - Dự đoán phần còn thiếu của văn bản để học biểu diễn ngôn ngữ
  - Input : Một chuỗi tokens
  - Output: Dự đoán tokens tiếp theo (GPT) hoặc điền từ bị che (BERT)
  - Phương pháp:
    - Tự hồi quy(Autoregressive - GPT)
    - Masked Language Modeling (BERT)

Fine-tuning

- Định nghĩa:
  - Fine-Tuning điều chỉnh mô hình đã pretraining cho các tác vụ cụ thể.
- Nguyên lí hoạt động :
  - Input: Dữ liệu có nhãn
  - Output: Tối ưu hóa đầu ra cho tác vụ mục tiêu
  - Phương pháp:
    - Supervised Fine-Tuning (SFT):
      - Huấn luyện trên cặp input-output
    - Reinforcement Learning from Human Feedback (RLHF)
      - Tối ưu hóa phản ứng đánh giá của con người

Reinforcement Learning from Human Feedback (RLHF)

- **Định nghĩa:**
  - RLHF là một trường hợp đặc biệt của finetuning giúp tinh chỉnh LLM bằng cơ chế học tăng cường dựa trên phản hồi của con người, giúp mô hình tạo đầu ra phù hợp với mong muốn của người dùng
- **Mục đích :**
  - LLM có thể sinh ra nội dung độc hại, sai lệch, hoặc không phù hợp vậy nên RLHF căn chỉnh lại mô hình bằng phản hồi trực tiếp từ con người
- **Quy trình RLHF:**
  - Thu thập dữ liệu phàn hồi (Human Feedback)
  - Huấn luyện mô hình phần thưởng (Reward Model)
  - Tối ưu LLM bằng PPO

[AI]

04.

PROMPT

Prompt Engineering

- **Định nghĩa:** Là quá trình tối ưu hóa cách đưa ra yêu cầu cho AI để xác nhận được kết quả chính xác và hiệu quả nhất. Nó quan trọng vi:
  - Tận dụng tối đa khả năng của AI
  - Tiết kiệm thời gian khi prompt tốt sẽ giảm số lần chỉnh sửa và tương tác với AI
  - Giảm sai sót, tránh gây hiểu lầm cho AI.

- **Các yếu tố ảnh hưởng đến hiêu quả của prompt :**
  - Độ rõ ràng
  - Kích thước mô hình AI
  - Độ phức tạp của nhiệm vụ
  - Ngữ cảnh ràng buộc

Phân loại prompt

- **System prompt:**
  - Là những hướng dẫn ẩn hoặc ràng buộc do developer thiết lập trước khi AI trả lời người dùng.
  - Vai trò:
    - Định hình hành vi mặc định của AI
    - Giới hạn phạm vi trả lời
    - Tăng tính an toàn

- **User prompt:**
  - Là những yêu cầu trực tiếp mà người dùng nhập vào AI
  - Vai trò:
    - Quyết định ngữ cảnh trực tiếp của câu hỏi
    - Ảnh hưởng trực tiếp đến chất lượng câu hỏi

Các cấu trúc Prompt

- **Zero-shot Prompt**
  - **Khái niệm:** Yêu cầu mô hình thực hiện tác vụ mà không cung cấp bất kì ví dụ nào.
  - **Cấu trúc:**
    - [Yêu cầu/ nhiệm vụ] + [Thông tin đầu vào] + [Định dạng]
  - **Ưu điểm:**
    - Không cần ví dụ mẫu, tiết kiệm thời gian
    - Dễ sử dụng, phù hợp tác vụ đơn giản.
    - Hoạt động tốt với các mô hình lớn
  - **Nhược điểm:**
    - Độ chính xác thấp hơn Few-shot/Chain-of-thought
    - Không hiệu quả với nhiệm vụ phức tạp, đòi hỏi ngữ cảnh
    - Có thể gây hiểu nhầm nếu prompt mơ hồ

Các cấu trúc Prompt

- **Few-shot Prompt**
  - **Khái niệm:** Yêu cầu mô hình thực hiện tác vụ bằng cách cung cấp một vài ví dụ mẫu trước khi đưa ra yêu cầu chính
  - **Cấu trúc:**
    - [Ví dụ 1] + [Ví dụ 2] + … + [Yêu cầu chính].
  - **Ưu điểm:**
    - Hiệu quả hơn Zero-shot với các tác vụ phức tạp
    - Giảm sai sót nhờ học từ ví dụ
  - **Nhược điểm:**
    - Tốn token (tăng chi phí)
    - Quá nhiều ví dụ có thể gây nhiễu

Các cấu trúc Prompt

- **Chain of thought (CoT)**
  - **Khái niệm:** Phương pháp khiến AI suy nghĩ trước khi đưa ra kết quả nhằm cải thiện độ chính xác với các tác vụ phức tạp.
  - **Cấu trúc :**
    - [Bài toán] + “Hãy giải từng bước”
  - **Các biến thể của CoT:**
    - Self-Consistency CoT
    - Automatic CoT
  - **Sử dụng Chain of Thought khi:**
    - Bài toán cần đến tính toán, logic
    - Câu hỏi suy luận phức tạp
    - Kiểm tra tính hợp lí
  - Sử dụng kết hợp với Few-shot prompt để vừa kết hợp ngữ cảnh và vừa yêu cầu suy luận

[AI]

05.

NGUỒN

NGUỒN

- https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- https://vinbigdata.com/cong-nghe-giong-noi/lam-the-nao-de-dao-tao-large-language-models.html?gad_source=1&gad_campaignid=22427790121&gbraid=0AAAAAp9MqYFE8TqSBcE-5Dgu8iO8UXPZ_&gclid=CjwKCAjwl_XBBhAUEiwAWK2hzmh_GqmiFOxTCgjyM0b-jQtgCIzDj9FmAVxBOpAv7EI8h-SHUZdcmxoC-vEQAvD_BwE
- https://chatgpt.com/
- https://chat.deepseek.com/
- https://tinhte.vn/thread/huong-dan-prompt-tu-co-ban-den-nang-cao-p1-zero-shot-va-few-shot-prompting.4011566/
- https://tinhte.vn/thread/huong-dan-prompt-tu-co-ban-den-nang-cao-p3-step-back-prompting-va-chain-of-thought-cot.4012561/
- https://www.viettelidc.com.vn/tin-tuc/llm-la-gi#:~:text=LLM%2C%20hay%20m%C3%B4%20h%C3%ACnh%20ng%C3%B4n,nhi%E1%BB%81u%20l%C4%A9nh%20v%E1%BB%B1c%20kh%C3%A1c%20nhau.
- https://vnptai.io/vi/blog/detail/llm-la-gi#:~:text=Large%20Language%20Model%20(LLM)%20%E2%80%93,LLM%20l%C3%A0%20g%C3%AC?
- https://tinhte.vn/thread/huong-dan-prompt-tu-co-ban-den-nang-cao-p1-zero-shot-va-few-shot-prompting.4011566/
- https://tinhte.vn/thread/huong-dan-prompt-tu-co-ban-den-nang-cao-p3-step-back-prompting-va-chain-of-thought-cot.4012561/

ARTIFICIAL INTELLIGENCE (AI)