**I. Cách hoạt động của GPT ( Generative Pre-trained Transformer)**

GPT là một mô hình **AI tạo sinh ( Generative AI )** dựa trên kiến trúc **Transformer** do Open AI phát triển

**Bước 1 : Tiền huấn luyện ( Pre - training)**

- GPT được huấn luyện **trên khối lương văn bản khổng lồ** từ internet , sách , báo , mã nguồn ,...

-Nhiệm vụ : **dự đoán từ tiếp theo** trong câu ( Next Token Prediction) .

( Ví dụ : " Hôm nay trời ..." --> mô hình dự đoán từ có khả năng cao nhất là “ đẹp” )

-GPT học được :

+ Cấu trúc ngữ pháp

+ Mối quan hệ ngữ nghĩa

+ Lượng kiến thức / thông tin khổng lồ

**Bước 2 : Tinh chỉnh ( Fine-tuning)**

-Mô hình được điều chỉnh trên dữ liệu được gắn nhãn bởi con người ( human feedback) để phù hợp hơn với mục tiêu .

-Dùng kỹ thuật RLHF( Reinforcement Learning from Human Feedback):

+AI tạo nhiều câu trả lời

+ Con người xếp hạng các câu trả lời

+ AI học theo thứ tự ưu tiên để xếp hạng câu trloi

**Bước 3 : Sinh văn bản ( Generation)**

-Khi người dung nhập câu hỏi , AI xử lý đầu vào thành vector ( embedding)

-Dựa trên xác suất , mô hình sinh ra từng từ tiếp theo cho tới khi hoàn thành câu trloi .

-Cơ chế **attention** trong Transformer giúp mô hình tập trung vào các từ liên quan để giữ ngữ cảnh

=> GPT như 1 người đọc hang triệu cuốn sách rồi đoán từ tiếp theo sao cho hợp lý ngữ cảnh .

**II. Ứng dụng của GPT**

GPT có thể ứng dụng rộng rãi trong nhiều lĩnh vực :  
+ Văn bản & Nội dung : Viết bài , tóm tắt, dịch , sáng tác truyên

+ Giáo dục : Giải thích khái niệm , tạo bài tập , hỗ trợ đa ngôn ngữ

+ Lập trình : Sinh code , sửa code , giải thích thuật toán

+ Kinh doanh : Viết mail , tạo kế hoạch / ý tưởng , phân tích thị trường

+ Chăm sóc khách hang : Chatbot 24/7 , trloi tự động

+ Nghiên cứu : Tìm kiếm thông tin , phân tích tài liệu khoa học

**III. Giới hạn và thách thức của GPT**

-Về kiến thức :   
+ Không thật sự “ hiểu” : Toàn bộ là kiến thức mà mô hình học từ bên ngoài chứ không có ý thức hay hiểu biết sâu .

+ Kiến thức bị giới hạn ở thời điểm huấn luyện : GPT được huấn luyện trên dữ liệu thu thập được tại mốc thời gian nhất định ( nếu được huấn luyện tại 2021 thì sẽ k cập nhật được kiến thức của 2022)=> thông tin sau đó có thể k chính xác ( Lí do : Huấn luyện GPT đòi hỏi **khối lượng dữ liệu cực lớn** và **chi phí rất cao** (hàng triệu USD, hàng trăm GPU\*\*Không thể “cập nhật liên tục” từng ngày như tìm kiếm Google, vì mỗi lần huấn luyện lại gần như là “đào tạo lại cả mô hình”.=> có thể cập nhật mô hình trên tập dữ liệu nhỏ hơn ( Fine-tuning) hoặc là tra cứu công cụ tìm kiếm thời gian thực ( Google))

-Về độ chính xác :

+ AI có thể bịa ra thông tin nghe có vẻ thuyết phục nhưng không đúng sự thật , phải kiểm chứng lại nhiều nguồn khác nhau .

+ Không đảm bảo độ chính xác 100% cho các dữ liệu chuyên ngành , học thuật ( y tế , pháp luật )

-Về đạo đức và bảo mật :  
+ Có thể bị lạm dụng để tạo thông tin sai lệch ( Do AI học từ RLHF)

+Có thể bị lộ dữ liệu nhạy cảm nếu người dung để lộ thông tin cá nhân

-Về kỹ thuật :

+ Tốn tài nguyên khi tính toán , cần GPU mạnh để huấn luyện .

+ Đầu ra phụ thuộc vào prompt ( câu lệnh đầu vào ) => prompt kém -> kết quả kém