

OpenAI API for Text

Mục tiêu bài học

Trong bài học này, bạn sẽ:

- Hiểu cách hoạt động cơ bản của **OpenAI API**
- Biết cách tạo **ứng dụng tạo văn bản** bằng `chat.completions` endpoint
- Nắm vai trò quan trọng của **system prompt**
- Cài đặt và kết nối với API để bắt đầu tạo phản hồi giống như ChatGPT

1. Tổng quan về OpenAI API

OpenAI cung cấp một bộ công cụ mạnh mẽ cho nhà phát triển, đặc biệt là:

- Dựa trên các **mô hình ngôn ngữ lớn (LLMs)** như GPT-4.0, GPT-4-turbo, GPT-3.5
- Có thể sinh:
 - Văn bản tự nhiên
 - Dữ liệu có cấu trúc
 - Đoạn mã code
 - Trả lời hội thoại (Chat)

API chính được dùng: `chat.completions.create`

Đây là endpoint mô phỏng giống **ChatGPT**.

2. Cài đặt và kết nối API

Bước 1: Nhận API Key

- Truy cập <https://platform.openai.com/account/api-keys>
- Tạo một **API Key**
- Nạp tối thiểu **\$5** (rất rẻ, ví dụ: dịch 1.200 đoạn văn bản chỉ mất **\$0.04**)

Bước 2: Cài đặt thư viện

Trong terminal hoặc Google Colab:

```
pip install openai
```

Bước 3: Kết nối API trong Python

```
from openai import OpenAI

client = OpenAI(
    api_key="sk-..." # Thay bằng API key của bạn
)
```

3. Tạo phản hồi văn bản với `chat.completions`

Cấu trúc cơ bản:

```
response = client.chat.completions.create(
    model="gpt-4", # hoặc "gpt-4-turbo", "gpt-3.5-turbo"
    messages=[
        {"role": "system", "content": "You are a helpful assistant."},
        {"role": "user", "content": "Write a haiku about recursion in programming."}
    ]
)

print(response.choices[0].message.content)
```

3 phần quan trọng:

- `model`: tên mô hình
- `messages`: danh sách hội thoại
 - `system`: điều chỉnh **tính cách và hành vi**
 - `user`: prompt nhập từ người dùng
 - `assistant`: phản hồi (có thể thêm nếu muốn giữ trạng thái hội thoại)
- `choices[0].message.content`: kết quả đầu ra

4. System Prompt là gì?

System Prompt là phần đầu tiên trong `messages[]`, giúp điều chỉnh **hành vi của AI**.

Ví dụ:

```
{"role": "system", "content": "You are an expert in software engineering and provide concise technical answers."}
```

Giúp AI:

- Trả lời ngắn gọn
- Mang tính kỹ thuật
- Phù hợp cho ngữ cảnh cụ thể

Lưu ý:

- **ChatGPT (trên web)**: không thể chỉnh `system prompt`
- **OpenAI API**: bạn **có thể** chỉnh được — tạo được trợ lý tùy biến hơn nhiều

So sánh giữa GPT-4 và GPT-4-Turbo

Mô hình	Chi phí	Tốc độ	Dung lượng context
GPT-4	Đắt hơn	Ổn định	8k hoặc 32k
GPT-4-Turbo	Rẻ hơn 5x	Nhanh hơn	128k
GPT-3.5-Turbo	Rất rẻ	Nhanh	16k

Kết luận

Nội dung chính	Ghi nhớ
<code>chat.completions.create</code> là công cụ chính để tạo phản hồi văn bản	
<code>system prompt</code> giúp định hình tính cách của AI	
Cài đặt bằng <code>pip install openai</code> và dùng <code>client.chat.completions.create()</code>	
Có thể dùng các model khác nhau để cân bằng giữa chi phí và chất lượng	
ChatGPT không cho chỉnh system prompt, API thì có thể	

Tác giả: **Đỗ Ngọc Tú**
Công Ty Phần Mềm **VHTSoft**