

Sử Dụng OpenAI API Để Hiểu, Phân Tích Và Mô Tả Hình Ảnh

Trong phần này, chúng ta sẽ khám phá cách sử dụng **OpenAI API với hình ảnh**, một tính năng mạnh mẽ và linh hoạt, cho phép bạn:

- Hiểu nội dung hình ảnh
- Mô tả hoặc phân tích hình ảnh
- Trích xuất văn bản từ hình ảnh (OCR)
- Kết hợp giữa hình ảnh và văn bản để đưa ra câu trả lời chính xác

Tổng Quan: Tại Sao API Này Lại Quan Trọng?

Thông thường, các mô hình ngôn ngữ chỉ làm việc với văn bản. Tuy nhiên, với khả năng **multimodal** (đa phương thức), OpenAI API hiện cho phép xử lý **hình ảnh** cùng với văn bản, mở ra rất nhiều ứng dụng thực tế như:

- Phân tích hóa đơn, biên lai, biểu đồ
- Trích xuất văn bản từ hình ảnh chụp màn hình
- Mô tả ảnh cho người khiếm thị
- Nhận diện nội dung ảnh trong hệ thống kiểm duyệt

Cách Gửi Hình Ảnh Cho API Phân Tích

Để gửi một yêu cầu đến API, bạn cần cấu trúc dữ liệu `messages` như sau:

```
{
  "model": "gpt-4-vision-preview",
  "messages": [
    {
      "role": "user",
      "content": [
        {
          "type": "text",
          "text": "Hãy mô tả những gì đang diễn ra trong hình ảnh sau."
        },
        {
          "type": "image_url",
          "image_url": {
            "url": "https://upload.wikimedia.org/wikipedia/commons/thumb/e/e9/Big_Ben_2012-06-23.jpg/800px-Big_Ben_2012-06-23.jpg"
          }
        }
      ]
    }
  ]
}
```

- `type: text`: Đây là đoạn văn bản mô tả yêu cầu (prompt).
- `type: image_url`: Đây là phần chứa hình ảnh (dưới dạng đường dẫn hoặc base64).
- `image_url.url`: Phải là **liên kết trực tiếp** tới ảnh (.jpg, .png,...), không phải trang HTML.

Các Tác Vụ Bạn Có Thể Làm Với Ảnh

1. Mô tả nội dung ảnh

```
"text": "Hãy mô tả bức ảnh này một cách chi tiết."
```



Kết quả sẽ là mô tả như: “Bức ảnh chụp Tháp đồng hồ Big Ben tại London vào một ngày nắng, với bầu trời trong xanh...”

2. Trích xuất nội dung văn bản trong ảnh (OCR)

"text": "Hãy đọc và trích xuất toàn bộ văn bản có trong hình ảnh sau."

“ Kết quả: Toàn bộ chữ có trong ảnh sẽ được AI trả lại, rất hữu ích khi xử lý biên lai, tài liệu scan, v.v.

Thay vì URL: Dùng Base64 Cho Ảnh Riêng Tư

Nếu ảnh bạn muốn phân tích **không được lưu công khai trên internet**, bạn có thể **chuyển ảnh sang dạng base64** và nhúng trực tiếp vào yêu cầu:

```
{
  "type": "image_url",
  "image_url": {
    "url": "..."
  }
}
```

“ Lợi ích của base64:

- Không cần tải ảnh lên server trung gian
- Bảo mật cao hơn (ảnh không công khai)
- Hoạt động trong các API nội bộ, offline hoặc giới hạn truy cập

Các Lưu Ý Khi Dùng API Với Ảnh

Yêu cầu	Mô tả
Ảnh rõ nét	Tránh ảnh mờ, tối hoặc có nhiễu
Định dạng ảnh	Hỗ trợ: <code>.jpg</code> , <code>.png</code> , <code>.gif</code>

Yêu cầu	Mô tả
Dung lượng ảnh	Không quá lớn (thường < 20MB)
Kết hợp text + image	Cung cấp prompt rõ ràng cùng với ảnh để tăng độ chính xác
Sử dụng hệ thống prompt	(Optional) Có thể dùng để hướng dẫn vai trò hoặc mục tiêu của AI

Ví Dụ: Phân Tích Biểu Đồ

```
"text": "Đây là biểu đồ. Hãy mô tả xu hướng chính và kết luận rút ra từ nó."
```

Với hình ảnh kèm theo là biểu đồ đường, biểu đồ cột, AI sẽ trả lời như một nhà phân tích dữ liệu.

Điểm chính	Nội dung
Đa năng	OpenAI API không chỉ làm việc với văn bản, mà còn hiểu được hình ảnh
Mô tả & Đọc ảnh	Có thể mô tả cảnh vật hoặc trích xuất chữ từ ảnh
Hỗ trợ base64	Hữu ích cho ảnh nhạy cảm, không công khai
Ứng dụng thực tế	Từ kiểm tra tài liệu, hỗ trợ người khiếm thị đến phân tích hình ảnh trong hệ thống doanh nghiệp

Dưới đây là **một ví dụ thực tế hoàn chỉnh** về cách **trích xuất thông tin từ hóa đơn** bằng cách sử dụng **OpenAI API với hình ảnh**, kết hợp cùng đoạn mã Python để thực thi yêu cầu này.

Tình huống giả định:

Bạn có một **ảnh chụp hóa đơn** và muốn tự động trích xuất các thông tin sau:

- Tên công ty
- Ngày xuất hóa đơn
- Tổng tiền
- Mã số thuế
- Số hóa đơn

Yêu cầu gửi đến OpenAI

Ví dụ về `messages` gửi tới GPT-4 API:

```
"messages": [  
  {  
    "role": "user",  
    "content": [  
      {
```

```
"type": "text",
"text": "Hãy trích xuất các thông tin sau từ hóa đơn trong hình ảnh: Tên công ty, Mã số thuế, Ngày lập hóa đơn, Số hóa đơn, và Tổng tiền thanh toán. Vui lòng trả về kết quả dưới dạng JSON."
},
{
  "type": "image_url",
  "image_url": {
    "url": "https://example.com/hoa_don_mau.jpg"
  }
}
]
}
```

Kết quả GPT trả về có thể là:

```
{
  "company_name": "Công Ty TNHH ABC Việt Nam",
  "tax_code": "0301234567",
  "invoice_date": "01/04/2024",
  "invoice_number": "0000123",
  "total_amount": "5.500.000 VND"
}
```

Đoạn mã Python thực tế

Bạn có thể dùng đoạn mã sau để gửi ảnh hóa đơn đến GPT-4 và nhận kết quả JSON:

```
import openai

openai.api_key = "YOUR_API_KEY"

response = openai.ChatCompletion.create(
  model="gpt-4-vision-preview",
  messages=[
    {
      "role": "user",
      "content": [
        {
          "type": "text",
```

```
"text": "Hãy trích xuất các thông tin sau từ hóa đơn trong hình ảnh: Tên công ty, Mã số thuế, Ngày  
lập hóa đơn, Số hóa đơn, và Tổng tiền thanh toán. Vui lòng trả về kết quả dưới dạng JSON."
```

```
    },  
    {  
      "type": "image_url",  
      "image_url": {  
        "url": "https://example.com/hoa_don_mau.jpg" # Thay bằng URL ảnh thật  
      }  
    }  
  ]  
}  
],  
max_tokens=500  
)  
  
print(response['choices'][0]['message']['content'])
```

Nếu bạn dùng ảnh nội bộ (Base64):

Bạn có thể chuyển ảnh hóa đơn thành Base64 và gửi đi:

```
import base64  
  
with open("hoa_don_mau.jpg", "rb") as img:  
    b64 = base64.b64encode(img.read()).decode("utf-8")  
  
image_url = f"data:image/jpeg;base64,{b64}"
```

Sau đó thay dòng:

```
"url": "https://example.com/hoa_don_mau.jpg"
```

bằng:

```
"url": image_url
```

Ứng dụng thực tế

Bạn có thể áp dụng kỹ thuật này để:

- Tự động lưu hóa đơn vào hệ thống kế toán

- Gửi cảnh báo nếu tổng tiền vượt ngưỡng
- Kiểm tra hóa đơn trùng
- Phân loại hóa đơn theo nhà cung cấp

Tác giả: Đỗ Ngọc Tú
Công Ty Phần Mềm VHTSoft

Phiên bản #1

Được tạo 5 tháng 5 2025 08:08:38 bởi Đỗ Ngọc Tú

Được cập nhật 6 tháng 5 2025 09:52:34 bởi Đỗ Ngọc Tú