

AI

- [Giới thiệu về Ollama](#)

Giới thiệu về Ollama

Ollama là gì?

Ollama là một công cụ mã nguồn mở cho phép chạy các mô hình ngôn ngữ lớn (LLM) trên máy tính cá nhân. Nó giúp người dùng có thể:

- Chạy các mô hình AI như Llama 2, Mistral và nhiều mô hình khác một cách cục bộ
- Không cần kết nối internet để sử dụng AI
- Bảo mật dữ liệu tốt hơn vì mọi thứ được xử lý trên máy local

Ưu điểm chính của Ollama

- Dễ dàng cài đặt và sử dụng
- Hỗ trợ nhiều mô hình AI khác nhau
- Miễn phí và mã nguồn mở
- Có thể tích hợp vào các ứng dụng khác thông qua API

Ollama đặc biệt hữu ích cho các nhà phát triển và người dùng muốn thử nghiệm với AI mà không phụ thuộc vào các dịch vụ đám mây.

Tính năng chính của Ollama

- **Chạy LLMs cục bộ:** Không cần kết nối internet, giúp đảm bảo quyền riêng tư.
- **Tích hợp đơn giản:** Dễ dàng sử dụng qua dòng lệnh hoặc tích hợp với các ứng dụng.
- **Tùy chỉnh mô hình:** Cho phép tinh chỉnh và mở rộng mô hình theo nhu cầu.
- **Hỗ trợ nhiều nền tảng:** Hoạt động trên Windows, macOS và Linux.

Cách sử dụng Ollama

1. **Cài đặt Ollama:** Tải về từ ollama.com và làm theo hướng dẫn.
2. **Chạy mô hình AI:** Ví dụ, để chạy Llama 2, bạn chỉ cần dùng lệnh:

```
ollama run llama2
```

3. **Tùy chỉnh mô hình:** Bạn có thể huấn luyện mô hình riêng bằng cách sử dụng lệnh

```
ollama create.
```

Các lệnh cơ bản của Ollama

- **ollama run [model]** - Chạy một mô hình cụ thể
- **ollama pull [model]** - Tải xuống một mô hình mới
- **ollama list** - Xem danh sách các mô hình đã cài đặt
- **ollama rm [model]** - Xóa một mô hình

- **ollama serve** - Khởi động máy chủ Ollama

Các lệnh này có thể được thực thi trong terminal hoặc command prompt sau khi đã cài đặt Ollama thành công.

Ví dụ sử dụng

```
# Tải và chạy mô hình Llama 2
ollama pull llama2
ollama run llama2

# Kiểm tra các mô hình đã cài đặt
ollama list

# Xóa một mô hình không cần thiết
ollama rm llama2
```