

Ollama vs. LangChain

Ollama và LangChain đều hỗ trợ làm việc với **mô hình ngôn ngữ lớn (LLM)**, nhưng **mục đích sử dụng** rất khác nhau.

1. Ollama - Chạy LLM cục bộ trên máy

📌 Tóm tắt:

- Chạy **mô hình LLM** ngay trên máy tính, **không cần internet**.
- Hỗ trợ nhiều mô hình (**Llama, Mistral, DeepSeek, Gemma, v.v.**).
- Đơn giản, dễ dùng, **chỉ cần 1 lệnh để chạy AI**.

📌 Ví dụ sử dụng:

Cài đặt Ollama:

```
curl -fsSL https://ollama.com/install.sh | sh
```

Chạy mô hình Llama 3:

```
ollama run llama3
```

Chạy mô hình DeepSeek Coder:

```
ollama run deepseek-coder
```

📌 Ưu điểm của Ollama:

- ☐ **Dễ sử dụng:** Chỉ cần `ollama run model_name` là chạy được AI.
- ☐ **Không cần internet:** Phù hợp để chạy AI **ngoại tuyến**.
- ☐ **Tối ưu GPU/CPU:** Dễ dàng chạy trên máy tính cá nhân.
- ☐ **Hỗ trợ nhiều mô hình:** Tải và chạy **Llama, DeepSeek, Mistral, Gemma, v.v.**

📌 Nhược điểm:

- ☐ **Không hỗ trợ tự động kết nối nhiều mô hình AI.**
 - ☐ **Không có pipeline** để kết hợp AI với dữ liệu từ nhiều nguồn.
-

2. LangChain - Xây dựng ứng dụng AI phức tạp

Tóm tắt:

- Là **framework** giúp kết hợp **LLM** với dữ liệu.
- Dùng để **tạo chatbot, tìm kiếm tài liệu (RAG), tự động hóa AI**.
- Kết nối với nhiều nền tảng như **OpenAI, Ollama, Google Gemini, v.v.**

Ví dụ sử dụng LangChain với Ollama:

```
from langchain_community.llms import Ollama

llm = Ollama(model="mistral")
response = llm.invoke("Giải thích Định lý Pythagoras.")
print(response)
```

Ưu điểm của LangChain:

- **Linh hoạt:** Kết nối nhiều AI khác nhau (GPT-4, Llama, Claude, v.v.).
- **Xử lý dữ liệu từ nhiều nguồn:** Tích hợp với **PDF, website, cơ sở dữ liệu**.
- **Tạo pipeline AI phức tạp:** Xây dựng ứng dụng như ChatGPT với dữ liệu riêng.

Nhược điểm:

- **Không thể chạy AI cục bộ** (cần kết nối Ollama hoặc OpenAI).
- **Cấu hình phức tạp hơn Ollama**.

Chọn cái nào?

Tính năng	Ollama	LangChain
Chạy mô hình AI cục bộ	☑ Có	☐ Không
Hỗ trợ nhiều mô hình AI	☑ Có	☑ Có
Xử lý dữ liệu từ PDF, DB	☐ Không	☑ Có
Tạo chatbot, pipeline AI	☐ Không	☑ Có
Dễ sử dụng	☑ Rất dễ	☐ Phức tạp hơn
Không cần Internet	☑ Có	☐ Không

Kết luận:

- **Nếu muốn chạy AI cục bộ trên máy → [👉 Ollama](#)**
 - **Nếu muốn xây dựng ứng dụng AI phức tạp → [👉 LangChain](#)**
 - **Muốn kết hợp cả hai? Dùng **LangChain + Ollama** để vừa có **AI cục bộ**, vừa có **pipeline AI mạnh mẽ**. [👉](#)**
-

Phiên bản #1

Được tạo 13 tháng 3 2025 07:56:17 bởi Đỗ Ngọc Tú

Được cập nhật 13 tháng 3 2025 07:57:55 bởi Đỗ Ngọc Tú