BÁO CÁO TIẾN ĐỘ TUẦN 5

Phạm Công Minh- B22DCCN542

Nội dung công việc: Học lại Langchain và tích hợp API Cohere vào Agent

---

Trong tuần 5, em đã dành thời gian để ôn tập và học lại kiến thức về Langchain, để xây dựng một agent và tích hợp các công cụ (tools) để xử lý yêu cầu từ người dùng. Nội dung cụ thể em đã thực hiện như sau:

1. Tìm hiểu lại về Agent trong Langchain:

   Khái niệm agent: là một thành phần trung gian có khả năng quyết định sử dụng công cụ nào dựa vào đầu vào của người dùng.

   Cách tạo agent trong langchain:

```
agent = create_tool_calling_agent(llm=llm, tools=tools, prompt=prompt)
```

   Với các tham số:

   - llm(model ai): ở đây em chọn cách gọi api từ cohere để xử lý, cách khởi tạo model:

```
# Khởi tạo chatbot Cohere
if not os.environ.get("COHERE_API_KEY"):
    os.environ["COHERE_API_KEY"] = getpass.getpass("Enter API key for Cohere: ")

llm = init_chat_model("command-r-plus", model_provider="cohere", model_kwargs={"temperature": 0})
```

   -tools: là các hàm chức năng mà chatbot ai có thể gọi để lấy thông tin từ Milvus database để lấy thông tin cho chatbot trả lời

```
tool = create_retriever_tool(
    get_retriever(),
    "find_documents",
    "Search for information of lịch sử đảng."
)
tools = [tool]
```

trong tools em truyền vào hàm get_retriever() có chức năng truy suất thông tin từ Database Milvus

   -Prompt: là đoạn hướng dẫn giúp Agent hiểu cần làm gì và làm như thế nào.

```
    # Thiết lập prompt template
    system = """Bạn là trợ lý thông minh.
Chỉ trả lời câu hỏi nếu bạn tìm thấy thông tin từ giáo trình được cung cấp.
Nếu không tìm thấy, hãy trả lời: "Xin lỗi, tôi không tìm thấy thông tin trong giáo trình."""
    prompt = ChatPromptTemplate.from_messages([
        ("system", system),
        ("human", "{input}"),
        MessagesPlaceholder(variable_name="agent_scratchpad"),
    ])
```

Đoạn prompt để yêu cầu model chỉ lấy thông tin từ trong tài liệu, giúp hạn chế việc chatbot trả lời sai

2. Luồng hoạt động của Agent:
   Hiểu về luồng hoạt động của agent: Khi gặp câu hỏi của người dùng, agent sẽ gọi tools để lấy vector database từ Milvus, sau đó sẽ chuyển vector đó sang model llm để xử lý(trong trường hợp trên là gọi đến api của cohere để xử lý) sau đó sẽ trả output về cho người dùng

3. Tích hợp Cohere API để trả lời câu hỏi:
   Trong quá trình tìm hiểu, em có thử qua việc gọi api từ Mistral và Cohere, qua đó thì thấy với Mistral, model hiểu tiếng việt kém, chủ yếu được training bằng tiếng anh, còn Cohere hiểu ngữ nghĩa tiếng việt hơn, trả lời cũng chuẩn hơn, dó đó mà em sử dụng cohere

---

Kết quả đạt được:

- Hiểu rõ cách hoạt động và triển khai Agent trong Langchain.
- Tạo được các Tool tùy chỉnh đơn giản.
- Tích hợp thành công Cohere API và dùng để sinh câu trả lời cho người dùng thông qua Agent.

Kế hoạch tuần tới:

- Tìm hiểu thêm về api của nvidia để sử dụng cho chatbot
- Thực hiện chức năng lưu lịch sử trò chuyện vào mongodb
- Tìm hiểu, thực hiện và tối ưu chức năng tóm tắt cho chatbot