**I. MÔ TẢ BÀI TOÁN**

1. **Định nghĩa LangChain và mục đích**
   * **Khái niệm về LangChain như framework kết nối LLM với dữ liệu bên ngoài**
   * **Ứng dụng của LangChain trong xây dựng trợ lý ảo đọc hiểu tài liệu định dạng hỗn hợp**
2. **Vấn đề cốt lõi LangChain giải quyết trong bài toán**
   * **Thách thức tích hợp LLM với dữ liệu từ tài liệu không có cấu trúc (PDF)**
   * **Khó khăn trong kết nối LLM với dữ liệu có cấu trúc (CSV)**
   * **Nhu cầu tạo hệ thống hỏi đáp thông minh trên tài liệu đa dạng dựa trên LangChain**
3. **Sơ lược về hệ sinh thái LangChain**
   * **LangChain Framework cho xử lý tài liệu và câu hỏi**
   * **LangSmith cho debug và monitor**
   * **LangServe cho API deployment**
   * **LangGraph cho multi-agent orchestration**

**II. MÔ TẢ GIẢI THUẬT**

1. **Kiến trúc cốt lõi LangChain**
   * **Mô hình thành phần hoá của LangChain**
   * **Cách kết nối các thành phần trong LangChain**
2. **Giải thuật Chains cho xử lý tài liệu PDF**
   * **Document loaders cho PDF**
   * **Text splitters và semantic chunking**
   * **Sequential chains cho document processing**
3. **Giải thuật Chains cho xử lý dữ liệu CSV**
   * **CSV loaders và parsers**
   * **Data transformation với chains**
   * **Integration với pandas**
4. **Giải thuật RAG (Retrieval-Augmented Generation)**
   * **Document processing với LangChain**
   * **Vector stores và embeddings**
   * **Retrieval chains**
5. **Giải thuật Question Answering với LangChain**
   * **Question parsing chains**
   * **Router chains cho định tuyến đến nguồn dữ liệu phù hợp (PDF/CSV)**
   * **RetrievalQA chains**
6. **Giải thuật Agents**
   * **PDF và CSV Tools**
   * **ReAct pattern (Reasoning + Acting)**
   * **Multi-agent orchestration với LangGraph**

**III. LƯU ĐỒ XỬ LÝ**

1. **Lưu đồ xử lý Chain cơ bản trong LangChain**
   * **Input → Prompt Template → LLM → Output Parser → Response**
2. **Lưu đồ xử lý tài liệu PDF với LangChain**
   * **PDF → DocumentLoader → TextSplitter → Embedding → VectorStore → RetrievalChain**
3. **Lưu đồ xử lý dữ liệu CSV với LangChain**
   * **CSV → CSVLoader → DataFrameLoader → TabularLLM Chain → Structured Output**
4. **Lưu đồ xử lý RAG trong LangChain**
   * **Query → Embedding → Vector Search → Document Retrieval → Context Augmentation → LLM → Response**
5. **Lưu đồ xử lý Agent trong LangChain**
   * **User Question → Agent Planning → Tool Selection (PDF Tool/CSV Tool) → Tool Execution → Result Evaluation → Next Action/Final Response**

**IV. MÔ TẢ THƯ VIỆN SỬ DỤNG**

1. **Thư viện cốt lõi LangChain**
   * **langchain và langchain-core**
   * **langchain-community**
   * **LCEL (LangChain Expression Language)**
2. **Thư viện document loaders và processors**
   * **langchain-document-loaders cho PDF**
   * **CSV loaders và pandas integration**
   * **langchain-text-splitters**
3. **Thư viện vector database và embeddings**
   * **langchain-chroma/FAISS/Pinecone**
   * **OpenAI Embeddings/HuggingFace Embeddings**
4. **Thư viện LLM và Integration**
   * **langchain-openai cho OpenAI API**
   * **Anthropic Claude integration**
   * **HuggingFace Hub integration**
5. **Thư viện dev tools và deployment**
   * **LangSmith cho tracing và debug**
   * **LangServe cho API deployment**
   * **LangGraph cho multi-agent orchestration**

**V. CHI TIẾT INPUT-OUTPUT CỦA CÁC FUNCTION/METHODS**

1. **LangChain Document Loaders**
   * **Input: PDF file path / CSV file path**
   * **Output: Document objects với text và metadata**
2. **LangChain Text Splitters**
   * **Input: Document objects**
   * **Output: Chunked documents cho PDF và CSV**
3. **LangChain Embedding Models**
   * **Input: Document chunks**
   * **Output: Vector embeddings**
4. **LangChain Vector Stores**
   * **Input: Embeddings và documents**
   * **Output: Indexed vector database**
5. **LangChain RetrievalQA Chain**
   * **Input: User query, Vector store**
   * **Output: Generated answer dựa trên context từ tài liệu**
6. **LangChain CSV Agent**
   * **Input: CSV file, User query**
   * **Output: Phân tích dữ liệu và trả lời**
7. **LangChain Multi-Document Agent**
   * **Input: PDF và CSV files, User query**
   * **Output: Phân tích và trả lời dựa trên nhiều nguồn**

**VI. CODE VÍ DỤ TỪNG FUNCTION**

1. **Cài đặt và Import Libraries**

pip install langchain-community langchain-openai langchain-core langchain faiss-cpu pandas

1. **PDF Document Loading và Processing với LangChain**

from langchain\_community.document\_loaders import PyPDFLoader

from langchain.text\_splitter import RecursiveCharacterTextSplitter

# Code ví dụ PDF loading với LangChain

1. **CSV Document Loading và Processing với LangChain**

from langchain\_community.document\_loaders.csv\_loader import CSVLoader

from langchain.chains import create\_pandas\_dataframe\_agent

# Code ví dụ CSV loading và agent

1. **Tạo Embedding và Vector Database**

from langchain\_openai import OpenAIEmbeddings

from langchain\_community.vectorstores import FAISS

# Code ví dụ tạo embedding và vector database

1. **Triển khai RAG Pipeline với LangChain**

from langchain.chains import RetrievalQA

from langchain\_openai import ChatOpenAI

# Code ví dụ RAG pipeline

1. **LangChain Agent cho PDF và CSV**

from langchain.agents import create\_structured\_parser\_agent

from langchain\_community.tools.document\_loader import load\_pdf\_or\_csv

# Code ví dụ agent xử lý nhiều tài liệu

1. **LangChain Memory và Conversation**

from langchain.memory import ConversationBufferMemory

from langchain.chains import ConversationalRetrievalChain

# Code ví dụ conversation với memory

1. **Triển khai với LangServe**

from langserve import RemoteRunnable, RemoteClient

# Code ví dụ triển khai API

**VII. BỘ CÂU HỎI TÌNH HUỐNG CHO CÁC NHÓM (6 slides)**

1. **Câu hỏi về thiết kế LangChain cho xử lý tài liệu (5 câu)**
   * **Làm thế nào để thiết kế LangChain chains xử lý tài liệu PDF có định dạng phức tạp (bảng, biểu đồ)?**
   * **Khi nào nên sử dụng RecursiveCharacterTextSplitter thay vì MarkdownHeaderTextSplitter?**
2. **Câu hỏi về LangChain cho phân tích dữ liệu CSV (5 câu)**
   * **Làm thế nào để kết hợp thông tin từ PDF và CSV trong cùng một LangChain chain?**
   * **Chiến lược nào để tạo LangChain agent xử lý các cột dữ liệu thiếu trong CSV?**
3. **Câu hỏi về RAG với LangChain (5 câu)**
   * **Làm thế nào để giảm thiểu hiện tượng hallucination trong RAG pipeline của LangChain?**
   * **So sánh hiệu suất của các vector stores khác nhau trong LangChain?**
4. **Câu hỏi về LangChain Agents (5 câu)**
   * **Khi nào nên sử dụng Agent thay vì Chain trong LangChain?**
   * **Làm thế nào để thiết kế custom tools trong LangChain cho xử lý tài liệu đa ngôn ngữ?**

**VIII. NỘI DUNG BỔ SUNG (TỔNG KẾT)**

1. **Tóm tắt các điểm chính**
2. **Demo và use cases**
   * **Ứng dụng trong phân tích báo cáo tài chính**
   * **Ứng dụng trong tổng hợp tài liệu học thuật**
   * **Ứng dụng trong phân tích dữ liệu kinh doanh**
3. **Danh sách bộ câu hỏi trắc nghiệm (10-20 câu)**
4. **Tài liệu tham khảo và nguồn học thêm**
5. **Hướng phát triển tương lai**
   * **Tích hợp với các dạng tài liệu khác (docx, pptx, html)**
   * **Phân tích trực quan (charts, graphs) trong tài liệu**
   * **Real-time collaborative document analysis**