

**TÀI LIỆU ĐẶC TẢ HỆ THỐNG  
GIA SƯ AI**

**Hà Nội – 2026**

## Mục lục

1.	Tổng quan hệ thống.....	1
1.1	Giới thiệu chung.....	1
1.2	Phạm vi hệ thống.....	1
2.	Kiến trúc hệ thống.....	2
3.	Các chức năng của hệ thống .....	3
3.1.	Upload và xử lý tài liệu học tập.....	3
3.2.	Tóm tắt nội dung tài liệu.....	5
3.3.	Lộ trình học tập.....	7
3.4.	Sinh Flashcards học tập .....	9
3.5.	Tạo đề ôn tập tự động .....	11
3.6.	Hỏi đáp với AI .....	13
3.7.	Sinh đề kiểm tra dựa trên ma trận đề .....	15
3.8.	Giải bài tập.....	17
3.9.	Học bài.....	19
4.	Cơ sở dữ liệu .....	21
5.	Giao diện (Wireframe) .....	21

## **Mục lục hình ảnh**

Hình 2.1: Kiến trúc hệ thống .....	2
Hình 3.1: Luồng Tải tài liệu .....	3
Hình 3.2: Luồng Tóm tắt tài liệu .....	5
Hình 3.3: Luồng Sinh lô trình học tập .....	7
Hình 3.4: Luồng Tạo Flashcards .....	9
Hình 3.5: Luồng Tạo đề ôn tập .....	11
Hình 3.6: Luồng Hỏi chat AI .....	13
Hình 3.7: Luồng Tạo đề kiểm tra .....	15
Hình 3.8: Luồng Giải bài tập .....	17
Hình 3.9: Luồng Học bài .....	19
Hình 4.1: Luồng dữ liệu.....	21

# 1. Tổng quan hệ thống

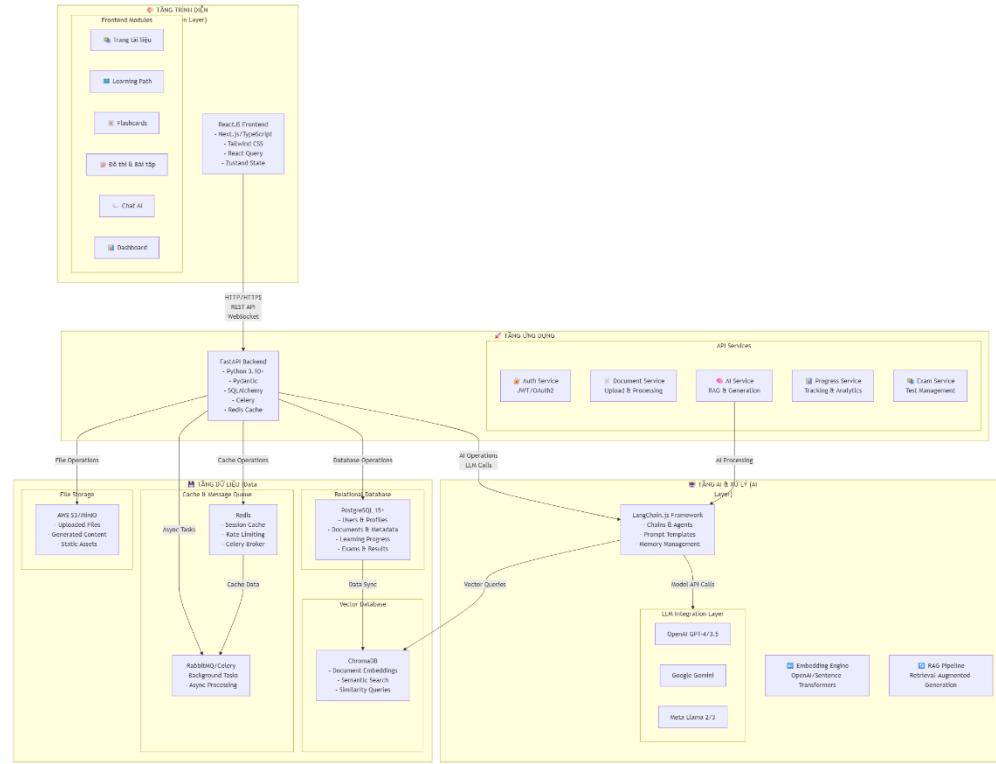
## 1.1 Giới thiệu chung

Hệ thống “Gia sư AI” được thiết kế để hỗ trợ người học tự động hóa nhiều khâu trong quá trình ôn luyện: từ tải lên tài liệu học, tóm tắt nội dung, tạo bài tập và đề thi đến hỏi đáp trực tiếp với trợ lý ảo. Mục tiêu là tận dụng sức mạnh của các mô hình ngôn ngữ lớn (LLMs) hiện đại để nâng cao hiệu quả học tập, kết hợp với cơ sở dữ liệu chuẩn và vector store cho RAG.

## 1.2 Phạm vi hệ thống

- In-scope:
  - Xử lý các định dạng tài liệu: PDF, DOCX.
  - Tóm tắt tài liệu.
  - Sinh câu hỏi và Flashcard từ văn bản.
  - Chatbot hỏi đáp dựa trên nội dung tài liệu (RAG).
  - Tạo đề kiểm tra.
  - Giải bài tập.
  - Học theo lộ trình học tập.
- Out-scope:
  - Xử lý video.
  - Tính năng mạng xã hội/diễn đàn phức tạp.

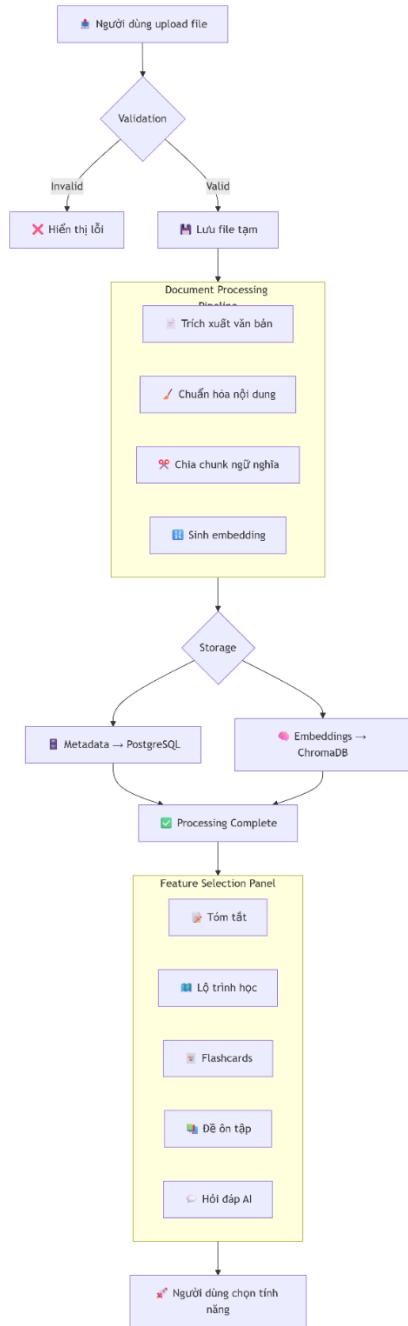
## 2. Kiến trúc hệ thống



Hình 2.1: Kiến trúc hệ thống

### 3. Các chức năng của hệ thống

#### 3.1. Upload và xử lý tài liệu học tập



Hình 3.1: Luồng Tải tài liệu

a) Mô tả chức năng

Chuyển tài liệu học tập thô (PDF, DOCX) thành **dữ liệu có cấu trúc + embedding ngữ nghĩa**, sẵn sàng cho các chức năng AI như tóm tắt, flashcards, hỏi đáp (RAG), tạo đề.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Người dùng chọn chức năng “Tải tài liệu”.

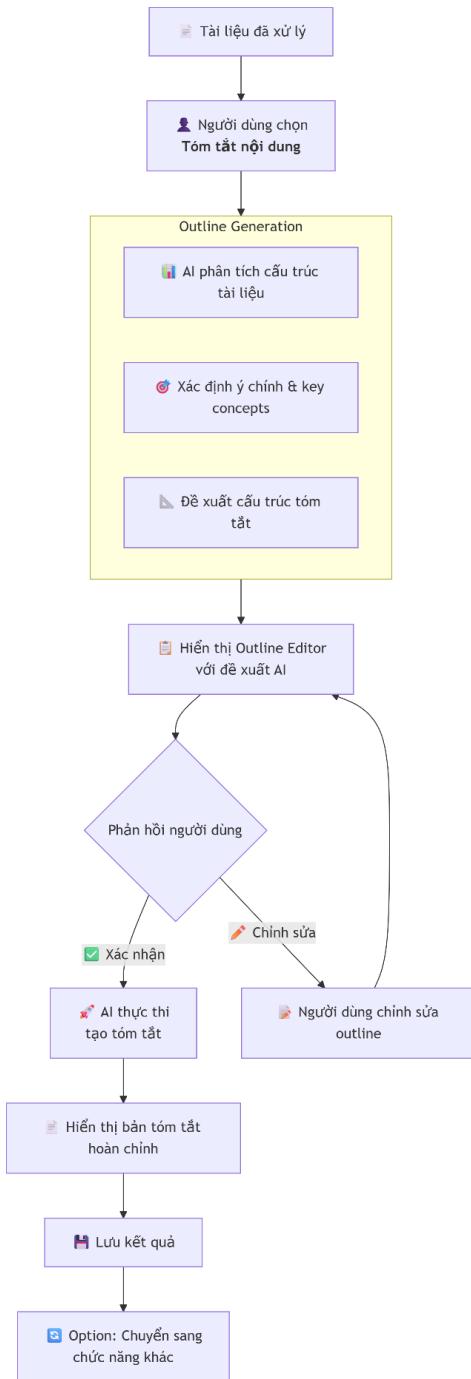
c) Luồng:

1. Người dùng upload tài liệu (PDF, DOCX)
2. Hệ thống:
  - Trích xuất văn bản
  - Chuẩn hóa nội dung
  - Chia nhỏ văn bản (chunking)
  - Sinh embedding
3. Lưu:
  - Văn bản & metadata vào CSDL
  - Embedding vào Vector Database
4. Hệ thống hiển thị các tùy chọn xử lý tiếp theo:
  - Tóm tắt nội dung (mặc định)
  - Lộ trình học tập
  - Sinh flashcard ôn tập
  - Tạo đề ôn tập
  - Hỏi đáp AI về tài liệu

d) Kỹ thuật áp dụng

- Semantic embedding
- Chunking
- Vector database
- RAG preparation

### 3.2. Tóm tắt nội dung tài liệu



Hình 3.2: Luồng Tóm tắt tài liệu

a) Mô tả

Chức năng Tóm tắt tài liệu cho phép hệ thống tự động tạo bản tóm tắt nội dung chính của tài liệu học tập đã được người dùng tải lên. Hỗ trợ người học nhanh chóng nắm bắt các ý quan trọng, làm nền tảng cho việc ôn tập và khai thác kiến thức hiệu quả.

Chức năng này được thiết kế theo mô hình **Human-in-the-Loop**, trong đó người dùng có thể tham gia vào quá trình xác nhận hoặc chỉnh sửa cấu trúc tóm tắt trước khi nội dung cuối cùng được tạo.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Tài liệu học tập đã được tải lên và xử lý thành công.
- Người dùng chọn chức năng “Tóm tắt tài liệu”.

c) Luồng

1. Người dùng chọn Tóm tắt nội dung
2. Hệ thống phân tích tài liệu và lập kế hoạch tóm tắt (Outline)
3. Hiển thị outline cho người dùng:
  - Các ý chính
  - Cấu trúc tóm tắt
4. Người dùng:
  - Xác nhận outline
  - Hoặc chỉnh sửa outline
5. AI sinh nội dung tóm tắt dựa trên outline đã xác nhận
6. Lưu kết quả

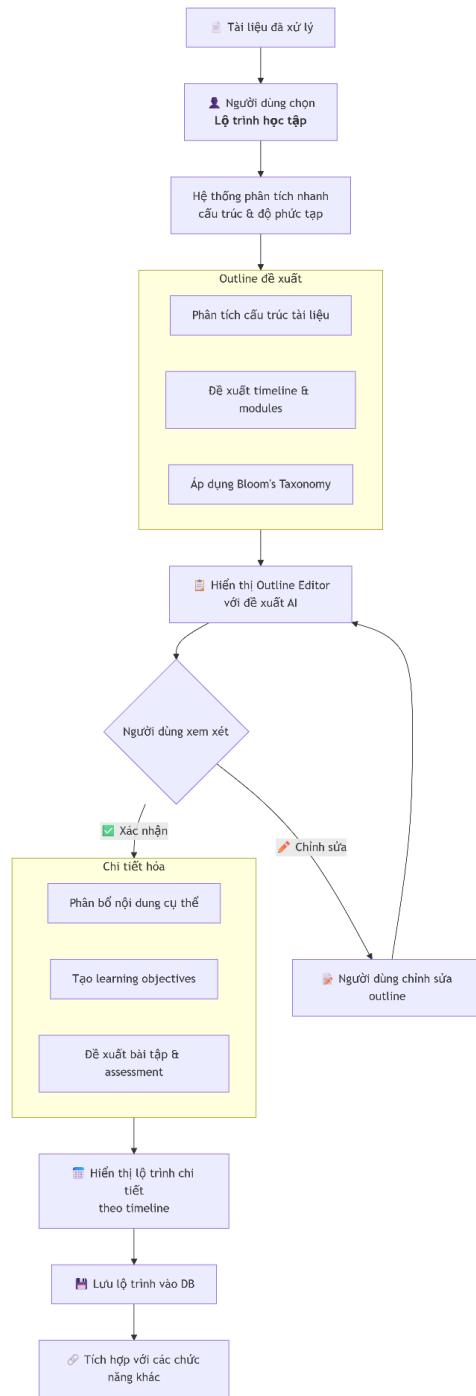
d) Công nghệ

- AI Framework: LangChain.js
- LLM: GPT-3.5

e) Kỹ thuật áp dụng

- AI Planning + Outline Generation
- Human-in-the-Loop
- Abstractive summarization

### 3.3. Lộ trình học tập



Hình 3.3: Luồng Sinh lộ trình học tập

a) Mô tả

Chức năng Sinh lô trình học tập cho phép hệ thống tự động tạo ra một lô trình học có cấu trúc từ nội dung tài liệu học tập mà người dùng đã tải lên.

Lô trình được tổ chức theo giai đoạn và bài học, giúp người học tiếp cận kiến thức một cách có hệ thống và phù hợp với tiến trình học tập.

Chức năng này được thiết kế theo mô hình Human-in-the-Loop, cho phép người dùng tham gia xác nhận hoặc điều chỉnh cấu trúc lô trình trước khi hệ thống sinh nội dung chi tiết.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Tài liệu học tập đã được tải lên và xử lý thành công.
- Người dùng chọn chức năng “Tạo lô trình học tập”.

c) Luồng

1. Người dùng chọn chức năng **Lộ trình học tập** từ danh sách các chức năng AI liên quan đến tài liệu.
2. Hệ thống phân tích nội dung tài liệu để xác định các chủ đề, khái niệm và mối quan hệ giữa chúng.
3. Hệ thống đề xuất một outline lô trình học tập, bao gồm:
  - Các giai đoạn học tập
  - Thứ tự kiến thức và mức độ tiếp cận
4. Outline được hiển thị cho người dùng để xem xét.
5. Người dùng chỉnh sửa hoặc xác nhận outline theo mục tiêu học tập cá nhân.
6. Sau khi outline được xác nhận, hệ thống sinh ra lô trình học tập chi tiết.
7. Lô trình học tập được lưu trữ và hiển thị cho người dùng.

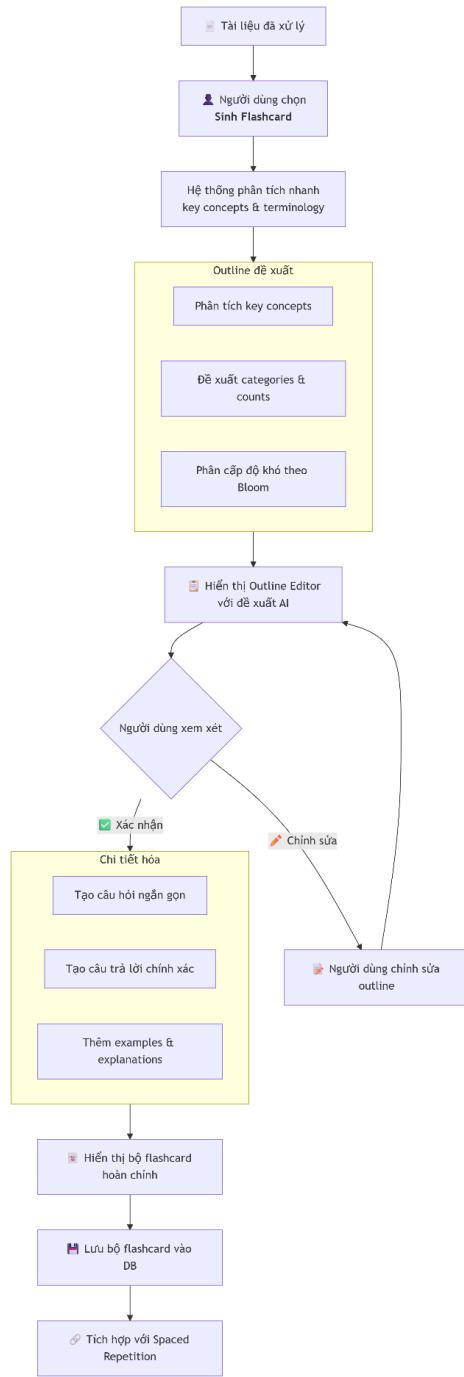
d) Công nghệ

- AI Framework: LangChain.js
- LLM: GPT-3.5

e) Kỹ thuật áp dụng

- Curriculum Generation
- Prompt theo Bloom's Taxonomy
- Human-in-the-loop

### 3.4. Sinh Flashcards học tập



Hình 3.4: Luồng Tạo Flashcards

a) Mô tả

Chức năng **Sinh flashcards ôn tập** cho phép hệ thống tự động tạo ra các thẻ học (flashcards) dưới dạng câu hỏi – trả lời (Q&A) từ nội dung tài liệu học tập đã được tải lên. Các flashcards này hỗ trợ người học trong quá trình ghi nhớ, ôn tập và củng cố kiến thức một cách hiệu quả.

Tương tự các chức năng AI khác trong hệ thống, chức năng này được thiết kế theo mô hình **Human-in-the-Loop**, cho phép người dùng tham gia điều chỉnh cấu trúc flashcards trước khi nội dung cuối cùng được sinh ra.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Tài liệu học tập đã được tải lên và xử lý thành công.
- Người dùng chọn chức năng “Sinh flashcards”.

c) Luồng

1. Người dùng chọn chức năng **Sinh flashcards** từ danh sách các chức năng AI của tài liệu.
2. Hệ thống phân tích nội dung tài liệu để xác định các chủ đề và khái niệm chính.
3. Hệ thống đề xuất một **outline flashcards**, bao gồm:
  - Các chủ đề kiến thức
  - Số lượng flashcards cho mỗi chủ đề
  - Mức độ nhận thức của câu hỏi
4. Outline flashcards được hiển thị cho người dùng để chỉnh sửa hoặc xác nhận.
5. Sau khi outline được xác nhận, hệ thống sinh các flashcards hoàn chỉnh dưới dạng câu hỏi – trả lời.
6. Các flashcards được lưu trữ trong hệ thống và hiển thị cho người dùng.

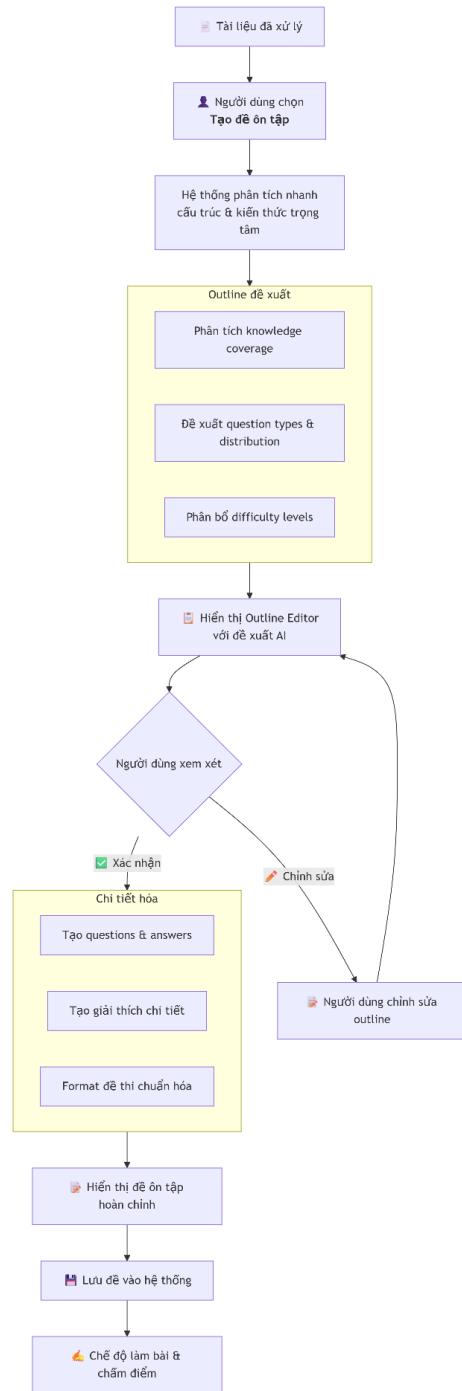
d) Công nghệ

- AI Framework: LangChain.js
- LLM: GPT-3.5

e) Kỹ thuật áp dụng

- Automatic Question Generation
- Prompt phân cấp độ nhận thức
- Human-in-the-loop

### 3.5. Tạo đề ôn tập tự động



Hình 3.5: Luồng Tạo đề ôn tập

a) Mô tả

Chức năng Tạo đề ôn tập cho phép hệ thống tự động sinh ra một bộ câu hỏi ôn tập từ nội dung tài liệu học tập đã được tải lên.

Đề ôn tập có thể bao gồm các câu hỏi trắc nghiệm và/hoặc tự luận, được thiết kế nhằm giúp người học luyện tập và tự đánh giá mức độ hiểu bài sau khi học.

Chức năng được triển khai theo mô hình Human-in-the-Loop, trong đó người dùng có thể tham gia xác nhận hoặc điều chỉnh cấu trúc đề trước khi hệ thống sinh nội dung hoàn chỉnh.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Tài liệu học tập đã được tải lên và xử lý thành công.
- Người dùng chọn chức năng “Tạo đề ôn tập”.

c) Luồng

1. Người dùng chọn chức năng **Tạo đề ôn tập** từ danh sách các chức năng AI của tài liệu.
2. Hệ thống phân tích nội dung tài liệu để xác định phạm vi kiến thức phù hợp cho đề ôn tập.
3. Hệ thống đề xuất một **outline đề ôn tập**, bao gồm:
  - Loại câu hỏi (trắc nghiệm, tự luận)
  - Số lượng câu hỏi
  - Độ khó của từng nhóm câu hỏi
4. Outline đề ôn tập được hiển thị cho người dùng để chỉnh sửa hoặc xác nhận.
5. Sau khi outline được xác nhận, hệ thống sinh ra đề ôn tập hoàn chỉnh.
6. Đề ôn tập được lưu trữ trong hệ thống.
7. Người dùng có thể thực hiện làm bài và xem kết quả sau khi hoàn thành.

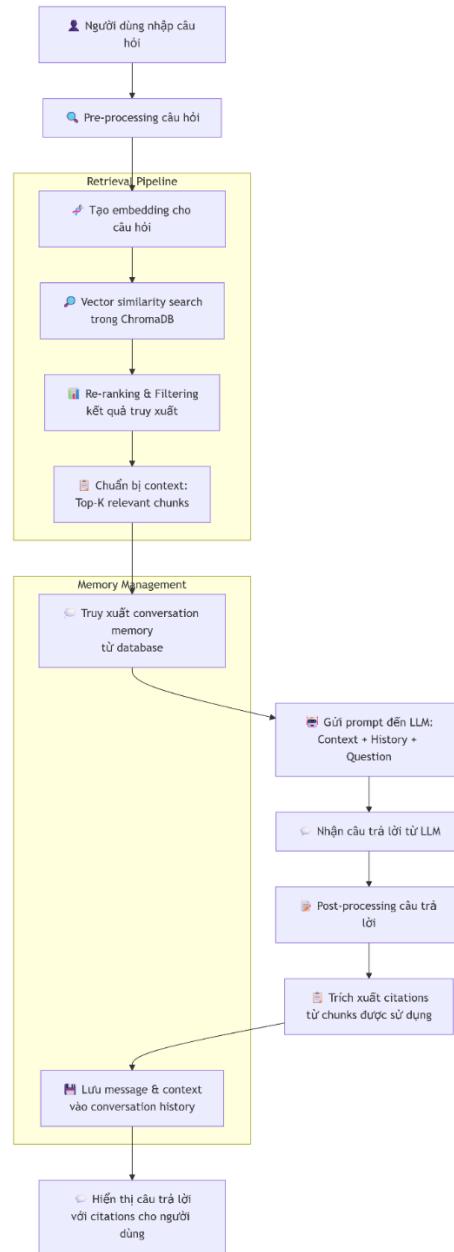
d) Công nghệ

- AI Framework: LangChain.js
- LLM: GPT-3.5/Gemini

e) Kỹ thuật áp dụng

- Automatic Question Generation
- Human-in-the-loop

### 3.6. Hỏi đáp với AI



Hình 3.6: Luồng Hỏi chat AI

#### a) Mô tả

Chức năng Hỏi đáp AI về tài liệu cho phép người dùng đặt câu hỏi trực tiếp liên quan đến nội dung của tài liệu học tập đã được tải lên hệ thống.

Hệ thống sử dụng mô hình chatbot tích hợp AI để trả lời câu hỏi dựa trên ngữ cảnh trích xuất từ chính tài liệu của người dùng, thay vì trả lời theo kiến thức chung.

Chức năng này được xây dựng dựa trên mô hình RAG nhằm đảm bảo câu trả lời có cẩn cứ, giảm hiện tượng suy diễn sai và tăng độ chính xác.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Tài liệu học tập đã được tải lên và xử lý thành công.
- Người dùng chọn chức năng “Hỏi đáp AI”.

c) Luồng

1. Người dùng nhập câu hỏi liên quan đến nội dung tài liệu.
2. Hệ thống sinh embedding cho câu hỏi của người dùng.
3. Embedding của câu hỏi được sử dụng để truy vấn Vector Database.
4. Hệ thống lấy ra các đoạn nội dung trong tài liệu có mức độ liên quan cao.
5. Các đoạn nội dung liên quan được kết hợp với câu hỏi và gửi tới mô hình ngôn ngữ.
6. Mô hình AI sinh câu trả lời dựa trên ngữ cảnh được cung cấp.
7. Câu hỏi và câu trả lời được lưu lại dưới dạng hội thoại để phục vụ việc theo dõi và tham chiếu.

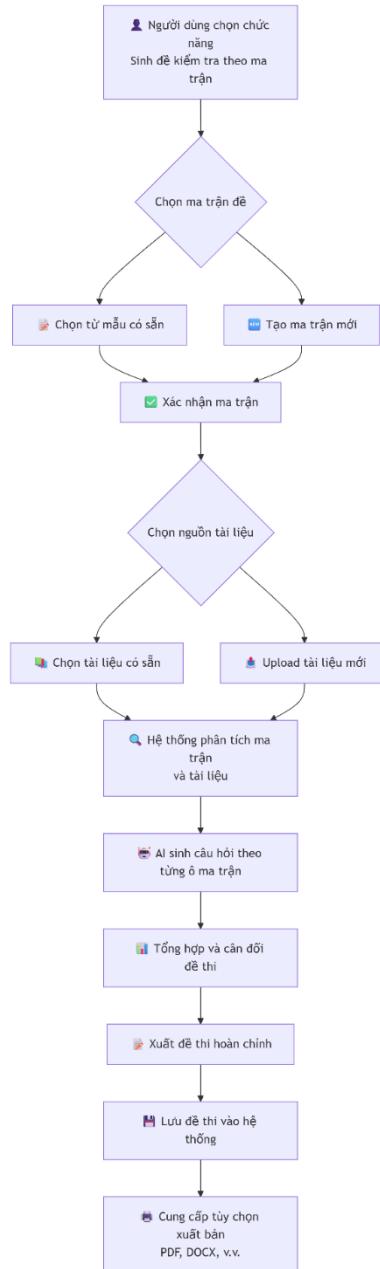
d) Công nghệ

- AI Framework: LangChain.js
- Vector database: ChromaDB
- LLM: GPT-3.5/Gemini/Llama

e) Kỹ thuật áp dụng

- Retrieval-Augmented Generation (RAG): Kết hợp truy vấn ngữ nghĩa và sinh câu trả lời từ mô hình ngôn ngữ.
- Conversation Memory
- Context Grounding

### 3.7. Sinh đè kiểm tra dựa trên ma trận đè



Hình 3.7: Luồng Tạo đè kiểm tra

#### a) Mô tả

Chức năng Sinh đè kiểm tra theo ma trận đè cho phép hệ thống tự động tạo ra một đè kiểm tra hoàn chỉnh dựa trên ma trận đè do người dùng lựa chọn hoặc xây dựng.

Chức năng này hướng đến việc hỗ trợ người học tạo đề kiểm tra có tính hệ thống, cân bằng và phù hợp mục tiêu đánh giá.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Người dùng đã chọn hoặc tạo một ma trận để hợp lệ.
- Người dùng đã chọn nguồn nội dung để sinh đề:
  - Sử dụng tài liệu đã có trong hệ thống
  - Upload tài liệu mới để sử dụng cho việc sinh đề. → Xử lý tài liệu upload.
- Nội dung tài liệu đã sẵn sàng để phục vụ cho các chức năng AI.

c) Luồng

1. Người dùng chọn hoặc tạo một ma trận để trong hệ thống.
2. Người dùng lựa chọn nguồn tài liệu
3. Hệ thống phân tích cấu trúc ma trận để, xác định các ô ma trận tương ứng với:
  - Chủ đề nội dung
  - Mức độ nhận thức
  - Số lượng câu hỏi
4. Hệ thống lần lượt sinh câu hỏi cho từng ô trong ma trận để theo các ràng buộc đã xác định.
5. Các câu hỏi được tổng hợp và sắp xếp để tạo thành một đề kiểm tra hoàn chỉnh.
6. Đề kiểm tra được lưu trữ trong hệ thống và sẵn sàng cho việc sử dụng. Cho phép tải xuống các định dạng PDF.

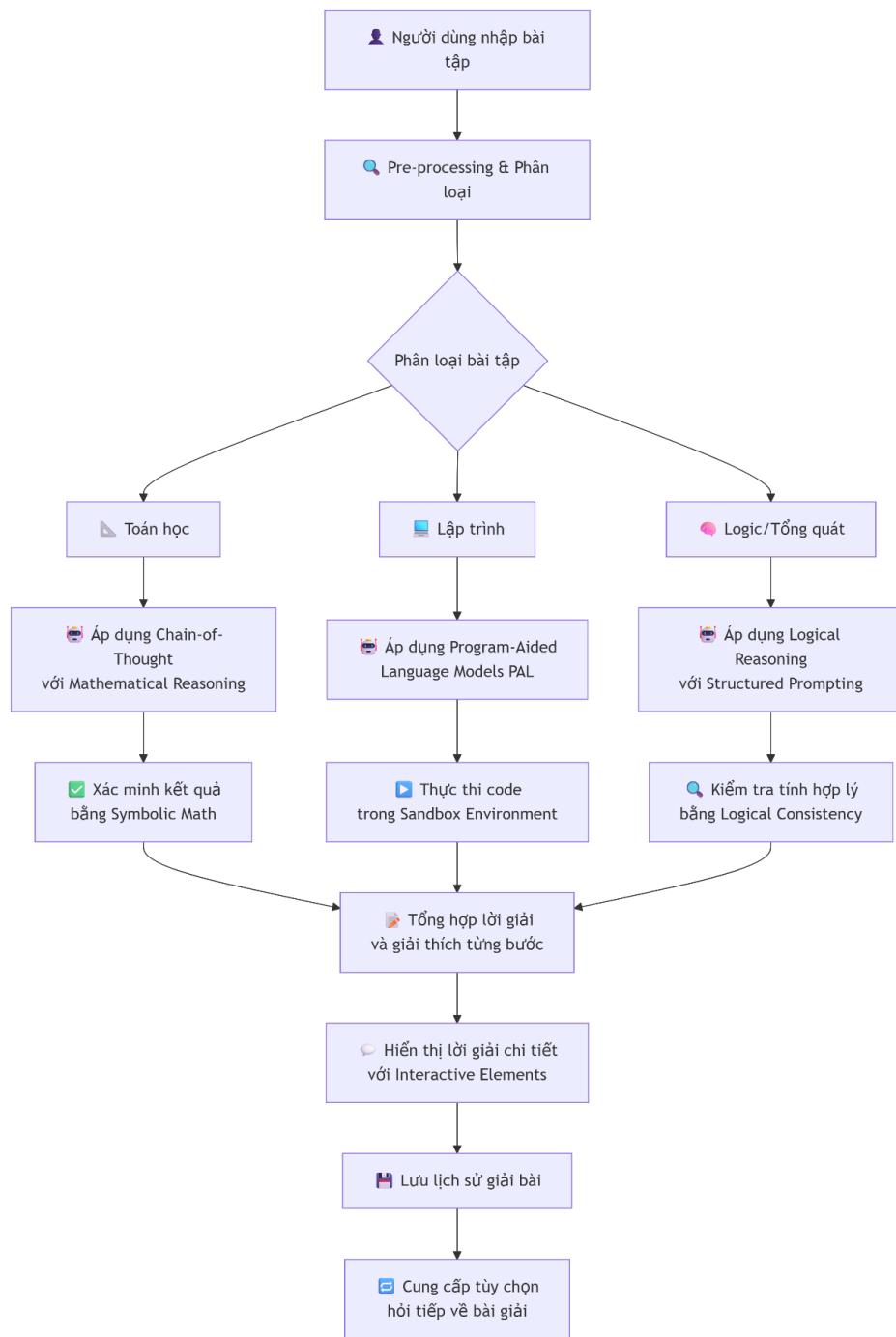
d) Công nghệ

- AI Framework: LangChain.js
- LLM: GPT-3.5/Gemini

e) Kỹ thuật áp dụng

- Blueprint-based Test Generation: Sinh đề kiểm tra dựa trên cấu trúc ma trận để đã xác định.
- Rule-based + LLM Hybrid: Kết hợp các ràng buộc định sẵn với mô hình ngôn ngữ để đảm bảo đề kiểm tra tuân thủ yêu cầu.

### 3.8. Giải bài tập



Hình 3.8: Luồng Giải bài tập

a) Mô tả

Chức năng Giải bài tập cho phép hệ thống hỗ trợ người dùng giải bài tập ở các môn học ở nhiều lĩnh vực khác nhau. Hệ thống sinh ra lời giải chi tiết theo từng bước, giúp người học hiểu rõ quá trình suy luận và cách tiếp cận bài toán.

Chức năng này tập trung vào việc hỗ trợ tư duy và giải thích, đóng vai trò như một trợ giảng AI trong quá trình học tập.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Người dùng có thể nhập bài tập dưới dạng văn bản (toán học, logic, lập trình).

c) Luồng

1. Người dùng nhập nội dung bài tập cần giải.
2. Backend tiếp nhận nội dung bài tập và gửi yêu cầu tới mô hình ngôn ngữ.
3. Mô hình AI sinh ra lời giải chi tiết, bao gồm các bước suy luận và kết quả.
4. Backend nhận kết quả từ mô hình AI và trả về cho người dùng.
5. Hệ thống hiển thị lời giải và các bước giải một cách rõ ràng, dễ theo dõi.

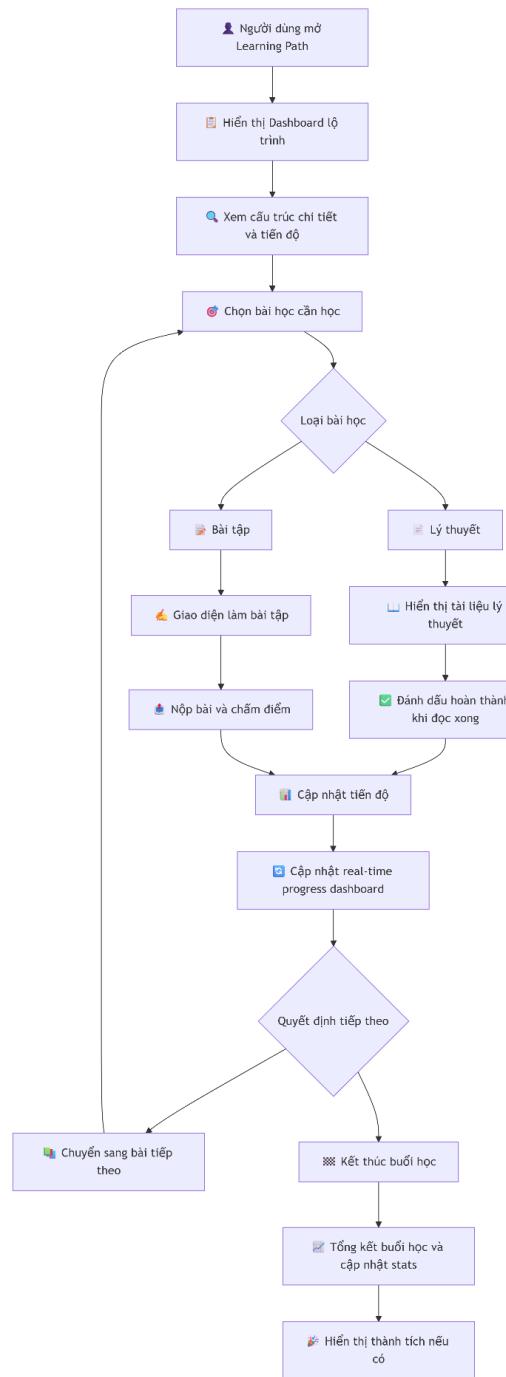
d) Công nghệ

- AI Framework: LangChain.js
- LLM: GPT-3.5/Gemini
- Kỹ thuật prompting

e) Kỹ thuật áp dụng

- Chain-of-Thought Prompting: Hướng mô hình sinh lời giải theo từng bước suy luận logic.
- Explainable AI (XAI ở mức ứng dụng): Trình bày lời giải theo cách người học có thể hiểu và theo dõi được quá trình suy luận

### 3.9. Học bài



Hình 3.9: Luồng Học bài

a) Mô tả

Chức năng **Học bài** theo lộ trình học tập cho phép người dùng học nội dung theo lộ trình/đề ôn tập đã được hệ thống sinh ra.

Hệ thống hỗ trợ ghi nhận trạng thái hoàn thành của từng nội dung học hoặc điểm số, đồng thời cập nhật tiến độ học tập tổng thể của người dùng.

Chức năng này giúp người học theo dõi quá trình học tập một cách có hệ thống và tạo động lực học tập liên tục.

b) Tiền điều kiện

- Người dùng đã đăng nhập vào hệ thống.
- Lộ trình học tập, đề ôn tập đã được sinh và lưu trữ.
- Người dùng truy cập vào một lộ trình học tập.

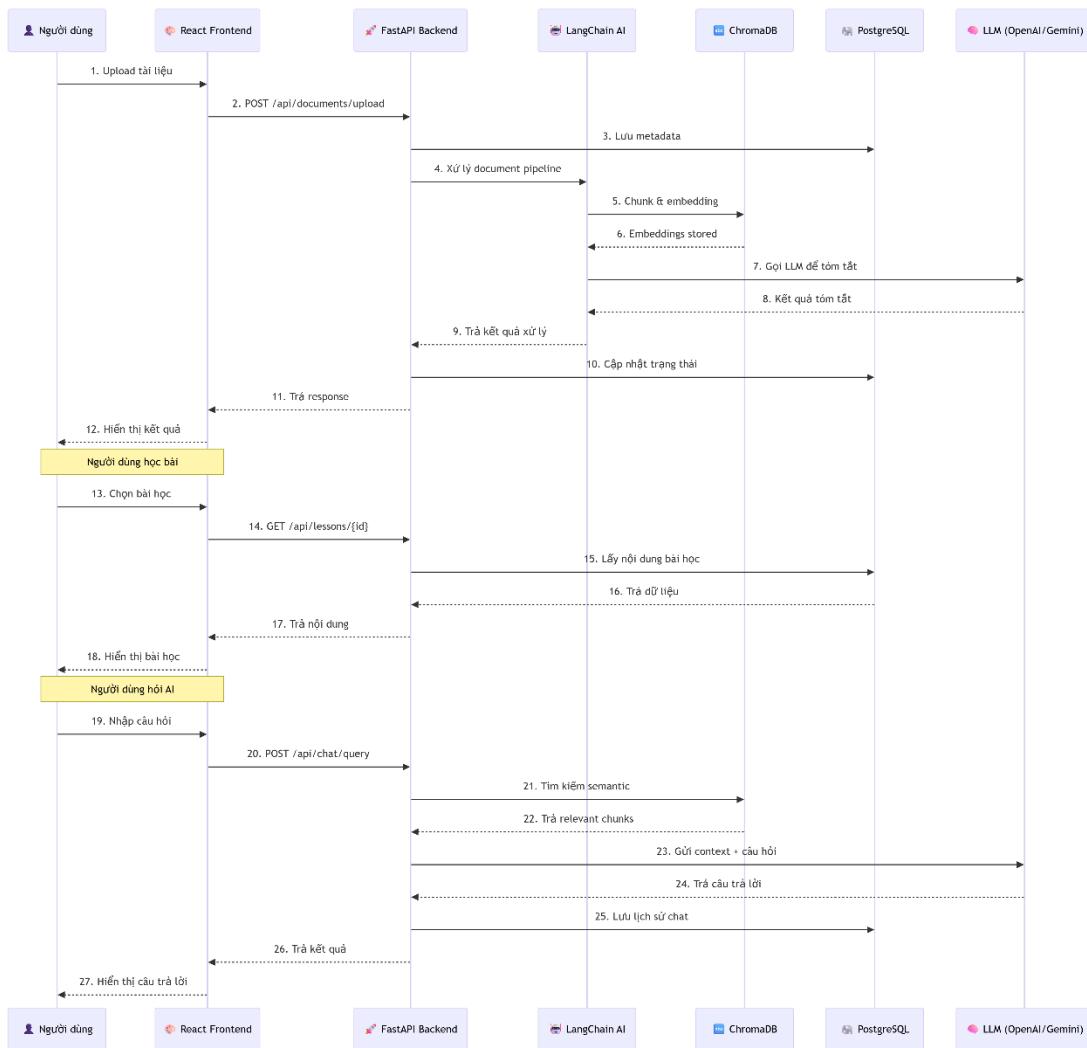
c) Luồng

1. Người dùng chọn một lộ trình học tập đã được tạo.
2. Hệ thống hiển thị danh sách các bài học theo thứ tự trong lộ trình.
3. Người dùng chọn một bài học để bắt đầu học.
4. Trong quá trình học, người dùng có thể:
  - Đánh dấu bài học đã hoàn thành
  - Chuyển sang bài học tiếp theo
5. Hệ thống ghi nhận trạng thái hoàn thành của từng bài học.
6. Tiến độ học tập tổng thể của lộ trình được cập nhật và hiển thị cho người dùng.

d) Kỹ thuật áp dụng

- Progress Tracking: Theo dõi và cập nhật tiến độ học tập.
- State Management: Quản lý trạng thái hoàn thành của bài học.
- Result Tracking: Ghi nhận kết quả làm bài.
- Basic Assessment Logic: Đánh giá và tổng hợp kết quả.

## 4. Cơ sở dữ liệu



Hình 4.1: Luồng dữ liệu

## 5. Giao diện (Wireframe)