

TÀI LIỆU ĐẶC TẢ YÊU CẦU PHẦN MỀM (SRS)

HỆ THỐNG STUDY BUDDY - CHATBOT AI HỖ TRỢ HỌC TẬP

Nhóm thực hiện: Nhóm 6

Lớp: Kỹ thuật phần mềm-1-3-24(COUR01.LT7)

Giảng viên: Mai Xuân Tráng

Ngày: 18/06/2025

TÀI LIỆU ĐẶC TẢ YÊU CẦU PHẦN MỀM (SRS)	1
HỆ THỐNG STUDY BUDDY - CHATBOT AI HỖ TRỢ HỌC TẬP	1
1. GIỚI THIỆU	4
1.1 Mục đích	4
1.2 Phạm vi	4
1.3 Từ điển thuật ngữ	4
1.4 Tổng quát	5
2. CÁC YÊU CẦU CHỨC NĂNG	6
2.1 Các tác nhân	6
2.1.1 Khách (Guest)	6
2.1.2 Người dùng đã đăng nhập (Authenticated User)	6
2.1.3 Hệ thống AI (AI System)	6
2.1.4 Hệ thống cơ sở dữ liệu (Database System)	6
2.2 Các chức năng của hệ thống	6
2.3 Biểu đồ use-case tổng quát	8
2.4 Biểu đồ use-case phân rã	8
2.4.1 Biểu đồ use-case cho Khách	8
2.4.2 Biểu đồ use-case cho Người dùng đã đăng nhập	9
2.5 Quy trình nghiệp vụ	10
2.5.1 Quy trình đăng nhập	10
2.5.2 Quy trình xử lý tài liệu	11
2.5.3 Quy trình chat với AI	12
2.5.4 Quy trình chat theo trang PDF	13
2.5.5 Quy trình quản lý sessions	14
2.6 Đặc tả use-case	14
UC001: Đăng nhập	14
UC002: Upload và xử lý tài liệu	16
UC003: Chat với AI sử dụng RAG	17
UC004: Chat theo trang PDF cụ thể	18
UC005: Quản lý chat sessions	20
UC006: Đăng ký tài khoản	21
UC007: Đăng xuất	22
3. CÁC YÊU CẦU PHI CHỨC NĂNG	23
3.1 Các yêu cầu về hiệu năng	23
3.2 Yêu cầu về bảo mật	24
3.3 Yêu cầu về giao diện	24
3.4 Ràng buộc	24
4. PHÂN CHIA CÔNG VIỆC NHÓM	25

4.1 Tổng quan phân chia	25
4.2 Thành viên 1: Nguyễn Ngọc Công Anh- Frontend Development & UI/UX	25
4.3 Thành viên 2: Trần Long Khánh - Backend Logic & AI Integration	27
4.4 Thành viên 3: Trần Đức Việt - Database & Authentication	28
4.5 Timeline và Milestone	29
4.6 Communication và Collaboration	30
4.7 Quality Assurance	30
5. KẾT LUẬN	31
5.1 Tổng kết dự án	31
5.2 Điểm mạnh của hệ thống	31
Khả năng mở rộng: Kiến trúc cho phép mở rộng thêm nhiều tính năng trong tương lai.	
31	
5.3 Rủi ro và giải pháp	31

1. GIỚI THIỆU

1.1 Mục đích

Tài liệu Đặc tả Yêu cầu Phần mềm (SRS) này được xây dựng nhằm xác định một cách chi tiết và đầy đủ các yêu cầu chức năng và phi chức năng của hệ thống Study Buddy - một ứng dụng chatbot AI hỗ trợ học tập thông minh. Tài liệu này đóng vai trò là cầu nối giao tiếp chính giữa nhóm phát triển, giảng viên hướng dẫn và các bên liên quan khác trong quá trình phát triển dự án.

Mục đích chính của tài liệu bao gồm: cung cấp hướng dẫn cụ thể cho nhóm phát triển trong việc thiết kế và xây dựng hệ thống, đảm bảo sản phẩm cuối cùng đáp ứng đúng nhu cầu người dùng và tiêu chuẩn kỹ thuật đã đề ra, tạo cơ sở cho việc kiểm thử và đánh giá chất lượng hệ thống.

1.2 Phạm vi

Hệ thống Study Buddy là một ứng dụng web-based chatbot AI được thiết kế để hỗ trợ học tập và nghiên cứu thông qua khả năng xử lý tài liệu thông minh và tương tác tự nhiên. Hệ thống cung cấp khả năng chat với AI sử dụng Local LLM (thông qua LM Studio), xử lý và phân tích các định dạng tài liệu phổ biến (PDF, DOCX, TXT, MD) với công nghệ OCR tiên tiến, tạo tóm tắt tự động và câu hỏi gợi ý từ nội dung tài liệu, triển khai RAG (Retrieval-Augmented Generation) để trả lời câu hỏi dựa trên tài liệu, và quản lý lịch sử chat với khả năng tìm kiếm và phân loại.

Hệ thống phục vụ ba nhóm người dùng chính: Khách (người dùng chưa đăng nhập) có thể xem giao diện và đăng ký tài khoản, Người dùng đã đăng nhập có thể sử dụng đầy đủ các tính năng chat và xử lý tài liệu, và Hệ thống AI đóng vai trò tác nhân tự động xử lý và phản hồi các yêu cầu từ người dùng.

Hệ thống không hỗ trợ: thanh toán trực tuyến hoặc thương mại điện tử, chỉnh sửa hoặc tạo mới tài liệu gốc, chia sẻ tài liệu giữa người dùng, và tích hợp với các hệ thống quản lý học tập khác (LMS).

1.3 Từ điển thuật ngữ

Thuật ngữ	Định nghĩa
AI (Artificial Intelligence)	Trí tuệ nhân tạo, công nghệ cho phép máy tính thực hiện các tác vụ thông minh như con người

Thuật ngữ	Định nghĩa
Chatbot	Chương trình máy tính được thiết kế để mô phỏng cuộc trò chuyện với người dùng thông qua text
LLM (Large Language Model)	Mô hình ngôn ngữ lớn được huấn luyện để hiểu và tạo ra văn bản tự nhiên
LM Studio	Công cụ cho phép chạy các mô hình AI ngôn ngữ lớn trên máy tính cá nhân
OCR (Optical Character Recognition)	Công nghệ nhận dạng ký tự quang học, chuyển đổi hình ảnh text thành text có thể chỉnh sửa
RAG (Retrieval-Augmented Generation)	Kỹ thuật kết hợp tìm kiếm thông tin và tạo sinh text để cải thiện độ chính xác
API (Application Programming Interface)	Giao diện lập trình ứng dụng, cho phép các phần mềm giao tiếp với nhau
Session	Phiên làm việc, một cuộc trò chuyện liên tục giữa người dùng và hệ thống
Streamlit	Framework Python để xây dựng ứng dụng web tương tác nhanh chóng
Supabase	Nền tảng cơ sở dữ liệu mã nguồn mở cung cấp database và authentication

1.4 Tổng quát

Tài liệu SRS này được tổ chức thành ba phần chính một cách logic và có hệ thống. Phần đầu tiên là Giới thiệu, cung cấp cái nhìn tổng quan về mục đích, phạm vi và bối cảnh của dự án, đồng thời giải thích các thuật ngữ kỹ thuật quan trọng sẽ được sử dụng xuyên suốt tài liệu. Phần thứ hai là Các yêu cầu chức năng, mô tả chi tiết các tác nhân trong hệ thống, danh sách đầy đủ các chức năng, các biểu đồ use-case từ tổng quát đến cụ thể, quy trình nghiệp vụ và đặc tả use-case chi tiết. Phần thứ ba là Các yêu cầu phi chức năng, trình bày các yêu cầu về hiệu năng, bảo mật, giao diện người dùng và các ràng buộc kỹ thuật. Cuối cùng là phần Phân chia công việc nhóm, mô tả chi tiết trách nhiệm và nhiệm vụ cụ thể của từng thành viên trong nhóm phát triển.

2. CÁC YÊU CẦU CHỨC NĂNG

2.1 Các tác nhân

Hệ thống Study Buddy bao gồm các tác nhân sau:

2.1.1 Khách (Guest)

Người dùng chưa có tài khoản hoặc chưa đăng nhập vào hệ thống. Khách có quyền xem giao diện chính của ứng dụng, đọc thông tin giới thiệu về các tính năng, thực hiện đăng ký tài khoản mới, và truy cập vào form đăng nhập. Tuy nhiên, khách không thể sử dụng các tính năng chat hoặc upload tài liệu.

2.1.2 Người dùng đã đăng nhập (Authenticated User)

Người dùng đã có tài khoản và đăng nhập thành công vào hệ thống. Đây là tác nhân chính có quyền truy cập đầy đủ các tính năng: thực hiện chat với AI, upload và xử lý tài liệu, xem tóm tắt và câu hỏi gợi ý, quản lý lịch sử chat sessions, chat với trang PDF cụ thể, và đăng xuất khỏi hệ thống.

2.1.3 Hệ thống AI (AI System)

Tác nhân tự động bao gồm các dịch vụ AI và xử lý backend. Hệ thống AI chịu trách nhiệm xử lý các yêu cầu chat từ người dùng thông qua Local LLM, thực hiện OCR và trích xuất nội dung từ tài liệu, tạo tóm tắt tự động và câu hỏi gợi ý, triển khai RAG để trả lời câu hỏi dựa trên tài liệu, và lưu trữ dữ liệu vào cơ sở dữ liệu.

2.1.4 Hệ thống cơ sở dữ liệu (Database System)

Tác nhân quản lý dữ liệu sử dụng Supabase, chịu trách nhiệm lưu trữ thông tin người dùng và xác thực, quản lý các chat sessions và messages, lưu trữ metadata của tài liệu đã upload, và cung cấp dữ liệu theo yêu cầu từ ứng dụng.

2.2 Các chức năng của hệ thống

Hệ thống Study Buddy cung cấp các chức năng chính sau:

Đăng ký tài khoản: Cho phép khách tạo tài khoản mới với email và mật khẩu, bao gồm xác thực email và validation dữ liệu đầu vào.

Đăng nhập/Đăng xuất: Xác thực người dùng để truy cập hệ thống và quản lý phiên làm việc một cách an toàn.

Chat với AI: Giao tiếp tự nhiên với AI assistant sử dụng Local LLM, hỗ trợ nhiều ngôn ngữ với khả năng hiểu context và duy trì cuộc hội thoại.

Upload và xử lý tài liệu: Hỗ trợ upload các định dạng PDF, DOCX, TXT, MD với khả năng OCR cho PDF có hình ảnh và trích xuất text chính xác.

Tóm tắt tự động: Sử dụng AI để tạo tóm tắt ngắn gọn, súc tích từ nội dung tài liệu dài, giúp người dùng nắm bắt nhanh ý chính.

Tạo câu hỏi gợi ý: Tự động phân tích nội dung và đề xuất các câu hỏi có thể quan tâm để hỗ trợ việc học tập và nghiên cứu.

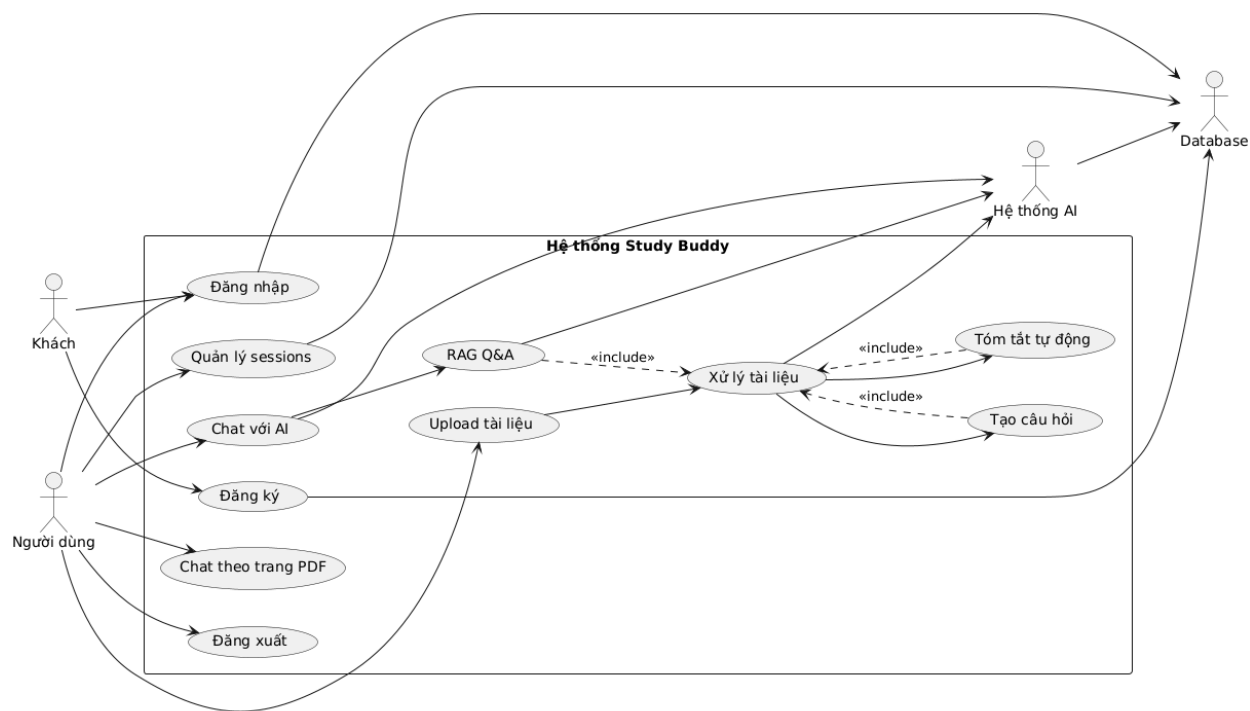
RAG (Retrieval-Augmented Generation): Trả lời câu hỏi chính xác dựa trên nội dung tài liệu đã upload, kết hợp tìm kiếm thông tin và tạo sinh câu trả lời.

Quản lý chat sessions: Lưu trữ, phân loại và tìm kiếm lịch sử các cuộc trò chuyện với khả năng tạo title thông minh và preview.

Chat theo trang PDF: Tính năng độc đáo cho phép chat với từng trang cụ thể của tài liệu PDF, hiển thị ảnh trang và nội dung tương ứng.

Quản lý tài liệu cá nhân: Lưu trữ và tổ chức các tài liệu đã upload với metadata đầy đủ và khả năng xóa khi không cần thiết.

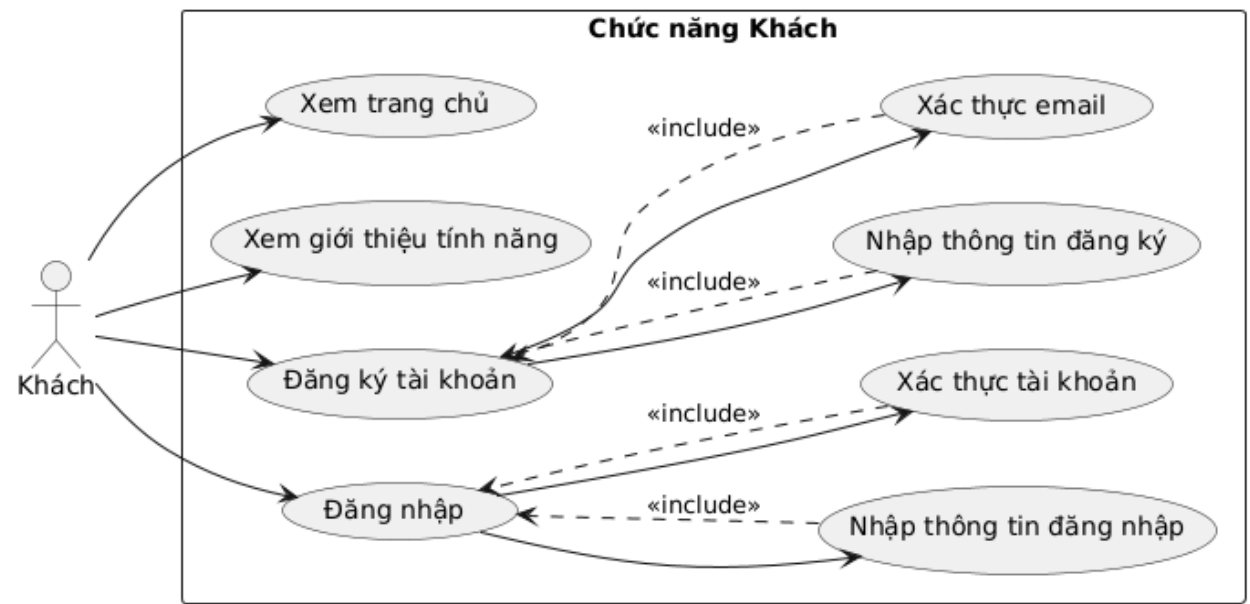
2.3 Biểu đồ use-case tổng quát



Hình 2-1: Biểu đồ use-case tổng quát của hệ thống Study Buddy

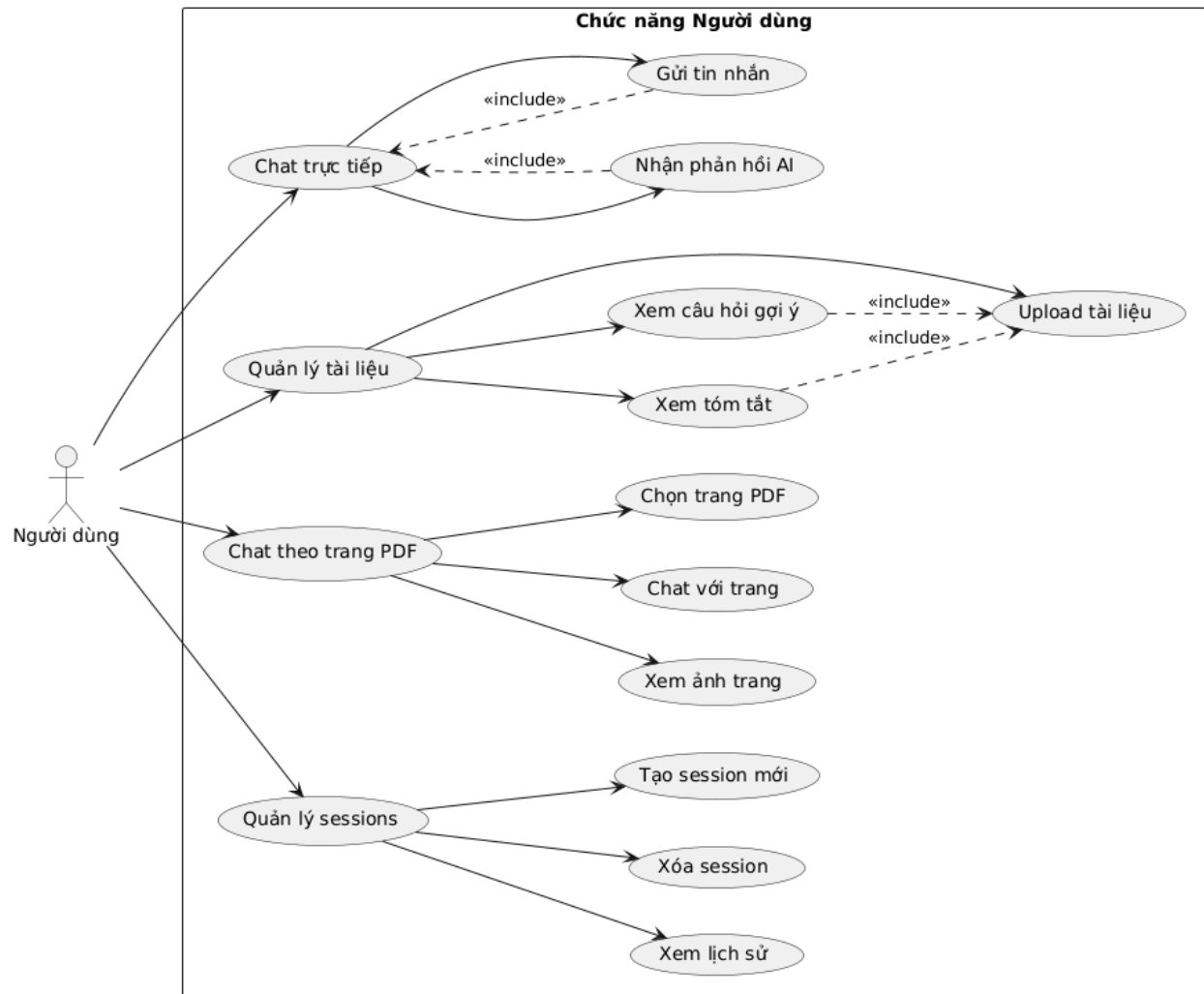
2.4 Biểu đồ use-case phân rã

2.4.1 Biểu đồ use-case cho Khách



Hình 2-2: Biểu đồ use-case chi tiết cho tác nhân Khách

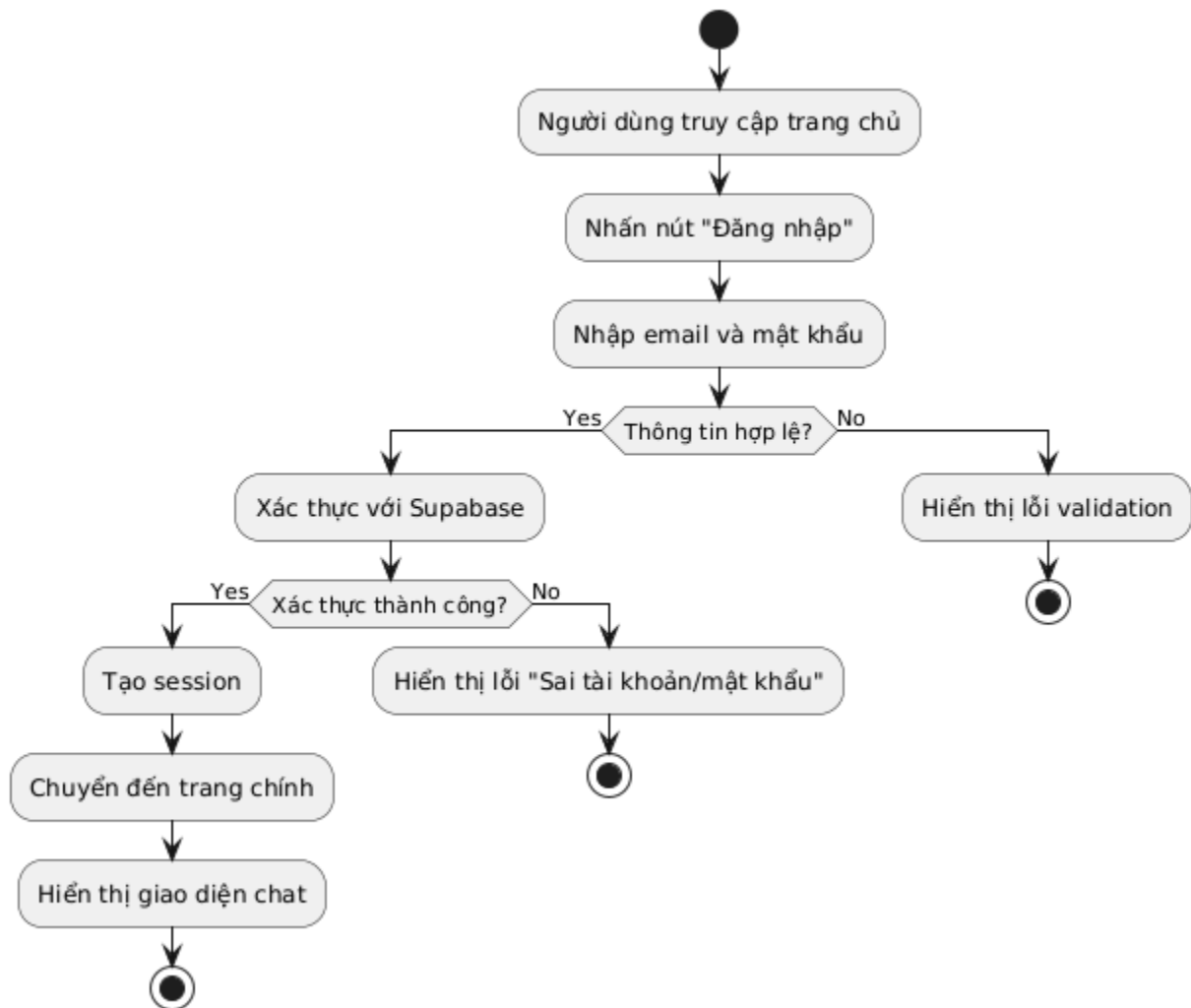
2.4.2 Biểu đồ use-case cho Người dùng đã đăng nhập



Hình 2-3: Biểu đồ use-case chi tiết cho Người dùng đã đăng nhập

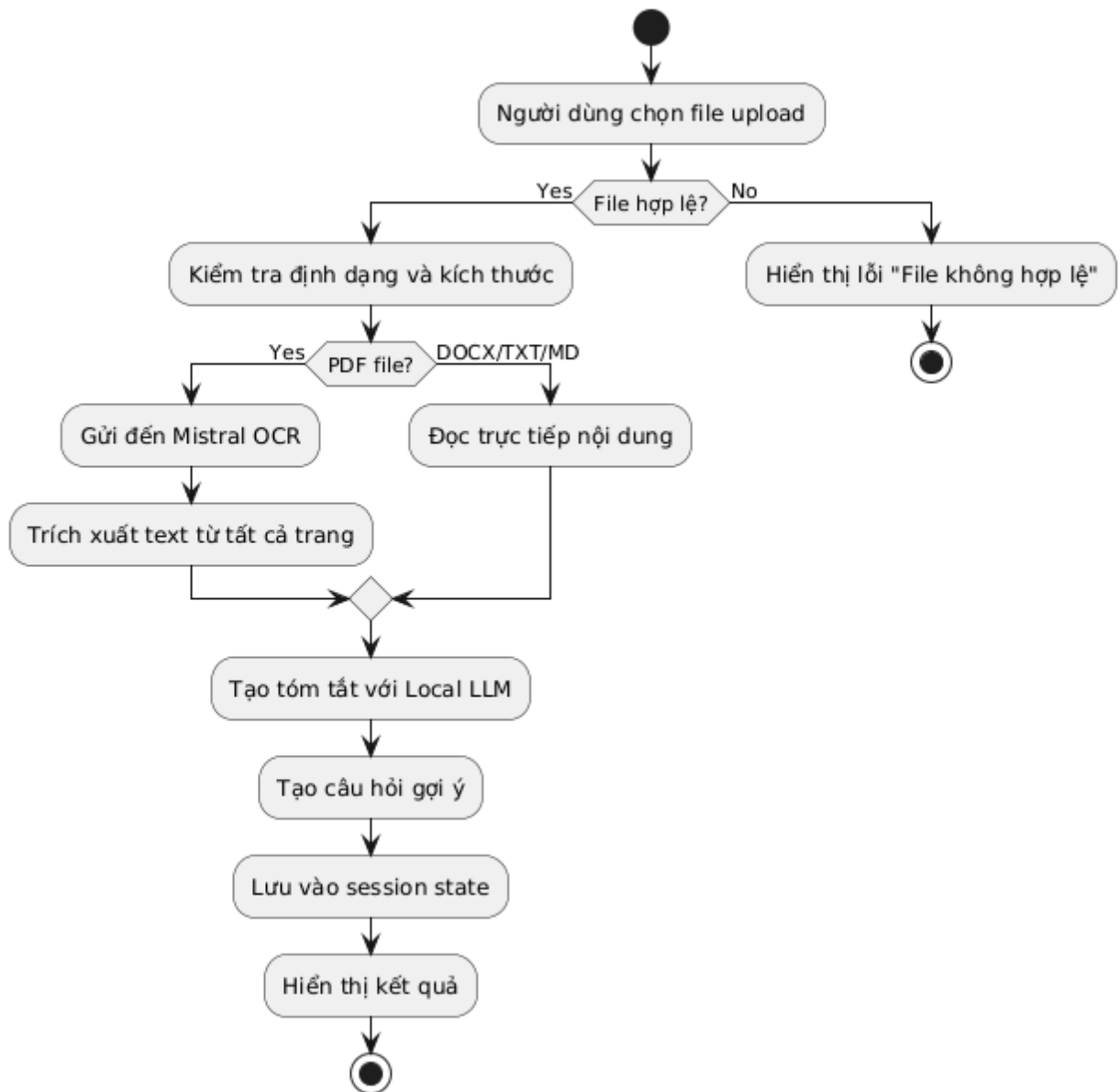
2.5 Quy trình nghiệp vụ

2.5.1 Quy trình đăng nhập



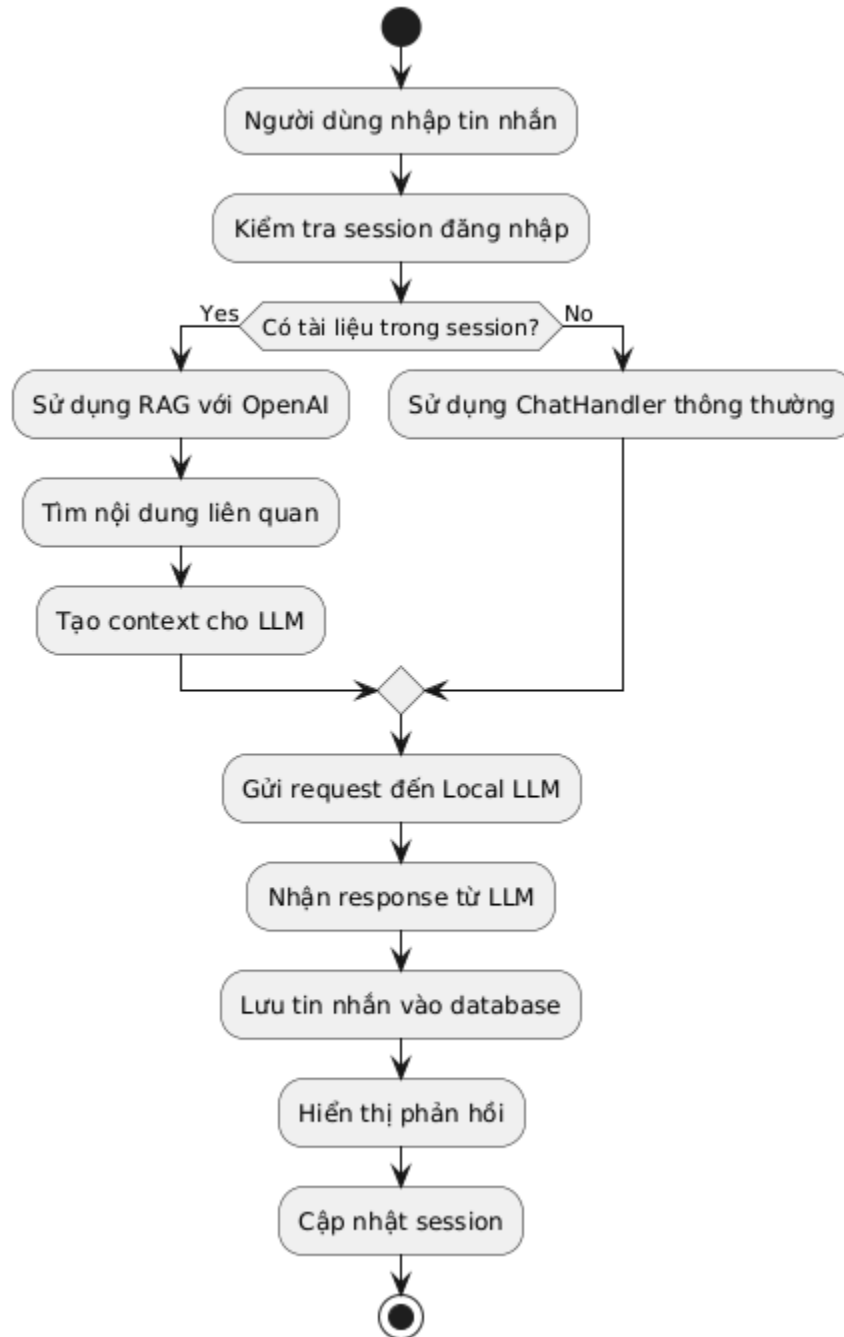
Hình 2-4: Biểu đồ hoạt động cho quy trình đăng nhập

2.5.2 Quy trình xử lý tài liệu



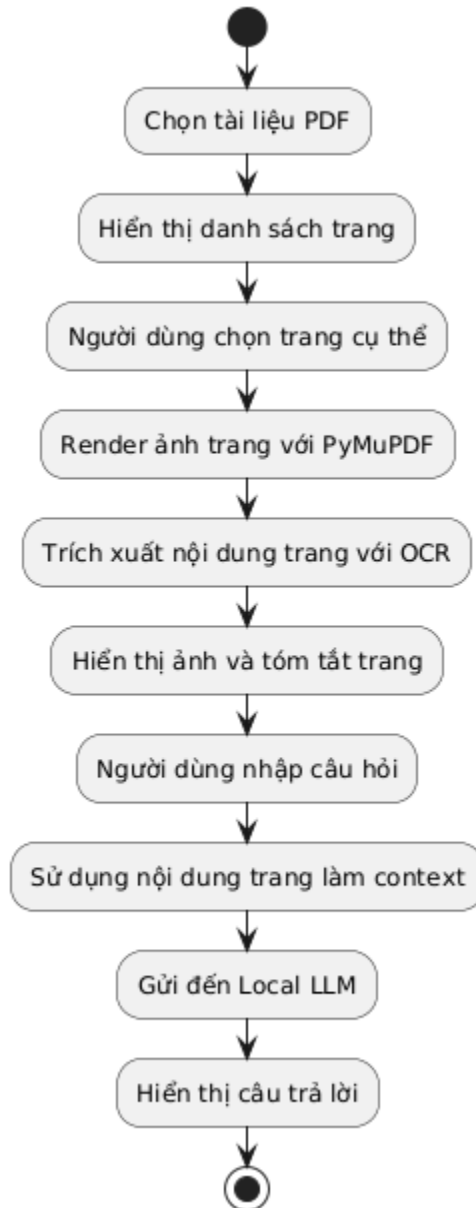
Hình 2-5: Biểu đồ hoạt động cho quy trình xử lý tài liệu

2.5.3 Quy trình chat với AI



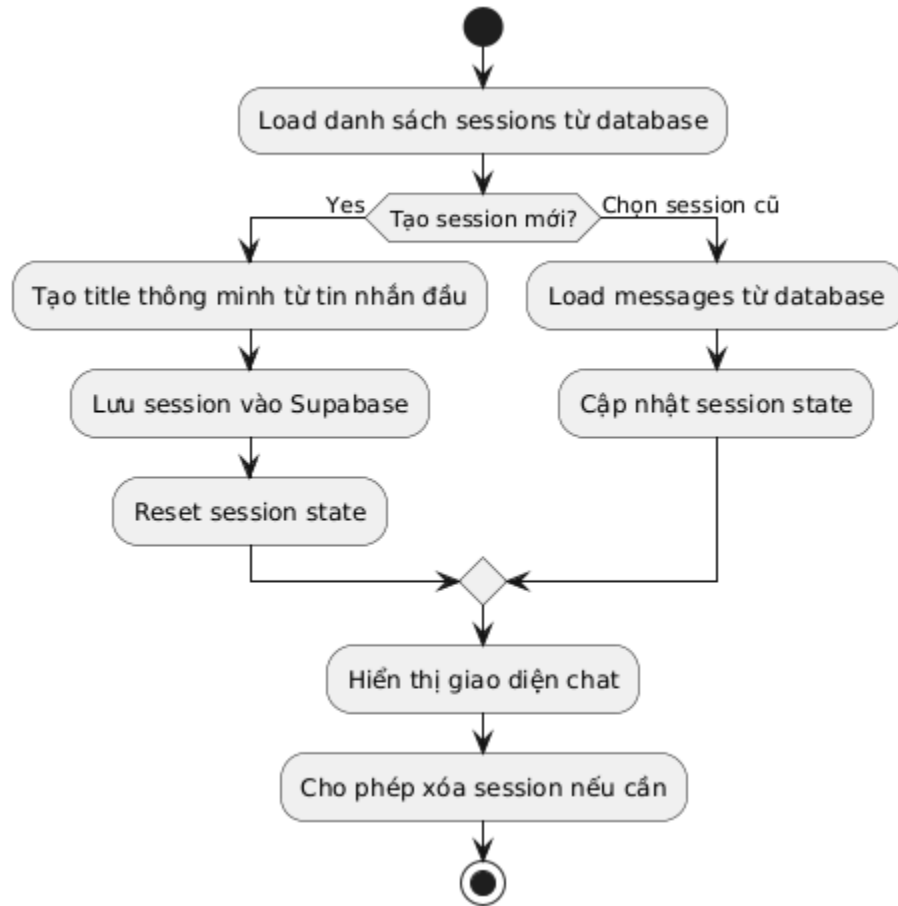
Hình 2-6: Biểu đồ hoạt động cho quy trình chat với AI

2.5.4 Quy trình chat theo trang PDF



Hình 2-7: Biểu đồ hoạt động cho quy trình chat theo trang PDF

2.5.5 Quy trình quản lý sessions



Hình 2-8: Biểu đồ hoạt động cho quy trình quản lý sessions

2.6 Đặc tả use-case

UC001: Đăng nhập

Trường	Mô tả
Mã Use-case	UC001
Tên Use-case	Đăng nhập vào hệ thống
Tác nhân	Khách
Mô tả	Xác thực người dùng để truy cập các tính năng của hệ thống
Sự kiện kích hoạt	Người dùng nhấn nút “Đăng nhập” trên giao diện
Tiền điều kiện	- Hệ thống đang hoạt động

Trường	Mô tả
	- Người dùng đã có tài khoản
Luồng sự kiện chính	<ol style="list-style-type: none">1. Hệ thống hiển thị form đăng nhập2. Người dùng nhập email và mật khẩu3. Người dùng nhấn nút “Đăng nhập”4. Hệ thống validate thông tin đầu vào5. Hệ thống gửi request xác thực đến Supabase6. Supabase xác thực thành công7. Hệ thống tạo session cho người dùng8. Chuyển hướng đến trang chính với quyền đầy đủ
Luồng sự kiện thay thế	<p>A1: Email hoặc mật khẩu trống (tại bước 4)</p> <ol style="list-style-type: none">4a. Hệ thống hiển thị thông báo lỗi “Vui lòng nhập đầy đủ thông tin”4b. Quay lại bước 2 <p>A2: Thông tin đăng nhập sai (tại bước 6)</p> <ol style="list-style-type: none">6a. Supabase trả về lỗi xác thực6b. Hệ thống hiển thị “Email hoặc mật khẩu không chính xác”6c. Quay lại bước 2 <p>A3: Lỗi kết nối database (tại bước 5)</p> <ol style="list-style-type: none">5a. Hệ thống hiển thị “Lỗi kết nối, vui lòng thử lại”5b. Quay lại bước 2
Hậu điều kiện	<ul style="list-style-type: none">- Người dùng được xác thực và có quyền truy cập đầy đủ- Session được tạo và lưu trữ- Giao diện chính được hiển thị

Bảng dữ liệu đầu vào:

Tên trường	Mô tả	Bắt buộc	Điều kiện hợp lệ	Ví dụ
Email	Địa chỉ email của người dùng	Có	Định dạng email hợp lệ	user@example.com

Tên trường	Mô tả	Bắt buộc	Điều kiện hợp lệ	Ví dụ
Password	Mật khẩu của người dùng	Có	Độ dài tối thiểu 6 ký tự	mypassword123

UC002: Upload và xử lý tài liệu

Trường	Mô tả
Mã Use-case	UC002
Tên Use-case	Upload và xử lý tài liệu
Tác nhân	Người dùng đã đăng nhập
Mô tả	Upload tài liệu và xử lý để trích xuất nội dung, tạo tóm tắt và câu hỏi gợi ý
Sự kiện kích hoạt	Người dùng chọn file trong file uploader
Tiền điều kiện	<ul style="list-style-type: none"> - Người dùng đã đăng nhập - File có định dạng hỗ trợ (PDF, DOCX, TXT, MD)
Luồng sự kiện chính	<ol style="list-style-type: none"> 1. Người dùng chọn file từ máy tính 2. Hệ thống validate file (định dạng, kích thước) 3. Hệ thống hiển thị progress “Đang xử lý tài liệu” 4. Nếu là PDF: gửi đến Mistral OCR để trích xuất text 5. Nếu là DOCX/TXT/MD: đọc trực tiếp nội dung 6. Hệ thống tạo tóm tắt với Local LLM 7. Hệ thống tạo câu hỏi gợi ý 8. Lưu tài liệu vào session state 9. Hiển thị tóm tắt và câu hỏi gợi ý 10. Thông báo “Đã xử lý xong”
Luồng sự kiện thay thế	<p>A1: File không hợp lệ (tại bước 2)</p> <ol style="list-style-type: none"> 2a. Hệ thống hiển thị “File không được hỗ trợ” 2b. Kết thúc use case <p>A2: File quá lớn (tại bước 2)</p>

Trường	Mô tả
	2a. Hệ thống hiển thị “File vượt quá 50MB” 2b. Kết thúc use case A3: Lỗi OCR (tại bước 4) 4a. Mistral OCR trả về lỗi 4b. Hệ thống hiển thị “Không thể xử lý PDF” 4c. Kết thúc use case A4: Lỗi tạo tóm tắt (tại bước 6) 6a. Local LLM không phản hồi 6b. Sử dụng tóm tắt mặc định “Không thể tóm tắt” 6c. Tiếp tục bước 7
Hậu điều kiện	- Tài liệu được lưu trong session - Nội dung được trích xuất và cache - Tóm tắt và câu hỏi được hiển thị- Sẵn sàng cho chat RAG

Bảng dữ liệu đầu vào:

Tên trường	Mô tả	Bắt buộc	Điều kiện hợp lệ	Ví dụ
File	Tài liệu cần upload	Có	PDF, DOCX, TXT, MD, ≤50MB	document.pdf

UC003: Chat với AI sử dụng RAG

Trường	Mô tả
Mã Use-case	UC003
Tên Use-case	Chat với AI sử dụng RAG
Tác nhân	Người dùng đã đăng nhập
Mô tả	Đặt câu hỏi và nhận câu trả lời từ AI dựa trên tài liệu đã upload
Sự kiện kích hoạt	Người dùng nhập tin nhắn và nhấn Enter
Tiền điều kiện	- Người dùng đã đăng nhập - Có ít nhất một tài liệu đã được xử lý

Trường	Mô tả
Luồng sự kiện chính	<ol style="list-style-type: none"> 1. Người dùng nhập câu hỏi vào chat input 2. Hệ thống lưu tin nhắn người dùng 3. Hiển thị tin nhắn người dùng ngay lập tức 4. Tìm nội dung liên quan trong tài liệu 5. Tạo context từ nội dung liên quan 6. Gửi request đến Local LLM với context 7. Nhận response từ LLM 8. Lưu response vào database 9. Hiển thị câu trả lời của AI 10. Cập nhật session title nếu cần
Luồng sự kiện thay thế	<p>A1: Không có tài liệu (tại bước 4)</p> <ol style="list-style-type: none"> 4a. Sử dụng ChatHandler thông thường 4b. Gửi trực tiếp đến LLM không có context 4c. Tiếp tục bước 7 <p>A2: Local LLM không phản hồi (tại bước 6)</p> <ol style="list-style-type: none"> 6a. Timeout sau 30 giây 6b. Hiển thị “Không thể kết nối LLM” 6c. Kết thúc use case <p>A3: Lỗi lưu database (tại bước 8)</p> <ol style="list-style-type: none"> 8a. Supabase trả về lỗi 8b. Vẫn hiển thị câu trả lời cho người dùng 8c. Log lỗi để xử lý sau
Hậu điều kiện	<ul style="list-style-type: none"> - Tin nhắn được lưu vào database - Câu trả lời được hiển thị- Lịch sử chat được cập nhật - Session vẫn duy trì

UC004: Chat theo trang PDF cụ thể

Trường	Mô tả
Mã Use-case	UC004
Tên Use-case	Chat theo trang PDF cụ thể

Trường	Mô tả
Tác nhân	Người dùng đã đăng nhập
Mô tả	Chọn trang cụ thể của PDF và chat về nội dung trang đó
Sự kiện kích hoạt	Người dùng chuyển sang tab “Chat với trang PDF”
Tiền điều kiện	<ul style="list-style-type: none"> - Người dùng đã đăng nhập - Có ít nhất một file PDF đã upload
Luồng sự kiện chính	<ol style="list-style-type: none"> 1. Hiển thị danh sách tài liệu PDF 2. Người dùng chọn tài liệu PDF 3. Hiển thị dropdown chọn trang 4. Người dùng chọn số trang 5. Render ảnh trang bằng PyMuPDF 6. Trích xuất nội dung trang bằng OCR 7. Hiển thị ảnh trang và tóm tắt nội dung 8. Người dùng nhập câu hỏi về trang 9. Sử dụng nội dung trang làm context 10. Gửi đến Local LLM 11. Hiển thị câu trả lời liên quan đến trang
Luồng sự kiện thay thế	<p>A1: Không có PDF (tại bước 1)</p> <ol style="list-style-type: none"> 1a. Hiển thị “Chưa có tài liệu PDF” 1b. Kết thúc use case <p>A2: Lỗi render trang (tại bước 5)</p> <ol style="list-style-type: none"> 5a. PyMuPDF báo lỗi 5b. Hiển thị “Không thể hiển thị trang” 5c. Vẫn cho phép chat với nội dung text <p>A3: Trang không tồn tại (tại bước 4)</p> <ol style="list-style-type: none"> 4a. Kiểm tra số trang hợp lệ 4b. Hiển thị “Trang không tồn tại” 4c. Reset về trang 1
Hậu điều kiện	<ul style="list-style-type: none"> - Ảnh trang được hiển thị

Trường	Mô tả
	<ul style="list-style-type: none"> - Nội dung trang được cache - Chat riêng biệt cho trang được tạo - Có thể chuyển trang khác

UC005: Quản lý chat sessions

Trường	Mô tả
Mã Use-case	UC005
Tên Use-case	Quản lý chat sessions
Tác nhân	Người dùng đã đăng nhập
Mô tả	Tạo, xem, tải và xóa các session chat
Sự kiện kích hoạt	Người dùng thao tác với sidebar sessions
Tiền điều kiện	<ul style="list-style-type: none"> - Người dùng đã đăng nhập - Database Supabase hoạt động
Luồng sự kiện chính	<ol style="list-style-type: none"> 1. Load danh sách sessions từ database 2. Hiển thị danh sách với title và preview 3. Người dùng chọn một session 4. Load messages từ database 5. Cập nhật giao diện chat 6. Hiển thị “Đã tải session”
Luồng sự kiện thay thế	<p>A1: Tạo session mới</p> <ol style="list-style-type: none"> 1a. Người dùng nhấn “Chat mới” 1b. Reset session state 1c. Chuẩn bị cho cuộc chat mới <p>A2: Xóa session</p> <ol style="list-style-type: none"> 2a. Người dùng nhấn nút xóa 2b. Xác nhận xóa từ database 2c. Refresh danh sách sessions 2d. Nếu đang ở session bị xóa, reset state

Trường	Mô tả
	A3: Lỗi load session 4a. Database trả về lỗi 4b. Hiện thị “Không thể tải session” 4c. Giữ nguyên session hiện tại
Hậu điều kiện	- Session được load/tạo/xóa thành công - Giao diện cập nhật phù hợp - Lịch sử được đồng bộ

UC006: Đăng ký tài khoản

Trường	Mô tả
Mã Use-case	UC006
Tên Use-case	Đăng ký tài khoản mới
Tác nhân	Khách
Mô tả	Tạo tài khoản mới để sử dụng hệ thống
Sự kiện kích hoạt	Người dùng nhấn “Đăng ký” trên form login
Tiền điều kiện	- Hệ thống hoạt động - Email chưa được đăng ký
Luồng sự kiện chính	1. Hiện thị form đăng ký 2. Người dùng nhập email, mật khẩu, xác nhận mật khẩu 3. Validate thông tin đầu vào 4. Kiểm tra email chưa tồn tại 5. Tạo tài khoản trong Supabase 6. Gửi email xác thực (nếu cần) 7. Hiện thị thông báo thành công 8. Chuyển về form đăng nhập
Luồng sự kiện thay thế	A1: Thông tin không hợp lệ 3a. Email không đúng định dạng 3b. Mật khẩu quá ngắn

Trường	Mô tả
	3c. Mật khẩu xác nhận không khớp 3d. Hiển thị lỗi tương ứng A2: Email đã tồn tại 4a. Supabase báo email đã được sử dụng 4b. Hiển thị “Email đã được đăng ký” 4c. Quay lại bước 2
Hậu điều kiện	- Tài khoản mới được tạo - Sẵn sàng để đăng nhập

Bảng dữ liệu đầu vào:

Tên trường	Mô tả	Bắt buộc	Điều kiện hợp lệ	Ví dụ
Email	Địa chỉ email để đăng ký	Có	Định dạng email, chưa tồn tại	newuser@example.com
Password	Mật khẩu cho tài khoản	Có	Tối thiểu 6 ký tự	newpassword123
Confirm Password	Xác nhận mật khẩu	Có	Phải giống với Password	newpassword123

UC007: Đăng xuất

Trường	Mô tả
Mã Use-case	UC007
Tên Use-case	Đăng xuất khỏi hệ thống
Tác nhân	Người dùng đã đăng nhập
Mô tả	Kết thúc phiên làm việc và xóa thông tin xác thực
Sự kiện kích hoạt	Người dùng nhấn nút “Đăng xuất”
Tiền điều kiện	- Người dùng đã đăng nhập
Luồng sự kiện chính	1. Người dùng nhấn nút “Đăng xuất”

Trường	Mô tả
	2. Hệ thống xóa session hiện tại 3. Reset tất cả state variables 4. Chuyển về giao diện đăng nhập 5. Hiện thị “Đã đăng xuất thành công”
Luồng sự kiện thay thế	Không có luồng thay thế
Hậu điều kiện	<ul style="list-style-type: none"> - Session bị xóa - Người dùng trở về trạng thái chưa đăng nhập - Tất cả dữ liệu nhạy cảm bị xóa khỏi bộ nhớ

3. CÁC YÊU CẦU PHI CHỨC NĂNG

3.1 Các yêu cầu về hiệu năng

Hệ thống Study Buddy phải đảm bảo các yêu cầu hiệu năng sau để mang lại trải nghiệm người dùng tốt nhất:

Thời gian phản hồi chat: Hệ thống phải trả lời tin nhắn chat thông thường trong vòng 3-5 giây, tính từ khi người dùng gửi tin nhắn đến khi hiển thị câu trả lời của AI. Đối với câu hỏi phức tạp có sử dụng RAG, thời gian phản hồi có thể lên đến 9-15 giây.

Thời gian xử lý tài liệu: Xử lý file PDF có OCR hoàn thành trong vòng 10-30 giây tùy thuộc vào số trang (1-10 trang: $\leq 10s$, 11-50 trang: $\leq 30s$). Xử lý file DOCX, TXT, MD phải hoàn thành trong vòng 2-5 giây.

Tải trang và giao diện: Thời gian tải trang chính không quá 2 giây với kết nối internet ổn định. Chuyển đổi giữa các tab và tương tác giao diện phải mượt mà, không có độ trễ đáng kể.

Khả năng xử lý đồng thời: Hệ thống hỗ trợ ít nhất 5-10 người dùng đồng thời mà không ảnh hưởng đáng kể đến hiệu năng, phù hợp với quy mô demo và thử nghiệm.

Render ảnh PDF: Chuyển đổi trang PDF thành ảnh hoàn thành trong 2-3 giây với độ phân giải 150 DPI, đảm bảo chất lượng ảnh rõ nét để đọc được text.

3.2 Yêu cầu về bảo mật

Hệ thống phải tuân thủ các nguyên tắc bảo mật cơ bản để bảo vệ thông tin người dùng:

Mã hóa truyền tải: Tất cả dữ liệu được truyền giữa client và server phải được mã hóa, bao gồm thông tin đăng nhập, nội dung chat và tài liệu upload.

Xác thực và phân quyền: Sử dụng Supabase Auth để quản lý xác thực với mã hóa mật khẩu theo chuẩn bcrypt. Người dùng chỉ có quyền truy cập dữ liệu của chính mình, không thể xem hoặc chỉnh sửa dữ liệu của người khác.

Bảo mật API: Các API keys (Mistral, Local LLM) được lưu trữ an toàn trong biến môi trường, không được mã hóa cứng trong mã nguồn.

Validation dữ liệu đầu vào: Tất cả dữ liệu từ người dùng được kiểm tra nghiêm ngặt để tránh các cuộc tấn công injection. File upload được kiểm tra định dạng và kích thước để tránh các tệp độc hại.

Session management: Session hết hạn sau 24 giờ không hoạt động. Đăng xuất sẽ xóa hoàn toàn session và không cho phép sử dụng lại.

Bảo vệ dữ liệu nhạy cảm: Nội dung tài liệu và lịch sử chat được lưu trữ an toàn trong Supabase với bảo mật cấp hàng. Không lưu nhật ký hoặc bộ nhớ đệm thông tin nhạy cảm ở phía client.

3.3 Yêu cầu về giao diện

Giao diện người dùng phải thân thiện, trực quan và dễ sử dụng:

Thiết kế responsive: Giao diện hoạt động tốt trên kích thước màn hình máy tính.

Bố cục giao diện: Giao diện gồm phần đầu trang chứa logo và thông tin người dùng, thanh bên trái cho điều hướng và quản lý sessions, khu vực nội dung chính cho giao diện chat, và phần chân trang chứa thông tin bổ sung. Sử dụng tab để phân tách “Chat trực tiếp” và “Chat theo trang PDF”.

Phản hồi trực quan: Các chỉ báo tải rõ ràng khi xử lý tài liệu hoặc chờ AI phản hồi. Thanh tiến trình cho các tác vụ dài như OCR. Thông báo nổi cho thông báo thành công/lỗi.

3.4 Ràng buộc

Hệ thống phải hoạt động trong các ràng buộc kỹ thuật và môi trường sau:

Công nghệ frontend: Sử dụng Streamlit framework cho việc xây dựng giao diện web. Không sử dụng React, Vue hoặc framework frontend khác.

Công nghệ backend: Python 3.8+ với các thư viện chính: streamlit, supabase, openai, mistralai, PyMuPDF, python-docx. Không sử dụng framework backend riêng biệt như FastAPI hay Django.

Cơ sở dữ liệu: Bắt buộc sử dụng Supabase (PostgreSQL) để lưu trữ dữ liệu người dùng, sessions và messages. Không sử dụng SQLite hoặc database local khác.

Yêu cầu hệ thống: Máy chủ cần có tối thiểu 8GB RAM để chạy Local LLM, kết nối internet ổn định để truy cập Mistral OCR API và Supabase. Hỗ trợ GPU là lợi thế nhưng không bắt buộc.

Yêu cầu client: Trình duyệt hỗ trợ JavaScript ES6+, kết nối internet tối thiểu 5 Mbps để upload tài liệu và chat mượt mà. Không yêu cầu cài đặt plugin hay extension đặc biệt.

Kích thước file: Giới hạn upload file tối đa 50MB per file, hỗ trợ tối đa 5 files cùng lúc trong một session. PDF tối đa 100 trang để đảm bảo hiệu năng OCR.

Dependency và license: Tất cả dependencies phải có license tương thích (MIT, Apache 2.0, BSD). Không sử dụng các thư viện có license GPL hoặc commercial license.

Deployment: Có thể deploy trên Streamlit Cloud, Heroku, hoặc chạy local. Không yêu cầu Docker container nhưng hỗ trợ nếu cần.

4. PHÂN CHIA CÔNG VIỆC NHÓM

4.1 Tổng quan phân chia

Dự án Study Buddy được chia thành 3 module chính tương ứng với 3 thành viên, mỗi người chịu trách nhiệm chính cho một lĩnh vực và hỗ trợ các module khác khi cần thiết.

4.2 Thành viên 1: Nguyễn Ngọc Công Anh- Frontend Development & UI/UX

Vai trò chính: Frontend Developer & UI/UX Designer

Trách nhiệm chính:

- Phát triển giao diện người dùng: Thiết kế và triển khai toàn bộ giao diện Streamlit, bao gồm bố cục, thành phần giao diện, và tùy chỉnh giao diện với CSS.

- Tối ưu trải nghiệm người dùng: Đảm bảo giao diện thân thiện với người dùng, dễ sử dụng và tuân thủ các nguyên tắc thiết kế UX/UI.
- CSS và giao diện: Phát triển và duy trì tệp `assets/styles/style.css` để tạo giao diện đẹp và nhất quán.
- Kiểm thử giao diện: Kiểm thử giao diện.

File phụ trách:

- `src/pages/home_page.py` - Giao diện trang chính
- `src/pages/login_page.py` - Giao diện đăng nhập
- `src/utils/ui_components.py` - Các thành phần UI
- `assets/styles/style.css` - Tùy chỉnh CSS
- `app.py` - Cấu trúc ứng dụng chính

Tasks cụ thể:

1. Thiết kế mô phỏng và khung dây cho toàn bộ ứng dụng.
2. Triển khai bố cục responsive cho máy tính để bàn và điện thoại.
3. Tạo các thành phần Streamlit tùy chỉnh cho giao diện chat.
4. Phát triển bằng chuyên cho câu hỏi gợi ý.
5. Tối ưu trạng thái tải và xử lý lỗi giao diện.
6. Triển khai chế độ sáng/tối (nếu có thời gian).
7. Kiểm thử tương thích trình duyệt và sửa lỗi.
8. Viết tài liệu cho các thành phần UI.

Deliverables:

- Giao diện hoàn chỉnh và responsive.
- Khung CSS tùy chỉnh.
- Thư viện thành phần có thể tái sử dụng.
- Tài liệu UI/UX.
- Báo cáo tương thích trình duyệt.

4.3 Thành viên 2: Trần Long Khánh - Backend Logic & AI Integration

Vai trò chính: Backend Developer & AI Engineer

Trách nhiệm chính:

- Tích hợp AI: Tích hợp và tối ưu các dịch vụ AI (Local LLM, Mistral OCR, OpenAI-compatible APIs).
- Xử lý tài liệu: Phát triển module xử lý tài liệu với OCR, trích xuất văn bản và phân tích nội dung.
- Triển khai RAG: Xây dựng hệ thống Retrieval-Augmented Generation cho câu hỏi và trả lời thông minh.
- Logic chat: Phát triển bộ xử lý chat và quản lý hội thoại.
- Xử lý lỗi: Triển khai hệ thống xử lý lỗi toàn diện và hệ thống ghi nhật ký.

File phụ trách:

- src/utils/document_processor.py - Xử lý tài liệu và OCR
- src/utils/chat_handler.py - Logic chat với AI
- src/utils/error_handler.py - Hệ thống xử lý lỗi
- src/utils/validators.py - Kiểm tra dữ liệu
- src/config/constants.py - Tham số cấu hình

Tasks cụ thể:

- Thiết lập và cấu hình tích hợp Local LLM với LM Studio.
- Triển khai Mistral OCR cho xử lý PDF.
- Phát triển pipeline RAG với vector embeddings.
- Tối ưu thời gian phản hồi cho các truy vấn AI.
- Triển khai phân đoạn tài liệu và tìm kiếm tương tự.
- Xây dựng khung xử lý lỗi.
- Tối ưu hiệu năng cho các hoạt động AI.
- Kiểm thử đơn vị cho các hàm AI.

Deliverables:

- Module tích hợp AI hoàn chỉnh.
- Pipeline xử lý tài liệu.
- Hệ thống RAG với độ chính xác cao.
- Khung xử lý lỗi.

- Báo cáo tối ưu hiệu năng.
- Bộ kiểm thử AI

4.4 Thành viên 3: Trần Đức Việt - Database & Authentication

Vai trò chính: Database Administrator & Security Engineer

Trách nhiệm chính:

- Thiết kế cơ sở dữ liệu: Thiết kế và triển khai lược đồ cơ sở dữ liệu trên Supabase.
- Hệ thống xác thực: Phát triển hệ thống đăng nhập/đăng ký an toàn.
- Quản lý session: Quản lý phiên người dùng và lịch sử chat.
- Lưu trữ dữ liệu: Triển khai logic lưu trữ và truy xuất dữ liệu.
- Bảo mật: Đảm bảo bảo mật dữ liệu và quyền riêng tư của người dùng.

File phụ trách:

- src/utils/chat_persistence.py - Thao tác cơ sở dữ liệu
- database_schema.sql - Thiết kế lược đồ cơ sở dữ liệu

Tasks cụ thể:

- Thiết kế lược đồ cơ sở dữ liệu cho người dùng, phiên, tin nhắn, tài liệu.
- Thiết lập dự án Supabase với xác thực.
- Triển khai chính sách bảo mật cấp hàng.
- Phát triển thao tác CRUD cho tất cả thực thể.
- Tối ưu truy vấn cơ sở dữ liệu và lập chỉ mục.
- Triển khai quy trình sao lưu và khôi phục.
- Quản lý phiên và xử lý hết hạn.
- Kịch bản di chuyển và khởi tạo dữ liệu.
- Kiểm tra bảo mật và kiểm thử xâm nhập.
- Giám sát và tối ưu hiệu năng.

Deliverables:

- Lược đồ cơ sở dữ liệu hoàn chỉnh với tài liệu.
- Hệ thống xác thực an toàn.
- Khung quản lý phiên.

- Báo cáo kiểm tra bảo mật.
- Báo cáo tối ưu hiệu năng.
- Quy trình sao lưu/khôi phục cơ sở dữ liệu

4.5 Timeline và Milestone

Giai đoạn 1 (Tuần 1-2): Thiết lập và Nền tảng

- Thành viên 1: Thiết kế mockup và bố cục cơ bản
- Thành viên 2: Thiết lập AI integrations và xử lý tài liệu cơ bản
- Thành viên 3: Thiết lập lược đồ cơ sở dữ liệu và xác thực

Giai đoạn 2 (Tuần 3-4): Phát triển cốt lõi

- Thành viên 1: Phát triển các thành phần giao diện chính và thiết kế responsive
- Thành viên 2: Triển khai RAG system và logic chat
- Thành viên 3: Hoàn thiện CRUD operations và quản lý phiên

Giai đoạn 3 (Tuần 5-6): Tích hợp và Kiểm thử

- Tất cả thành viên: Kiểm thử tích hợp và sửa lỗi
- Thành viên 1: Tinh chỉnh UI/UX và kiểm thử cross-browser
- Thành viên 2: Tối ưu hiệu năng và xử lý lỗi
- Thành viên 3: Kiểm thử bảo mật và kiểm tra dữ liệu

Giai đoạn 4 (Tuần 7-8): Hoàn thiện cuối cùng và Triển khai

- Kiểm tra mã nguồn và tái cấu trúc
- Hoàn thiện tài liệu
- Chuẩn bị triển khai
- Chuẩn bị trình diễn

4.6 Communication và Collaboration

Daily Standups: Họp online 15 phút mỗi ngày để đồng bộ tiến độ và vấn đề.

Code Review: Mỗi thành viên review code của 2 thành viên còn lại trước khi merge

Tài nguyên chung:

- Kho GitHub cho quản lý phiên bản.
- Discord/Slack cho giao tiếp.
- Notion/Trello cho quản lý nhiệm vụ.
- Google Drive cho tài liệu chia sẻ.

Điểm tích hợp:

- Thành viên 1 & 2: Các thành phần giao diện gọi hàm AI.
- Thành viên 1 & 3: Luồng xác thực giao diện.
- Thành viên 2 & 3: Phản hồi AI lưu vào cơ sở dữ liệu.

4.7 Quality Assurance

Trách nhiệm kiểm thử:

- Thành viên 1: Kiểm thử giao diện và thiết kế responsive.
- Thành viên 2: Kiểm thử đơn vị cho hàm AI, kiểm thử tích hợp.
- Thành viên 3: Kiểm thử cơ sở dữ liệu, kiểm thử bảo mật.

Tiêu chuẩn mã nguồn:

- Tuân thủ Python PEP 8.
- Gợi ý kiểu dữ liệu cho tất cả hàm.
- Chuỗi tài liệu toàn diện.
- Xử lý lỗi ở mọi cấp độ.

Yêu cầu tài liệu:

- README với hướng dẫn thiết lập.
- Tài liệu API cho các hàm nội bộ.
- Hướng dẫn sử dụng cho người dùng cuối.
- Hướng dẫn triển khai.

5. KẾT LUẬN

5.1 Tổng kết dự án

Hệ thống Study Buddy là một ứng dụng chatbot AI hỗ trợ học tập tiên tiến, tích hợp nhiều công nghệ hiện đại như Local LLM, OCR, và RAG để mang lại trải nghiệm học tập thông minh và hiệu quả. Dự án được thiết kế với kiến trúc modular, dễ bảo trì và mở rộng, phù hợp với quy mô nhóm 3 người và thời gian phát triển 7 tuần.

5.2 Điểm mạnh của hệ thống

Tính độc đáo: Chat theo trang PDF cụ thể là tính năng độc đáo, chưa có ở nhiều sản phẩm tương tự.

Tập trung vào quyền riêng tư: Sử dụng Local LLM đảm bảo dữ liệu không bị rò rỉ ra bên ngoài.

Toàn diện: Hỗ trợ đầy đủ pipeline từ upload tài liệu đến chat thông minh.

Khả năng mở rộng: Kiến trúc cho phép mở rộng thêm nhiều tính năng trong tương lai.

5.3 Rủi ro và giải pháp

Rủi ro kỹ thuật:

- Local LLM có thể chậm: Chuyển sang API đám mây nếu cần.
- Độ chính xác OCR không cao: Kết hợp nhiều công cụ OCR.
- Hiệu năng cơ sở dữ liệu: Triển khai bộ nhớ đệm và tối ưu hóa.

Rủi ro tiến độ:

- Phạm vi mở rộng: Quản lý phạm vi nghiêm ngặt và ưu tiên.
- Khó khăn kỹ thuật: Thời gian đệm và giải pháp dự phòng.
- Phối hợp nhóm: Họp đồng bộ thường xuyên và giao tiếp rõ ràng.