

ĐỀ XUẤT DỰ ÁN CÔNG NGHỆ THÔNG TIN NHÓM 6

Thành viên:

Nguyễn Ngọc Công Anh -22010201

Trần Long Khánh - 22010449

Trần Đức Việt - 22010032

1. Tên dự án

Ứng dụng hỗ trợ học tập thông minh: Đọc và hiểu nội dung tài liệu PDF bằng mô hình ngôn ngữ lớn (LLM)

2. Đối tượng sử dụng

- Sinh viên, giảng viên và những người cần học hoặc ôn lại kiến thức nhanh chóng.
- Phù hợp với người dùng muốn học tập tập trung và hiệu quả hơn.

3. Vấn đề giải quyết

- Giảm thời gian đọc toàn bộ tài liệu dài và phức tạp.
- Cô đọng kiến thức, giúp người dùng nắm bắt nội dung cốt lõi.
- Hỗ trợ sinh ra các câu hỏi mới từ tài liệu sẵn có để củng cố kiến thức.

4. Lý do sản phẩm hấp dẫn và đáng xây dựng

- **Nhu cầu thực tế:**
 - Nhiều tài liệu kỹ thuật, học thuật hoặc văn bản pháp lý có nội dung dài, khó hiểu.
 - Người dùng thường mất nhiều thời gian để đọc và tìm kiếm thông tin cụ thể.
- **Sức mạnh công nghệ:**
 - Các mô hình ngôn ngữ lớn (LLM) như ChatGPT, Grok, Claude... hỗ trợ hiệu quả trong việc trích xuất, tóm tắt và giải thích nội dung tài liệu.
- **Tiềm năng thị trường:**

- Thị trường ứng dụng AI hỗ trợ học tập và công việc văn phòng đang phát triển mạnh mẽ.

5. Sản phẩm khả dụng tối thiểu (MVP)

- Chức năng cốt lõi: Đọc và tóm tắt nội dung tài liệu PDF theo yêu cầu của người dùng.
- Giao diện đơn giản, dễ sử dụng, tập trung vào trải nghiệm người dùng.

6. Công nghệ và tài nguyên sử dụng

- **Backend:**
 - Python, FastAPI và Streamlit để xử lý file PDF và tích hợp mô hình LLM.
- **Công nghệ AI:**
 - OCR để trích xuất văn bản từ PDF, kể cả tài liệu scan.
 - API LLM (OpenAI hoặc Grok) để tóm tắt và trả lời câu hỏi.
 - RAG (Retrieval-Augmented Generation) pipeline để đảm bảo đầu ra chính xác, dựa trên nội dung tài liệu.
- **Tài nguyên:**
 - Hosting giá rẻ (Heroku) cho phiên bản demo.