BÁO CÁO TIẾN ĐỘ DỰ ÁN

**Xây dựng hệ thống Hỏi–Đáp đa phương thức dựa trên RAG trong lĩnh vực Giao thông - Văn hoá và Du lịch**

**Thành viên thực hiện:** Ngô Anh Tuấn, Trần Ngọc Phát

**Giảng viên hướng dẫn:** Trần Thanh Phước

**Tuần 1 – Giai đoạn 1 – Ngày 27/09/2025**

**Tóm tắt nội dung**

Hoàn thành xác định yêu cầu sơ bộ, nghiên cứu kỹ thuật cốt lõi của RAG đa phương thức và dựng kiến trúc sơ bộ. Chọn giải pháp mô hình và cơ sở hạ tầng phù hợp với giới hạn tài nguyên.

# Tổng quan

- Xác định phạm vi và yêu cầu chức năng: input (text, ảnh), output (câu trả lời văn bản có tham chiếu)

+ Target: TP.HCM

- Nghiên cứu về RAG, VLM, retrieval vs generation, và tích hợp KG.

- Chuẩn bị danh sách nguồn dữ liệu sơ bộ cho TP.HCM: cổng dữ liệu mở chính phủ, trang du lịch, báo chí, OpenStreetMap, ảnh địa danh.

- Sửa, chuẩn hóa báo cáo nội bộ: thêm mô tả pipeline, caption ảnh, tham chiếu nguồn.

# Tài liệu yêu cầu

- Mục tiêu: phạm vi TP. Hồ Chí Minh

- Stakeholder: người dùng chính, giảng viên và nhóm phát triển

- User stories / Use cases:

+ input: text only, image-only và text + image

+ output: danh sách POI/docs + câu trả lời tổng hợp

- Yêu cầu chức năng (F-req): truy vấn QA, trả về source citations, upload ảnh, filter theo chủ đề, API endpoints cơ bản.

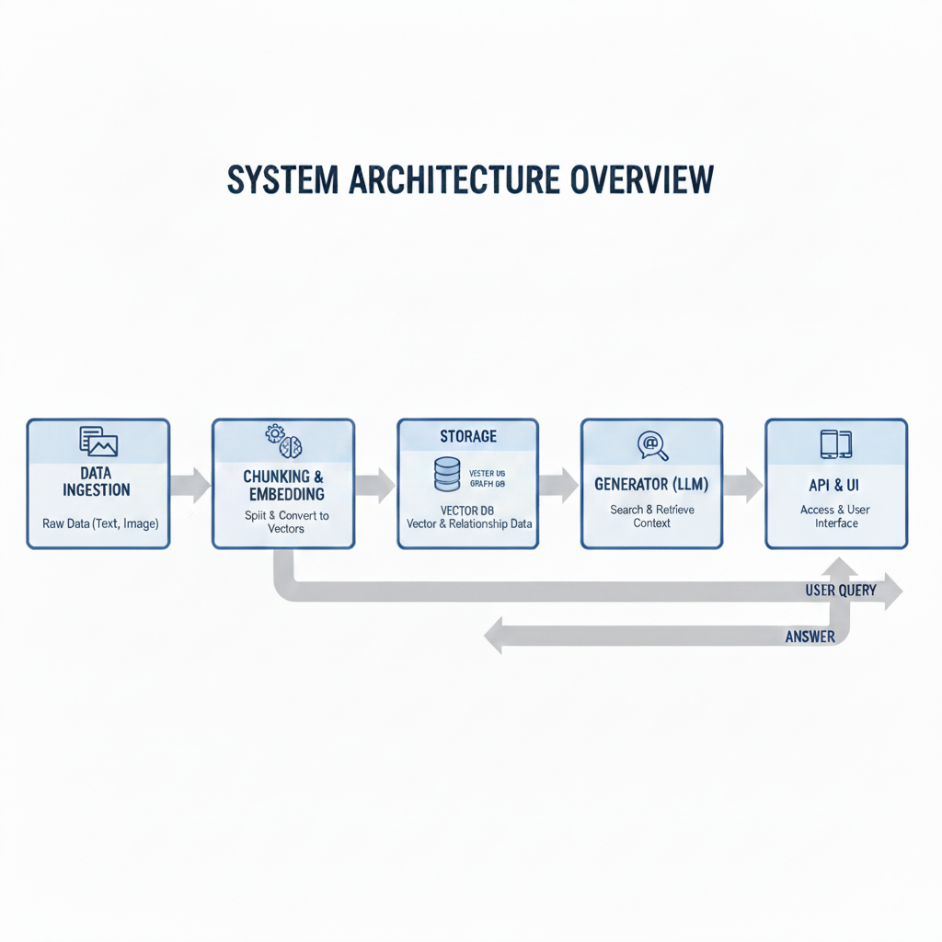
- Yêu cầu phi chức năng (NF-req): latency p95 < X s, throughput, availability, lưu trữ, bảo mật, giới hạn kích thước file.

- Dữ liệu & nguồn (data spec): định dạng, kích thước ước tính, nguồn (cổng dữ liệu mở, báo, OSM, ảnh), điều kiện license.

- Định nghĩa dữ liệu mẫu: schema document, metadata fields (title, date, geo, tags, source, lang).

- Chỉ số đánh giá (metrics): retrieval@k, EM/F1 cho QA, human rating, latency, precision of citations.

# Kiến trúc sơ bộ



# Kế hoạch tuần 2 (27/09 - 05/10)

- Hoàn thiện tài liệu yêu cầu và xác nhận với giảng viên.

- Bắt đầu thu thập dataset mẫu (text + ảnh) cho 3 chủ đề nhỏ: giao thông, ẩm thực/điểm du lịch, văn hóa.

- Thiết lập môi trường dev: container hóa (Docker) + repo mẫu.

- Kiểm thử embedding pipeline; sửa lỗi mã hóa chunk.

- Đánh giá nhanh 2 option retriever (FAISS vs Chroma) trên dataset mẫu.

- Ghi nhật ký test và cập nhật báo cáo tiến độ.