

Chatbot:

Instructor: PhD. Nguyen Khac Tiep

Presented by:

Tran Dinh Khanh Dang

Le Minh Nhut

Thai Ngoc Quan

Department of Computer Science
University of Information Technology

Ngày 15 tháng 4 năm 2025

1. Introduction

Chatbot tra cứu thông tin giao thông, biển báo, luật đường bộ

- Xây dựng một hệ thống chatbot sử dụng trí tuệ nhân tạo.
- Cho phép người dùng hỏi về các thông tin liên quan đến giao thông đường bộ tại Việt Nam.
- Hỗ trợ tra cứu biển báo, quy định luật giao thông, mức phạt vi phạm, v.v.

Định nghĩa bài toán

Mục tiêu: Xây dựng một chatbot có khả năng trả lời tự động các câu hỏi liên quan đến giao thông đường bộ tại Việt Nam.

Các thành phần chính:

- **Xử lý ngôn ngữ tự nhiên (NLP):** Phân tích ý nghĩa câu hỏi người dùng.
- **Tìm kiếm thông tin (IR):** Truy xuất các đoạn văn bản phù hợp từ cơ sở dữ liệu luật và biển báo.
- **Sinh câu trả lời (Answer Generation):** Sinh câu trả lời rõ ràng, trích dẫn đúng luật.

- **Ngôn ngữ lập trình:** Python.
- **Framework NLP:**
 - **Transformers (HuggingFace):** Sử dụng các mô hình pre-trained như PhoBERT, BERT-Vi để hiểu tiếng Việt.
 - **RAG (Retrieval-Augmented Generation):** Truy xuất nội dung và sinh câu trả lời.
- **Cơ sở dữ liệu:** Vector DB (FAISS hoặc Chroma) lưu trữ và truy xuất đoạn luật/biến báo liên quan.
- **Giao diện người dùng:** Ứng dụng web đơn giản (Streamlit hoặc Gradio).

Input:

- Câu hỏi dạng văn bản tiếng Việt.
- Ví dụ: “Mức phạt khi không đội mũ bảo hiểm là bao nhiêu?”

Output:

- Câu trả lời văn bản phù hợp, đúng luật.
- Có thể bao gồm trích dẫn điều luật hoặc mô tả ngắn gọn.
- Ví dụ: “Phạt tiền từ 400.000 đến 600.000 đồng theo Điều 7, Khoản 2 Nghị định 168/2024/NĐ-CP”

- Nhu cầu tìm kiếm thông tin giao thông ngày càng phổ biến, đặc biệt với người mới học lái xe hoặc người tham gia giao thông hàng ngày.
- Các văn bản luật và tài liệu hiện hành thường khó tiếp cận, gây khó khăn cho người dân trong việc tra cứu.
- Chatbot giúp đơn giản hóa quá trình tìm kiếm, hỗ trợ người dùng nhanh chóng, chính xác và tiện lợi.

Dữ liệu thực nghiệm dự kiến

- Tập dữ liệu các câu hỏi - câu trả lời từ đề thi lý thuyết GPLX.
- Văn bản Luật Giao thông đường bộ Việt Nam.
- Dữ liệu biển báo giao thông được phân loại theo nhóm.
- Câu hỏi thu thập từ người dùng thực tế trên các diễn đàn, mạng xã hội.