

ĐỀ CƯƠNG CHI TIẾT

TÊN ĐỀ TÀI TIẾNG VIỆT: Xây dựng hệ thống luật sư ảo hỗ trợ tra cứu pháp luật
TÊN ĐỀ TÀI TIẾNG ANH: Building a Virtual Legal Assistant for Law Query
Cán bộ hướng dẫn: ThS. Đặng Việt Dũng
Thời gian thực hiện: Từ ngày 01/09/2025 đến ngày 21/12/2025
Sinh viên thực hiện: Phạm Công Minh - 22520882
Nội dung đề tài: 1. Lý do chọn đề tài <ul style="list-style-type: none">Trong bối cảnh xã hội hiện nay, khói lượng văn bản pháp luật tại Việt Nam ngày càng nhiều và phức tạp. Các bộ luật, nghị định, thông tư liên tục được ban hành, sửa đổi hoặc bổ sung. Người dân, sinh viên luật cũng như doanh nghiệp gặp nhiều khó khăn trong việc tra cứu nhanh chóng và chính xác các quy định cần thiết. Thực tế cho thấy, đa số người dân khi có thắc mắc pháp lý thường phải tìm đến luật sư, hoặc tìm kiếm trên mạng nhưng thông tin lại phân tán, khó kiểm chứng độ chính xác.Các hệ thống tra cứu pháp luật hiện tại thường hoạt động ở mức từ khóa, chưa đủ khả năng hiểu ngữ nghĩa câu hỏi. Người dùng phải nhập đúng từ khóa hoặc điều khoản thì mới tìm ra kết quả, dẫn đến trải nghiệm không thân thiện và không phù hợp với số đông người dùng phổ thông.Sự phát triển của trí tuệ nhân tạo, đặc biệt là xử lý ngôn ngữ tự nhiên (NLP), nhận diện giọng nói (STT), tổng hợp giọng nói (TTS), và các mô hình ngôn ngữ lớn (LLM) đã mở ra cơ hội để xây dựng một hệ thống ‘luật sư ảo’. Hệ thống này có thể hiểu câu hỏi tự nhiên bằng tiếng Việt, tìm căn cứ pháp luật trong cơ sở dữ liệu, và trả lời bằng văn bản hoặc giọng nói như một trợ lý ảo chuyên biệt.

- Đề tài này vì vậy mang ý nghĩa thực tiễn cao: giúp người dân tiếp cận pháp luật dễ dàng, hỗ trợ luật sư và sinh viên luật tra cứu nhanh chóng, đồng thời là bước đi quan trọng trong việc áp dụng AI vào lĩnh vực pháp lý tại Việt Nam.

2. Mục tiêu đề tài

- Mục tiêu tổng quát: Xây dựng một hệ thống hỗ trợ tra cứu pháp luật tự động, có khả năng hiểu và trả lời câu hỏi tự nhiên của người dùng, kèm theo căn cứ pháp lý rõ ràng.
- Mục tiêu cụ thể:
 - Xây dựng Ontology & Knowledge Graph tối ưu cho lĩnh vực pháp luật.
 - Thiết kế query engine dựa trên subgraph matching (star decomposition) kết hợp LLM để hiểu rõ ý chính câu hỏi cũng như nâng cao khả năng trả lời.
 - Tích hợp chức năng nhập câu hỏi bằng giọng nói (STT) và trả lời bằng giọng nói (TTS).
 - Kết hợp LLM và RAG để sinh câu trả lời tự nhiên nhưng đảm bảo luôn trích dẫn căn cứ pháp luật.
 - Phát triển ứng dụng demo chạy trên web và mobile.

3. Khảo sát

- Các nghiên cứu trong và ngoài nước cho thấy việc áp dụng Ontology và Knowledge Graph vào lĩnh vực pháp luật mang lại hiệu quả rõ rệt. Ví dụ, nghiên cứu “A Practical Approach to Leverage Knowledge Graphs for Legal Query” đề xuất chuẩn hóa triple, loại bỏ triple vô nghĩa, gộp triple đồng nghĩa, sau đó truy vấn bằng star decomposition. Trong khi đó, nghiên cứu “Information Retrieval from Legal Documents with Ontology and Graph Embeddings Approach” sử dụng Legal-Onto ontology kết hợp embedding (PhoBERT) để tìm kiếm ngữ nghĩa.
- Hiện tại, một số chatbot pháp luật ở Việt Nam chủ yếu dừng ở mức tìm kiếm từ khóa, chưa có khả năng phân tích ngữ nghĩa và trả lời tự nhiên. Trên thế giới, các sản phẩm như ROSS Intelligence hay DoNotPay đã thử nghiệm AI trong lĩnh vực luật nhưng chủ yếu phục vụ hệ thống common law, chưa có giải pháp cho luật Việt Nam.
- Khoảng trống nghiên cứu và ứng dụng chính là cơ hội để đề tài này phát triển một hệ thống luật sư ảo phù hợp với văn bản pháp luật Việt Nam.

4. Phạm vi đề tài

- Phạm vi dữ liệu: Tập trung trước vào lĩnh vực giao thông đường bộ (Luật GT, Nghị định xử phạt). Sau khi thành công có thể mở rộng sang đất đai, lao động.
- Phạm vi kỹ thuật: Backend bằng Go + PostgreSQL, frontend bằng ReactJS + React Native, tích hợp STT/TTS và LLM.
- Phạm vi triển khai: xây dựng ứng dụng demo, phục vụ mục đích nghiên cứu và minh chứng.

5. Đối tượng sử dụng

- Người dân: tra cứu nhanh mức phạt, định nghĩa khái niệm.
- Sinh viên luật: tham khảo căn cứ pháp lý cho bài tập và nghiên cứu.
- Luật sư: công cụ hỗ trợ tra cứu nhanh khi hành nghề.
- Doanh nghiệp/tổ chức: kiểm tra tính hợp pháp của hoạt động, văn bản.

6. Phương pháp thực hiện

- Quy trình thực hiện đề tài gồm các bước:
 - Chuẩn hóa dữ liệu pháp luật: trích xuất văn bản thành Điều/Khoản/Điểm.
 - Xây dựng Ontology và Knowledge Graph từ dữ liệu chuẩn hóa.
 - Phát triển query engine: phân loại câu hỏi, tạo query graph, star decomposition và subgraph matching.
 - Tích hợp chức năng giọng nói: STT cho nhập liệu, TTS cho trả lời.
 - Kết hợp LLM RAG để tạo câu trả lời tự nhiên.
 - Triển khai demo ứng dụng web và mobile.

7. Kết quả mong đợi

- Hệ thống luật sư ảo có thể tiếp nhận câu hỏi bằng văn bản hoặc giọng nói.
- Trả lời câu hỏi bằng văn bản và giọng nói, kèm theo căn cứ pháp lý rõ ràng.
- Độ chính xác tối thiểu 75% với câu hỏi khái niệm và vi phạm trong luật giao thông.
- Trải nghiệm người dùng thân thiện, dễ sử dụng cho người dân phổ thông.

Kế hoạch thực hiện:

Giai Đoạn	Thời Gian	Nội Dung
Phân tích yêu cầu	01/09 – 07/09	Xác định phạm vi, kiến trúc, công nghệ, thiết kế BRD
Chuẩn hóa dữ liệu	08/09 – 22/09	Xây dựng schema & nhập dữ liệu mẫu
Xây dựng backend	15/09 - 20/09	Thiết kế giao diện người dùng trên figma
Query engine	22/09 - 24/09	Thiết kế cơ sở dữ liệu chi tiết dựa trên những gì đã thống nhất.
Frontend	01/11 – 15/11	Web + Mobile giao diện query
Voice layer	16/11 – 25/11	Tích hợp STT/TTS
LLM layer	26/11 – 05/12	Tích hợp RAG sinh câu trả lời tự nhiên
Kiểm thử & triển khai	06/12 – 15/12	Test end-to-end, sửa lỗi, deploy
Hoàn thiện tài liệu	16/12 – 31/12	Viết báo cáo, chuẩn bị bảo vệ

TP. HCM, ngày ... tháng ... năm 2025

Xác nhận của CBHD (Ký tên và ghi rõ họ tên)	Sinh viên 1 (Ký tên và ghi rõ họ tên)	Sinh viên 2 (Ký tên và ghi rõ họ tên)
Đặng Việt Dũng	Phạm Công Minh	