

# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://youtu.be/i5T52tJz3X0>
- Link slides (dạng .pdf đặt trên Github của nhóm):  
[PPNCKH-CS2205.CH183/Anh Nguyễn Trần Ngọc - CS2205.NOV2024.DeCuong.FinalReport.Template.Slide.pdf at main · nana-sociu/PPNCKH-CS2205.CH183](https://github.com/nana-sociu/PPNCKH-CS2205.CH183/blob/main/CS2205.NOV2024.DeCuong.FinalReport.Template.Slide.pdf)
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*
- *Lớp Cao học, mỗi nhóm một thành viên*

- Họ và Tên: Nguyễn Trần Ngọc Anh
- MSSV: 240104002



- Lớp: CS2205.NOV2024
- Tự đánh giá (điểm tổng kết môn): 9/10
- Số buổi vắng: 1
- Số câu hỏi QT cá nhân:
- Số câu hỏi QT của cả nhóm:
- Link Github:  
<https://github.com/nana-sociu/PPNCKH-CS2205.CH183>

# **ĐỀ CƯƠNG NGHIÊN CỨU**

## **TÊN ĐỀ TÀI (IN HOA)**

**ỨNG DỤNG MÔ HÌNH NGÔN NGỮ LỚN CHO VIỆC HỎI VÀ TRẢ LỜI TRONG  
LĨNH VỰC PHÁP LÝ VIỆT NAM**

## **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**

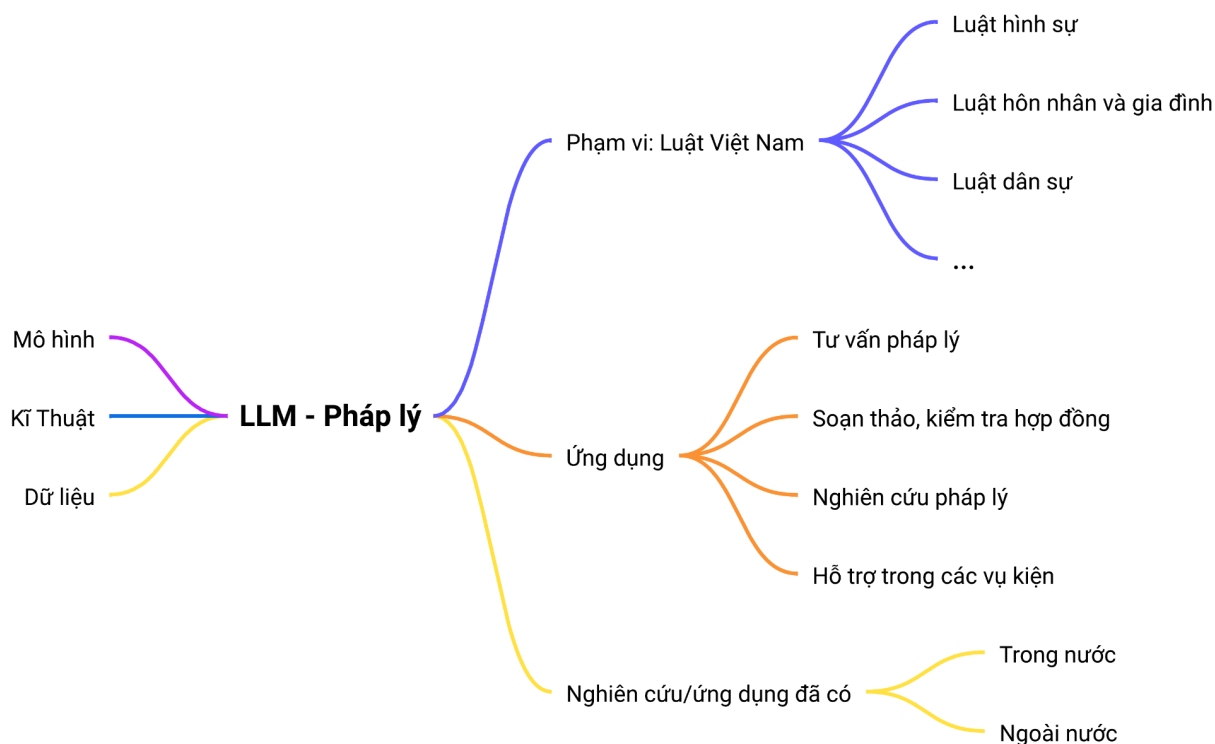
**APPLYING LARGE LANGUAGE MODEL FOR QUESTION ANSWERING IN  
THE VIETNAM LEGAL DOMAIN**

## **TÓM TẮT**

Ngày nay với sự phát triển mạnh mẽ của các mô hình ngôn ngữ lớn, con người như có thêm sức mạnh để thực hiện hóa mọi điều không thể thành có thể. Với nhiều ứng dụng đời sống, pháp luật là một khía cạnh cần được khai thác nhiều hơn nữa, giúp con người hiểu biết, nhận thức để bảo vệ bản thân khỏi thế giới rộng lớn này. Tôi đi đến với một ý tưởng ứng dụng hỏi đáp về lĩnh vực luật pháp. Tuy không mới mẻ ở nước ngoài nhưng lại mới mẻ với Việt Nam và luật Việt Nam. Nghiên cứu này mở ra nhằm giúp mọi người gần gũi với luật pháp hơn và tiếp cận pháp luật một cách nhanh chóng.

## **GIỚI THIỆU**

Xã hội ngày càng phát triển, hiện đại, con người cũng dễ bị đánh lừa hơn bởi sự tinh vi. Các vụ lừa đảo diễn ra thường xuyên vào ngày nay, lừa tiền qua nhiều hình thức ngân hàng, bất động sản, tình cảm, mạng xã hội,... Với mong muốn mọi người có hiểu biết hơn về luật pháp, giúp giảm thiểu những mất mát không đáng có, nghiên cứu về mô hình ngôn ngữ lớn chuyên tiếng Việt, giúp hỏi đáp các vấn đề pháp lý Việt Nam như luật hình sự, luật hôn nhân và gia đình, luật dân sự,... Ứng dụng hỏi đáp về lĩnh vực pháp lý Việt Nam giúp người dùng tư vấn pháp lý miễn phí, soạn thảo, kiểm tra hợp đồng, nghiên cứu pháp lý, và hỗ trợ làm hồ sơ trong các vụ kiện. Áp dụng kỹ thuật thể hệ tăng cường truy xuất (RAG) giúp mô hình truy cập thông tin cập nhật và chính xác nhất.



## MỤC TIÊU

Nghiên cứu ứng dụng hỏi đáp về lĩnh vực pháp lý Việt Nam giúp:

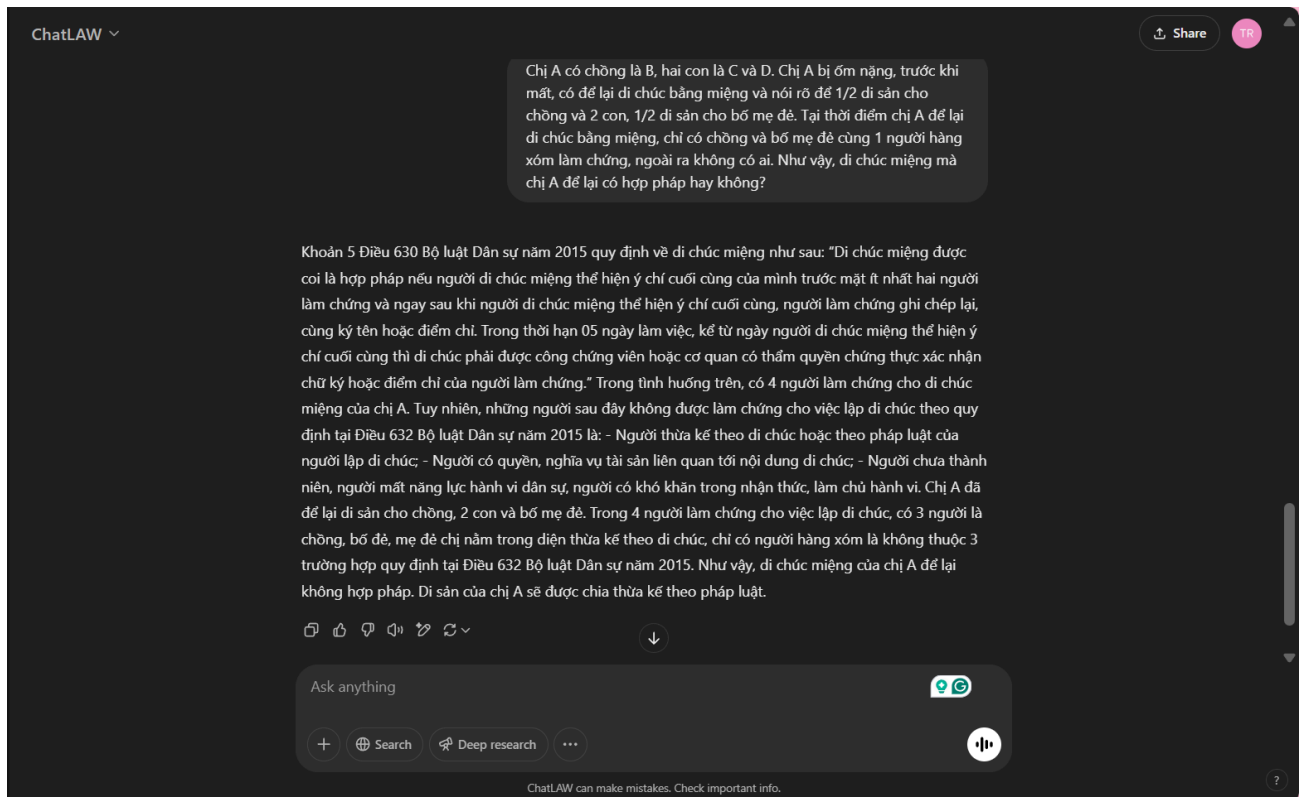
- Người dùng (luật sư, tư vấn viên,...) nâng cao hiệu suất công việc
- Cập nhật thông tin luật pháp mới nhất, chính xác nhất
- Hỗ trợ người dùng (người dân) trả lời các vấn đề liên quan đến luật pháp

## NỘI DUNG VÀ PHƯƠNG PHÁP

Bài toán sử dụng mô hình tinh chỉnh của VinBigData ([Link](#)), được xếp hạng điểm chuẩn 36.98 trên VMLU ([Link](#)). Điều này khiến bài toán có thể hiểu được tiếng Việt tốt hơn mà không quá máy móc như dịch từ tiếng Anh sang tiếng Việt như các mô hình ChatGPT, Llama đang hoạt động. Kết hợp với kỹ thuật thể hệ tăng cường truy xuất (RAG) để mô hình có thêm kiến thức về luật pháp, các câu hỏi được cung cấp với thời gian mới nhất là 2024, nội dung cung cấp bao gồm câu hỏi và câu trả lời về luật pháp Việt Nam như luật hình sự, luật hôn nhân, luật dân sự và cùng nhiều các bộ luật khác. Triển khai kỹ thuật phân đoạn ngữ nghĩa (semantic chunking) giúp mô hình hiểu ngữ cảnh câu hỏi một cách gần nhất.

## KẾT QUẢ MONG ĐỢI

Mô hình có thể hoạt động tốt, trả lời các câu hỏi một cách chính xác, câu trả lời của mô hình sát với câu trả lời thật sự.



Các điểm đánh giá định lượng, định tính đạt kết quả cao hoặc tương đối (sẽ tìm cách cải tiến để có kết quả tốt hơn)

## TÀI LIỆU THAM KHẢO

- [1] S. K. Nigam, S. K. Mishra, A. K. Mishra, N. Shallum, and A. Bhattacharya, "Legal Question-Answering in the Indian Context: Efficacy, Challenges, and Potential of Modern AI Models," arXiv preprint arXiv:2309.14735, 2023.
- [2] J. Li, R. Bhambhoria, S. Dahan, and X. Zhu, "Experimenting with Legal AI Solutions: The Case of Question-Answering for Access to Justice," arXiv preprint arXiv:2409.07713, 2024.
- [3] N. Wiratunga, R. Abeyratne, L. Jayawardena, K. Martin, S. Massie, I. Nkisi-Orji, R. Weerasinghe, A. Liret, and B. Fleisch, "CBR-RAG: Case-Based Reasoning for Retrieval Augmented Generation in LLMs for Legal Question Answering," in International Conference on Case-Based Reasoning (ICCBR), Springer, 2024, pp.

445–460.

[4] A. Louis, G. van Dijck, and G. Spanakis, "Interpretable Long-Form Legal Question Answering with Retrieval-Augmented Large Language Models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 20, 2024, pp. 22266–22275.

[5] B. Padiu, R. Iacob, T. Rebedea, and M. Dascalu, "To What Extent Have LLMs Reshaped the Legal Domain So Far? A Scoping Literature Review," *Information*, vol. 15, no. 11, p. 662, 2024.