

# Nghiên cứu và xây dựng Chatbot tư vấn Luật Đất đai bằng Hybrid RAG và kiểm chứng pháp lý

Design and development of a Land Law consultation Chatbot using a Hybrid RAG framework with legal validity verification

> **KS NGUYỄN HỮU LUÂN\***, **TS NGUYỄN QUỐC TUẤN**

Khoa Công nghệ Thông tin, Trường Đại học Giao thông vận tải

\*Email: luannh@utc.edu.vn

## TÓM TẮT

Hệ thống pháp luật Việt Nam, đặc biệt là Luật Đất đai có tính phức tạp cao, gây khó khăn cho việc tra cứu và tư vấn. Việc áp dụng các Mô hình Ngôn ngữ lớn (LLMs) trong lĩnh vực này đối mặt với thách thức lớn về tính chính xác pháp lý và khả năng duy trì ngữ cảnh dài. Nghiên cứu này đề xuất LMentor-HybridRAG, một framework chatbot chuyên biệt cho tư vấn Luật Đất đai Việt Nam, kết hợp LLMs với cơ chế Retrieval-Augmented Generation nâng cao. Các cải tiến chính bao gồm: Truy xuất lai (kết hợp truy xuất ngữ nghĩa và từ khóa), kỹ thuật tăng cường ngữ cảnh phân cấp (Hierarchical Contextual Augmentation - HCA) để tổng hợp thông tin từ nhiều nguồn luật và cơ chế kiểm chứng pháp lý tự động. Mô hình đã được huấn luyện và đánh giá trên bộ dữ liệu Vietnamese Land Law QA Dataset (VLL-QA), cho thấy kết quả khả quan về độ chính xác và giảm thiểu tỷ lệ ảo giác, có tiềm năng ứng dụng thực tiễn trong lĩnh vực tư vấn pháp luật tại Việt Nam.

**Từ khóa:** Chatbot; Luật Đất đai; Hybrid RAG; kiểm chứng pháp lý; xử lý ngôn ngữ tự nhiên; truy xuất thông tin.

## ABSTRACT

Vietnam's legal system, particularly the Land Law, is highly complex, making information retrieval and legal consultation challenging. Applying Large Language Models (LLMs) in this domain is further hindered by risks of hallucination and limitations in long-context reasoning. This study introduces LMentor-HybridRAG, a specialized Land Law consultation framework that integrates LLMs with an enhanced RAG architecture. Its key features include hybrid retrieval, Hierarchical Contextual Augmentation (HCA) and automated Legal Fact-Checking. Evaluated on the VLL-QA dataset, the model demonstrates improved accuracy and reduced hallucination rates, indicating strong potential for real-world legal advisory use in Vietnam.

**Keywords:** Chatbot; Land law; Hybrid RAG; legal fact-checking; natural language processing (NLP); information retrieval.

## 1. ĐẶT VẤN ĐỀ

Nhu cầu tiếp cận thông tin pháp luật một cách nhanh chóng và chính xác là một yêu cầu cấp thiết trong bối cảnh hiện nay. Hệ thống pháp luật Việt Nam, đặc biệt là Luật Đất đai và các văn bản liên quan có cấu trúc đa tầng, thường xuyên cập nhật và mang tính chuyên môn cao. Điều này gây không ít khó khăn cho người dân và doanh nghiệp trong việc tra cứu, diễn giải các quy định về quyền sử dụng đất, thủ tục hành chính hay giải quyết tranh chấp.

Sự phát triển của Trí tuệ nhân tạo tạo sinh (Generative AI) và các Mô hình ngôn ngữ lớn (LLMs) vốn dựa trên kiến trúc Transformer [3], như GPT-4 [1] đã mở ra tiềm năng xây dựng các hệ thống tư vấn pháp lý tự động. Tuy nhiên, việc áp dụng LLMs vào lĩnh vực pháp

luật phải đối mặt với thách thức cố hữu là hiện tượng "ảo giác" (hallucination) - một vấn đề đã được khảo sát chi tiết [13] - khi mô hình có thể tạo ra những thông tin nghe hợp lý nhưng lại sai lệch so với văn bản gốc.

Phương pháp Retrieval-Augmented Generation (RAG) [2] ra đời như một giải pháp, bằng cách kết hợp khả năng sinh ngôn ngữ với việc truy xuất thông tin từ một kho tri thức chuyên biệt. Dù vậy, RAG truyền thống vẫn bộc lộ hạn chế khi áp dụng cho lĩnh vực phức tạp như Luật Đất đai, như đã được chỉ ra trong nhiều khảo sát gần đây [11]. Các vấn đề chính bao gồm: (1) Khó khăn trong việc truy xuất chính xác các điều khoản do sự mơ hồ của ngôn ngữ tự nhiên; (2) Thất bại trong việc tổng hợp thông tin từ nhiều nguồn văn bản

để trả lời một câu hỏi phức tạp; (3) Thiếu cơ chế kiểm chứng tự động để đảm bảo tính xác thực của câu trả lời.

Để giải quyết các vấn đề trên, nghiên cứu này đề xuất LMentor-HybridRAG, một framework RAG được thiết kế cho bối cảnh pháp luật Việt Nam. Các đóng góp chính của nghiên cứu bao gồm: (1) Đề xuất kiến trúc LMentor-HybridRAG tích hợp truy xuất lai, kỹ thuật tăng cường ngữ cảnh phân cấp (HCA) và cơ chế kiểm chứng pháp lý; (2) Xây dựng bộ dữ liệu Vietnamese Land Law QA Dataset (VLL-QA) với hơn 500 câu hỏi thực tế; (3) Cung cấp đánh giá thực nghiệm toàn diện, cho thấy hiệu quả của mô hình trong việc tăng độ chính xác và giảm tỷ lệ ảo giác.

## 2. CƠ SỞ LÝ THUYẾT

Lĩnh vực chatbot pháp lý có thể được phân loại thành 3 nhóm chính. Các hệ thống dựa trên quy tắc (rule-based) và dựa trên truy xuất (retrieval-based) hoạt động hiệu quả với các kịch bản cố định nhưng thiếu khả năng diễn giải sâu. Nhóm chatbot tạo sinh (generative chatbot) dựa trên LLMs mang lại tính linh hoạt cao nhưng phải đối mặt với vấn đề ảo giác. Để khắc phục ảo giác, các nghiên cứu gần đây đã tập trung vào cải tiến RAG [11]. Một ví dụ là truy xuất lai (Hybrid Retrieval) [4], kết hợp truy xuất ngữ nghĩa (semantic) và truy xuất từ khóa (lexical). Nhiều kỹ thuật tiên tiến khác cũng được đề xuất như RAG tự điều chỉnh (Self-RAG) [7] hay Corrective RAG (CRAG) [8]. Bên cạnh đó còn có thêm kỹ thuật tăng cường ngữ cảnh phân cấp (Hierarchical Contextual Augmentation - HCA) [5] hỗ trợ duy trì ngữ cảnh dài. Các công trình tập trung vào lĩnh vực pháp lý [10] hay các hệ thống QA pháp lý [9] nhấn mạnh tầm quan trọng của kiểm chứng pháp lý (Legal Fact-Checking) [6].

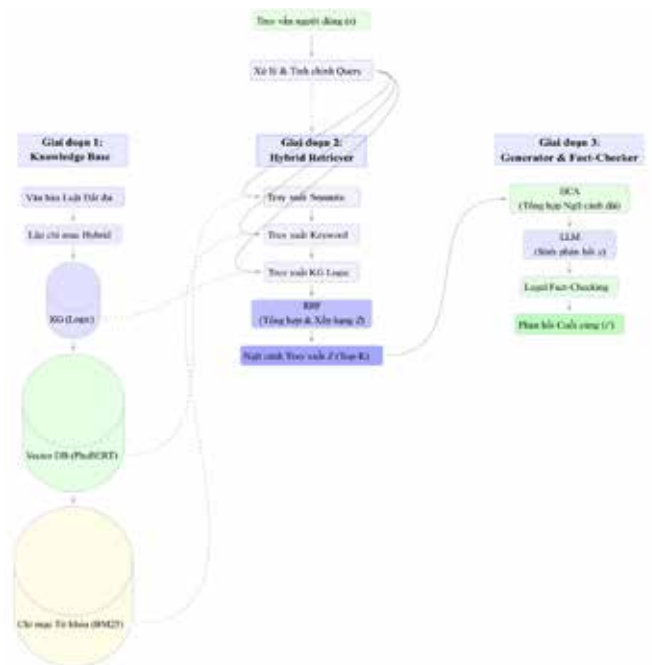
Hầu hết các công trình tiên tiến trên thế giới chỉ tập trung vào ngữ cảnh tiếng Anh và hệ thống pháp luật phương Tây. Tại Việt Nam, tính bản địa hóa là thách thức lớn, đặc biệt khi áp dụng các LLM cho tiếng Việt. Một số công trình đã có những bước đi đầu tiên trong ứng dụng NLP vào lĩnh vực dịch vụ công và pháp luật, ví dụ như ứng dụng NLP trong xây dựng chatbot dịch vụ công [15] hay đánh giá hiệu quả của chatbot trong tư vấn pháp luật [17]. Bên cạnh đó, các nghiên cứu về thách thức và cơ hội của AI trong pháp luật tại Việt Nam [14] cũng như tiềm năng ứng dụng AI trong quản lý đất đai [16] đã được thực hiện. Các công trình tại Việt Nam trước đây chủ yếu tập trung vào các hệ thống hỏi đáp cơ bản hoặc chatbot dịch vụ công với khả năng hạn chế. Chưa có nhiều nghiên cứu đề xuất một kiến trúc RAG nâng cao, được thiết kế đặc biệt để giải quyết đồng thời vấn đề ngữ cảnh dài và kiểm chứng pháp lý cho các văn bản luật phức tạp như Luật Đất đai. Đây chính là khoảng trống nghiên cứu mà framework LMentor-HybridRAG hướng tới giải quyết.

Để giải quyết những thách thức đã nêu, chúng tôi đề xuất kiến trúc LMentor-HybridRAG - một hệ thống RAG đa tầng được thiết kế chuyên biệt cho lĩnh vực pháp luật Việt Nam. Kiến trúc này được minh họa trong Hình 1, bao gồm 3 giai đoạn chính:

- **Giai đoạn 1 - Xây dựng cơ sở tri thức (Knowledge Base):** Giai đoạn nền tảng, tập trung vào việc xử lý và cấu trúc hóa nguồn văn bản luật thô thành một kho tri thức có khả năng truy vấn hiệu quả.

- **Giai đoạn 2 - Truy xuất lai (Hybrid Retriever):** Giai đoạn này chịu trách nhiệm tìm kiếm và trích xuất những thông tin pháp lý liên quan nhất từ cơ sở tri thức để trả lời cho một truy vấn cụ thể của người dùng.

- **Giai đoạn 3 - Sinh và kiểm chứng phân hồi (Generator & Fact-Checker):** Giai đoạn cuối cùng, sử dụng thông tin đã truy xuất để tạo ra một câu trả lời tự nhiên, chính xác và được kiểm chứng về mặt pháp lý trước khi gửi đến người dùng.



Hình 1. Kiến trúc tổng thể của hệ thống LMentor-HybridRAG

### 2.1. Giai đoạn Knowledge Base: Xây dựng tri thức và lập chỉ mục Hybrid

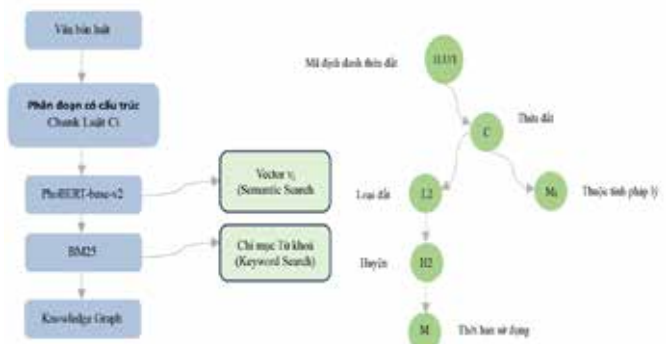
Giai đoạn này tập trung vào việc xây dựng kho tri thức Luật Đất đai có cấu trúc, nhằm tối ưu hóa quá trình truy xuất. Văn bản luật được thu thập, chuẩn hóa và phân đoạn thành các chunks (C) theo cấu trúc logic (chương, điều, khoản). Việc lựa chọn kích thước và chiến lược phân đoạn (chunking strategy) phù hợp là rất quan trọng. Sau đó, dữ liệu được lập chỉ mục kép:

- **Vector Indexing:** Mỗi chunk  $C_i$  được mã hóa thành một vector ngữ nghĩa  $v_i$  thông qua một mô hình nhúng ngôn ngữ  $E$ . Quá trình này được biểu diễn bởi công thức:  $v_i = E(C_i)$ .

Các vector này được lưu trữ trong một cơ sở dữ liệu vector, cho phép tìm kiếm dựa trên sự tương đồng về ngữ nghĩa.

- **Keyword Indexing:** Đồng thời, các từ khóa quan trọng trong mỗi chunk được lập chỉ mục bằng các thuật toán truy xuất thông tin cổ điển như BM25. Điều này đảm bảo rằng các thuật ngữ pháp lý cụ thể có thể được tìm thấy một cách chính xác.

Ngoài ra, để tăng cường khả năng suy luận, các thực thể pháp lý (ví dụ: loại đất, thời hạn sử dụng) và quan hệ giữa chúng được trích xuất để tạo thành một đồ thị tri thức (Knowledge Graph - KG) dưới dạng các bộ ba (subject, predicate, object), giúp giải quyết các truy vấn yêu cầu tham chiếu chéo. Quy trình này được minh họa trong Hình 2.



Hình 2. Quy trình xây dựng và biểu diễn Knowledge Base pháp lý

**2.2. Giai đoạn Hybrid Retriever: Truy xuất đa chiến lược**

Giai đoạn này nhận truy vấn  $x$  từ người dùng và trả về tập hợp các đoạn trích liên quan  $Z$  bằng cách kết hợp 3 luồng truy xuất song song: Truy xuất ngữ nghĩa, truy xuất từ khóa và truy xuất dựa trên đồ thị tri thức. Kết quả từ 3 luồng này ( $Z_{sem}, Z_{key}, Z_{KG}$ ) được hợp nhất và xếp hạng lại bằng thuật toán Reciprocal Rank Fusion (RRF) để chọn ra tập hợp  $Z$  cuối cùng. RRF tính điểm cho mỗi đoạn trích dựa trên thứ hạng của nó trong các danh sách kết quả khác nhau, giúp ưu tiên các kết quả xuất hiện ở vị trí cao trên nhiều luồng. Điểm RRF cho một đoạn trích  $d$  được tính như sau:

$$RRF\_score(d) = \sum_{r \in R} \left( \frac{1}{k + rank_r(d)} \right) \tag{1}$$

Trong đó:  $R$  - Tập hợp các danh sách xếp hạng;  $rank_r(d)$  - Thứ hạng của  $d$  trong danh sách  $r$ ;  $k$  - Một hằng số (thường là 60) để giảm ảnh hưởng của các thứ hạng quá cao. Cơ chế này được minh họa trong Hình 3.

**2.3. Giai đoạn Generator và Legal Fact-Checking: Tổng hợp và kiểm chứng**

Giai đoạn này sử dụng LLM để sinh ra câu trả lời  $y$  dựa trên truy vấn  $x$  và ngữ cảnh  $Z$ . Để giải quyết vấn đề ngữ cảnh dài, kỹ thuật Hierarchical Contextual Augmentation (HCA) [5] được áp dụng để tổng hợp  $Z$  thành ngữ cảnh cô đọng  $Z_{HCA}$ . LLM sau đó tạo ra một câu trả lời sơ bộ  $y' = LLM(x, Z_{HCA})$ .

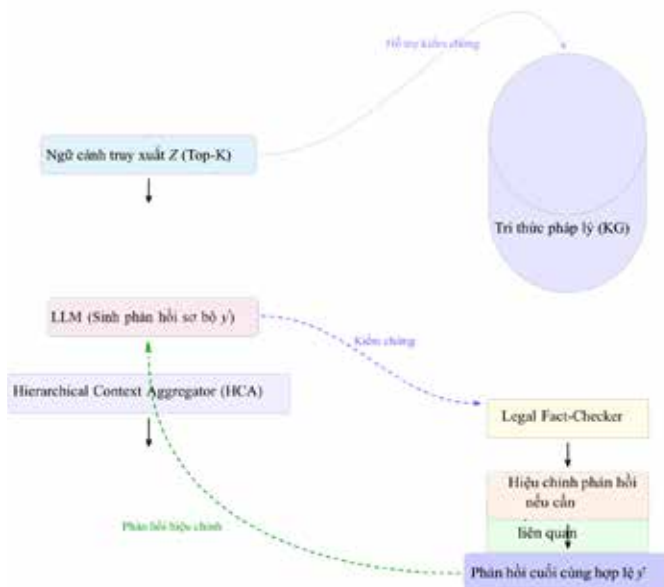
Một cải tiến trong kiến trúc này là cơ chế Legal Fact-Checking [6]. Một mô hình xác minh  $\delta$  được sử dụng để kiểm tra tính nhất quán và sự thật của  $y'$  so với các văn bản luật gốc trong  $Z$ . Phản hồi cuối cùng  $y$  chỉ được chấp nhận nếu vượt qua kiểm chứng:

$$\hat{y} = \begin{cases} y', & \text{if } \delta(y', Z) = \text{True} \\ \text{re-generate or flag,} & \text{otherwise} \end{cases} \tag{2}$$

Quy trình này đảm bảo rằng thông tin do chatbot cung cấp có căn cứ pháp lý rõ ràng, giúp giảm thiểu đáng kể hiện tượng ảo giác.



Hình 3. Cơ chế hoạt động của Hybrid Retriever



Hình 4. Quy trình sinh phản hồi với HCA và kiểm chứng pháp lý

### 3. TRIỂN KHAI MÔ HÌNH ĐỀ XUẤT

Trong nghiên cứu này, chúng tôi đánh giá hiệu suất của LMentor-HybridRAG so với các mô hình nền tảng trên bộ dữ liệu VLL-QA.

#### 3.1. Thiết lập thí nghiệm

Thí nghiệm được thực hiện trên bộ dữ liệu VLL-QA (Vietnamese Land Law QA Dataset) do chúng tôi xây dựng. Bộ dữ liệu này chứa 520 cặp câu hỏi-trả lời liên quan đến Luật Đất đai Việt Nam. Các câu hỏi được thu thập từ các nguồn công khai như diễn đàn pháp luật, trang web tư vấn, phản ánh các tình huống và thắc mắc thực tế của người dân. Mỗi cặp dữ liệu bao gồm: (1) Câu hỏi gốc của người dùng; (2) Câu trả lời được soạn thảo cẩn thận, diễn giải rõ ràng các quy định liên quan; (3) Trích dẫn điều khoản luật cụ thể làm căn cứ cho câu trả lời. Cấu trúc này cho phép đánh giá cả khả năng truy xuất đúng điều luật và tính chính xác của nội dung phản hồi do mô hình tạo ra. Một số ví dụ về câu hỏi trong bộ dữ liệu:

- "Thủ tục xin cấp Giấy chứng nhận quyền sử dụng đất lần đầu cần những giấy tờ gì?"

- "Đất nông nghiệp có được chuyển nhượng cho người không trực tiếp sản xuất nông nghiệp không?"

- "Thời hạn sử dụng đất ở tại đô thị là bao lâu?"

\* Các chỉ số đánh giá: Việc đánh giá hiệu suất được thực hiện dựa trên các chỉ số chuẩn:

- Độ chính xác truy xuất: Sử dụng chỉ số Hit@k (công thức (4)), đo lường tỷ lệ các truy vấn mà điều luật liên quan nằm trong top k kết quả trả về của retriever.

- Chất lượng phản hồi sinh ra: Sử dụng F1-score, ROUGE-L, và BERTScore (F1) để đo độ tương đồng về mặt từ vựng và ngữ nghĩa giữa câu trả lời của mô hình và câu trả lời tham chiếu trong bộ dữ liệu.

- Tỷ lệ ảo giác (Hallucination Rate): Đo lường phần trăm các câu trả lời do mô hình sinh ra có chứa thông tin không chính xác hoặc không có căn cứ trong các văn bản luật đã truy xuất. Tỷ lệ này được đánh giá thủ công trên một tập mẫu đại diện.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

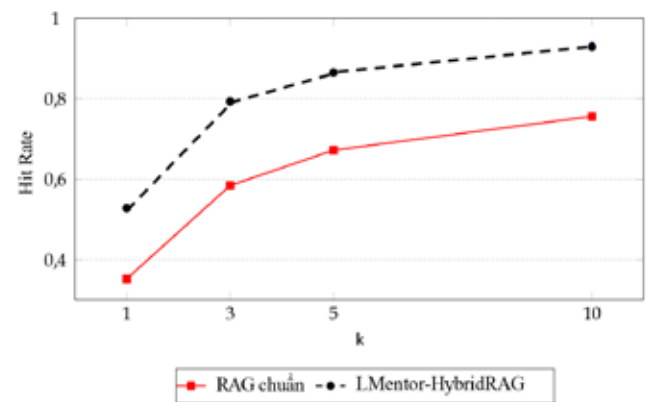
$$\text{Hit}@k = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \mathbb{1}(\text{"tìm thấy câu trả lời trong top k đoạn trích cho truy vấn } i\text{"}) \quad (4)$$

### 3.2. Kết quả thực nghiệm

- **Đánh giá khả năng truy xuất:** Bảng 1 và Hình 5 so sánh hiệu suất truy xuất của RAG chuẩn (chỉ dùng Semantic Search) và phương pháp đề xuất (Hybrid Retrieval). Trong Hình 5, đường nét đứt (LMentor-HybridRAG) cho thấy hiệu suất cao hơn rõ rệt so với đường màu đỏ (RAG chuẩn) ở mọi điểm k. Đặc biệt, tại k = 1, mô hình này đạt 0,528 (so với 0,352), nghĩa là có khả năng tìm thấy đúng điều luật ngay trong kết quả đầu tiên cao hơn khoảng 17,6%. Sự cải thiện này cho thấy việc kết hợp truy xuất từ khóa (BM25) và ngữ nghĩa (vector) là cần thiết để xử lý hiệu quả các thuật ngữ pháp lý chuyên ngành.

Bảng 1. So sánh hiệu suất truy xuất (Hit@k)

Phương pháp	Hit@1	Hit@3	Hit@5	Hit@10
RAG chuẩn (Semantic Search)	0,352	0,584	0,672	0,756
<b>LMentor-HybridRAG</b>	<b>0,528</b>	<b>0,793</b>	<b>0,865</b>	<b>0,932</b>



Hình 5. Đồ thị so sánh hiệu suất truy xuất (Hit@k)

- **Đánh giá độ chính xác sinh phản hồi:** Bảng 2 cho thấy kết quả đánh giá cuối cùng về chất lượng phản hồi. LMentor-HybridRAG đạt F1-score 0,87, cao hơn so với RAG chuẩn. Đặc biệt, tỷ lệ ảo giác giảm đáng kể xuống còn 5% là minh chứng cho hiệu quả của cơ chế kiểm chứng pháp lý.

Bảng 2. Đánh giá độ chính xác và độ tin cậy sinh phản hồi

Mô hình	F1-score	BERTScore (F1)	ROUGE-L	Hallucination
PhoBERT	0,74	0,81	0,69	N/A
RAG chuẩn	0,79	0,84	0,73	18%
<b>LMentor-HybridRAG</b>	<b>0,87</b>	<b>0,90</b>	<b>0,82</b>	<b>5%</b>

Kết quả thực nghiệm cho thấy hiệu quả của mô hình đề xuất, chủ yếu đến từ khả năng kết hợp đa tầng. Truy xuất lai đảm bảo rằng các điều khoản cụ thể được truy xuất chính xác hơn, trong khi kỹ thuật HCA giúp xử lý vấn đề ngữ cảnh dài. Một đóng góp chính của bài báo là cơ chế kiểm chứng pháp lý, giúp mô hình đạt tỷ lệ ảo giác thấp (5%), đáp ứng yêu cầu về độ tin cậy trong lĩnh vực pháp luật. Khi đối chiếu với một số nghiên cứu trong nước, ví dụ như công trình về ứng dụng AI trong pháp luật [14] hay các hệ thống chatbot dịch vụ công [15], kết quả F1-score 0,87 của LMentor-HybridRAG cho thấy tiềm năng so với các hệ thống trước đây. Điều này cho thấy LMentor-HybridRAG là một

hướng tiếp cận có tiềm năng trong việc ứng dụng AI một cách tin cậy vào lĩnh vực pháp luật tại Việt Nam.

Mặc dù đạt kết quả khả quan, mô hình vẫn còn một số hạn chế. Độ chính xác có thể chưa tối ưu đối với các câu hỏi yêu cầu suy luận pháp lý đa bước phức tạp, ví dụ như các tình huống liên quan đến tranh chấp hoặc chống chéo nhiều quy định. Việc giải quyết các trường hợp này đòi hỏi sự tích hợp sâu hơn của Đồ thị tri thức (Knowledge Graph). Hơn nữa, việc triển khai và vận hành các LLM chất lượng cao cùng cơ chế HCA/Fact - Checking vẫn đòi hỏi tài nguyên tính toán đáng kể, là một thách thức cho việc ứng dụng rộng rãi.

#### 4. KẾT LUẬN

Nghiên cứu này đã trình bày và đánh giá LMentor-HybridRAG, một kiến trúc RAG nâng cao cho chatbot tư vấn Luật Đất đai Việt Nam. Bằng cách tích hợp các kỹ thuật truy xuất lại, tổng hợp ngữ cảnh phân cấp và kiểm chứng pháp lý, mô hình đã đạt được sự cân bằng giữa khả năng diễn giải ngôn ngữ tự nhiên của LLM và yêu cầu về tính chính xác trong lĩnh vực pháp luật. Kết quả thực nghiệm với F1-score 0,87 và tỷ lệ ảo giác giảm xuống còn 5% đã minh chứng cho hiệu quả và tiềm năng của phương pháp đề xuất.

Trong tương lai, các hướng nghiên cứu có thể tập trung vào việc mở rộng và làm sâu sắc hơn Đồ thị tri thức (KG) để hỗ trợ suy luận đa bước, giải quyết các trường hợp pháp lý phức tạp hơn, có thể tham khảo các phương pháp tích hợp KG vào RAG [12]. Bên cạnh đó, việc phát triển cơ chế kiểm chứng pháp lý dựa trên các mô hình chuyên gia, được huấn luyện trên án lệ và ý kiến từ các luật sư, sẽ là một hướng đi hứa hẹn. Cuối cùng, việc mở rộng bộ dữ liệu sang các lĩnh vực luật khác như dân sự, hôn nhân và gia đình là cần thiết để hướng tới một hệ thống tư vấn pháp lý toàn diện và tin cậy cho người dân Việt Nam.

**Lời cảm ơn:** Nghiên cứu này được tài trợ bởi Trường Đại học Giao thông vận tải trong Đề tài mã số T2025-CN-001.

#### TÀI LIỆU THAM KHẢO

- [1]. Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei (2020), "Language Models are Few-Shot Learners", in *Adv. Neural Inf. Process. Syst.*, vol.33, pp.1877-1901. [Online]. Available: <https://arxiv.org/abs/2005.14165>.
- [2]. Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, Douwe Kiela (2020), "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *arXiv preprint, arXiv:2005.11401*. [Online]. Available: <https://arxiv.org/abs/2005.11401>.
- [3]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin (2017), "Attention is All You Need", in *Adv. Neural Inf. Process. Syst.*, pp.5998-6008. [Online]. Available: <https://arxiv.org/abs/1706.03762>.
- [4]. Tao Zhang, Wu-Jun Li (2022), "Hybrid Contrastive Learning for Legal Text Retrieval", *Proc. AAAI Conf. Artif. Intell.*, vol.36, no.10, pp.11210-11218. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/21376>.
- [5]. Jinhuk Kim, Min-Keyong Kim, Gyu-Ri Kim, Sang-Wook Kim (2023), "Hierarchical Contextual Augmentation for Long-Document Question Answering", *Proc. ACL*. [Online]. Available: <https://arxiv.org/abs/2305.11899>.
- [6]. Max S. W. Zhang, Bodhisattwa Prasad Majumder, Julian McAuley (2023), "Fact-Checking with Factuality Prediction for Generative Models", *Proc. EMNLP*. [Online]. Available:

<https://arxiv.org/abs/2310.19852>.

[7]. Ojas Trivedi, Hongjin Ahn, Hwaran Lee, W. Ronny Huang, Yixu Song, Sung-Ju Lee, Min-Yen Kan (2023), "Self-RAG: Learning to Retrieve, Generate, and Critique", *arXiv preprint, arXiv:2310.11511*. [Online]. Available: <https://arxiv.org/abs/2310.11511>.

[8]. Shi-Qi Yan, Jia-Chen Gu, Yun-Zhe Lin, Zhen-Hua Ling, Quan-Sen Yang (2024), "Corrective Retrieval Augmented Generation", *arXiv preprint, arXiv:2401.15884*. [Online]. Available: <https://arxiv.org/abs/2401.15884>.

[9]. Yixuan Wang, Yutao Guo, Zhaowei Wang, Yijun Liu (2023), "LegalQA: A Question Answering benchmark for the legal domain", *Proc. Findings of ACL*. [Online]. Available: <https://arxiv.org/abs/2305.02174>.

[10]. Jiarui Ma, Weiming Zhang, Zihan Li, Ting Liu, Nian Liu (2023), "Challenges and Opportunities in Applying Large Language Models to the Legal Domain", *arXiv preprint, arXiv:2311.16421*. [Online]. Available: <https://arxiv.org/abs/2311.16421>.

[11]. Yunfan Gao, Yuncong Hu, Yanfeng Wang, Hanjun Dai, Zhirui Zhang, Wei Liu, Peng Li, Yang Liu, Siwei Dong, Ruofei Zhang, Yichong Xu, Xin Zhao, Ji-Rong Wen (2023), "Retrieval-Augmented Generation for Large Language Models: A Survey", *arXiv preprint, arXiv:2312.10997*. [Online]. Available: <https://arxiv.org/abs/2312.10997>.

[12]. Xuefeng He, Zheheng Luo, Zhaowei Wang (2024), "Knowledge Graph Enhanced Retrieval-Augmented Generation", *arXiv preprint, arXiv:2403.11183*. [Online]. Available: <https://arxiv.org/abs/2403.11183>.

[13]. Mengzhang Zhang, Yijiang Liu, Yufan Wang, Yixin Liu, Yijia Zhang, Yichi Zhang, Zhirui Zhang, Heng Zhang, Chao Zhang, Yuxiao Dong, Jie Tang (2023), "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges and Open Questions", *arXiv preprint, arXiv:2311.05232*. [Online]. Available: <https://arxiv.org/abs/2311.05232>.

[14]. N. T. H. Nhung (2022), "Thách thức và cơ hội của trí tuệ nhân tạo trong lĩnh vực pháp luật", *Tạp chí Luật học*. [Online]. Available: <https://tapchi.hlu.edu.vn/SubNews/Details/21804>.

[15]. N. Q. Huy (2021), "Ứng dụng xử lý ngôn ngữ tự nhiên trong xây dựng chatbot dịch vụ công tại Việt Nam", *Proceedings of ICT National Conference*. [Online]. Available: <https://vjol.info.vn/index.php/ict/article/view/65059>.

[16]. L. T. Trung (2020), "Tiềm năng ứng dụng trí tuệ nhân tạo trong quản lý đất đai", *Tạp chí Địa chính Việt Nam*. [Online]. Available: <https://tcdy.gov.vn/ai-quan-ly-dat-dai>.

[17]. P. Q. Viet (2023), "Hiệu quả chatbot trong tư vấn pháp luật và giải đáp thủ tục hành chính", *Tạp chí KH&CN Việt Nam*. [Online]. Available: <https://vjol.info.vn/index.php/jst/article/view/76432>.