ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH

CÁC VẤN ĐỀ NGHIÊN CỨU VÀ ỨNG DỤNG TRONG KHMT (CS529)

BÁO CÁO ĐỒ ÁN

RoadLawQA - Hệ thống hỏi đáp về Luật giao thông đường bộ

Giảng viên hướng dẫn: ThS. Đỗ Văn Tiến
Sinh viên thực hiện: Nguyễn Thanh Tùng 23521745
Mai Lê Bá Vương 23521821
Võ Hoàng Minh 23520961

TP. Hồ Chí Minh, Tháng 1 năm 2026

Mục lục

1 Câu 1

Câu hỏi : Trình bày các bước thường thực hiện trong một quy trình nghiên cứu khoa học và quy trình xây dựng một ứng dụng có sử dụng AI

Trả lời:

Quy trình này thường được áp dụng khi bạn viết báo khoa học (paper), làm khóa luận, hoặc tìm ra phương pháp giải quyết vấn đề mới.

- Xác định vấn đề và Câu hỏi nghiên cứu (Problem Definition):
  - Tìm kiếm một vấn đề chưa được giải quyết tốt hoặc một khía cạnh mới của vấn đề cũ.
  - Ví dụ: "Làm thế nào để cải thiện độ chính xác của mô hình phân tích cảm xúc tiếng Việt khi dữ liệu bị mất cân bằng?"
- Tổng quan tài liệu (Literature Review):
  - Đọc các bài báo (papers) trước đó để hiểu State-of-the-Art (SOTA) hiện tại là gì.
  - Tìm ra "Research Gap" (khoảng trống nghiên cứu) – điểm mà các phương pháp hiện tại chưa làm được.
- Đề xuất phương pháp (Methodology Proposal):
  - Xây dựng giả thuyết khoa học.
  - Thiết kế kiến trúc mô hình mới, hàm loss mới, hoặc quy trình xử lý dữ liệu mới.
  - Đây là bước quan trọng nhất để đánh giá tính đóng góp của nghiên cứu.
- Thực nghiệm (Experimentation):
  - Thu thập dữ liệu (thường là các bộ dataset chuẩn - benchmark datasets).
  - Huấn luyện mô hình đề xuất và các mô hình cơ sở (baselines) để so sánh.
  - Sử dụng các độ đo khắt khe (Accuracy, F1-Score, AUC, v.v.) để chứng minh giả thuyết.
- Phân tích kết quả và Kết luận:
  - Giải thích tại sao mô hình hoạt động tốt (hoặc không tốt).
  - Trực quan hóa dữ liệu/kết quả (Biểu đồ, Confusion Matrix)
- Công bố (Publication):
  - Viết báo cáo/paper và gửi đi hội nghị hoặc tạp chí khoa học.

Phần 2: Quy trình Xây dựng Ứng dụng AI (AI Application Lifecycle) Quy trình này tập trung vào việc đưa mô hình vào phục vụ người dùng cuối (End-user), thường gọi là MLOps.

- Xác định yêu cầu bài toán (Business Understanding):

  - Mục tiêu không phải là "độ chính xác cao nhất" mà là "giải quyết nỗi đau của người dùng"
  - Xác định các ràng buộc: Thời gian phản hồi (latency), chi phí phần cứng, khả năng chịu tải.
- Thu thập và Xử lý dữ liệu (Data Engineering):
  - Khác với nghiên cứu (dùng dataset sạch có sẵn), bước này phải xây dựng pipeline để lấy dữ liệu từ thực tế (Log, User input, API).
  - Làm sạch, gán nhãn (Labeling) và lưu trữ dữ liệu (Data Lake/Warehouse).
- Xây dựng và Huấn luyện mô hình (Model Development):
  - Chọn mô hình phù hợp (đôi khi mô hình đơn giản như Logistic Regression hay XGBoost lại tốt hơn Deep Learning vì nó nhanh và nhẹ).
  - Fine-tune mô hình.
  - Đóng gói mô hình
- Tích hợp hệ thống (Integration và Backend):
  - Xây dựng API (dùng FastAPI, Flask) để các phần mềm khác gọi vào mô hình.
  - Xây dựng giao diện người dùng (Frontend) để người dùng tương tác.
- Triển khai (Deployment):
  - Đưa ứng dụng lên môi trường sản phẩm (Production) dùng Docker, Kubernetes hoặc Cloud (AWS/GCP)..
  - Tối ưu hóa tốc độ suy luận (Inference speed).
- Giám sát và Bảo trì (Monitoring và Maintenance):
  - Theo dõi Data Drift (dữ liệu thực tế thay đổi khác với lúc train).
  - Retrain (huấn luyện lại) mô hình định kỳ khi có dữ liệu mới.

2 Tổng quan về đề tài

2.1 Tóm tắt

Trong bối cảnh hệ thống văn bản pháp luật về giao thông đường bộ tại Việt Nam ngày càng phức tạp và thường xuyên cập nhật, việc tra cứu thông tin chính xác, kịp thời là một thách thức lớn đối với người dân. Các mô hình ngôn ngữ lớn hoặc chatbot phổ thông hiện nay thường thiếu kiến thức pháp luật chuyên sâu, đặc biệt là các quy định mới ban hành, đồng thời dễ gặp tình trạng ảo giác (hallucination), gây rủi ro cao trong các ứng dụng tư vấn pháp lý.

Trong đồ án này, bọn em đề xuất xây dựng *RoadLawQA – hệ thống hỏi đáp chuyên biệt về Luật giao thông đường bộ Việt Nam*, áp dụng kỹ thuật Retrieval-Augmented Generation (RAG) nhằm đảm bảo câu trả lời bám sát văn bản gốc và có căn cứ pháp lý minh bạch. Phạm vi dữ liệu của hệ thống tập trung vào ba văn bản luật mới nhất, bao gồm: Luật Đường bộ (35/2024/QH15), Luật Trật tự, an toàn giao thông đường bộ (36/2024/QH15) và Nghị định 168/2024/NĐ-CP.

Về mặt kỹ thuật, RoadLawQA tiếp nhận truy vấn ngôn ngữ tự nhiên, thực hiện truy xuất thông tin bằng cơ chế hybrid search kết hợp với kỹ thuật reranking để tối ưu hóa độ liên quan của dữ liệu đầu vào cho mô hình sinh. Bên cạnh khả năng giải đáp thắc mắc kèm trích dẫn nguồn, hệ thống còn tích hợp tính năng tạo câu hỏi trắc nghiệm (quiz) tự động, hỗ trợ người dùng ôn tập và củng cố kiến thức.

Hệ thống được đánh giá trên tập dữ liệu kiểm thử (50 câu hỏi) được xây dựng thủ công, gồm các dạng câu hỏi: tra cứu sự thật, tra cứu mức phạt (chế tài) và xử lý tình huống/suy luận. Kết quả thực nghiệm cho thấy RoadLawQA đạt hiệu quả cao trong việc truy xuất thông tin và sinh câu trả lời đúng trọng tâm, hạn chế bịa đặt, qua đó chứng minh tính khả thi và độ tin cậy trong việc hỗ trợ người dùng tiếp cận pháp luật giao thông.

2.2 Đặt vấn đề

Hệ thống văn bản pháp luật tại Việt Nam nói chung và lĩnh vực giao thông đường bộ nói riêng có đặc thù là khối lượng dữ liệu rất lớn, phức tạp và thường xuyên được cập nhật, sửa đổi. Điều này tạo ra rào cản lớn cho người dân trong việc tiếp cận và tra cứu thông tin chính xác, đặc biệt là khi cần xác định các mức xử phạt cụ thể cho từng hành vi vi phạm.

Hiện nay, sự bùng nổ của các mô hình ngôn ngữ lớn (LLM) và chatbot AI đã mang lại công cụ hỗ trợ tra cứu thông tin mạnh mẽ. Tuy nhiên, các hệ thống chatbot phổ thông thường gặp phải hai hạn chế lớn khi áp dụng vào lĩnh vực pháp lý:

- **Thiếu kiến thức cập nhật:** Dữ liệu huấn luyện thường bị giới hạn tại một thời điểm trong quá khứ, không bao gồm các văn bản luật mới nhất vừa được ban hành.
- **Ảo giác (Hallucination):** Mô hình có xu hướng tự sinh ra thông tin sai lệch hoặc bịa đặt các điều khoản không có thật nhưng với văn phong rất thuyết phục, gây ra sự nguy hiểm và rủi ro pháp lý cho người sử dụng.

Xuất phát từ thực tế đó, việc xây dựng một hệ thống hỏi đáp chuyên biệt, có khả năng truy xuất chính xác và trích dẫn nguồn luật cụ thể là vô cùng cấp thiết.

2.3 Mục tiêu đề tài

Mục tiêu chính của đề tài là xây dựng và đánh giá một hệ thống hỏi đáp chuyên biệt về luật giao thông đường bộ Việt Nam với các mục tiêu cụ thể như sau:

- Xây dựng hệ thống hỏi đáp RoadLawQA cho phép đặt câu hỏi bằng ngôn ngữ tự nhiên.
- Sinh câu trả lời dựa trên văn bản luật gốc, hạn chế hiện tượng bịa đặt thông tin.
- Cung cấp trích dẫn pháp lý rõ ràng cho mỗi câu trả lời.
- Hỗ trợ thêm chức năng tạo câu hỏi trắc nghiệm nhằm giúp người dùng ôn tập và củng cố kiến thức
- Đánh giá hiệu quả của hệ thống thông qua các độ đo phù hợp cho bài toán truy xuất và sinh câu trả lời.

2.4 Phạm vi nghiên cứu

Trong khuôn khổ đề tài này, hệ thống RoadLawQA tập trung vào các văn bản pháp luật mới nhất liên quan đến giao thông đường bộ tại Việt Nam (nguồn lấy từ CSDL Quốc gia về văn bản pháp luật - vbpl.vn), bao gồm:

1. Luật Đường bộ (Luật số 35/2024/QH15).
2. Luật Trật tự, an toàn giao thông đường bộ (Luật số 36/2024/QH15).
3. Nghị định 168/2024/NĐ-CP quy định xử phạt vi phạm hành chính.

Phạm vi chức năng của hệ thống bao gồm hỏi đáp dựa trên văn bản luật và sinh câu hỏi trắc nghiệm phục vụ học tập. Đề tài không hướng đến việc tư vấn pháp lý chuyên sâu hoặc xử lý các tình huống pháp lý vượt ra ngoài nội dung của các văn bản đã được xem xét.

2.5 Mô tả bài toán

Hệ thống RoadLawQA được xây dựng nhằm giải quyết bài toán chính là hỏi đáp về luật giao thông đường bộ Việt Nam, đồng thời có thêm một chức năng phụ là tạo câu hỏi trắc nghiệm phục vụ mục đích ôn tập kiến thức pháp luật. Bài toán được đặc tả thông qua các quy định cụ thể về đầu vào và đầu ra như sau:

- **Chức năng Hỏi đáp (Question Answering):**
  - *Input:* Một câu hỏi (query) dưới dạng ngôn ngữ tự nhiên từ người dùng.
    Ví dụ: "Không đội mũ bảo hiểm khi đi xe máy bị xử phạt như thế nào?".
  - *Output:* Câu trả lời bằng ngôn ngữ tự nhiên, chính xác, bám sát nội dung văn bản luật gốc và kèm theo trích dẫn (citation) nguồn luật.
- **Chức năng Tạo câu hỏi trắc nghiệm (Quiz Generation):**
  - *Input:* Người dùng nhập chủ đề cụ thể hoặc yêu cầu hệ thống tạo ngẫu nhiên.
  - *Output:* Một bộ câu hỏi trắc nghiệm bao gồm: nội dung câu hỏi, các phương án lựa chọn (A, B, C, D) và đáp án đúng, được sinh ra tự động từ cơ sở dữ liệu pháp luật của hệ thống.

3 Các nghiên cứu liên quan

Nhóm nghiên cứu cũng đã tìm hiểu về các phương pháp xử lý ngôn ngữ tự nhiên trong lĩnh vực pháp lý, sự phát triển của các hệ thống hỏi đáp (Question Answering - QA). Qua đó, đồ án phân tích các hạn chế của những nghiên cứu trước đây để làm nổi bật tính cấp thiết và hướng tiếp cận của hệ thống RoadLawQA.

3.1 Xử lý ngôn ngữ tự nhiên trong lĩnh vực pháp lý (Legal NLP)

Lĩnh vực pháp lý đặt ra nhiều thách thức đặc thù cho các bài toán Xử lý ngôn ngữ tự nhiên (NLP). Khác với văn bản thông thường, văn bản quy phạm pháp luật thường có cấu trúc phân tầng chặt chẽ (Chương, Mục, Điều, Khoản, Điểm), câu văn dài, sử dụng nhiều thuật ngữ chuyên ngành và đòi hỏi độ chính xác tuyệt đối về mặt ngữ nghĩa.

Các nghiên cứu sớm trong lĩnh vực Legal NLP chủ yếu tập trung vào các bài toán như: phân loại văn bản luật, trích xuất thực thể (như tên luật, mức phạt), và tìm kiếm văn bản liên quan. Tại Việt Nam, sự ra đời của các mô hình ngôn ngữ tiền huấn luyện (Pre-trained Language Models) như PhoBERT hay ViBERT đã cải thiện đáng kể khả năng hiểu tiếng Việt, tạo tiền đề cho các ứng dụng pháp lý phức tạp hơn. Tuy nhiên, việc áp dụng trực tiếp các mô hình này vào bài toán hỏi đáp vẫn gặp khó khăn do sự chồng chéo giữa các văn bản luật và sự thay đổi thường xuyên của các quy định.

3.2 Các phương pháp tiếp cận trong hệ thống Hỏi-Đáp

Sự phát triển của các hệ thống chatbot và hỏi đáp pháp luật có thể được chia thành hai hướng tiếp cận chính:

3.2.1 Phương pháp dựa trên luật và từ khóa (Rule-based & Keyword Matching)

Các hệ thống truyền thống thường sử dụng việc khớp từ khóa (keyword matching) hoặc các luật "If-Then" được định nghĩa thủ công.

- **Ưu điểm:** Hoạt động ổn định trong phạm vi hẹp, dễ kiểm soát câu trả lời.
- **Nhược điểm:** Thiếu linh hoạt, không hiểu được ngữ cảnh hoặc các câu hỏi diễn đạt theo ngôn ngữ tự nhiên phức tạp. Khi văn bản luật thay đổi (ví dụ: thay đổi mức phạt), hệ thống yêu cầu nỗ lực cập nhật thủ công rất lớn.

3.2.2 Phương pháp dựa trên Mô hình ngôn ngữ lớn (LLMs)

Sự xuất hiện của các LLM như GPT-3, GPT-4, Llama đã thay đổi hoàn toàn cách tương tác giữa người và máy.

- **Ưu điểm:** Khả năng sinh ngữ tự nhiên trôi chảy, hiểu ý định người dùng tốt.
- **Nhược điểm (Vấn đề ảo giác):** Như đã đề cập ở phần Đặt vấn đề, LLM hoạt động theo cơ chế xác suất dự đoán từ tiếp theo. Trong lĩnh vực luật, LLM thường xuyên gặp hiện tượng "ảo giác" (hallucination) – tự tin đưa ra các điều luật hoặc mức phạt không có thật. Hơn nữa, tri thức của LLM bị giới hạn bởi thời điểm cắt dữ liệu (knowledge cutoff), do đó không thể trả lời về các luật mới như Luật Đường bộ 2024 hay Luật Trật tự, ATGT 2024 nếu không được tinh chỉnh (fine-tune) lại.

3.3 Kỹ thuật Retrieval-Augmented Generation (RAG)

Để giải quyết nhược điểm của hai phương pháp trên, kỹ thuật Retrieval-Augmented Generation (RAG) đã được giới thiệu và nhanh chóng trở thành tiêu chuẩn cho các hệ thống hỏi đáp dựa trên tri thức (Knowledge-based QA).

RAG kết hợp sức mạnh của một bộ truy xuất thông tin (Retriever) và một mô hình sinh văn bản (Generator). Quy trình hoạt động tổng quát bao gồm:

1. **Truy xuất (Retrieve):** Khi nhận câu hỏi, hệ thống tìm kiếm các đoạn văn bản liên quan nhất từ cơ sở dữ liệu tri thức (Vector Database) bên ngoài.
2. **Tăng cường (Augment):** Các đoạn văn bản tìm được (context) được ghép cùng với câu hỏi gốc để tạo thành một prompt hoàn chỉnh.
3. **Sinh (Generate):** LLM sử dụng prompt này để sinh ra câu trả lời, đảm bảo nội dung dựa trên thông tin vừa truy xuất được.

Ưu điểm vượt trội của RAG trong bài toán pháp lý là tính minh bạch (có thể trích dẫn nguồn) và tính cập nhật (chỉ cần cập nhật cơ sở dữ liệu văn bản luật mà không cần huấn luyện lại mô hình).

4 Nội dung chính - Phương pháp

Hệ thống RoadLawQA được xây dựng dựa trên kiến trúc Retrieval-Augmented Generation (RAG), cho phép kết hợp khả năng hiểu ngôn ngữ tự nhiên vượt trội của các mô hình ngôn ngữ lớn (LLM) với độ chính xác và tính cập nhật của cơ sở dữ liệu pháp luật hiện hành. Cách tiếp cận này giúp khắc phục nhược điểm "ảo giác" (hallucination) thường gặp ở các LLM truyền thống khi trả lời các câu hỏi chuyên ngành hẹp như luật pháp.

4.1 Tổng quan hệ thống

Kiến trúc tổng thể của hệ thống RoadLawQA được thiết kế gồm hai giai đoạn xử lý chính tương tác chặt chẽ với nhau: Giai đoạn Ngoại tuyến (Offline Pipeline) và Giai đoạn Trực tuyến (Online Pipeline).

- **Giai đoạn Ngoại tuyến (Indexing):** Chịu trách nhiệm thu thập, tiền xử lý, phân đoạn và xây dựng chỉ mục vector từ các văn bản quy phạm pháp luật. Mục tiêu của giai đoạn này là chuyển đổi dữ liệu luật thô thành một kho dữ liệu luật đã được đánh chỉ mục, sẵn sàng cho quá trình truy xuất ngữ nghĩa.
- **Giai đoạn Trực tuyến (Retrieval & Generation):** Được kích hoạt khi người dùng gửi câu hỏi. Hệ thống thực hiện chuỗi tác vụ bao gồm: phân tích câu hỏi, truy xuất hỗn hợp (Hybrid Retrieval), xếp hạng lại kết quả (Reranking) để tìm ra các điều luật liên quan nhất, và cuối cùng là sinh câu trả lời (Generation) thông qua mô hình ngôn ngữ lớn.

Các hình trên minh họa khả năng tương tác thực tế của hệ thống, bao gồm sinh câu trả lời có trích dẫn luật và tạo câu hỏi trắc nghiệm phục vụ ôn tập kiến thức.

4.2 Xử lý dữ liệu và Đánh chỉ mục (Indexing)

Chất lượng của một hệ thống RAG phụ thuộc phần lớn vào khâu chuẩn bị dữ liệu. Quy trình này bao gồm các bước thu thập, phân đoạn và mã hóa vector.

4.2.1 Nguồn dữ liệu và Tiền xử lý

Hệ thống tập trung vào dữ liệu pháp lý trong lĩnh vực giao thông đường bộ. Nguồn dữ liệu đầu vào bao gồm 3 văn bản pháp luật mới nhất và quan trọng nhất hiện nay, được thu thập trực tiếp từ Cổng thông tin quốc gia *vbpl.vn*:

1. Luật Đường bộ (Luật số 35/2024/QH15).
2. Luật Trật tự, an toàn giao thông đường bộ (Luật số 36/2024/QH15).
3. Nghị định 168/2024/NĐ-CP quy định về xử phạt vi phạm hành chính.

Dữ liệu thô sau khi thu thập sẽ trải qua bước làm sạch (cleaning) để loại bỏ các ký tự nhiễu, chuẩn hóa định dạng văn bản trước khi đưa vào quy trình phân đoạn.

4.2.2 Chiến lược Phân đoạn (Chunking Strategy)

Để giải quyết bài toán cấu trúc phân cấp của văn bản luật, hệ thống RoadLawQA áp dụng chiến lược **Hierarchical Leaf-Only Chunking with Context Enrichment**.

Cụ thể, văn bản luật được phân tích và biểu diễn dưới dạng một cây phân cấp nhiều cấp. Hệ thống chỉ thực hiện đánh chỉ mục các **nút lá (leaf nodes)**, tương ứng với các đơn vị nhỏ nhất như gạch đầu dòng (bullet), điểm hoặc khoản (trong trường hợp không tồn tại cấp con). Thay vì lưu trữ toàn bộ các node trong cây (gây dư thừa dữ liệu), cách tiếp cận này giúp tối ưu hóa không gian lưu trữ và tránh trùng lặp thông tin so với việc đánh chỉ mục cả các node cha.

Tuy nhiên, nút lá khi đứng độc lập thường thiếu ngữ cảnh (ví dụ: "phạt tiền từ..." mà không biết hành vi nào). Để giải quyết vấn đề này, trước khi thực hiện mã hóa vector và đánh chỉ mục, nội dung của các cấp cha được bổ sung ngược vào mỗi chunk thông qua các tag cấu trúc đặc biệt:

```
[CHAPTER] ...  [ARTICLE] ...  [CLAUSE] ...  [POINT] ...  [Nội dung nút lá]
```

Phiên bản văn bản sau khi làm giàu ngữ cảnh này được sử dụng làm đầu vào cho mô hình Embedding và được lưu trữ trong Vector Database. Nhờ đó, mỗi đơn vị truy xuất (chunk) có kích thước nhỏ gọn nhưng vẫn bảo toàn được ý nghĩa pháp lý đầy đủ, giúp mô hình Embedding và Reranker hiểu chính xác ngữ cảnh.

Đối với các trường hợp nút lá có nội dung quá dài (vượt quá 1500 tokens), hệ thống áp dụng kỹ thuật *sliding window* với kích thước cửa sổ 900 tokens và độ chồng lấp (overlap) 300 tokens. Chiến lược này giúp cân bằng giữa độ chi tiết của thông tin, hiệu quả truy xuất và độ chính xác của câu trả lời được sinh ra.

4.2.3 Mô hình Embedding và Vector Database

Để biểu diễn văn bản pháp luật trong không gian vector ngữ nghĩa, nhóm em tiến hành khảo sát và so sánh thực nghiệm một số mô hình embedding đa ngôn ngữ phổ biến. Việc đánh giá được thực hiện trên tập câu hỏi luật giao thông, sử dụng các độ đo truy xuất như Recall@5 và MRR@5. Dựa trên kết quả thực nghiệm (trình bày chi tiết trong Phần Kết quả đánh giá), mô hình `Alibaba-NLP/gte-multilingual-base` được lựa chọn cho hệ thống RoadLawQA.

Mô hình này đạt hiệu năng vượt trội trên tập dữ liệu đánh giá, với Recall@5 đạt **0.880** và MRR@5 đạt **0.760**, cao hơn so với các mô hình được khảo sát khác. Ngoài ra, với kích thước embedding ở mức vừa phải, mô hình đảm bảo cân bằng tốt giữa chất lượng biểu diễn ngữ nghĩa và hiệu năng truy xuất, phù hợp cho triển khai hệ thống hỏi đáp pháp luật.

Vector embedding được tính toán trên **phiên bản chunk đã được làm giàu ngữ cảnh pháp lý**, thay vì chỉ sử dụng nội dung nút lá đơn lẻ. Nhờ đó, embedding không chỉ phản ánh nội dung của quy định cụ thể mà còn mã hóa được ngữ cảnh phân cấp (chương, điều, khoản, điểm) của đoạn văn bản trong toàn bộ văn bản luật.

Các vector embedding cùng với metadata pháp lý tương ứng được lưu trữ và quản lý trong **Weaviate**, đóng vai trò là Vector Database trung tâm của hệ thống. Weaviate hỗ trợ hiệu quả cho việc tìm kiếm tương đồng và truy xuất thông tin dựa trên vector, đồng thời cho phép lọc theo metadata pháp lý nhằm phục vụ việc trích dẫn và hiển thị nguồn luật trong các bước xử lý tiếp theo.

4.3 Chiến lược Truy xuất và Xếp hạng

Để truy xuất chính xác các điều khoản pháp luật dựa trên câu hỏi ngôn ngữ tự nhiên của người dùng, hệ thống RoadLawQA áp dụng kiến trúc truy xuất hai giai đoạn (two-stage retrieval). Giai đoạn đầu tiên tập trung vào việc sinh tập ứng viên bằng chiến lược truy xuất hỗn hợp, trong khi giai đoạn thứ hai thực hiện xếp hạng lại nhằm tinh chỉnh độ liên quan của kết quả.

4.3.1 Truy xuất hỗn hợp (Hybrid Retrieval)

Tìm kiếm vector thuần túy (dense retrieval) có thể bỏ sót các từ khóa mang tính định danh quan trọng trong văn bản luật, chẳng hạn như số điều, mức xử phạt hoặc các giá trị định lượng cụ thể. Do đó, hệ thống RoadLawQA kết hợp hai phương pháp truy xuất bổ trợ cho nhau:

- **Sparse Retrieval (BM25):** Giúp bắt chính xác các từ khóa và thuật ngữ pháp lý mang tính định danh trong văn bản luật.

- **Dense Retrieval (Vector Search):** Khai thác biểu diễn ngữ nghĩa của câu hỏi và văn bản luật để xử lý các cách diễn đạt đa dạng của người dùng.

Trong hệ thống, thành phần sparse retrieval được xây dựng dưới dạng một chỉ mục BM25 riêng biệt, sử dụng cơ chế tokenization tiếng Việt nhằm đảm bảo xử lý chính xác các từ ghép và thuật ngữ pháp lý. Song song đó, vector database được sử dụng để thực hiện truy xuất ngữ nghĩa. Hai tập kết quả được kết hợp để tạo ra danh sách ứng viên ban đầu cho bước xếp hạng tiếp theo.

Chiến lược truy xuất hỗn hợp giúp cải thiện đáng kể độ bao phủ thông tin, với chỉ số Recall@5 đạt **0.900**, cao hơn so với việc chỉ sử dụng một phương pháp truy xuất đơn lẻ.

4.3.2 Xếp hạng lại (Reranking)

Danh sách ứng viên thu được từ bước truy xuất ban đầu vẫn có thể chứa các đoạn văn bản chưa thực sự phù hợp với ngữ cảnh câu hỏi. Để tinh chỉnh độ chính xác, hệ thống áp dụng bước xếp hạng lại (reranking) bằng mô hình cross-encoder **BAAI/bge-reranker-v2-m3**.

Mô hình reranker đánh giá trực tiếp mức độ phù hợp ngữ nghĩa giữa câu hỏi và từng đoạn văn bản, từ đó sắp xếp lại danh sách kết quả và ưu tiên các điều khoản liên quan nhất. Việc tích hợp bước reranking giúp hệ thống đạt hiệu năng cao hơn, với Recall@5 đạt **0.920** và nDCG@5 đạt **0.827**.

4.4 Sinh nội dung dựa trên truy xuất

Sau khi hoàn tất quá trình truy xuất và xếp hạng, hệ thống RoadLawQA tiến hành sinh nội dung đầu ra dựa trên tập các đoạn văn bản luật liên quan nhất. Giai đoạn này nhằm chuyển đổi thông tin pháp lý được truy xuất thành câu trả lời mạch lạc, dễ hiểu đối với người dùng, đồng thời đảm bảo tính chính xác và khả năng kiểm chứng nguồn luật.

4.4.1 Sinh câu trả lời (Answer Generation)

Hệ thống RoadLawQA sử dụng **Gemini API (gemini-2.5-flash)** làm mô hình ngôn ngữ nền tảng cho tác vụ sinh câu trả lời. Đầu vào của mô hình bao gồm câu hỏi của người dùng và danh sách Top-k đoạn văn bản luật liên quan nhất (Context) đã được truy xuất và sắp xếp lại ở bước trước.

Để đảm bảo tính chính xác pháp lý (legal accuracy) và hạn chế hiện tượng sinh thông tin sai lệch, hệ thống áp dụng kỹ thuật **prompt engineering** với các ràng buộc rõ ràng. Cụ thể, mô hình được hướng dẫn đóng vai trò như một trợ lý pháp lý trung lập, chỉ tổng hợp và diễn giải thông tin dựa trên các đoạn văn bản luật được cung cấp trong Context, thay vì suy đoán hoặc bổ sung kiến thức ngoài phạm vi ngữ cảnh.

Bên cạnh đó, câu trả lời sinh ra bắt buộc phải đi kèm với các trích dẫn pháp lý cụ thể (ví dụ: điều, khoản và văn bản luật liên quan), được lấy trực tiếp từ metadata của các đoạn văn bản được sử dụng. Chiến lược này giúp nâng cao tính minh bạch và khả năng kiểm chứng của hệ thống, đồng thời góp phần cải thiện độ chính xác pháp lý.

4.4.2 Sinh câu hỏi trắc nghiệm (Quiz Generation)

Bên cạnh tính năng hỏi đáp, hệ thống còn tích hợp module sinh câu hỏi trắc nghiệm tự động nhằm hỗ trợ người dùng chủ động ôn tập và kiểm tra kiến thức pháp luật.Điểm đặc biệt của module này là khả năng cá nhân hóa trải nghiệm học tập thông qua hai chế độ sinh câu hỏi linh hoạt, cho phép người dùng tùy chọn số lượng câu hỏi mong muốn (ví dụ: 5, 10, hoặc 20 câu) trong mỗi lần kiểm tra.

1. **Chế độ Sinh ngẫu nhiên (Random Mode):** Hệ thống thực hiện lấy mẫu ngẫu nhiên (random sampling) các đoạn văn bản luật (chunks) từ cơ sở dữ liệu vector. Chế độ này giúp người dùng ôn tập kiến thức tổng quát, bao phủ nhiều khía cạnh khác nhau của Luật giao thông mà không bị giới hạn trong một chủ đề cụ thể.
2. **Chế độ Sinh theo chủ đề (Topic-based Mode):** Người dùng nhập một chủ đề cụ thể vào ô tìm kiếm (ví dụ: *"nồng độ cồn"*, *"đi ngược chiều"*). Hệ thống sẽ tái sử dụng quy trình *Retrieval* (như mô tả ở phần 3.3) để truy xuất các đoạn văn bản luật liên quan nhất đến từ khóa đó. Sau đó, các ngữ cảnh này được đưa vào mô hình để sinh ra bộ câu hỏi chuyên sâu xoay quanh chủ đề người dùng quan tâm.

Để đảm bảo giá trị giáo dục, đầu ra của module không chỉ là đáp án đúng/sai đơn thuần mà được cấu trúc với đầy đủ các thành phần:

- **Câu hỏi (Question):** Nội dung câu hỏi tình huống hoặc lý thuyết.
- **Các phương án (Options):** 4 lựa chọn (A, B, C, D) gây nhiễu hợp lý.
- **Đáp án đúng (Key):** Xác định phương án chính xác.
- **Giải thích (Explanation):** Lập luận chi tiết tại sao đáp án đó là đúng, giúp người dùng hiểu rõ bản chất vấn đề.
- **Trích dẫn nguồn (Citation):** Cung cấp căn cứ pháp lý cụ thể (Điều, Khoản, Văn bản luật) để người dùng có thể đối chiếu, đảm bảo tính minh bạch và chính xác của kiến thức.

5 Dữ liệu và Độ đo đánh giá

Để đánh giá hiệu quả của hệ thống RoadLawQA một cách khách quan và định lượng, chúng em đã thiết lập một quy trình kiểm thử nghiêm ngặt bao gồm việc xây dựng bộ dữ liệu chuẩn (Gold Standard) và áp dụng hệ thống đa độ đo cho từng thành phần của kiến trúc RAG.

5.1 Bộ dữ liệu đánh giá

Trong bối cảnh bài toán hỏi đáp luật giao thông tại Việt Nam hiện chưa có bộ dữ liệu chuẩn công khai (benchmark dataset), nhóm nghiên cứu đã tự xây dựng một tập dữ liệu kiểm thử (test set) chất lượng cao. Dữ liệu được trích xuất và xác minh thủ công dựa trên 3 văn bản gốc: Luật Đường bộ (35/2024/QH15), Luật Trật tự, an toàn giao thông đường bộ (36/2024/QH15) và Nghị định 168/2024/NĐ-CP.

5.1.1 Quy mô và Cấu trúc

Bộ dữ liệu bao gồm **50 cặp câu hỏi - câu trả lời (QA pairs)**. Mỗi cặp dữ liệu đều đi kèm với nhãn chuẩn (Ground Truth), bao gồm câu trả lời đúng và trích dẫn pháp lý chính xác (số hiệu Điều, Khoản, Điểm).

Nhóm công khai toàn bộ tập dữ liệu đánh giá này tại đây.

Loại câu hỏi	**Câu hỏi minh họa**
Tra cứu sự thật	Khi người điều khiển giao thông (Cảnh sát giao thông) giơ tay thẳng đứng, hiệu lệnh này có ý nghĩa gì đối với người tham gia giao thông?
Tra cứu chế tài	Người đi bộ vào đường cao tốc bị phạt bao nhiêu tiền?
Tình huống/Suy luận	Hôm qua tôi lái xe ô tô đi làm về lúc 6 rưỡi tối (18h30). Trời mùa hè vẫn còn sáng nên tôi quên không bật đèn xe. Trường hợp này tôi có bị Cảnh sát giao thông phạt không?

Bảng 1: Minh họa các loại câu hỏi trong bộ dữ liệu đánh giá

5.1.2 Phân loại câu hỏi

Để đánh giá toàn diện khả năng của mô hình ở các cấp độ nhận thức khác nhau, bộ dữ liệu được phân chia thành 3 loại hình câu hỏi đặc trưng (như minh họa tại Bảng 1):

- **Tra cứu sự thật:** Kiểm tra khả năng truy hồi kiến thức về định nghĩa, quy định chung. Ví dụ: *"Hiệu lệnh giơ tay thẳng đứng của Cảnh sát giao thông có ý nghĩa gì?"*.
- **Tra cứu chế tài:** Đây là loại câu hỏi phổ biến nhất trong thực tế, yêu cầu hệ thống phải truy xuất chính xác con số (mức phạt tiền) hoặc hình thức xử phạt bổ sung (tước giấy phép lái xe). Ví dụ: *"Người đi bộ vào đường cao tốc bị phạt bao nhiêu tiền?"*.
- **Tình huống và Suy luận:** Đây là nhóm câu hỏi có độ khó cao nhất, yêu cầu mô hình phải hiểu ngữ cảnh tình huống, kết nối nhiều điều khoản luật khác nhau để đưa ra kết luận đúng/sai hoặc tư vấn hướng xử lý.

5.2 Các độ đo đánh giá

Chúng em áp dụng các nhóm độ đo riêng biệt cho từng module để phân tích chi tiết hiệu năng của hệ thống.

5.2.1 Độ đo đánh giá truy xuất (Retrieval Metrics)

Module Truy xuất (Retriever) đóng vai trò quyết định trong việc tìm kiếm đúng văn bản luật. Chúng em sử dụng các chỉ số phổ biến trong lĩnh vực Information Retrieval tại $k = 5$ (Top-5 kết quả trả về):

- **Recall@5:** Tỷ lệ các văn bản luật liên quan được truy xuất trong top-5 kết quả, phản ánh mức độ bao phủ thông tin của hệ thống.
- **HitRate@5:** Tỷ lệ các câu hỏi mà hệ thống truy xuất được ít nhất một văn bản luật liên quan trong top-5 kết quả.
- **MRR@5:** Đo lường vị trí xuất hiện của văn bản luật liên quan đầu tiên trong danh sách kết quả; giá trị càng cao cho thấy kết quả đúng được xếp càng sớm.
- **nDCG@5:** Đánh giá chất lượng xếp hạng tổng thể bằng cách xét đồng thời mức độ liên quan và vị trí của các văn bản trong danh sách; đặc biệt phù hợp để đánh giá hiệu quả của bước reranking.

5.2.2 Độ đo đánh giá sinh câu trả lời (Answer Generation Metrics)

Các độ đo truyền thống trong NLP như BLEU hay ROUGE thường dựa trên sự trùng khớp từ ngữ (n-gram overlap), không phản ánh đúng tính chính xác về mặt pháp lý. Do đó, chúng em định nghĩa 3 độ đo chuyên biệt được đánh giá thủ công:

- **Legal Accuracy (Độ chính xác pháp lý):** Đánh giá mức độ đúng đắn của nội dung câu trả lời so với quy định pháp luật và ground truth. Độ đo này nhằm kiểm soát hiện tượng sinh thông tin sai lệch (hallucination).
- **Legal Citation (Độ chính xác trích dẫn):** Đánh giá tính chính xác của các trích dẫn pháp lý (điều, khoản, văn bản luật) được đưa ra trong câu trả lời. Một câu trả lời có nội dung đúng nhưng trích dẫn sai vẫn bị xem là không đạt yêu cầu trong bối cảnh pháp lý.
- **Relevance (Độ liên quan):** Đánh giá mức độ bám sát trọng tâm câu hỏi của câu trả lời, đảm bảo hệ thống không trả lời lan man hoặc lạc đề.

Mỗi tiêu chí được chấm theo thang điểm ba mức $\{0, 0.5, 1\}$, trong đó: 0 tương ứng với không đạt yêu cầu, 0.5 tương ứng với đạt một phần, và 1 tương ứng với đạt đầy đủ. Điểm cuối cùng cho mỗi câu trả lời được tính bằng giá trị trung bình của các tiêu chí trên.

5.2.3 Độ đo đánh giá sinh câu hỏi trắc nghiệm (Quiz Generation Metrics)

Đối với module sinh câu hỏi trắc nghiệm, ta sẽ đánh giá bằng 3 metrics sau:

- **Accuracy:** Sử dụng phương pháp *LLM-as-a-Judge*. Tiêu chí này đánh giá tính đúng đắn của đáp án đúng so với ngữ cảnh pháp lý đầu vào. Một mô hình ngôn ngữ lớn độc lập đóng vai trò giám khảo, nhận đầu vào gồm *Context* (ngữ cảnh pháp lý) và *Quiz* (câu hỏi + đáp án đúng) do hệ thống sinh ra để chấm điểm. Kết quả được tính theo thang nhị phân: điểm 1 nếu đáp án hoàn toàn phù hợp với quy định pháp luật trong Context; điểm 0 nếu đáp án sai lệch, bịa đặt hoặc không có căn cứ.
- **Topic Alignment:** Sử dụng phương pháp *LLM-as-a-Judge*. Tiêu chí này đánh giá mức độ bám sát chủ đề trong chế độ sinh câu hỏi theo chủ đề. Một mô hình ngôn ngữ lớn được sử dụng làm giám khảo, nhận đầu vào gồm *Topic* (chủ đề do người dùng nhập) và *Question*

(câu hỏi trắc nghiệm được sinh ra). Kết quả được chấm theo thang nhị phân: điểm 1 nếu nội dung câu hỏi phù hợp và nhất quán với chủ đề đã cho; điểm 0 nếu câu hỏi lạc đề hoặc chỉ liên quan một cách mơ hồ.

- **Format Compliance:** Đánh giá khả năng tích hợp của đầu ra thông qua các quy tắc kiểm tra tự động (Rule-based). Một mẫu câu hỏi được tính là đạt (điểm 1) nếu tuân thủ tuyệt đối cấu trúc JSON yêu cầu, bao gồm đầy đủ các trường thông tin bắt buộc: `question`, `options`, `correct_answer`. Ngược lại, điểm 0 được gán nếu sai định dạng hoặc thiếu trường dữ liệu.

6 Kết quả đánh giá

Thực nghiệm được tiến hành nhằm mục đích chọn lựa các thành phần tối ưu cho kiến trúc RAG, đồng thời đánh giá hiệu năng tổng thể của hệ thống RoadLawQA trên bộ dữ liệu kiểm thử đã xây dựng.

6.1 So sánh hiệu năng mô hình Embedding

Trong kiến trúc RAG, khả năng biểu diễn ngữ nghĩa của mô hình Embedding ảnh hưởng trực tiếp đến chất lượng tìm kiếm. Chúng em đã tiến hành so sánh 4 mô hình embedding phổ biến hiện nay bao gồm: *BAAI/bge-m3*, *Alibaba-NLP/gte-multilingual-base*, *hiieu/halong_ embedding* và *intfloat/multilingual-e5-large*.

Kết quả thực nghiệm, được trình bày trong Bảng 2, cho thấy mô hình **Alibaba-NLP/gte-multilingual-base** đạt hiệu năng tốt nhất trên hầu hết các độ đo truy xuất quan trọng. Cụ thể, mô hình này đạt Recall@5 = **0.762** và HitRate@5 = **0.880**, cho thấy hệ thống có khả năng truy xuất được ít nhất một văn bản luật liên quan trong phần lớn các trường hợp. Đồng thời, chỉ số nDCG@5 đạt **0.788**, cao hơn so với các mô hình còn lại, phản ánh chất lượng xếp hạng tốt hơn của các văn bản liên quan trong danh sách kết quả.

Đối với các mô hình khác, *BAAI/bge-m3* và *intfloat/multilingual-e5-large* cho hiệu năng ở mức khá nhưng chưa vượt trội trên tất cả các độ đo. Ngoài ra, mô hình *hiieu/halong_ embedding*, dù được tinh chỉnh cho tiếng Việt, cho kết quả thấp hơn trên tập dữ liệu pháp lý. Điều này cho thấy các mô hình embedding chuyên biệt có thể gặp hạn chế trong việc biểu diễn các thuật ngữ pháp lý phức tạp so với các mô hình đa ngôn ngữ quy mô lớn.

Dựa trên kết quả này, nhóm quyết định lựa chọn *Alibaba-NLP/gte-multilingual-base* làm mô hình embedding chính cho hệ thống.

Model	Recall@5	MRR@5	nDCG@5	HitRate@5
BAAI/bge-m3	0.740	0.720	0.746	0.840
Alibaba-NLP/gte-multilingual-base	**0.762**	**0.760**	**0.788**	**0.880**
hiieu/halong_embedding	0.658	0.630	0.664	0.780
intfloat/multilingual-e5-large	0.710	0.668	0.704	0.820

Bảng 2: Bảng so sánh hiệu năng giữa các mô hình Embedding

6.2 Nghiên cứu các chiến lược truy xuất

Để chứng minh hiệu quả của kiến trúc đề xuất (Hybrid Search kết hợp Reranking), chúng tôi thực hiện so sánh 4 chiến lược truy xuất khác nhau. Kết quả tại Bảng 3 cho thấy sự cải thiện rõ rệt qua từng bước nâng cấp:

- **BM25-only:** Cho hiệu năng thấp nhất với Recall@5 = 0.707, cho thấy hạn chế của truy xuất dựa thuần túy trên từ khóa trong bối cảnh câu hỏi ngôn ngữ tự nhiên.
- **Dense-only:** Truy xuất ngữ nghĩa giúp cải thiện Recall@5 lên 0.762 và nâng cao chất lượng xếp hạng, phản ánh khả năng nắm bắt ngữ cảnh tốt hơn so với BM25.
- **Hybrid:** Việc kết hợp BM25 và Dense Search giúp tăng đáng kể độ bao phủ thông tin, với Recall@5 đạt 0.817 và HitRate@5 đạt 0.900.
- **Hybrid + Reranker (bge-reranker-v2-m3):** Bước reranking không cải thiện Recall@5 nhưng cải thiện rõ rệt chất lượng xếp hạng, với MRR@5 và nDCG@5 đạt giá trị cao nhất. Trong đó, mô hình *bge-reranker-v2-m3* cho hiệu năng ổn định và vượt trội so với các phương án reranker khác.

Phương pháp	**Recall@5**	**MRR@5**	**nDCG@5**	**HitRate@5**
BM25-only	0.707	0.538	0.598	0.780
Dense-only	0.762	0.760	0.791	0.880
Hybrid	**0.817**	0.752	0.789	0.900
Hybrid + Reranker (bge-reranker-v2-m3)	**0.817**	**0.796**	**0.827**	**0.920**
Hybrid + Reranker (gte-multilingual-reranker)	0.727	0.703	0.748	0.880

Bảng 3: Kết quả so sánh các chiến lược truy xuất (Ablation Study)

6.3 Đánh giá chất lượng sinh câu trả lời (End-to-End)

Chất lượng sinh câu trả lời của hệ thống được đánh giá trên tập 50 câu hỏi kiểm thử. Kết quả tổng thể cho thấy hệ thống hoạt động ổn định với điểm trung bình **Legal Accuracy đạt 0.83**, **Legal Citation đạt 0.77** và **Relevance đạt 0.86**.

Phân tích chi tiết theo từng loại câu hỏi (Bảng 4):

- **Tra cứu sự thật:** Đạt độ chính xác cao nhất (**0.97**), chứng tỏ hệ thống rất mạnh trong việc truy xuất định nghĩa và quy định rõ ràng.
- **Tra cứu chế tài:** Độ chính xác ở mức khá tốt (**0.87**), đáp ứng tốt nhu cầu tra cứu mức phạt của người dùng.
- **Tình huống/Suy luận:** Độ chính xác có điểm số thấp nhất (**0.70**). Nguyên nhân là do các câu hỏi tình huống đòi hỏi khả năng suy luận logic đa bước và kết nối nhiều điều khoản luật, đây vẫn là thách thức chung của các hệ thống LLM hiện nay.

Loại câu hỏi	Số câu	Legal Accuracy	Legal Citation	Relevance
Tra cứu sự thật	15	0.97	0.87	0.93
Tra cứu chế tài	15	0.87	0.80	0.87
Tình huống/Suy luận	20	0.70	0.68	0.80
Tất cả (Overall)	**50**	**0.83**	**0.77**	**0.86**

Bảng 4: Kết quả đánh giá chất lượng sinh câu trả lời theo từng loại câu hỏi

6.4 Đánh giá chất dựa trên bộ câu hỏi lý thuyết sát hạch lái xe

Ngoài ra, để đánh giá tổng quan hơn về chất lượng của mô hình, nhóm nghiên cứu đã sử dụng 80 câu hỏi trắc nghiệm được trích từ bộ 200 câu hỏi lý thuyết sát hạch lái xe hạng A1 2025 và sử dụng độ đo accuracy để quan sát kết quả. Kết quả cho thấy hệ thống trả lời rất tốt với **Accuracy đạt 0.825**

6.5 Đánh giá sinh câu hỏi trắc nghiệm (Quiz Generation)

Kết quả đánh giá cho thấy module sinh câu hỏi trắc nghiệm (cho cả loại câu hỏi ngẫu nhiên và theo chủ đề) đều đạt **Accuracy = 1.0** (đánh giá bằng LLM-as-a-judge) và **Format Compliance = 1.0** (đánh giá rule-based) trên các kịch bản sinh quiz. Đối với chế độ sinh theo chủ đề, hệ thống cũng đạt **Topic Alignment = 1.0**, cho thấy các câu hỏi được tạo bám sát chủ đề do người dùng nhập vào. Các kết quả này cho thấy module sinh quiz đáp ứng tốt yêu cầu về tính đúng đắn pháp lý, mức độ liên quan theo chủ đề và định dạng đầu ra.

7 Kết luận

Dựa trên những hạn chế của các phiên bản thử nghiệm trước và yêu cầu thực tế của bài toán tra cứu pháp luật, nhóm nghiên cứu đã thực hiện các cải tiến quan trọng về cả tính năng, dữ liệu và hiệu năng hệ thống.

7.1 Mở rộng tính năng: Từ tra cứu thụ động sang học tập chủ động

Trong các phiên bản trước, hệ thống chỉ dừng lại ở việc trả lời câu hỏi (QA) - một hình thức tra cứu thụ động. Nhận thấy nhu cầu ôn thi giấy phép lái xe và tìm hiểu luật của người dùng là rất lớn, chúng em đã phát triển thêm module **Quiz Generation (Tạo đề trắc nghiệm)**, người dùng có thể tạo ngẫu nhiên hoặc tạo theo chủ đề mà người dùng nhập.

Cải tiến này chuyển đổi vai trò của hệ thống từ một "công cụ tìm kiếm" thành một "trợ lý học tập". Hệ thống có khả năng tự động phân tích văn bản luật để sinh ra các câu hỏi trắc nghiệm khách quan với cấu trúc chuẩn (Câu hỏi, 4 phương án, Đáp án đúng, Giải thích). Điều này giúp người dùng chủ động kiểm tra kiến thức và ghi nhớ luật hiệu quả hơn.

7.2 Nâng cấp và Chuẩn hóa Bộ dữ liệu đánh giá

Nhận thấy tầm quan trọng của dữ liệu kiểm thử đối với độ tin cậy của mô hình, nhóm đã nỗ lực mở rộng quy mô bộ dữ liệu đánh giá (Evaluation Dataset) chuyên biệt cho domain luật giao thông Việt Nam.

- **Tăng quy mô:** Số lượng cặp câu hỏi - đáp án (QA pairs) được tăng từ 30 lên **50 mẫu**. Việc tăng kích thước mẫu giúp các chỉ số đánh giá có ý nghĩa thống kê cao hơn và phản ánh chính xác hơn hiệu năng thực tế.
- **Tăng độ phủ và độ phân hóa:** Bộ dữ liệu mới bao phủ rộng hơn các tình huống pháp lý và được phân hóa rõ rệt thành 3 cấp độ nhận thức: Tra cứu sự thật (dễ), Tra cứu chế tài (trung bình) và Tình huống/suy luận (khó).