

Chatbot Tra cứu Luật Giao thông Đường bộ Việt Nam

Đồ án môn học CS431.P22

Thực hiện bởi:

Lê Minh Nhựt
Trần Đình Khánh Đăng
Thái Ngọc Quân

Giảng viên hướng dẫn: TS. Nguyễn Vinh Tiệp

¹Khoa Khoa học Máy tính
Trường Đại học Công nghệ Thông tin - ĐHQG TP.HCM

Báo cáo Cuối kỳ – Tháng 5, 2025

Mục lục Tổng quát

- 1 Giới thiệu
- 2 Dữ liệu
- 3 Quy trình xử lý
- 4 Độ đo đánh Giá

Nội dung

1 Giới thiệu

2 Dữ liệu

3 Quy trình xử lý

4 Độ đo đánh Giá

Giới thiệu: Thực trạng giao thông

- Theo **CSGT - Bộ Công an**, năm **2024**, toàn quốc xảy ra **21.532 vụ tai nạn giao thông**, làm chết **9.954 người**, bị thương **16.044 người**. Có **23 vụ đặc biệt nghiêm trọng**, khiến **76 người chết**, **17 người bị thương**.
- Trong đó:
 - **3.065 vụ** do **đi sai chiều/phần/làn đường**, gây **1.423 người chết**, **2.764 người bị thương**.
 - **360 vụ** do **vượt đèn đỏ**, làm **122 người chết**, **301 người bị thương**.
 - **143 vụ** do **đi ngược chiều**, khiến **38 người chết**, **141 người bị thương**.
- Nguyên nhân chủ yếu là do **vi phạm quy tắc giao thông, bất chấp pháp luật**, dẫn đến **tai nạn và cái chết thương tâm không đáng có**.

Giới thiệu: Luật giao thông đã nghiêm khắc hơn

Luật GTDB Việt Nam hiện nay

- Trong bối cảnh này, **pháp luật** càng phải thể hiện **vai trò quan trọng**, ảnh hưởng trực tiếp đến **an toàn & trật tự giao thông đường bộ**.
- Hàng loạt **văn bản pháp luật đường bộ mới**, đặc biệt là **Nghị định 168/2024/NĐ-CP** với các **mức phạt tăng lên** cho các hành vi vi phạm, vừa có hiệu lực **đầu năm 2025**.
- Mục tiêu của việc **cải cách luật giao thông** là nhằm **tăng tính răn đe** cho những người hay vi phạm giao thông, nhằm **giảm tải tối đa tai nạn giao thông**.

Giới thiệu: Ví dụ

CHÍNH PHỦ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Số: 168/2024/NĐ-CP

Hà Nội, ngày 26 tháng 12 năm 2024

NGHỊ ĐỊNH

QUY ĐỊNH XỬ PHẠT VI PHẠM HÀNH CHÍNH VỀ TRẬT TỰ, AN TOÀN GIAO THÔNG TRONG
LĨNH VỰC GIAO THÔNG ĐƯỜNG BỘ; TRỪ ĐIỂM, PHỤC HỒI ĐIỂM GIẤY PHÉP LÁI XE

Căn cứ Luật Tổ chức Chính phủ ngày 19 tháng 6 năm 2015; Luật sửa đổi, bổ sung một số điều của
Luật Tổ chức Chính phủ và Luật Tổ chức chính quyền địa phương ngày 22 tháng 11 năm 2019;

Căn cứ Luật Xử lý vi phạm hành chính ngày 20 tháng 6 năm 2012; Luật sửa đổi, bổ sung một số
điều của Luật Xử lý vi phạm hành chính ngày 15 tháng 11 năm 2020;

Căn cứ Luật Trật tự, an toàn giao thông đường bộ ngày 27 tháng 6 năm 2024;

Theo đề nghị của Bộ trưởng Bộ Công an;

Chính phủ ban hành Nghị định quy định xử phạt vi phạm hành chính về trật tự, an toàn giao thông
trong lĩnh vực giao thông đường bộ; trừ điểm, phục hồi điểm giấy phép lái xe.

Chương I

NHỮNG QUY ĐỊNH CHUNG

Điều 1. Phạm vi điều chỉnh

1. Nghị định này quy định về:

a) Xử phạt vi phạm hành chính về trật tự, an toàn giao thông trong lĩnh vực giao thông đường bộ
bao gồm: hành vi vi phạm hành chính; hình thức, mức xử phạt, biện pháp khắc phục hậu quả đối
với từng hành vi vi phạm hành chính; thẩm quyền lập biên bản, thẩm quyền xử phạt, mức phạt tiền
cụ thể theo từng chức danh đối với hành vi vi phạm hành chính về trật tự, an toàn giao thông trong
lĩnh vực giao thông đường bộ;

b) Mức trừ điểm giấy phép lái xe đối với từng hành vi vi phạm hành chính; trình tự, thủ tục, thẩm
quyền trừ điểm, phục hồi điểm giấy phép lái xe để quản lý việc chấp hành pháp luật về trật tự, an
toàn giao thông đường bộ của người lái xe.

2. Các hành vi vi phạm hành chính trong lĩnh vực quản lý nhà nước khác liên quan đến trật tự, an
toàn giao thông trong lĩnh vực giao thông đường bộ mà không quy định tại Nghị định này thì áp
dụng quy định tại các Nghị định quy định về xử phạt vi phạm hành chính trong các lĩnh vực đó để
xử phạt.

Hình 1: Nghị định 168/2024/NĐ-CP.

Giới thiệu: Động lực và Giải pháp

Động lực Thực hiện

- Hệ thống pháp luật hiện nay **phức tạp**, gồm nhiều loại văn bản (**Luật, Nghị định, Thông tư...**), thường **dài dòng, khó tra cứu**.
- **Luật mới, tuyên truyền còn hạn chế**, nhiều người tham gia giao thông **chưa nắm rõ**, dẫn đến **vi phạm nghiêm trọng**.
- Một bộ phận người dân **thiếu ý thức**, nếu hiểu rõ **mức phạt và tính răn đe**, sẽ **tuân thủ tốt hơn**, góp phần **giảm tai nạn**.
- **Thiếu công cụ tra cứu tập trung**, người dân gặp khó khăn khi cần **giải đáp nhanh các thắc mắc**.

Giải pháp đề xuất: Chatbot Luật GTĐB

- Xây dựng **Chatbot thông minh**, chuyên biệt cho **Luật Giao thông Đường bộ Việt Nam**.
- Cung cấp **kênh tra cứu tập trung, nhanh chóng, thân thiện và dễ sử dụng**.

Mục tiêu Phát triển Chatbot

- **Trả lời câu hỏi:** Về quy tắc, mức phạt, thủ tục cơ bản, thông tin, biển báo giao thông (có hình ảnh minh họa...).
- **Đảm bảo chính xác:** Dựa trên văn bản pháp luật hiện hành, nguồn tin chính thống, đáng tin cậy.
- **Thân thiện người dùng:** Giao diện tương tác tự nhiên, đơn giản, phản hồi mạch lạc, dễ hiểu.
- **Truy xuất nguồn:** Chỉ rõ điều, khoản, điểm, văn bản gốc, kèm trích dẫn nguồn đầy đủ.

Giới thiệu: Ví dụ



Xe máy vượt đèn đỏ phải chịu các mức phạt nào?



Người điều khiển xe máy vượt đèn đỏ sẽ phải chịu các mức phạt sau:

- Trường hợp không gây tai nạn giao thông:
 - Phạt tiền từ **4.000.000 đồng đến 6.000.000 đồng**. (Theo Điều 7, Khoản 7, Điểm c, Nghị định số 168_2024_NĐ-CP)
 - Bị trừ **04 điểm** vào giấy phép lái xe. (Theo Điều 7, Khoản 13, Điểm b, Nghị định số 168_2024_NĐ-CP)
- Trường hợp gây tai nạn giao thông:
 - Phạt tiền từ **10.000.000 đồng đến 14.000.000 đồng**. (Theo Điều 7, Khoản 10, Điểm b, Nghị định số 168_2024_NĐ-CP)
 - Bị trừ **10 điểm** vào giấy phép lái xe. (Theo Điều 7, Khoản 13, Điểm d, Nghị định số 168_2024_NĐ-CP)

Nguồn tham khảo (Văn bản pháp luật):

- [Nghị định số 168/2024/NĐ-CP](#)

Nhập câu hỏi của bạn về Luật GTĐB...



Hình 2: Hỏi đáp về mức phạt vượt đèn đỏ của xe máy.

Nội dung

1 Giới thiệu

2 **Dữ liệu**

3 Quy trình xử lý

4 Độ đo đánh Giá

Dữ liệu: Tầm quan trọng

Tầm quan trọng của Dữ liệu

- **Nền tảng cốt lõi:** Chất lượng và độ tin cậy dữ liệu quyết định độ chính xác của Chatbot.
- Nếu văn bản pháp luật sai/lỗi thời → dẫn tới tư vấn sai, gây nhầm lẫn kiến thức pháp luật.

Nguồn Dữ liệu Chính

- Do đó **Thư viện Pháp Luật:** là nguồn chính để tiếp cận các văn bản pháp luật về giao thông đường bộ mới nhất và chính xác nhất.

Phạm vi Dữ liệu

- Lĩnh vực: **Trật tự, An toàn Giao thông Đường bộ.**
- Số lượng: **45 văn bản** chọn lọc (**Luật, Nghị định, Thông tư...**) mới nhất, đang có hiệu lực thi hành hiện nay. Và khoảng **400 biển báo giao thông đường bộ** các loại
VD: **Luật 36/2024/QH15, Nghị định 168/2024/NĐ-CP, ...**
- Cơ quan ban hành: **Quốc hội, Chính phủ, Bộ Công an, Bộ GTVT,...**

Dữ liệu: Cấu trúc văn bản Pháp luật

Chương II

QUY TẮC GIAO THÔNG ĐƯỜNG BỘ

Điều 10. Quy tắc chung

1. Người tham gia giao thông đường bộ phải đi bên phải theo chiều đi của mình, đi đúng làn đường, phần đường quy định, chấp hành báo hiệu đường bộ và các quy tắc giao thông đường bộ khác.
2. Người lái xe và người được chở trên xe ô tô phải thắt dây đai an toàn tại những chỗ có trang bị dây đai an toàn khi tham gia giao thông đường bộ.
3. Khi chở trẻ em dưới 10 tuổi và chiều cao dưới 1,35 mét trên xe ô tô không được cho trẻ em ngồi cùng hàng ghế với người lái xe, trừ loại xe ô tô chỉ có một hàng ghế; người lái xe phải sử dụng, hướng dẫn sử dụng thiết bị an toàn phù hợp cho trẻ em.

Điều 11. Chấp hành báo hiệu đường bộ

1. Báo hiệu đường bộ bao gồm: hiệu lệnh của người điều khiển giao thông; đèn tín hiệu giao thông; biển báo hiệu đường bộ; vạch kẻ đường và các dấu hiệu khác trên mặt đường; cọc tiêu, tường bảo vệ, rào chắn, đỉnh phản quang, tiêu phản quang, cột Km, cọc H; thiết bị âm thanh báo hiệu đường bộ.
2. Người tham gia giao thông đường bộ phải chấp hành báo hiệu đường bộ theo thứ tự ưu tiên từ trên xuống dưới như sau:
 - a) Hiệu lệnh của người điều khiển giao thông;
 - b) Tín hiệu đèn giao thông;
 - c) Biển báo hiệu đường bộ;
 - d) Vạch kẻ đường và các dấu hiệu khác trên mặt đường;
 - đ) Cọc tiêu, tường bảo vệ, rào chắn, đỉnh phản quang, tiêu phản quang, cột Km, cọc H;
 - e) Thiết bị âm thanh báo hiệu đường bộ.

Hình 3: Cấu trúc của 1 văn bản Pháp luật

Dữ liệu: Thu thập và làm sạch dữ liệu

- Nội dung của mỗi văn bản pháp luật sẽ được lưu vào file text **một cách thủ công bằng cách copy trực tiếp** từ các văn bản gốc, cách này có thể thực hiện rất dễ dàng.
- Tuy nhiên **có 2 thách thức cần giải quyết**, nhằm đáp ứng một bộ dữ liệu rõ ràng cho chatbot là:
 - **Phải xử lý các Điều, Khoản, Điểm của Luật được bổ sung, sửa đổi, thay thế giữa các văn bản liên quan nhau.**
 - **Phải xử lý các Điều, Khoản, Điểm đề cập tới Điều, Khoản, Điểm khác chứ không nêu cụ thể đối tượng, hành vi vi phạm, ...**

⇒ Cách hiệu quả nhất để giải quyết là **đọc văn bản và thực hiện thay đổi thủ công** để tạo ra các đoạn có nội dung đầy đủ hơn.

ĐIỀU KHOẢN THI HÀNH

Điều 52. Sửa đổi, bổ sung một số điều của Nghị định số 100/2019/NĐ-CP ngày 30 tháng 12 năm 2019 của Chính phủ quy định xử phạt vi phạm hành chính trong lĩnh vực giao thông đường bộ và đường sắt đã được sửa đổi, bổ sung một số điều theo Nghị định số 123/2021/NĐ-CP ngày 28 tháng 12 năm 2021 của Chính phủ sửa đổi, bổ sung một số điều của các Nghị định quy định xử phạt vi phạm hành chính trong lĩnh vực hàng hải; giao thông đường bộ, đường sắt; hàng không dân dụng

1. Bổ sung khoản 2a vào sau [khoản 2 Điều 1](#) như sau:

"2a. Hình thức, mức xử phạt, biện pháp khắc phục hậu quả đối với từng hành vi vi phạm hành chính; thẩm quyền lập biên bản, thẩm quyền xử phạt, mức phạt tiền cụ thể đối với từng chức danh về trật tự, an toàn giao thông trong lĩnh vực giao thông đường bộ thì áp dụng quy định tại Nghị định quy định xử phạt vi phạm hành chính về trật tự, an toàn giao thông trong lĩnh vực giao thông đường bộ; trừ điểm, phục hồi điểm giấy phép lái xe".

Hình 4: Minh họa Luật cần sửa đổi bổ sung

- Chúng ta phải dựa vào thông tin của Điều 52, Khoản 1 trong *Hình 4* để cập nhật lại nội dung của Nghị định số 100/2019/NĐ-CP (vẫn còn hiệu lực) nếu chưa có văn bản hợp nhất.

Dữ liệu: Thu thập và làm sạch dữ liệu

- Ta phải tạo 1 đoạn mới cho Khoản 12, Điểm a (*Hình 5a*) dựa trên thông tin của Khoản 2, Điểm d (*Hình 5b*) như sau: a) Thực hiện hành vi **xe không được quyền ưu tiên lắp đặt sử dụng thiết bị phát tín hiệu của xe được quyền ưu tiên** còn bị tịch thu thiết bị phát tín hiệu ưu tiên lắp đặt, sử dụng trái quy định.

12. Ngoài việc bị áp dụng hình thức xử phạt chính, người điều khiển xe thực hiện hành vi vi phạm còn bị áp dụng các hình thức xử phạt bổ sung sau đây:

a) Thực hiện hành vi quy định tại điểm đ khoản 2 Điều này còn bị tịch thu thiết bị phát tín hiệu ưu tiên lắp đặt, sử dụng trái quy định;

b) Thực hiện hành vi quy định tại điểm a, điểm b, điểm h, điểm i, điểm k khoản 9 Điều này bị tước quyền sử dụng giấy phép lái xe từ 10 tháng đến 12 tháng;

Hình 5a: Minh họa khi Luật chỉ trích dẫn Điều, Khoản, Điểm khác

2. Phạt tiền từ 400.000 đồng đến 600.000 đồng đối với người điều khiển xe thực hiện một trong các hành vi vi phạm sau đây:

a) Dừng xe, đỗ xe trên phần đường xe chạy ở đoạn đường ngoài đô thị nơi có lề đường;

b) Điều khiển xe chạy quá tốc độ quy định từ 05 km/h đến dưới 10 km/h;

c) Điều khiển xe chạy tốc độ thấp mà không đi bên phải phần đường xe chạy gây cản trở giao thông;

d) Dừng xe, đỗ xe ở lòng đường gây cản trở giao thông; tụ tập từ 03 xe trở lên ở lòng đường, trong hầm đường bộ; đỗ, để xe ở lòng đường, vỉa hè trái phép;

đ) Xe không được quyền ưu tiên lắp đặt, sử dụng thiết bị phát tín hiệu của xe được quyền ưu tiên;

Hình 5b: Thông tin đầy đủ nằm ở Khoản 2, Điểm đ trước đó

Quy trình Tiền xử lý

- Các file text đã được xử lý chuyển đổi thành file JSON. Trong đó:
 - **Chia văn bản pháp luật thành các đoạn nhỏ** (chunks) dựa trên Chương, Mục, Điều, Khoản.
 - **Trích xuất Metadata:** Gắn thông tin về số hiệu VB; ngày hiệu lực; URL; vị trí: Chương, Mục, Điều, Khoản, Điểm; tên ảnh biển báo giao thông (nếu có).
 - Cách lưu này không chỉ giúp giữ được ngữ cảnh trong văn bản pháp luật, mà còn giúp cho các đoạn không vượt quá kích thước tối đa mà các mô hình dùng để xử lý có thể nhận (512 token).

Dữ liệu: Cấu trúc của một đoạn (chunk)

```
{  
  "id": "161_2024_NĐ-CP_chunk_7",  
  "text": "Chương I QUY ĐỊNH CHUNG\nĐiều 1. Phạm vi điều  
chỉnh\n6. Đối với vận chuyển hàng hóa nguy hiểm ...",  
  "metadata": {  
    "source": "Nghị định số 161_2024_NĐ-CP",  
    "effective_date": "1-1-2025",  
    "url": "https://thuvienphapluat.vn/van-ban/...",  
    "context": {  
      "khoan": "6",  
      "dieu": "1",  
      "chuong": "I"  
    }  
  }  
}
```

Dữ liệu: Cấu trúc của một đoạn (chunk) (có biển báo)

```
{
  "id": "41_2024_BGTVT_chunk_510",
  "text": "Phụ lục B\nB.2 Biển số P.102 “Cấm đi ngược chiều”\nna) Để báo đường cấm các loại xe (cơ giới và thô sơ)...",
  "metadata": {
    "source": "Quy chuẩn kỹ thuật số 41_2024_BGTVT",
    "effective_date": "1-1-2025",
    "url": "https://luatvietnam.vn/giao-thong/q...",
    "context": {
      "khoan": "a",
      "dieu": "B.2",
      "phu_luc": "B"
    }
  },
  "traffic_sign": ["P_102.png"]
}
```

Nội dung

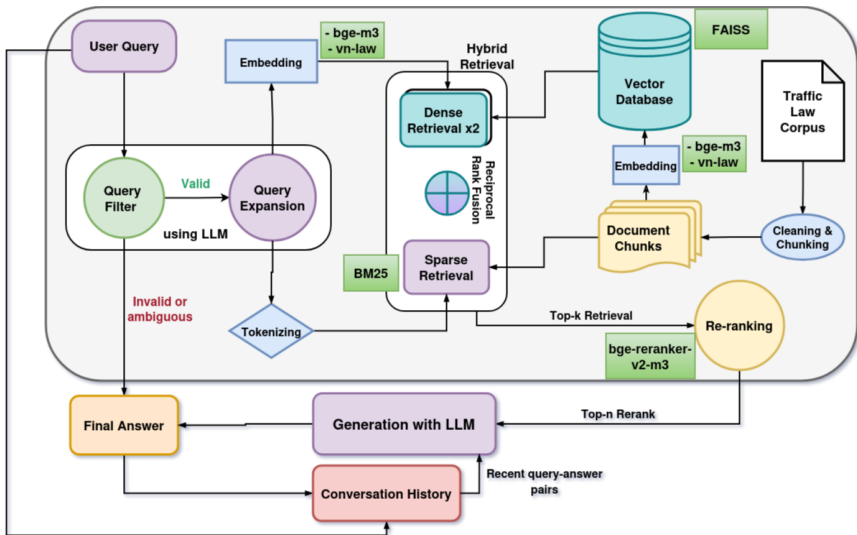
1 Giới thiệu

2 Dữ liệu

3 Quy trình xử lý

4 Độ đo đánh Giá

Quy trình: Pipeline



Hình 6: Pipeline chính của chatbot giao thông đường bộ

Quy trình: 1. Chuẩn bị Dữ liệu & Embedding

Mục tiêu: Chuyển đổi văn bản pháp luật thành vector để tìm kiếm hiệu quả hơn (thay vì so khớp từ khóa đơn thuần).

- **Đọc Dữ liệu Nguồn:**

- Tải file JSON chứa các đoạn văn bản (đã qua chunking)

- **Vector Hóa Ngữ Nghĩa (Embedding):**

- **Khái niệm:** Mỗi đoạn văn bản sẽ được chuyển thành một vector số thực, thể hiện ý nghĩa tổng quát của đoạn văn.
- **Ví dụ:**
 - “Điều 5. Xe không có biển số sẽ bị xử phạt” \rightarrow $[0.13, -0.02, \dots, 0.27]$
 - Các đoạn văn có ý nghĩa tương tự sẽ nằm gần nhau trong không gian vector.

Quy trình: 1. Chuẩn bị Dữ liệu & Embedding

- **Công cụ:** Sử dụng SentenceTransformer với ba mô hình thử nghiệm:
 - **vn-law-embedding** – một mô hình được huấn luyện chuyên biệt cho tiếng Việt và văn bản pháp lý (768 chiều)
 - **bge-m3** – mô hình mạnh mẽ đa ngôn ngữ, hỗ trợ cả văn bản ngắn và dài. Được huấn luyện trên nhiều nhiệm vụ (retrieval, classification, clustering), mang lại hiệu quả tốt trong các bài toán truy xuất thông tin (1024 chiều)
 - **multilingual-e5-large** – một mô hình embedding đa ngôn ngữ tiên tiến, được thiết kế tối ưu cho tác vụ truy xuất thông tin với câu truy vấn định dạng chuẩn “query: ...” và “passage: ...”. Hỗ trợ hơn 100 ngôn ngữ, bao gồm tiếng Việt (1024 chiều)

Quy trình: 1. Chuẩn bị Dữ liệu & Embedding

Bảng 1: So sánh hiệu suất các mô hình embedding

Độ đo	vn-law-embedding				bge-m3				multilingual-e5-large			
	@3	@5	@10	@15	@3	@5	@10	@15	@3	@5	@10	@15
Recall	0.4677	0.5381	0.5938	0.6124	0.5477	0.6308	0.7058	0.7346	0.5173	0.6124	0.6864	0.7126
MRR	0.6717	0.6812	0.6849	0.6862	0.7117	0.7203	0.7244	0.7260	0.6967	0.7083	0.7132	0.7148
nDCG	0.5450	0.5484	0.5633	0.5701	0.6150	0.6255	0.6490	0.6597	0.5919	0.6087	0.6306	0.6407

Quy trình: 2. Xây dựng Chỉ mục Tìm kiếm

• Dense Index:

- Dựa trên vector embedding đã tạo từ bước trước.
- Dùng thư viện FAISS để tìm kiếm gần đúng nhanh (Approximate Nearest Neighbors).
- Cấu hình: IndexFlatL2 – dùng khoảng cách Euclidean.
- Mỗi vector được gắn với thông tin chunk gốc (ID, tiêu đề, metadata).

• Sparse Index:

- Dựa trên mô hình truyền thống BM25 (TF-IDF cải tiến).
- Phân tách từ (tokenize), loại bỏ stopwords, tính độ quan trọng của từ.
- Hiệu quả tốt với truy vấn ngắn, từ khóa rõ ràng dựa trên các đoạn văn bản pháp luật đã được xử lý.

Hybrid Retriever:

- Kết hợp cả dense và sparse để tận dụng điểm mạnh của hai phương pháp.
 - sparse: truyền thống, phù hợp khi truy vấn rõ ràng.
 - dense: tốt hơn với câu truy vấn tự nhiên, ngữ cảnh mơ hồ.
 - Với phương pháp hiện dùng 2 mô hình cho truy vấn ngữ nghĩa (dense) kết hợp với 1 mô hình truy vấn từ khóa (sparse) để nhằm tăng độ đa dạng khi truy vấn.
 - hybrid: Xếp hạng kết quả từ các phương pháp bằng RRF (Reciprocal Rank Fusion).

Quy trình: 3. Tiếp nhận & Phân tích Truy vấn

1 Nhận Truy vấn:

- Giao diện người dùng nhập câu hỏi dạng tự nhiên (VD: “Lái xe ô tô không có giấy phép lái xe thì bị phạt bao nhiêu?”).

2 Phân tích Ngữ nghĩa & Ngữ cảnh:

- Dùng mô hình ngôn ngữ lớn (LLM - Gemini) hiểu mục đích, hành vi, và thực thể liên quan trong truy vấn.



Lái xe ô tô không có giấy phép lái xe thì bị phạt bao nhiêu?



Người điều khiển xe ô tô không có giấy phép lái xe sẽ bị phạt tiền từ 18.000.000 đồng đến 20.000.000 đồng (Theo Điều 18, Khoản 9, Điểm b, Nghị định số 168_2024_NĐ-CP).

Nguồn tham khảo (Văn bản pháp luật):

- [Nghị định số 168/2024/NĐ-CP](#)

Nhập câu hỏi của bạn về Luật GTĐB...



Hình 7: Giao diện truy vấn

Quy trình: 3. Tiếp nhận & Phân tích Truy vấn

3 Kiểm tra Liên quan (Query Filtering):

- Nếu truy vấn không liên quan đến luật hoặc ngoài phạm vi: trả lời từ chối hoặc chuyển hướng.
- Nếu hợp lệ: tiếp tục truy xuất thông tin.



Lái xe ô tô không có giấy phép lái xe thì bị phạt bao nhiêu?



Người điều khiển xe ô tô không có giấy phép lái xe sẽ bị phạt tiền từ **18.000.000 đồng đến 20.000.000 đồng** (Theo Điều 18, Khoản 9, Điểm b, Nghị định số 168/2024_NĐ-CP).

Nguồn tham khảo (Văn bản pháp luật):

- [Nghị định số 168/2024/NĐ-CP](#)

Nhập câu hỏi của bạn về Luật GTĐB...



Hình 8: Câu trả lời cho câu hỏi không liên quan

Quy trình: 3. Tiếp nhận & Phân tích Truy vấn

④ Mở rộng truy vấn (Query Expansion):

Mục tiêu: Bù đắp trường hợp “từ ngữ khác nhau” giữa câu hỏi và văn bản pháp lý. Hệ thống hỗ trợ 3 chế độ đặt câu hỏi:

- **Đơn giản:** Dùng trực tiếp câu gốc để truy vấn
- **Mở rộng:** Sinh thêm cụm từ hoặc paraphrase gần nghĩa, mở rộng thêm nhưng vẫn giữ nguyên câu từ của câu gốc.
- **Đa dạng:** Tạo ra các biến thể khác nhau (mặc định là 3) kết hợp với 1 câu gốc ban đầu để tăng sự đa dạng của việc tìm kiếm.

Ví dụ: Từ “lề đường” không được dùng trong văn bản pháp luật, thay vào đó là “vĩa hè”. Nếu mở rộng truy vấn ta có thể nắm bắt thêm từ khóa này, giúp tăng hiệu quả truy vấn.

Quy trình: 3. Tiếp nhận & Phân tích Truy vấn

Bảng 2: So sánh hiệu suất các mô hình embedding

Độ đo	Đơn giản (Gốc)				Đa dạng				Mở rộng			
	@3	@5	@10	@15	@3	@5	@10	@15	@3	@5	@10	@15
Recall	0.5771	0.6733	0.7630	0.7922	0.5808	0.6698	0.7576	0.7978	0.5675	0.6707	0.7582	0.7958
MRR	0.7211	0.7296	0.7349	0.7359	0.7483	0.7570	0.7630	0.7647	0.7306	0.7414	0.7465	0.7479
nDCG	0.6347	0.6559	0.6882	0.6998	0.6496	0.6604	0.6892	0.7036	0.6400	0.6571	0.6861	0.7004

- Truy vấn Đa dạng cho kết quả tốt nhất. Tuy nhiên, tốn thời gian và tài nguyên tìm kiếm hơn, nên hầu hết các thực nghiệm chúng em chọn truy vấn Mở rộng để cân bằng hiệu suất hơn.

Quy trình: 4. Truy xuất Thông tin

Tìm các đoạn văn bản (chunk) tiềm năng từ kho dữ liệu.

- **Lựa chọn Chiến lược Truy xuất chính:**

- Dựa trên **chế độ mở rộng** nhằm cân bằng hiệu suất. Các **chế độ đa dạng** để cho các câu hỏi phức tạp hơn.
- Thử nghiệm các chế độ **Dense** (vector), **Sparse** (BM25), hoặc **Hybrid**. Trong đó Hybrid cho kết quả tốt nhất (Bảng 3).

- **Thực hiện Truy xuất:**

- **Dense Search:** Dùng FAISS để tìm các vector gần nhất trong không gian embedding.
- **Sparse Search:** Dùng BM25 để truy vấn theo từ khóa.
- **Hybrid Search:** Kết hợp cả hai kết quả bằng kỹ thuật **Rank Fusion (RRF)**:
 - Gộp các bảng xếp hạng, tính điểm dựa vào thứ hạng.
 - Ưu tiên kết quả xuất hiện trong nhiều phương pháp.

- **Thu thập Kết quả Ứng viên:**

- Trả về danh sách các k-chunks tiềm năng cho bước Reranking hoặc sinh câu trả lời.

Quy trình: 4. Truy xuất Thông tin

Bảng 3: So sánh hiệu suất các mô hình theo chiến lược truy vấn.

Độ đo	Sparse				Dense				Hybrid (1 Dense, 1 Sparse)				Hybrid (2 Dense, 1 Sparse)			
	@3	@5	@10	@15	@3	@5	@10	@15	@3	@5	@10	@15	@3	@5	@10	@15
Recall	0.4926	0.5682	0.6312	0.6570	0.5477	0.6308	0.7058	0.7346	0.5657	0.6606	0.7455	0.7813	0.5675	0.6707	0.7582	0.7958
MRR	0.6428	0.6486	0.6516	0.6521	0.7117	0.7203	0.7244	0.7260	0.7189	0.7286	0.7332	0.7342	0.7306	0.7414	0.7465	0.7479
nDCG	0.5619	0.5705	0.5925	0.6021	0.6150	0.6255	0.6490	0.6597	0.6348	0.6490	0.6775	0.6904	0.6400	0.6571	0.6861	0.7004

Quy trình: 5. Xếp hạng lại (Reranking)

Tinh chỉnh thứ tự các đoạn văn bản ứng viên để nâng độ chính xác.

- **Chọn ứng viên ban đầu:**

- Lấy Top N chunk từ kết quả truy xuất ban đầu.

- **Áp dụng CrossEncoder để đánh giá lại:**

- Kết hợp từng cặp (**query, chunk**) làm đầu vào cho mô hình.
- Mô hình (CrossEncoder như namdp-ptit/ViRanker) trả về **score** phản ánh độ liên quan.
- Ưu điểm: chính xác hơn so với truy xuất thô.

- **Chọn kết quả cuối cùng:**

- Sắp xếp các chunk theo điểm số.
- Lấy Top K chunk (ví dụ $K = 10$) để đưa vào LLM sinh câu trả lời.

Nếu không dùng Reranker: Lấy trực tiếp Top K từ kết quả truy xuất.

Quy trình: 5. Xếp hạng lại (Reranking)

- **Công cụ:** Sử dụng các mô hình reranker để tinh chỉnh kết quả truy vấn ban đầu:
 - ViRanker – một mô hình reranker chuyên biệt cho tiếng Việt, được thiết kế để sắp xếp lại các kết quả tìm kiếm, ưu tiên các tài liệu liên quan nhất. Mô hình này tập trung vào sự phù hợp ngữ nghĩa trong bối cảnh tiếng Việt.
 - bge-reranker-v2-m3 – phiên bản cải tiến của mô hình reranker đa ngôn ngữ BGE, nổi bật với khả năng đánh giá độ liên quan giữa câu truy vấn và tài liệu một cách hiệu quả. Được huấn luyện trên lượng lớn dữ liệu đa ngôn ngữ, nó cung cấp khả năng reranking mạnh mẽ cho nhiều ngôn ngữ, bao gồm tiếng Việt.
 - Vietnamese_Reranker – một mô hình reranker được phát triển đặc biệt cho tiếng Việt, nhằm cải thiện độ chính xác của các kết quả truy xuất thông tin bằng cách tái sắp xếp các tài liệu dựa trên mức độ liên quan sâu hơn. Mô hình này thường tận dụng các đặc trưng ngôn ngữ riêng của tiếng Việt để đạt hiệu suất tối ưu.

Quy trình: 5. Xếp hạng lại (Reranking)

Bảng 4: So sánh hiệu suất các mô hình reranker.

Độ đo	Không sử dụng				ViRanker				bge-reranker-v2-m3				Vietnamese_Reranker			
	@3	@5	@10	@15	@3	@5	@10	@15	@3	@5	@10	@15	@3	@5	@10	@15
Recall	0.4837	0.5715	0.6881	0.7569	0.5448	0.6290	0.7439	0.7895	0.5675	0.6707	0.7582	0.7958	0.5661	0.6589	0.7495	0.7952
MRR	0.6017	0.6138	0.6242	0.6275	0.6956	0.7037	0.7128	0.7142	0.7306	0.7414	0.7465	0.7479	0.7128	0.7219	0.7279	0.7298
nDCG	0.5171	0.5333	0.5780	0.6030	0.5990	0.6126	0.6543	0.6717	0.6400	0.6571	0.6861	0.7004	0.6225	0.6369	0.6683	0.6851

Quy trình: 6. Tạo & Hoàn thiện Câu trả lời

1. Mục tiêu: Trả lời câu hỏi về luật giao thông dựa **duy nhất** vào "Ngữ cảnh được cung cấp".

2. Cấu trúc Prompt:

- **Lịch sử trò chuyện (tùy chọn):** Tóm tắt các tương tác trước đó (nếu có).
- **Ngữ cảnh cung cấp:** Chứa các đoạn văn bản pháp luật liên quan.
- **Câu hỏi hiện tại:** Do người dùng đặt ra.
- **Yêu cầu trả lời:**
 - Chỉ dùng ngữ cảnh, không suy đoán.
 - Trích dẫn nguồn rõ ràng, gộp nhóm nếu cùng văn bản.
 - Trình bày súc tích, mạch lạc.
 - Phân biệt từ ngữ pháp lý tương đương.
 - Nếu không có thông tin phù hợp → trả lời "Không tìm thấy".
 - Ưu tiên thứ tự điểm, khoản, điều khi so sánh.
 - Hiển thị biển báo (nếu có) bằng placeholder:
[DISPLAY_TRAFFIC_SIGN_INDEX_x]

2 phiên bản prompt: chi tiết và ngắn gọn tùy mục tiêu ứng dụng.

Quy trình: 6. Tạo & Hoàn thiện Câu trả lời

• Hậu xử lý (Post-processing):

- Trích xuất và định dạng trích dẫn (citation) → liên kết đến tài liệu gốc.
- Kiểm tra sự đầy đủ của câu trả lời (có đúng, rõ, nguồn không).
- Trích xuất tên ảnh chứa biển báo, để hiển thị.
- Gửi câu trả lời cùng link nguồn về giao diện người dùng.
- Ghi lại lịch sử truy vấn phản hồi

Quy trình: 7. Ghi nhớ ngữ cảnh

- Lưu lịch sử truy vấn, câu trả lời (lưu khoảng 4 query và answer gần nhất để giữ được độ ổn định của câu trả lời).
- Hỗ trợ đối thoại liên tục.

Hình ảnh biển báo cấm đi ngược chiều

Biển báo cấm đi ngược chiều là biển số P.102. (Theo Phụ lục II, B.2, Quy chuẩn kỹ thuật số 41_2024_BGTVT)



Nguồn tham khảo (Văn bản pháp luật):

- [Quy chuẩn kỹ thuật số 41/2024/BGTVT](#)

Xe máy nếu vi phạm biển này bị phạt bao nhiêu?

Đối với xe máy (xe mô tô, xe gắn máy) vi phạm biển cấm đi ngược chiều:

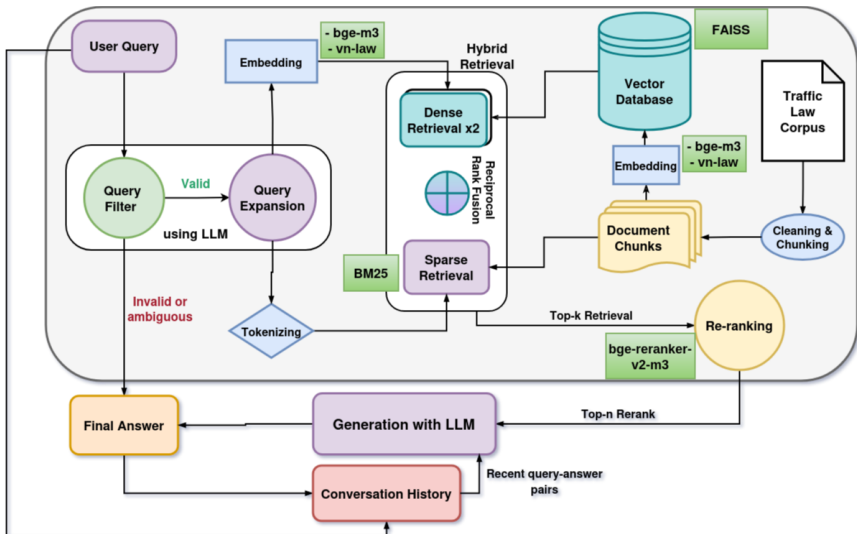
- Phạt tiền từ **4.000.000 đồng đến 6.000.000 đồng** và bị trừ **02 điểm** giấy phép lái xe (Theo Điều 7, Khoản 7, Điểm a và Khoản 13, Điểm a, Nghị định số 168_2024_NĐ-CP).
- Nếu hành vi này gây tai nạn giao thông, phạt tiền từ **10.000.000 đồng đến 14.000.000 đồng** và bị trừ **10 điểm** giấy phép lái xe (Theo Điều 7, Khoản 10, Điểm a và Khoản 13, Điểm d, Nghị định số 168_2024_NĐ-CP).

Nguồn tham khảo (Văn bản pháp luật):

- [Nghị định số 168/2024/NĐ-CP](#)

Hình 9. Ghi nhớ ngữ cảnh

Quy trình: Pipeline



Hình 6: Pipeline chính của chatbot giao thông đường bộ

Nội dung

- 1 Giới thiệu
- 2 Dữ liệu
- 3 Quy trình xử lý
- 4 Độ đo đánh Giá**

Mục tiêu Đánh giá

- Đo hiệu quả truy vấn trong RAG.
- Đo hiệu quả trả lời của chatbot.

Bộ dữ liệu Đánh giá

- 300 câu hỏi tổng hợp thủ công về Luật GTDB.
- Ground truth: các ID/thông tin của các chunk liên quan.

Cách thức đánh giá

- Đối với đánh giá truy vấn: dựa trên các độ đo Recall@k, MRR@k, nDCG@k.
- Đối với đánh giá câu trả lời: đánh giá bán tự động bằng Gemini 2.5.

Đánh giá truy vấn: Dữ liệu

Ví dụ dữ liệu đánh giá truy vấn

```
[  
  {  
    "query_id": "eval_001",  
    "query": "Mức phạt không đội mũ bảo hiểm xe máy ...?",  
    "relevant_chunk_ids": ["168_2024_NĐ-CP_161", ...]  
  },  
  {  
    "query_id": "eval_002",  
    "query": "Chạy quá tốc độ 10 km/h ô tô bị phạt ...",  
    "relevant_chunk_ids": [  
      "168_2024_NĐ-CP_chunk_103",  
      "168_2024_NĐ-CP_chunk_127"  
    ]  
  },  
  ...  
]
```

Đánh giá truy vấn: Độ đo Recall@k

$$\text{Recall@}k = \frac{|S_q \cap R_q|}{|R_q|}$$

- Đo lường **tỉ lệ** tài liệu liên quan (R_q) thực sự được trả về trong top- k kết quả (S_q).
- Quan trọng để đánh giá **độ bao phủ** (không bỏ sót) của hệ thống tìm kiếm.
- Thường tính ở các ngưỡng $k = 3, 5, 10$ để so sánh hiệu quả. Thêm $k=15$ vì có những điều luật có rất nhiều Điểm.

Đánh giá truy vấn: Độ đo MRR@k & nDCG@k

$$\text{MRR} = \frac{1}{|Q|} \sum_{q=1}^{|Q|} \frac{1}{\text{rank}_q} \quad , \quad \text{nDCG@k} = \frac{1}{\text{IDCG@k}} \sum_{i=1}^k \frac{\text{rel}_i}{\log_2(i+1)}$$

- **MRR:** Tập trung vào vị trí xuất hiện của kết quả liên quan đầu tiên. Giá trị càng cao \rightarrow kết quả đúng xuất hiện càng sớm.
- **nDCG@k:** Ưu tiên xếp những đoạn có *độ liên quan* cao ở đầu. Tính điểm giảm giá (discount) theo vị trí để phản ánh chất lượng ranking.
- Cả hai chỉ số giúp đánh giá *chất lượng thứ tự* của kết quả, không chỉ số lượng.

Đánh giá câu trả lời: Dữ liệu

Ví dụ dữ liệu đánh giá câu trả lời

```
[{
  "query_id": "eval_001",
  "query": "Mức phạt không đội mũ BH xe máy 2025?",
  "relevant_chunk_ids": {
    "text": "Chương I NHỮNG QUY ĐỊNH CHUNG\nĐiều 9. ...",
    "metadata": {
      "source": "Luật số 36_2024_QH15",
      "effective_date": "1-1-2025",
      "url": "https://thuvienphapluat.vn/van-ban/...",
      "context": { "khoan": "2", "dieu": "9", "chuong": "1"
    }
  }
},
...
]
```

Đánh giá câu trả lời: Quy trình (1)

- Câu trả lời từ chatbot giao thông đường bộ được kết hợp với các đoạn văn bản pháp lý liên quan (chunks) và **đưa vào một prompt chuẩn**.
- Việc chấm điểm được **thực hiện bán tự động** bằng mô hình **Gemini 2.5**.
- Prompt đánh giá **giả lập một chuyên gia pháp lý Việt Nam**, yêu cầu đánh giá **dựa trên các quy định hiện hành về giao thông đường bộ**.
- Các tiêu chí đánh giá **rõ ràng, cụ thể, bám sát nội dung pháp luật, không dựa vào suy đoán ngoài văn bản đã cung cấp**.

Đánh giá câu trả lời: Quy trình (2)

4 tiêu chí chấm điểm chính (thang điểm: 1 – rất kém, 5 – xuất sắc):

- ❶ **Tính chính xác và đầy đủ:** – Trả lời đúng theo luật, không thêm/sai thông tin.
- ❷ **Sự liên quan:** – Trả lời đúng trọng tâm, không lan man.
- ❸ **Cấu trúc và ngôn ngữ:** – Trình bày rõ ràng, dễ đọc, hợp lý.
- ❹ **Khả năng trích dẫn nguồn:** – Dẫn nguồn đúng định dạng, đúng nội dung pháp lý.

Cuối cùng, hệ thống đưa ra nhận xét ngắn gọn (khoảng 100 từ) giải thích lý do chấm điểm.

Đánh giá câu trả lời: Kết quả

Bảng 5: Điểm đánh giá trung bình các tiêu chí

Tiêu chí	Tính Chính Xác và Đầy Đủ	Sự Liên Quan	Cấu Trúc và Ngôn Ngữ	Khả năng trích dẫn nguồn
Điểm số	3.041	4.048	4.696	3.027

**Cảm ơn thầy và các bạn
đã lắng nghe!**

Q&A

- Luật Giao Thông 2025 và các nghị định thông tư hướng dẫn mới nhất.
- vn-law-embedding on Hugging Face.
- ViRanker on Hugging Face.
- LLM Engineers Handbook.
- AI Engineering Book.