

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN -
ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ
MINH

KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN CUỐI KỲ
HỆ THỐNG TƯ VẤN TUYỂN SINH THÔNG
MINH CHO HỌC SINH PHỔ THÔNG

Môn học: Nhập môn học máy

Giảng viên hướng dẫn

TS. Bùi Tiến Lên

Giảng viên Võ Nhật Tân

Giảng viên Ngô Đức Bảo

Sinh viên thực hiện

Họ và tên	MSSV
Hoàng Lê Minh Đăng	22127051
Lại Ngọc Phương Nam	22127283
Phạm Thành Nhân	22127307
Nguyễn Quang Sáng	22127364
Nguyễn Văn Tú	22127434

Ngày 9 tháng 5 năm 2025

Mục lục

1	Giới thiệu	3
1.1	Thông tin nhóm	3
1.2	Ý tưởng	3
1.3	Mục tiêu	4
1.4	Khó khăn - Thử thách	5
2	Phát triển mô hình	6
2.1	Thu thập và tiền xử lý dữ liệu	6
2.2	Thiết kế và triển khai mô hình	9
2.3	Fine-tune mô hình	10
2.4	Đánh giá hiệu suất mô hình	11
2.4.1	Mục tiêu	11
2.4.2	Kết quả	12
2.4.3	Phân tích chi tiết	12
2.4.4	Kết luận	12
3	Sản phẩm của chúng tôi	13
3.1	Mức độ tích hợp mô hình vào ứng dụng web	13
3.2	Giao diện và trải nghiệm của người dùng	14
3.2.1	Thành phần giao diện chính	14
3.2.2	Giao diện chính	15
3.2.3	Khả năng tiếp cận và phản hồi người dùng	18
3.2.4	Kết luận	18
4	Tổng kết	20
4.1	Tổng quan	20
4.2	Định hướng tương lai	21

Danh mục ký hiệu, viết tắt

AI trí tuệ nhân tạo - Artificial intelligence

AR thực tế tăng cường - Augmented Reality

LLM mô hình ngôn ngữ lớn

RoPE Rotary Position Embedding

THPT trung học phổ thông

VR thực tế ảo - Virtual Reality

WCAG Web Content Accessibility Guidelines

YaRN Yet another RoPE-based Network

Chương 1

Giới thiệu

1.1 Thông tin nhóm

STT	MSSV	Họ và tên	Mức đóng góp	Ghi chú
1	22127307	Hoàng Lê Minh Đăng	100%	Nhóm trưởng
2	22127283	Lại Ngọc Phương Nam	100%	
3	22127307	Phạm Thành Nhân	100%	
4	22127307	Nguyễn Quang Sáng	100%	
5	22127434	Nguyễn Văn Tú	100%	

Bảng 1.1: Mức đóng góp của các thành viên

1.2 Ý tưởng

Trong bối cảnh hệ thống giáo dục đại học Việt Nam đang phát triển mạnh mẽ, với sự gia tăng nhanh chóng về số lượng cơ sở đào tạo và sự đa dạng hóa các chương trình học, học sinh trung học phổ thông (THPT) đang đứng trước nhiều cơ hội nhưng cũng đối mặt với không ít thách thức trong việc định hướng học vấn.. Sự phong phú về lựa chọn trường học và, tuy mang lại nhiều cơ hội, lại đặt ra bài toán phức tạp khi học sinh cần đưa ra quyết định phù hợp với năng lực, sở thích cá nhân. Hiện nay, học sinh thường gặp khó khăn do thiếu thông tin đáng tin cậy, hoặc phải dựa vào các nguồn không chính thức như truyền miệng, mạng xã hội, hoặc các diễn đàn trực tuyến không được kiểm chứng. Điều này dẫn đến những quyết định mang tính chủ quan, thiếu cơ sở dữ liệu khoa học. Hệ quả của tình trạng này là không ít học sinh đưa ra lựa chọn ngành học và trường đại học không phù hợp, dẫn đến nguy cơ lãng phí thời gian, tài chính và nỗ lực của bản thân cũng như gia đình. Những quyết định sai lầm có thể gây ra tình trạng chán nản, bỏ học. Do đó, việc xây dựng một nền tảng tư vấn tuyển sinh thông minh, tích hợp trí tuệ nhân tạo - Artificial intelligence (AI) và dữ liệu đáng tin cậy, là một nhu cầu cấp thiết để hỗ trợ học sinh đưa ra quyết định chính xác và hiệu quả hơn.

Ý tưởng của dự án là tạo ra một hệ thống tư vấn tuyển sinh toàn diện, sử dụng công nghệ hiện đại để cung cấp thông tin chính xác, cá nhân hóa và dễ tiếp cận, từ đó giúp học sinh Việt Nam định hướng tương lai học vấn một cách khoa học và tự tin. Nền tảng này không chỉ là một công cụ tìm kiếm thông tin mà còn là một “người bạn đồng hành”

thông minh, hỗ trợ học sinh trong toàn bộ quá trình từ khám phá bản thân, tìm hiểu ngành học, đến lựa chọn trường đại học phù hợp.

1.3 Mục tiêu

Dự án hướng đến việc xây dựng một nền tảng tư vấn hướng nghiệp và chọn trường đại học thông minh, được thiết kế đặc biệt dành cho học sinh Việt Nam, đáp ứng nhu cầu định hướng giáo dục trong bối cảnh hệ thống đại học ngày càng đa dạng và phức tạp. Bằng cách tích hợp AI, công nghệ thu thập dữ liệu web (web crawling¹), và kỹ thuật truy xuất tri thức thông minh (LightRAG[2]), hệ thống mang lại một công cụ toàn diện, chính xác, và dễ sử dụng, với mục tiêu cụ thể như sau:

- **Cung cấp thông tin chi tiết và đáng tin cậy về trường đại học:** Xây dựng một cơ sở dữ liệu toàn diện về 163 trường đại học trên toàn quốc, bao gồm thông tin chi tiết như danh sách ngành đào tạo, điểm chuẩn qua các năm, mức học phí, điều kiện xét tuyển, chương trình học bổng, thông tin về cơ sở vật chất (ký túc xá, thư viện, phòng thí nghiệm), và các hoạt động ngoại khóa. Dữ liệu sẽ được thu thập từ các nguồn chính thức và cập nhật thường xuyên để đảm bảo tính chính xác, giúp học sinh dễ dàng so sánh và lựa chọn trường phù hợp.
- **Hỗ trợ tìm hiểu lộ trình học tập:** Cung cấp thông tin chi tiết về lộ trình học tập của từng ngành học, bao gồm các học phần bắt buộc và tự chọn theo từng năm, các kỹ năng chuyên môn cần trang bị, và các chứng chỉ bổ sung (nếu có).
- **Tích hợp chatbot AI thông minh và thân thiện:** Phát triển một chatbot AI có khả năng tương tác trực tiếp với học sinh, trả lời nhanh chóng và chính xác các câu hỏi liên quan đến tuyển sinh, ngành học, điểm chuẩn, học phí. Chatbot sẽ sử dụng kỹ thuật RAG để truy xuất thông tin từ cơ sở dữ liệu nội bộ và các nguồn đáng tin cậy, đảm bảo câu trả lời chính xác và phù hợp ngữ cảnh. Giao diện chatbot được thiết kế thân thiện, với tính năng lưu lịch sử trò chuyện và gợi ý các câu hỏi thường gặp, mang lại trải nghiệm giống như trò chuyện với một cố vấn cá nhân.
- **Tăng cường trải nghiệm cá nhân hóa:** Hệ thống sẽ sử dụng dữ liệu người dùng (như kết quả trắc nghiệm, điểm thi, và lịch sử tương tác) để cung cấp các gợi ý và khuyến nghị được cá nhân hóa. Ví dụ, nền tảng có thể đề xuất các trường đại học phù hợp dựa trên điểm thi THPT Quốc gia của học sinh, hoặc gợi ý các ngành học dựa trên sở thích và kỹ năng của họ. Tính năng này nhằm giúp học sinh cảm thấy được hỗ trợ một cách riêng biệt, từ đó tăng sự tự tin khi đưa ra quyết định.

Hệ thống không chỉ là một công cụ tìm kiếm thông tin mà còn hướng đến việc trở thành một nền tảng tư vấn toàn diện, kết nối học sinh với các cơ hội giáo dục một cách khoa học và hiệu quả. Bằng cách cung cấp thông tin đáng tin cậy, cá nhân hóa trải nghiệm người dùng, và tận dụng công nghệ AI tiên tiến, nền tảng sẽ giúp học sinh giảm thiểu rủi ro từ các quyết định không phù hợp, đồng thời xây dựng sự tự tin trong việc định hình tương lai học tập. Trong dài hạn, dự án hướng đến việc trở thành một giải pháp tiêu chuẩn cho tư vấn tuyển sinh tại Việt Nam, góp phần cải thiện chất lượng định hướng giáo dục và giảm thiểu khoảng cách giữa đào tạo và nhu cầu thị trường lao động.

¹<https://www.cloudflare.com/learning/bots/what-is-a-web-crawler/>

1.4 Khó khăn - Thử thách

- **Dữ liệu thiếu đồng bộ và khó thu thập:** Thông tin về các trường đại học, bao gồm điểm chuẩn, học phí, ngành học, và các yêu cầu xét tuyển, thường phân tán trên nhiều nguồn như trang web của trường, cổng thông tin của Bộ Giáo dục và Đào tạo, hoặc các diễn đàn không chính thức. Việc thu thập, tổng hợp và cập nhật liên tục dữ liệu này đòi hỏi nguồn lực lớn và công nghệ phức tạp.
Hạn chế: Dữ liệu không chính xác hoặc lỗi thời có thể làm giảm độ tin cậy của nền tảng, gây ảnh hưởng tiêu cực đến trải nghiệm người dùng và dẫn đến các khuyến nghị không phù hợp.
- **Hiệu suất và độ chính xác của chatbot AI:** Đảm bảo chatbot AI có khả năng hiểu ngữ cảnh tiếng Việt, xử lý các thuật ngữ chuyên ngành giáo dục, và trả lời chính xác các câu hỏi phức tạp về tuyển sinh là một thách thức lớn, đặc biệt khi ngôn ngữ tiếng Việt có nhiều sắc thái và cách diễn đạt đa dạng.
Hạn chế: Trả lời không chính xác, chung chung hoặc thiếu chiều sâu có thể khiến người dùng mất niềm tin, giảm mức độ tương tác và hiệu quả của nền tảng.
- **Thiếu lòng tin từ người dùng:** Tại Việt Nam, học sinh và phụ huynh thường có xu hướng tin tưởng vào tư vấn trực tiếp từ giáo viên, gia đình, hoặc bạn bè hơn là các nền tảng công nghệ, đặc biệt là AI. Điều này bắt nguồn từ thói quen văn hóa và sự thiếu quen thuộc với các giải pháp công nghệ trong giáo dục.
Hạn chế: Tỷ lệ chấp nhận và sử dụng nền tảng có thể thấp, đặc biệt ở các khu vực nông thôn hoặc với những người dùng không quen thuộc với công nghệ.
- **Giao diện responsive tổn tài nguyên:** Phát triển giao diện người dùng đáp ứng tốt trên cả thiết bị di động và máy tính đòi hỏi thiết kế phức tạp, kiểm tra kỹ lưỡng trên nhiều nền tảng, và tối ưu hóa hiệu suất. Việt Nam có tỷ lệ sử dụng điện thoại di động cao, khiến việc tối ưu hóa trải nghiệm di động trở thành yếu tố sống còn.
Hạn chế: Trải nghiệm người dùng kém trên thiết bị di động hoặc các thiết bị có cấu hình thấp có thể dẫn đến mất người dùng, làm giảm hiệu quả của nền tảng.
- **Bảo mật dữ liệu cá nhân:** Hệ thống cần thu thập và lưu trữ thông tin cá nhân như điểm thi, sở thích ngành học, và kết quả trắc nghiệm để cung cấp tư vấn cá nhân hóa. Việc đảm bảo an toàn dữ liệu và tuân thủ các quy định bảo mật (như luật bảo vệ dữ liệu tại Việt Nam hoặc các tiêu chuẩn quốc tế) là một thách thức lớn.
Hạn chế: Rò rỉ dữ liệu hoặc vi phạm quyền riêng tư có thể gây tổn hại nghiêm trọng đến uy tín của nền tảng, dẫn đến mất lòng tin từ người dùng và các vấn đề pháp lý tiềm tàng.

Chương 2

Phát triển mô hình

2.1 Thu thập và tiền xử lý dữ liệu

Nhóm chúng tôi thu thập và tổng đường dẫn các website chính thức và website tuyển sinh của hầu hết các trường đại học ở Việt Nam để đảm bảo tính chính xác nhất từ trang web này ¹ là một hướng tiếp cận rất hiệu quả và phổ biến trong việc trích xuất dữ liệu web. Nhóm quyết định sử dụng thư viện **Selenium**[10] vì nó là một công cụ mạnh mẽ cho những tác vụ như trên. **Selenium** là một bộ công cụ mã nguồn mở được sử dụng chủ yếu để tự động hóa các trình duyệt web. Ban đầu, nó được tạo ra cho mục đích kiểm thử tự động các ứng dụng web, nhưng với khả năng tương tác mạnh mẽ với các thành phần trên trang web, **Selenium** đã trở thành một lựa chọn hàng đầu cho việc thu thập dữ liệu web, đặc biệt là với các trang web động sử dụng nhiều JavaScript để tải nội dung. Ở đây quy trình này đã được tự động hóa bằng cách tự động hóa bằng các bước:

1. **Tự động hóa việc truy cập vào đường dẫn:** Thay vì phải mở trình duyệt và gõ địa chỉ *dsdaihoc.com* một cách thủ công, nhóm sử dụng **Selenium** để tự động mở trình duyệt và điều hướng đến trang web này.
2. **Tự động chuyển trang:** Nếu danh sách các trường đại học trên *dsdaihoc.com* được chia thành nhiều trang, **Selenium** có thể được lập trình để tự động nhấp vào các nút “Next”, “Previous” hoặc các số trang để duyệt qua tất cả các trang chứa thông tin. Điều này giúp đảm bảo rằng nhóm không bỏ sót bất kỳ trường đại học nào.
3. **Đợi dữ liệu tải lên:** Một trong những ưu điểm lớn nhất của **Selenium** so với các thư viện thuần túy chỉ gửi yêu cầu HTTP (như *requests* [9] trong Python) là khả năng xử lý các trang web động. Nhiều trang web hiện đại sử dụng JavaScript để tải nội dung một cách không đồng bộ. **Selenium** cung cấp các cơ chế “chờ đợi” (waits) thông minh, cho phép tạm dừng và chờ cho đến khi một yếu tố cụ thể trên trang xuất hiện hoặc tải xong trước khi tiếp tục thực thi. Điều này cực kỳ quan trọng để đảm bảo thu thập được dữ liệu đầy đủ và chính xác.
4. **Thu thập và lọc ra các đường dẫn:** Sau khi trang đã tải xong, **Selenium** cho phép tương tác với các phần tử HTML. Nhóm có thể sử dụng các “bộ chọn” (selectors) như ID, class name, XPath, hoặc CSS selectors để xác định vị trí

¹<https://dsdaihoc.com/>

```

1 https://dhannd.edu.vn
2 https://dhannd.edu.vn/Thong-tin-tuyen-sinh
3 https://hust.edu.vn
4 https://ts.hust.edu.vn
5 https://www.ctu.edu.vn
6 https://tuyensinh.ctu.edu.vn
7 https://vsat.ctu.edu.vn
8 https://duytan.edu.vn
9 https://duytan.edu.vn/tuyen-sinh
10 https://hueuni.edu.vn
11 https://tuyensinh.hueuni.edu.vn
12 https://tnu.edu.vn
13 https://tuyensinh.tnus.edu.vn
14 https://vinhuni.edu.vn
15 https://tuyensinh.vinhuni.edu.vn
16 https://www.vnam.edu.vn
17 https://www.vnam.edu.vn/NewsDetail.aspx?lang=\u0026ItemID=1887
18 https://hvannd.edu.vn
19 https://hvannd.edu.vn/bv/ct/16032/thong-bao-chi-tieu-tuyen-sinh-tuyen-moi-cao-tao-trinh-do-dai-hoc-nam-2025-tai-hoc-vien-an-ninh-nhan-dan
20 https://daotaobj.c.edu.vn
21 https://hvcnsd.edu.vn
22 https://hvcnsd.edu.vn/tuyen-sinh
23 https://dkxt.apd.edu.vn
24 https://n11thcm.edu.vn

```

Hình 2.1: Kết quả thu được sau khi thực hiện lấy url website các trường đại học

chính xác của các đường dẫn đến website chính thức và website tuyển sinh của từng trường đại học (Hình 2.1).

Sau khi triển khai **Crawl4AI**², nhóm đã thành công thu thập được một lượng lớn dữ liệu từ các website trường đại học dưới định dạng markdown. Như đã phân tích, **Crawl4AI** là một dự án mã nguồn mở tiềm năng, có sẵn trên GitHub, cung cấp một công cụ web crawler và scraper thân thiện với người dùng. Điểm mạnh nổi bật của công cụ này là thiết kế hướng đến sự tương thích liền mạch với các mô hình ngôn ngữ lớn (LLM), cụ thể là việc trả về dữ liệu đã trích xuất ở định dạng markdown. Định dạng này không chỉ giữ lại cấu trúc ngữ nghĩa cơ bản của văn bản mà còn nâng cao đáng kể khả năng sử dụng của dữ liệu, giúp việc xử lý và tích hợp vào các ứng dụng AI, đặc biệt là LLMs, trở nên dễ dàng hơn.

Tuy nhiên, ở bước này, mặc dù đã sở hữu dữ liệu đầu vào dưới dạng markdown – một định dạng thuận lợi để làm việc hiệu quả với LLMs nhằm mục tiêu sinh ra bộ dữ liệu (dataset) cho mô hình – nhóm nhận thấy rằng đây mới chỉ là thành công bước đầu. Dữ liệu thu thập được, dù đã qua một lớp xử lý của **Crawl4AI**, về bản chất vẫn là dữ liệu "thô" và chưa hoàn toàn "sạch". Nó có thể còn chứa đựng:

- **Các yếu tố không mong muốn:** Như các đoạn văn bản điều hướng còn sót lại, thông tin footer, header chưa được loại bỏ triệt để, các menu, hoặc thậm chí là các đoạn quảng cáo, bình luận không liên quan đến nội dung chính. Nhiều và định dạng không nhất quán, có thể có các ký tự đặc biệt, lỗi định dạng markdown nhỏ, hoặc sự không nhất quán trong cách trình bày thông tin giữa các website khác nhau.
- **Nội dung không phù hợp:** Một số trang con hoặc phần nội dung có thể không thực sự hữu ích hoặc không liên quan trực tiếp đến mục tiêu xây dựng dataset của mô hình. Do đó, để có thể tối đa hóa hiệu suất của LLM khi huấn luyện hoặc tinh chỉnh (fine-tuning) dựa trên bộ dữ liệu này, một bước xử lý và làm sạch dữ liệu chuyên sâu là vô cùng cần thiết. Bước này sẽ đảm bảo rằng dữ liệu đầu vào cho LLM là chất lượng nhất, loại bỏ các yếu tố gây nhiễu, chuẩn hóa định dạng và chỉ giữ lại những thông tin thực sự giá trị. Chỉ khi đó, mới có thể khai thác tối đa tiềm năng của dữ liệu đã thu thập và xây dựng được một bộ dataset chất lượng cao, góp phần nâng cao độ chính xác và hiệu quả của mô hình AI cuối cùng.

²<https://github.com/unclecode/crawl4ai>


```

1  Nhập từ khóa tìm kiếm
2  Đăng nhập
3  [![Đại học Bách khoa Hà Nội](https://hust.edu.vn/uploads/sys/logo-dhbk-1-02_130_191.png)](https://hust.edu.vn/vi/ "Đại học Bách khoa Hà Nội")
4  * [Đại học bách khoa Hà Nội](https://hust.edu.vn)
5  * [HaNoi University of science and technology](https://hust.edu.vn)
6
7
8  * [Đại học](https://hust.edu.vn)
9  * [Bách khoa Hà Nội](https://hust.edu.vn)
10 * [HaNoi University](https://hust.edu.vn)
11 * [of science and technology](https://hust.edu.vn)
12
13
14 [![Đại học Bách khoa Hà Nội](https://hust.edu.vn/uploads/sys/logo-dhbk-1-02_130_191.png)](https://hust.edu.vn/vi/ "Đại học Bách khoa Hà Nội")
15 # Đại học Bách khoa Hà Nội
16 ## Đại học Bách khoa Hà Nội - HaNoi University of science and technology
17 * [Tuyển dụng](https://tuyendung.hust.edu.vn/ "Tuyển dụng")
18 * [Tài nguyên số](https://hust.edu.vn/vi/tai-nguyen-so/tai-nguyen-so/tai-nguyen-so-1.html "Tài nguyên số")
19 * [Email](https://mail.hust.edu.vn/ "Email")
20 * [Lịch công tác](https://hust.edu.vn/vi/lich-lam-viec/Truong-dai-hoc-BKHN/ "Lịch công tác")
21 * [Liên hệ](https://hust.edu.vn/vi/contact/ "Liên hệ")
22
23
24 [](https://hust.edu.vn/#flsearchform)
25 [ ](https://hust.edu.vn/vi/ "Tiếng Việt") [ ](https://hust.edu.vn/en/ "English")
26 * [Giới thiệu](https://hust.edu.vn/vi/about/ "Giới thiệu")
27 * [Thông điệp của Giám đốc Đại học](https://www.hust.edu.vn/vi/about/thong-diep-cua-giam-doc-dai-hoc.html "Thông điệp của Giám đốc Đại học")
28 * [Tổng quan](https://hust.edu.vn/vi/about/tong-quan.html "Tổng quan")
29 * [Lịch sử](https://hust.edu.vn/vi/about/cac-giai-doan-lich-su-191736.html "Lịch sử")
30 * [Sứ mạng, Tầm nhìn, Giá trị cốt lõi và Chính sách chất lượng](https://hust.edu.vn/vi/about/su-mang-tam-nhin-gia-tri-cot-loi-va-chinh-sach-chat-luong.html "Sứ mạng, Tầm nhìn, Giá trị cốt lõi và Chính sách chất lượng")
31 * [Chiến lược phát triển](https://hust.edu.vn/vi/about/chien-luoc-phat-trien.html "Chiến lược phát triển")
32 * [Danh hiệu - Khen thưởng](https://hust.edu.vn/vi/about/danh-hieu-khen-thuong.html "Danh hiệu - Khen thưởng")
33 * [Bảo cáo thường niên](https://hust.edu.vn/vi/about/bao-cao-thuong-nien.html "Bảo cáo thường niên")
34 * [Cơ cấu tổ chức](https://hust.edu.vn/vi/co-cau-to-chuc-bai-viet/ "Cơ cấu tổ chức")
35 * [BCH Đảng bộ Đại học](https://www.hust.edu.vn/vi/co-cau-to-chuc-bai-viet/bch-dang-bo-dai-hoc.html "BCH Đảng bộ Đại học")
36 * [Hội đồng Đại học](https://www.hust.edu.vn/vi/co-cau-to-chuc-bai-viet/hoi-dong-dai-hoc.html "Hội đồng Đại học")
37 * [Ban Giám đốc Đại học](https://www.hust.edu.vn/vi/co-cau-to-chuc-bai-viet/ban-giam-doc-dai-hoc.html "Ban Giám đốc Đại học")

```

Hình 2.2: Dữ liệu mẫu thô thu được khi cào từ website của Đại học Bách khoa Hà Nội

Sau khi có được dữ liệu, nhóm tiếp tục sử dụng api của OpenAI (*GPT-4o mini* 2.3) để tổng hợp và dọn dẹp dữ liệu từ các file markdown, vì đây là một phiên bản "mini" trong dòng *GPT-4o*, mô hình này được tối ưu để cung cấp một sự cân bằng lý tưởng giữa sức mạnh xử lý ngôn ngữ tiên tiến và hiệu quả về mặt chi phí. Đối với việc xử lý hàng loạt file markdown từ nhiều trường đại học, yếu tố này trở nên cực kỳ quan trọng, cho phép nhóm thực hiện công việc ở quy mô lớn mà vẫn đảm bảo tính kinh tế. Theo format file markdown của nhóm gồm 5 mục chính: Giới thiệu, tin tức mới nhất, thông tin tuyển sinh, chương trình đào tạo, ngành học và lĩnh vực cho tất cả các trường đã tổng hợp được.

Sau quá trình làm sạch dữ liệu với *GPT-4o mini*, nhóm sẽ đưa toàn bộ lượng thông tin đã được làm sạch và cấu trúc hóa này vào lưu trữ trong hệ quản trị cơ sở dữ liệu PostgreSQL³. Việc lựa chọn PostgreSQL không chỉ vì tính ổn định, mạnh mẽ của nó trong việc quản lý dữ liệu có cấu trúc, mà còn vì khả năng mở rộng và hỗ trợ các tính năng nâng cao. Dữ liệu này sẽ phục vụ trực tiếp cho việc truy xuất thông tin nhanh chóng và hiệu quả trên website (hoặc ứng dụng) cuối của dự án.

Song song đó, một bước tiến quan trọng khác là dữ liệu sau khi được làm sạch này cũng sẽ được sử dụng để xây dựng nền tảng cho tác vụ Retrieval Augmented Generation (*LightRAG*[2]). Cụ thể, nhóm sẽ tiến hành nhúng toàn bộ nội dung đã làm sạch này thông qua một số mô hình embedding (ví dụ như các mô hình từ OpenAI[3], Sentence Transformers[8], hoặc các giải pháp *LightRAG* tích hợp sẵn) để tạo ra các vector đại diện. Điểm đặc biệt ở đây là nhóm tận dụng extension pgvector mạnh mẽ ngay trong PostgreSQL. Điều này cho phép PostgreSQL không chỉ đóng vai trò là một cơ sở dữ liệu quan hệ truyền thống mà còn hoạt động như một vector database hiệu suất cao. Các vector dữ liệu này sẽ là "bộ não" cho hệ thống RAG, cho phép LLM truy xuất các thông tin liên quan và cập nhật nhất từ chính cơ sở dữ liệu của nhóm để tạo ra các câu trả lời chính xác, ngữ cảnh và đáng tin cậy hơn cho người dùng cuối.

Quy trình này đảm bảo rằng dữ liệu không chỉ được thu thập và làm sạch một cách hiệu

³<https://www.postgresql.org/>

```

1  # Đại học An ninh nhân dân (ANS)
2  ## Giới thiệu
3  Trường Đại học An ninh nhân dân (ANS) là một trong những cơ sở đào tạo hàng đầu của ngành Công an, nổi tiếng về đào tạo các
  chuyên ngành về an ninh và điều tra. Các ngành mạnh và uy tín nhất của trường bao gồm: An ninh mạng và phòng chống tội phạm
  sử dụng công nghệ cao, Trinh sát an ninh, và Điều tra hình sự. Trường không được xếp hạng chính thức trong các bảng xếp
  hạng đại học do tính chất đặc thù của ngành đào tạo. Tuy nhiên, ANS được đánh giá cao về chất lượng đào tạo, với đội ngũ
  giảng viên có trình độ cao và cơ sở vật chất hiện đại. Nhiều sinh viên tốt nghiệp từ trường đã trở thành những cán bộ công
  an xuất sắc, đảm nhiệm các vị trí quan trọng trong lực lượng Công an nhân dân.
4  ## Tin Tức Mới Nhất
5  * Website chính thức của trường: <https://dhannd.edu.vn>
6  * Trang thông tin tuyển sinh: <https://dhannd.edu.vn/Thong-tin-tuyen-sinh>
7
8
9  ## 1. Thay đổi chính trong năm 2025
10 ### Chỉ tiêu tuyển sinh
11 * Tổng chỉ tiêu tuyển sinh đại học chính quy: 260 chỉ tiêu (234 nam, 26 nữ)
12 * Khu vực tuyển sinh: Phía Nam (từ Đà Nẵng đến Cà Mau)
13 * Phân bổ theo vùng tuyển sinh:
14   * Vùng 4 (các tỉnh miền Trung): 50 chỉ tiêu
15   * Vùng 5 (các tỉnh Tây Nguyên): 40 chỉ tiêu
16   * Vùng 6 (các tỉnh Đông Nam Bộ): 80 chỉ tiêu
17   * Vùng 7 (các tỉnh Tây Nam Bộ): 75 chỉ tiêu
18   * Vùng 8 (chiến sĩ nghĩa vụ): 15 chỉ tiêu
19
20
21 ### Phương thức tuyển sinh
22 * Bài thi đánh giá của Bộ Công an gồm 3 phần:
23   * Tự luận bắt buộc: 1 câu nghị luận xã hội (25 điểm)
24   * Trắc nghiệm bắt buộc:
25     * Toán: 35 câu (35 điểm)
26     * Lịch sử: 10 câu (10 điểm)
27     * Ngoại ngữ (Tiếng Anh): 20 câu (15 điểm)
28   * Trắc nghiệm tự chọn: 15 câu (15 điểm), chọn 1 trong 4 môn:
29     * Vật lý (CA1)
30     * Hóa học (CA2)
31     * Sinh học (CA3)
32     * Địa lý (CA4)

```

Hình 2.3: Dữ liệu sau khi được làm sạch

quả mà còn được lưu trữ và chuẩn bị một cách tối ưu cho cả việc truy xuất thông thường lẫn các ứng dụng AI tiên tiến như LightRAG.

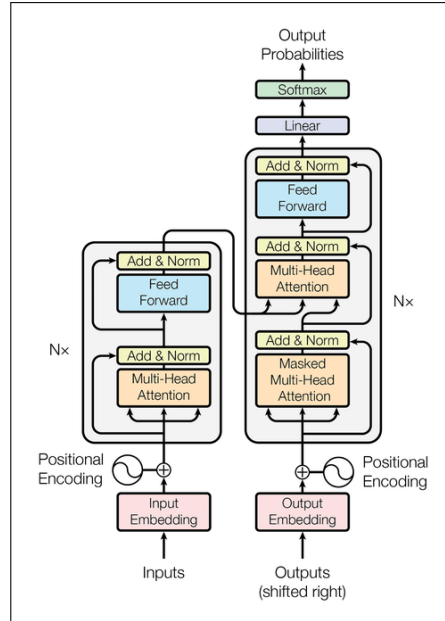
2.2 Thiết kế và triển khai mô hình

Nhóm sử dụng *Qwen 2.5 7B* [6] với kiến trúc Yet another RoPE-based Network (YaRN) [4] với nền tảng Transformer 2.4. YaRN, với kiến trúc Transformer cải tiến, tận dụng các thành phần như Rotary Position Embedding (RoPE) [12], *SwiGLU* [11], *RMSNorm* [15], *Attention QKV bias* [14], và *tied word embeddings* [5], mang lại hiệu suất vượt trội trong xử lý ngữ cảnh dài và các tác vụ ngôn ngữ phức tạp so với các mô hình LLMs open-source khác. Phần này trình bày chi tiết quá trình tích hợp mô hình này vào ứng dụng web, bao gồm các khía cạnh kỹ thuật, lợi ích và thách thức.

Kiến trúc Transformer là nền tảng của các mô hình ngôn ngữ hiện đại, bao gồm YaRN. Transformer sử dụng cơ chế self-attention để xử lý các chuỗi văn bản song song, cho phép mô hình nắm bắt mối quan hệ giữa các token bất kể khoảng cách. Trong YaRN, Transformer hoạt động dưới dạng decoder-only, tối ưu hóa cho việc tạo văn bản, như trả lời câu hỏi hoặc đưa ra khuyến nghị trong tư vấn tuyển sinh.

Với khả năng mở rộng cửa sổ ngữ cảnh của YaRN, mô hình có thể duy trì các cuộc hội thoại dài, ghi nhớ lịch sử tương tác của người dùng và cung cấp các phản hồi phù hợp với ngữ cảnh cụ thể. Các thành phần chính của kiến trúc YaRN được tận dụng như sau:

- **RoPE:** Được sử dụng để mã hóa vị trí của các token trong chuỗi, RoPE cho phép mô hình xử lý hiệu quả các câu hỏi dài hoặc các chuỗi thông tin phức tạp, chẳng hạn như khi học sinh yêu cầu tư vấn chi tiết về nhiều trường đại học hoặc so sánh các ngành học. Điều này đặc biệt hữu ích trong việc cung cấp câu trả lời mạch lạc và liên quan khi người dùng đặt các câu hỏi nối tiếp.
- **SwiGLU:** Hàm kích hoạt *SwiGLU* cải thiện khả năng học phi tuyến của mô



Hình 2.4: Kiến trúc Transformer

hình, giúp chatbot hiểu sâu hơn các sắc thái trong ngôn ngữ tiếng Việt, bao gồm các thuật ngữ giáo dục chuyên biệt như “điểm chuẩn” hay “ngành đào tạo”. Điều này đảm bảo rằng các phản hồi của chatbot không chỉ chính xác mà còn tự nhiên và dễ hiểu.

- **RMSNorm:** *RMSNorm* được sử dụng để chuẩn hóa các lớp trong mô hình, tăng cường sự ổn định và hiệu quả trong quá trình Fine-tune và suy luận. Điều này giúp hệ thống duy trì hiệu suất ổn định ngay cả khi xử lý khối lượng lớn yêu cầu từ người dùng đồng thời.
- **Attention QKV bias:** *Cơ chế Attention QKV bias* tăng cường khả năng tập trung của mô hình vào các phần quan trọng của ngữ cảnh, chẳng hạn như thông tin về điểm thi hoặc sở thích ngành học của học sinh, từ đó cung cấp các khuyến nghị cá nhân hóa chính xác hơn.
- **Tied word embeddings:** Việc sử dụng *tied word embeddings* giúp giảm kích thước mô hình và tối ưu hóa hiệu suất, đặc biệt quan trọng khi triển khai trên môi trường web với tài nguyên hạn chế, đảm bảo thời gian phản hồi nhanh chóng cho người dùng.

2.3 Fine-tune mô hình

Framework: Nhóm sử dụng thư viện PEFT để tinh chỉnh mô hình, với tài nguyên là GPU RTX3090 (Thuê server).

Tinh chỉnh với LoRA: Vì tiếng Việt chưa được hỗ trợ chính thức bởi Code Gemma 2B, chúng tôi áp dụng kỹ thuật LoRA (Low-Rank Adaptation) để tinh chỉnh mô hình cho nhu cầu cụ thể. LoRA là một phương pháp tinh chỉnh hiệu quả, cho phép điều chỉnh các mô hình ngôn ngữ lớn với tài nguyên tính toán tối thiểu.

Lý do chọn LoRA bao gồm:

- **Tập huấn luyện nhỏ:** LoRA chỉ thêm một số lượng nhỏ tham số (ma trận hạng thấp) vào mô hình, phù hợp với các trường hợp có ít dữ liệu huấn luyện.
- **Giữ nguyên mô hình đã huấn luyện:** LoRA giữ nguyên trọng số của mô hình gốc và chỉ thêm các ma trận hạng thấp để học các điều chỉnh đặc thù cho tác vụ, đảm bảo kiến thức nền tảng của mô hình không bị mất đi.
- **Giảm chi phí tính toán/lưu trữ:** Với số lượng tham số bổ sung rất nhỏ, LoRA giúp giảm đáng kể chi phí tính toán và lưu trữ.

Chi tiết huấn luyện:

- Tổng số tham số: 7,133,084,160 (16.33 GB)
- Số tham số có thể huấn luyện: 8,322,744 (31.34 MB)
- LoRA Rank: 32

Cấu hình cho quá trình train

- Tốc độ học (Learning Rate): 5×10^{-5} để đảm bảo cập nhật từ từ và tránh lỗi lớn
- Weight Decay: 0.01 nhằm chính quy hóa mô hình và tránh overfitting bằng cách phạt các trọng số lớn
- Clipnorm: 1.0 để tránh gradient quá lớn, tăng độ ổn định trong huấn luyện
- Các tham số loại trừ: Tham số chuẩn hóa tầng (Layer normalization) và hệ số lệch (bias) được loại trừ khỏi weight decay vì chúng thường ổn định và không cần chính quy hóa

2.4 Đánh giá hiệu suất mô hình

2.4.1 Mục tiêu

Mục tiêu của phần này là so sánh hiệu quả giữa hai mô hình:

- **Base Model:** Mô hình gốc chưa được tinh chỉnh.
- **LoRA Fine-Tuned Model:** Mô hình đã được tinh chỉnh bằng kỹ thuật Low-Rank Adaptation (LoRA).

Việc đánh giá được thực hiện dựa trên tập 10 mẫu kiểm thử, sử dụng các chỉ số BLEU và ROUGE (bao gồm ROUGE-1, ROUGE-2, ROUGE-L) để đo lường chất lượng sinh văn bản so với các đáp án tham chiếu.

Bảng 2.1: So sánh chỉ số giữa Base Model và LoRA Fine-Tuned Model

Chỉ số	Base Model	LoRA Fine-Tuned	Chênh lệch
BLEU	0.1225	0.1342	+0.0117
ROUGE-1	0.4488	0.4625	+0.0137
ROUGE-2	0.2628	0.2709	+0.0081
ROUGE-L	0.3542	0.3628	+0.0086

2.4.2 Kết quả

2.4.3 Phân tích chi tiết

- **BLEU:** Chỉ số BLEU tăng nhẹ (+0.0117), cho thấy mô hình fine-tuned tạo ra các cụm từ gần hơn với đáp án tham chiếu, phản ánh khả năng học được các mẫu biểu đạt chính xác hơn sau tinh chỉnh.
- **ROUGE-1:** Tăng vừa phải (+0.0137), phản ánh độ bao phủ từ đơn (unigram recall) đã được cải thiện. Mô hình học được cách dùng từ sát ngữ cảnh hơn mà vẫn giữ được độ đa dạng.
- **ROUGE-2:** Tăng nhẹ (+0.0081), thể hiện khả năng tái tạo tốt hơn các cụm hai từ liên tiếp, góp phần làm câu mạch lạc và tự nhiên hơn.
- **ROUGE-L:** Tăng ổn định (+0.0086), cho thấy sự cải thiện về mặt cấu trúc ngôn ngữ trong đầu ra của mô hình sau tinh chỉnh.

2.4.4 Kết luận

Từ các kết quả trên, có thể đưa ra các nhận định sau:

- Việc áp dụng kỹ thuật LoRA đã mang lại cải thiện đáng kể về chất lượng sinh văn bản của mô hình, cả về độ chính xác và cấu trúc.
- Tất cả các chỉ số đánh giá đều tăng nhẹ nhưng đồng nhất, cho thấy quá trình fine-tuning bằng LoRA là hiệu quả và ổn định.
- Điều này chứng minh rằng LoRA có thể giúp mô hình thích nghi tốt hơn với các ngữ cảnh cụ thể mà không làm giảm chất lượng ngôn ngữ tổng thể.

Chương 3

Sản phẩm của chúng tôi

3.1 Mức độ tích hợp mô hình vào ứng dụng web

Lợi ích của việc tích hợp YaRN[4]: Sự tích hợp của mô hình YaRN mang lại nhiều lợi ích cho ứng dụng web:

- **Xử lý ngữ cảnh dài:** Nhờ khả năng mở rộng cửa sổ ngữ cảnh, YaRN cho phép chatbot duy trì các cuộc hội thoại dài, ghi nhớ các chi tiết từ các tương tác trước đó, chẳng hạn như điểm thi hoặc sở thích của học sinh, để cung cấp tư vấn liên tục và nhất quán.
- **Cá nhân hóa nâng cao:** Mô hình sử dụng dữ liệu người dùng (như kết quả trắc nghiệm, điểm thi, hoặc lịch sử tìm kiếm) để đưa ra các gợi ý ngành học và trường đại học phù hợp, giúp học sinh cảm thấy được hỗ trợ một cách riêng biệt.
- **Hiệu suất tối ưu:** Các thành phần như SwiGLU và RMSNorm giúp mô hình hoạt động hiệu quả hơn, giảm thời gian phản hồi và đảm bảo trải nghiệm người dùng mượt mà, ngay cả khi có nhiều người dùng truy cập đồng thời.
- **Khả năng hiểu tiếng Việt:** YaRN được tinh chỉnh để xử lý ngôn ngữ tiếng Việt một cách tự nhiên, bao gồm các thuật ngữ giáo dục và sắc thái văn hóa, giúp chatbot giao tiếp thân thiện và dễ hiểu như một cố vấn thực thụ.

Thách thức trong tích hợp: Mặc dù YaRN mang lại nhiều lợi ích, việc tích hợp mô hình này vào ứng dụng web cũng đối mặt với một số thách thức:

- **Tài nguyên tính toán lớn:** Mô hình YaRN, với kiến trúc phức tạp và khả năng xử lý ngữ cảnh dài, yêu cầu tài nguyên tính toán đáng kể, đặc biệt khi triển khai trên quy mô lớn.
Hạn chế: Chi phí vận hành cao và thời gian phản hồi có thể tăng nếu không có cơ sở hạ tầng mạnh mẽ.
- **Tinh chỉnh ngôn ngữ tiếng Việt:** Mặc dù YaRN có khả năng xử lý ngôn ngữ tốt, việc tinh chỉnh để hiểu các sắc thái tiếng Việt và thuật ngữ giáo dục đòi hỏi dữ liệu huấn luyện chuyên biệt và nguồn lực lớn.
Hạn chế: Nếu không được tinh chỉnh tốt, chatbot có thể trả lời không chính xác hoặc thiếu tự nhiên, làm giảm trải nghiệm người dùng.

- **Tích hợp API phức tạp:** Việc kết nối mô hình YaRN với frontend của ứng dụng web qua API đòi hỏi thiết kế kỹ lưỡng để đảm bảo tốc độ, độ tin cậy và bảo mật.
Hạn chế: Lỗi tích hợp hoặc độ trễ API có thể ảnh hưởng đến hiệu suất tổng thể của hệ thống.
- **Bảo mật dữ liệu người dùng:** Dữ liệu cá nhân của học sinh (như điểm thi, sở thích) được sử dụng để cá nhân hóa tư vấn cần được bảo vệ nghiêm ngặt khi truyền qua API.
Hạn chế: Rò rỉ dữ liệu hoặc vi phạm bảo mật có thể gây mất lòng tin và dẫn đến các vấn đề pháp lý.

Để tận dụng tối đa tiềm năng của YaRN, hệ thống cần được triển khai trên cơ sở hạ tầng đám mây có khả năng mở rộng, kết hợp với các kỹ thuật tối ưu hóa như caching và nén dữ liệu để giảm độ trễ. Đồng thời, việc tinh chỉnh mô hình với dữ liệu giáo dục Việt Nam và thiết kế giao diện chatbot thân thiện sẽ đảm bảo rằng nền tảng không chỉ cung cấp thông tin chính xác mà còn mang lại trải nghiệm người dùng mượt mà và đáng tin cậy.

3.2 Giao diện và trải nghiệm của người dùng

Giao diện người dùng (UI) và trải nghiệm người dùng (UX) là yếu tố then chốt để đảm bảo nền tảng tư vấn tuyển sinh AI trở nên dễ sử dụng, thân thiện và hiệu quả đối với học sinh THPT tại Việt Nam. Với mục tiêu hỗ trợ định hướng học vấn và nghề nghiệp một cách toàn diện, giao diện được thiết kế hiện đại, trực quan và tối ưu hóa cho cả thiết bị di động lẫn máy tính, đảm bảo tính linh hoạt và khả năng tiếp cận ở mọi bối cảnh sử dụng. Sử dụng công nghệ React[7], TypeScript[13] và ChakraUI[1], nền tảng không chỉ đạt được tính responsive cao, tốc độ tải trang nhanh mà còn cung cấp giao diện linh hoạt, dễ dàng tùy chỉnh để đáp ứng nhu cầu đa dạng của người dùng. Phần này mô tả chi tiết thiết kế giao diện, các tính năng chính, cách chúng được triển khai trên các trang cụ thể và đóng góp của chúng vào việc nâng cao trải nghiệm người dùng.

3.2.1 Thành phần giao diện chính

- **Thanh điều hướng:** Thanh điều hướng cố định ở đầu trang, bao gồm logo “TuyensinhAI” với phong cách thiết kế tinh tế, các mục menu (Trang chủ, Danh sách trường, Tư vấn AI, Về chúng tôi) và nút đăng nhập/đăng ký tùy thuộc vào trạng thái người dùng. Thiết kế sử dụng tông màu xanh dương chủ đạo kết hợp với trắng, tạo cảm giác chuyên nghiệp, đáng tin cậy và thân thiện với lĩnh vực giáo dục.
- **Banner trang chủ:** Trang chủ nổi bật với một banner động được thiết kế dưới dạng carousel, hiển thị thông điệp hấp dẫn “Khám phá tương lai học vấn của bạn với AI” cùng hình ảnh minh họa trường đại học được chọn lọc kỹ lưỡng. Banner tích hợp các nút kêu gọi hành động (call-to-action) như “Bắt đầu tư vấn” và “Xem danh sách trường”, được đặt ở vị trí nổi bật để khuyến khích người dùng tương tác ngay lập tức, tăng cường sự tham gia và khám phá.
- **Danh sách trường đại học:** Giao diện danh sách trường được trình bày dưới dạng lưới các thẻ (card-based layout), mỗi thẻ chứa thông tin cơ bản như tên trường, địa điểm, hình ảnh đại diện và nút “Chi tiết” để chuyển hướng. Người dùng có thể

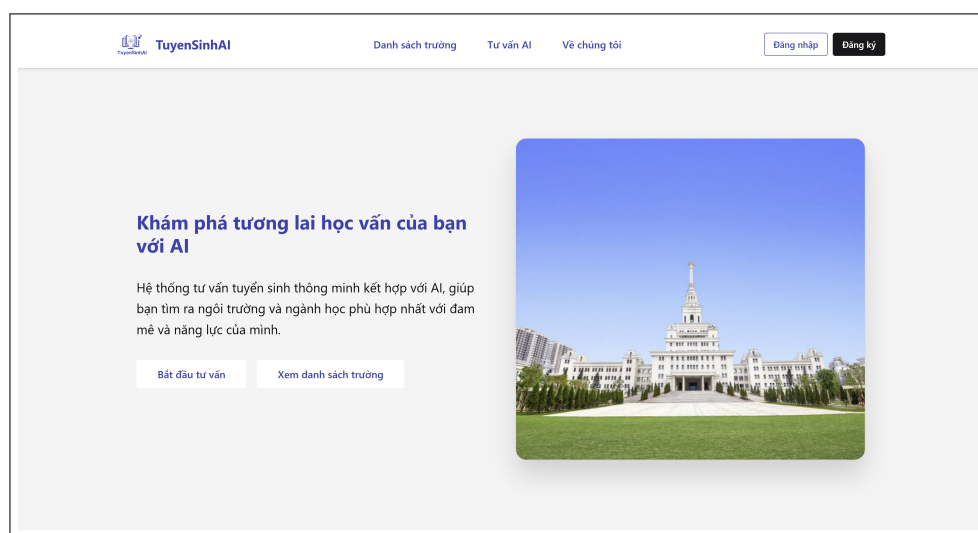
sử dụng các bộ lọc nâng cao (theo khu vực, ngành học, điểm chuẩn) và thanh tìm kiếm để cá nhân hóa kết quả, mang lại trải nghiệm tìm kiếm linh hoạt và hiệu quả.

- **Giao diện chatbot AI:** Giao diện chatbot được thiết kế với phong cách tối giản, tập trung vào khung trò chuyện ở trung tâm, hiển thị tin nhắn chào hỏi từ chatbot (ví dụ: “Xin chào! Tôi có thể giúp gì cho bạn về tuyển sinh đại học?”) cùng ô nhập liệu và các gợi ý câu hỏi nhanh. Sử dụng bảng màu xanh nhạt và trắng, giao diện tạo cảm giác thân thiện, dễ tiếp cận, đồng thời hỗ trợ lịch sử hội thoại để người dùng theo dõi dễ dàng.
- **Chân trang:** Chân trang được bố trí gọn gàng với các thông tin liên hệ (email, số điện thoại), liên kết đến các nền tảng mạng xã hội (Facebook, Instagram, LinkedIn), và thông tin bản quyền. Thiết kế này không chỉ hỗ trợ người dùng trong việc tìm kiếm trợ giúp mà còn tăng cường sự kết nối với cộng đồng, đồng thời đảm bảo tính chuyên nghiệp và dễ điều hướng.

3.2.2 Giao diện chính

Giao diện chính của website bao gồm 7 trang chính, mỗi trang được thiết kế để phục vụ các nhu cầu cụ thể của người dùng, từ đăng nhập, đăng ký đến khám phá thông tin trường học và nhận tư vấn. Dưới đây là chi tiết các trang:

- **Trang chủ:** Trang chủ là cổng vào chính với banner động, thông tin nổi bật về các tính năng (tư vấn AI, danh sách trường, lộ trình nghề nghiệp), và nút kêu gọi hành động. Thiết kế nhấn mạnh sự năng động, khuyến khích người dùng khám phá ngay lập tức.



Hình 3.1: Trang chủ của website với banner động và các nút kêu gọi hành động.

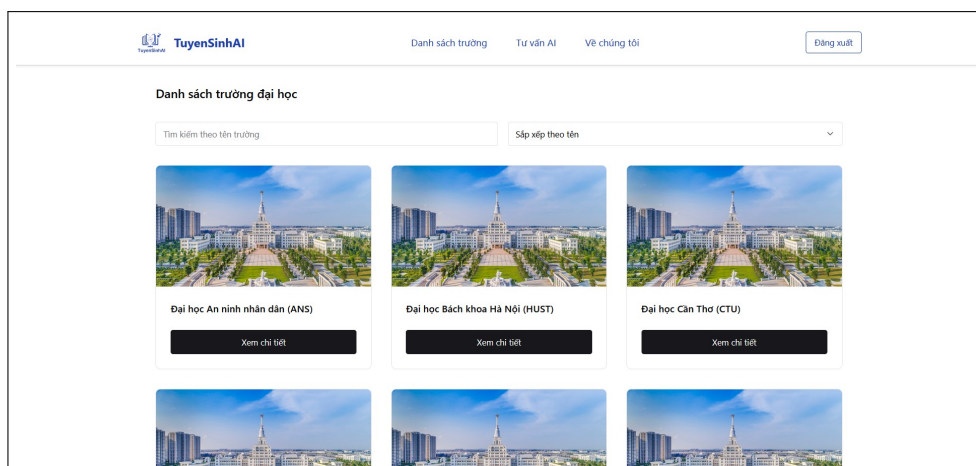
- **Trang "Đăng nhập":** Giao diện đăng nhập được thiết kế với form đơn giản, gồm các trường nhập liệu cho email và mật khẩu, nút “Đăng nhập” màu đen nổi bật, và liên kết “Đã có tài khoản? Đăng nhập” để chuyển hướng. Thiết kế tập trung vào sự tối giản, sử dụng nền trắng với viền nhẹ, giúp người dùng nhanh chóng truy cập hệ thống mà không bị phân tâm.

Hình 3.2: Giao diện trang đăng nhập với form đơn giản và nút “Đăng nhập” nổi bật.

- **Trang "Đăng ký":** Giao diện đăng ký mở rộng hơn, bao gồm các trường nhập liệu cho họ và tên, email, mật khẩu, xác nhận mật khẩu, và giới tính (Nam/Nữ) với tùy chọn radio button. Nút “Đăng ký” màu đen nổi bật, cùng với chú thích bắt buộc (*) và liên kết “Đã có tài khoản? Đăng nhập” để hỗ trợ người dùng mới. Bố cục rõ ràng, dễ hiểu, đảm bảo quy trình đăng ký thuận tiện.

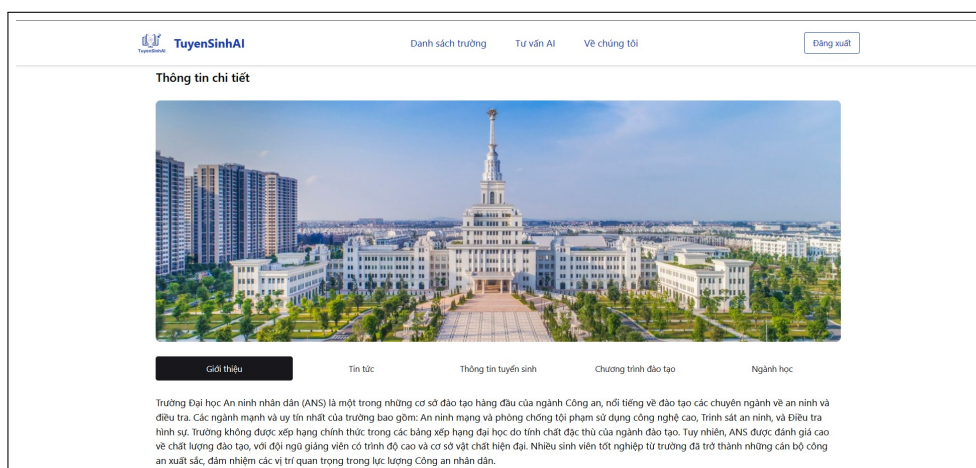
Hình 3.3: Giao diện trang đăng ký với các trường nhập liệu chi tiết và nút “Đăng ký”.

- **Trang "Danh sách các trường đại học":** Giao diện hiển thị danh sách trường dưới dạng lưới các thẻ, mỗi thẻ chứa hình ảnh trường, tên trường (ví dụ: Đại học An ninh nhân dân, Đại học Bách khoa Hà Nội, Đại học Cần Thơ), và nút “Xem chi tiết”. Người dùng có thể sử dụng thanh tìm kiếm “Tìm kiếm theo tên trường” và bộ lọc “Sắp xếp theo tên” hoặc điểm chuẩn, với thiết kế responsive để hiển thị mượt mà trên mọi thiết bị.
- **Trang "Thông tin chi tiết trường đại học":** Giao diện chi tiết bao gồm hình ảnh trường lớn, các tab điều hướng (Giới thiệu, Tin tức, Thông tin tuyển sinh,



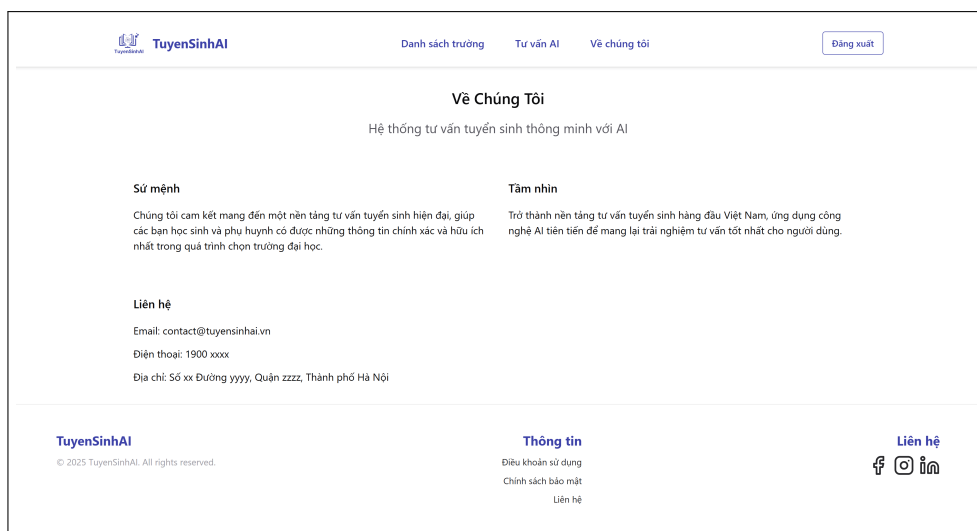
Hình 3.4: Giao diện danh sách trường đại học với lưới thẻ và bộ lọc tìm kiếm.

Chương trình đào tạo, Ngành học) và nội dung chi tiết như địa chỉ, học phí, điểm chuẩn. Thiết kế sử dụng văn bản rõ ràng, hình ảnh minh họa chất lượng cao.

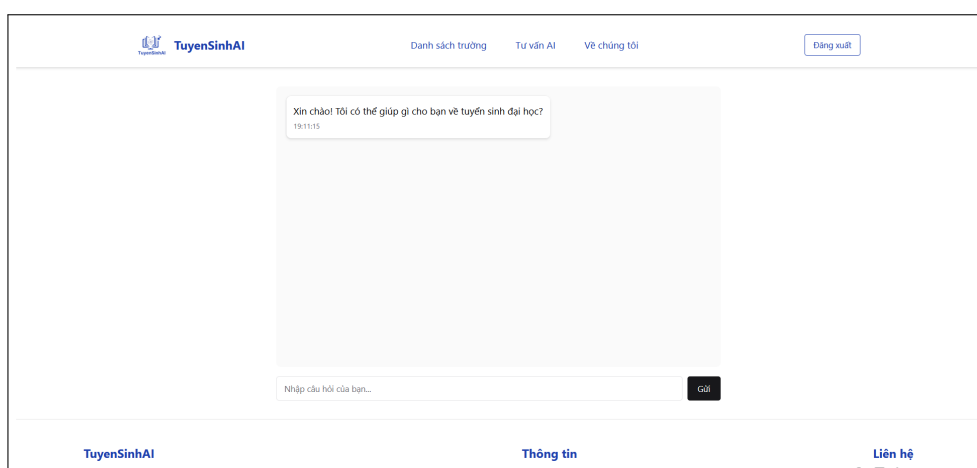


Hình 3.5: Giao diện thông tin chi tiết trường đại học với tab điều hướng và hình ảnh minh họa.

- **Trang "Về chúng tôi":** Giao diện bao gồm tiêu đề “Về Chúng Tôi”, phần sứ mệnh (“Hỗ trợ tư vấn tuyển sinh thông minh với AI”), tầm nhìn (“Trở thành nền tảng tư vấn tuyển sinh và hướng nghiệp hàng đầu tại Việt Nam”), và thông tin liên hệ. Phần chân trang chứa thông tin bản quyền, biểu tượng mạng xã hội, và nút “Liên hệ”, tạo sự chuyên nghiệp và dễ tiếp cận.
- **Trang "Tư vấn AI":** Giao diện này tập trung vào khung chatbot với các tính năng như trò chuyện thời gian thực, lịch sử hội thoại, và gợi ý câu hỏi (ví dụ: “Hỏi về điểm chuẩn”, “Tư vấn ngành học”). Thiết kế tối ưu cho cả văn bản và tương lai có thể mở rộng cho giọng nói, với giao diện thân thiện và phản hồi nhanh.



Hình 3.6: Giao diện trang “Về chúng tôi” với sứ mệnh, tầm nhìn và thông tin liên hệ.



Hình 3.7: Giao diện chatbot AI với tin nhắn chào hỏi, ô nhập liệu, và gợi ý câu hỏi, tạo cảm giác thân thiện và dễ sử dụng.

3.2.3 Khả năng tiếp cận và phản hồi người dùng

Nền tảng tuân thủ tiêu chuẩn Web Content Accessibility Guidelines (WCAG)¹ để hỗ trợ người dùng khuyết tật, bao gồm việc đảm bảo tương phản màu sắc phù hợp (tối thiểu 4.5:1), tích hợp tính năng đọc màn hình, và hỗ trợ điều hướng bằng bàn phím. Hệ thống tích hợp khảo sát nhanh sau khi sử dụng chatbot và các trang chính (đăng nhập, danh sách trường) để thu thập phản hồi từ người dùng, giúp cải tiến liên tục giao diện, tính năng, và trải nghiệm. Phản hồi được phân tích định kỳ để tối ưu hóa trải nghiệm, đảm bảo nền tảng đáp ứng tốt nhất nhu cầu của học sinh.

3.2.4 Kết luận

Giao diện và trải nghiệm người dùng của nền tảng được thiết kế để mang lại sự dễ dàng, đáng tin cậy và cá nhân hóa, giúp học sinh THPT tại Việt Nam định hướng học vấn và nghề nghiệp một cách hiệu quả. Bằng cách kết hợp thiết kế hiện đại với công nghệ tiên

¹<https://www.w3.org/WAI/standards-guidelines/wcag/>

tiến như React, TypeScript, và ChakraUI, cùng với việc tập trung vào nhu cầu người dùng và tuân thủ các tiêu chuẩn tiếp cận, nền tảng tạo điều kiện thuận lợi cho học sinh trong hành trình lựa chọn tương lai học tập và sự nghiệp.

Chương 4

Tổng kết

4.1 Tổng quan

Dự án phát triển website tư vấn tuyển sinh AI nhằm đáp ứng nhu cầu cấp thiết về một nền tảng định hướng học vấn và nghề nghiệp dành cho học sinh THPT tại Việt Nam. Với sự phát triển nhanh chóng của hệ thống giáo dục đại học, bao gồm sự gia tăng số lượng trường và sự đa dạng hóa các ngành học, học sinh thường gặp khó khăn trong việc đưa ra quyết định phù hợp do thiếu thông tin đáng tin cậy. Họ thường dựa vào các nguồn không chính thức như truyền miệng, mạng xã hội, hoặc các diễn đàn trực tuyến thiếu kiểm chứng, dẫn đến các lựa chọn mang tính chủ quan, thiếu cơ sở khoa học. Điều này có thể gây lãng phí thời gian, tài chính và ảnh hưởng tiêu cực đến tương lai nghề nghiệp của học sinh.

Nền tảng được xây dựng dựa trên sự kết hợp giữa AI, công nghệ thu thập dữ liệu web, và kỹ thuật truy xuất tri thức thông minh (**MiniRAG**). Đặc biệt, việc tích hợp mô hình ngôn ngữ lớn YaRN, với kiến trúc transformers cải tiến bao gồm RoPE, SwiGLU, RMSNorm, Attention QKV bias, và tied word embeddings, cho phép hệ thống xử lý ngữ cảnh dài, cung cấp tư vấn chính xác và cá nhân hóa. Hệ thống cung cấp thông tin chi tiết về 163 trường đại học trên toàn quốc, hỗ trợ học sinh tìm hiểu lộ trình học tập, triển vọng nghề nghiệp, và tương tác với chatbot AI thân thiện, mang lại trải nghiệm tư vấn toàn diện và khoa học.

Các tính năng chính của nền tảng bao gồm:

- Tư vấn cá nhân hóa dựa trên thông tin điểm thi, sở thích và mục tiêu nghề nghiệp.
- Cơ sở dữ liệu chi tiết về trường đại học, bao gồm điểm chuẩn, học phí, và học bổng.
- Thông tin về lộ trình học tập và triển vọng nghề nghiệp.
- Chatbot AI tích hợp YaRN, hỗ trợ tư vấn nhanh chóng và cá nhân hóa.

Dự án đối mặt với các thách thức như dữ liệu thiếu đồng bộ, hiệu suất chatbot AI, thiếu lòng tin từ người dùng, yêu cầu giao diện responsive, và bảo mật dữ liệu cá nhân. Những thách thức này đòi hỏi đầu tư vào công nghệ, dữ liệu, và chiến lược truyền thông để đáp ứng nhu cầu của học sinh Việt Nam.

4.2 Định hướng tương lai

Trong tương lai, dự án hướng đến trở thành nền tảng tư vấn tuyển sinh và hướng nghiệp hàng đầu tại Việt Nam, không chỉ phục vụ học sinh THPT mà còn mở rộng đến sinh viên, phụ huynh, và các tổ chức giáo dục. Các định hướng phát triển cụ thể bao gồm:

- **Tăng cường độ chính xác của dữ liệu:** Cập nhật và mở rộng cơ sở dữ liệu để bao gồm thông tin về các trường cao đẳng, học viện, và chương trình đào tạo quốc tế tại Việt Nam. Tích hợp dữ liệu thị trường lao động, như nhu cầu tuyển dụng và kỹ năng cần thiết, để cung cấp tư vấn toàn diện về triển vọng nghề nghiệp.
- **Cải tiến chatbot AI:** Nâng cấp chatbot YaRN[4] thông qua tích hợp với dữ liệu giáo dục Việt Nam, cải thiện khả năng xử lý ngôn ngữ tiếng Việt, và tích hợp các phương thức tư vấn đa dạng như văn bản và giọng nói. Khám phá khả năng kết hợp các mô hình AI bổ sung để tăng độ chính xác và tính cá nhân hóa của tư vấn.
- **Xây dựng ứng dụng di động:** Phát triển ứng dụng di động để tăng khả năng tiếp cận, tận dụng tỷ lệ sử dụng điện thoại cao tại Việt Nam. Ứng dụng sẽ cung cấp các tính năng như thông báo thời gian thực về hạn nộp hồ sơ, cập nhật điểm chuẩn, và tư vấn ngoại tuyến.
- **Tạo hệ sinh thái cộng đồng giáo dục:** Xây dựng một cộng đồng trực tuyến nơi học sinh, sinh viên, và cựu sinh viên chia sẻ kinh nghiệm, đánh giá trường học, và thảo luận về ngành nghề. Điều này sẽ tăng tính tương tác và cung cấp thông tin thực tế, nâng cao độ tin cậy của nền tảng.
- **Đảm bảo bảo mật dữ liệu:** Áp dụng các giải pháp bảo mật tiên tiến, như mã hóa đầu cuối và xác thực đa yếu tố, để bảo vệ thông tin cá nhân. Công khai minh bạch chính sách quyền riêng tư nhằm xây dựng lòng tin từ người dùng.
- **Tích hợp công nghệ tiên tiến:** Khám phá việc sử dụng thực tế ảo - Virtual Reality (VR) và thực tế tăng cường - Augmented Reality (AR) để cung cấp trải nghiệm tham quan trường đại học ảo, giúp học sinh hình dung rõ môi trường học tập và cơ sở vật chất.
- **Tăng cường tư vấn cá nhân hóa:** Phát triển các thuật toán AI nâng cao để phân tích dữ liệu người dùng (điểm thi, sở thích, lịch sử tương tác) nhằm cung cấp tư vấn phù hợp hơn, giúp học sinh đưa ra quyết định dựa trên thông tin định lượng và logic.

Với các định hướng này, dự án không chỉ giải quyết các thách thức hiện tại mà còn hướng đến xây dựng một hệ sinh thái giáo dục toàn diện, kết nối học sinh với cơ hội học tập và nghề nghiệp một cách hiệu quả. Bằng cách tận dụng công nghệ AI tiên tiến như YaRN và liên tục cải tiến, nền tảng sẽ góp phần nâng cao chất lượng định hướng giáo dục, thu hẹp khoảng cách giữa đào tạo và nhu cầu thị trường lao động, và hỗ trợ thế hệ trẻ Việt Nam xây dựng tương lai vững chắc.

Tài liệu tham khảo

- [1] Chakra UI. Chakra ui: Simple, modular and accessible ui components for react applications, n.d. Accessed: 2025-05-07.
- [2] Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. Lightrag: Simple and fast retrieval-augmented generation, 2025.
- [3] OpenAI. Api reference - embeddings. OpenAI Platform Documentation, n.d. Accessed: 7 May 2025.
- [4] Bowen Peng, Jeffrey Quesnelle, Honglu Fan, and Enrico Shippole. Yarn: Efficient context window extension of large language models, 2023.
- [5] Ofir Press and Lior Wolf. Using the output embedding to improve language models, 2017.
- [6] Qwen Team. Qwen2.5-7b, 2024. Accessed: 2025-05-07.
- [7] React Contributors. React – a javascript library for building user interfaces, n.d. Accessed: 2025-05-07.
- [8] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks, 2019.
- [9] Kenneth Reitz. Requests: Http for humans. <https://docs.python-requests.org/>, 2023. Accessed: 2025-05-07.
- [10] Selenium Project. Selenium webdriver. <https://www.selenium.dev/>, 2023. Accessed: 2025-05-07.
- [11] Noam Shazeer. Glue variants improve transformer, 2020.
- [12] Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding, 2023.
- [13] TypeScript Team. Typescript: Typed javascript at any scale, n.d. Accessed: 2025-05-07.
- [14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.
- [15] Biao Zhang and Rico Sennrich. Root mean square layer normalization, 2019.