

64GBBỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN
CHUYÊN NGÀNH TRÍ TUỆ NHÂN TẠO

2020



NGUYỄN THỊ PHÚ - 21110600

BÙI QUANG THIỆN - 21110656

Đề Tài :

**CHATBOT TƯ VẤN SINH VIÊN TRƯỜNG
ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

TIÊU LUẬN CHUYÊN NGÀNH TRÍ TUỆ NHÂN TẠO

GIÁO VIÊN HƯỚNG DẪN

THS. LÊ MINH TÂN

KHÓA 2021 – 2025

PHIẾU NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

Họ và tên Sinh viên 1: Nguyễn Thị Phú

MSSV: 21110600

Họ và tên Sinh viên 2: Bùi Quang Thiện

MSSV: 21110656

Ngành: Công nghệ thông tin

Tên đề tài: CHATBOT TỰ VÂN SINH VIÊN TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM

Họ và tên Giáo viên hướng dẫn: ThS. Lê Minh Tân

NHẬN XÉT

1. Về nội dung đề tài & khối lượng thực hiện:

2. Ưu điểm:

3. Khuyết điểm:

4. Đề nghị cho bảo vệ hay không ?

5. Đánh giá loại :

6. Điểm :

Tp. Hồ Chí Minh, ngày tháng năm 20

Giáo viên hướng dẫn

(Ký & ghi rõ họ tên)

PHIẾU NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN

Họ và tên Sinh viên 1: Nguyễn Thị Phú

MSSV: 21110600

Họ và tên Sinh viên 2: Bùi Quang Thiện

MSSV: 21110656

Ngành: Công nghệ thông tin

Tên đề tài: CHATBOT TƯ VẤN SINH VIÊN TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM

Họ và tên Giáo viên phản biện: TS. Phan Thị Huyền Trang

NHẬN XÉT

1. Về nội dung đề tài & khối lượng thực hiện:

2. Ưu điểm:

3. Khuyết điểm:

4. Đề nghị cho bảo vệ hay không ?

5. Đánh giá loại :

6. Điểm :

Tp. Hồ Chí Minh, ngày tháng năm 20

Giáo viên phản biện

(Ký & ghi rõ họ tên)

LỜI CẢM ƠN

Để hoàn thành tốt đề tài này, chúng em xin gửi lời cảm ơn chân thành đến các quý thầy cô trong khoa Công Nghệ Thông Tin đã luôn tận tình truyền đạt những kiến thức, kỹ năng bổ ích trong suốt quá trình học tập. Bên cạnh đó, chúng em xin cảm ơn trường Đại học Sư phạm Kỹ thuật TP.HCM đã tạo điều kiện về cơ sở vật chất cũng như các điều kiện cần thiết khác giúp chúng em có môi trường học tập và thực hiện đề tài một cách thuận lợi nhất.

Đặc biệt, chúng em xin gửi lời cảm ơn chân thành đến giảng viên, Thạc sĩ Lê Minh Tân, người đã trực tiếp hỗ trợ chúng em trong suốt quá trình thực hiện đề tài. Chúng em cảm ơn thầy đã đưa ra những lời khuyên, gợi ý từ kinh nghiệm thực tiễn của mình để định hướng cho nhóm đi đúng với yêu cầu của đề tài. Bên cạnh đó, chúng em cảm ơn thầy đã luôn giải đáp thắc mắc và đưa ra những góp ý, chỉnh sửa kịp thời giúp chúng em khắc phục nhược điểm và hoàn thành đúng thời hạn đã đề ra.

Đề tài được chúng em thực hiện trong khoảng thời gian không quá dài, với những kiến thức còn hạn chế về mặt kỹ thuật và kinh nghiệm trong việc triển khai một dự án Trí tuệ nhân tạo. Do đó, trong quá trình thực hiện đề tài có những thiếu sót là điều không thể tránh khỏi nên chúng em rất mong nhận được những ý kiến đóng góp quý báu từ quý thầy cô để có thể cải thiện về sau.

Cuối lời, chúng em kính chúc quý thầy cô luôn dồi dào sức khỏe và thành công hơn nữa trong sự nghiệp trồng người. Một lần nữa chúng em xin chân thành cảm ơn!

Tp. Hồ Chí Minh, ngày 07 tháng 12 năm 2024

Nhóm Sinh viên thực hiện

ĐỀ CƯƠNG TIỂU LUẬN CHUYÊN NGÀNH

Họ và Tên SV thực hiện 1: Nguyễn Thị Phú Mã số SV: 21110600

Họ và Tên SV thực hiện 2: Bùi Quang Thiện Mã số SV: 21110656

Thời gian làm tiểu luận: từ: 22/08/2024 Đến: 07/12/2024

Chuyên ngành: Trí tuệ nhân tạo

Tên đề tài: CHATBOT TỰ VẤN SINH VIÊN TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM

Giáo viên hướng dẫn: ThS. Lê Minh Tân

Nhiệm vụ của đề tài :

1. Nghiên cứu cơ sở lý thuyết: Tìm hiểu khái niệm về mô hình ngôn ngữ lớn, và phương pháp Retrieval Augmented Generation cho mô hình ngôn ngữ lớn và các công nghệ để triển khai phương pháp này, cũng như các kỹ thuật khác để tăng mức độ trải nghiệm người dùng như Cache, Học tăng cường.
2. Xây dựng Cơ sở dữ liệu: Thu thập, tiền xử lý dữ liệu Sổ tay Sinh viên năm 2023 và các câu hỏi Sinh viên thường hay gặp phải trên trang Tư vấn Sinh viên.
3. Thiết kế hệ thống Chatbot: Đề xuất kiến trúc tổng quan và xây dựng giao diện để minh họa cách người dùng sử dụng hệ thống.
4. Xây dựng, kiểm thử và tối ưu: Xây dựng, kiểm thử hệ thống với dữ liệu thực tế, tối ưu hiệu năng và độ chính xác của Chatbot.
5. Đưa ra giải pháp khác nếu phản hồi từ hệ thống chưa thỏa mãn yêu cầu người dùng: Xây dựng thêm trang đăng bài viết để người dùng có thể đăng tải thắc mắc trong trường hợp Chatbot không thể đưa ra câu trả lời hoặc thắc mắc quá chuyên sâu, chi tiết.

Đề cương viết đề tài :

MỤC LỤC

Chương 1: Tổng quan

- 1.1. Tính cấp thiết của đề tài
- 1.2. Mục tiêu của đề tài
- 1.3. Cách tiếp cận và phương pháp nghiên cứu
 - 1.3.1. Đối tượng nghiên cứu
 - 1.3.2. Phạm vi nghiên cứu
 - 1.3.3. Nội dung nghiên cứu
- 1.4. Phương pháp nghiên cứu

Chương 2: Cơ sở lý thuyết

- 2.1. Hiện thực hóa dự án
- 2.2. Giới thiệu về mô hình ngôn ngữ lớn
 - 2.2.1. Định nghĩa mô hình ngôn ngữ lớn
 - 2.2.2. Sự quan trọng của mô hình ngôn ngữ lớn
 - 2.2.3. Cách thức hoạt động của mô hình ngôn ngữ lớn
 - 2.2.4. Huấn luyện mô hình ngôn ngữ lớn
- 2.3. Giới thiệu về LangChain
 - 2.3.1. Định nghĩa LangChain
 - 2.3.2 Các module của LangChain
- 2.4. Giới thiệu về PostgreSQL
 - 2.4.1 Định nghĩa PostgreSQL
 - 2.4.2 Các tính năng của PostgreSQL
- 2.5 Giới thiệu về SQL Server
 - 2.5.1 Định nghĩa SQL Server
 - 2.5.2 Các tính năng của SQL Server
- 2.6 Giới thiệu về Chroma
 - 2.6.1. Cơ sở dữ liệu vector
 - 2.6.2. Định nghĩa Chroma
- 2.7. Giới thiệu về Python Django

- 2.7.1. Định nghĩa Django
- 2.7.2. Những điểm nổi bật của Django
- 2.7.2. Cách hoạt động của Django
- 2.8 Giới thiệu về Python Flask
 - 2.8.1. Định nghĩa Flask
 - 2.8.2. Các tính năng của Flask
- 2.9 Giới thiệu về SQLAlchemy
 - 2.9.1. Định nghĩa SQLAlchemy
 - 2.9.2. Các tính năng nổi bật của SQLAlchemy
- 2.10 Giới thiệu về RAG
 - 2.10.1. Định nghĩa RAG
 - 2.10.2. Thành phần của RAG
 - 2.10.3. Quy trình hoạt động của RAG
- 2.11 Giới thiệu về Cache
 - 2.11.1. Định nghĩa Cache
 - 2.11.2. Đặc điểm của Cache
 - 2.11.3. Phân loại Cache
 - 2.11.4. Tối ưu hóa phản hồi hệ thống bằng Cache cho câu hỏi lặp lại
 - 2.11.5. Quy trình truy vấn Cache của Chatbot
- 2.12 Giới thiệu về học tăng cường
- 2.13 Giới thiệu về Q-Learning
 - 2.13.1. Định nghĩa Q-Learning
 - 2.13.2. Các thành phần của Q-Learning
 - 2.13.3. Công thức của Q-Learning
- 2.14 Giới thiệu về Deep Q-Learning
 - 2.14.1 Nguyên lý hoạt động
 - 2.14.2 Kỹ thuật tăng cường cho Deep Q-Learning
 - 2.14.3 Công thức cập nhật

Chương 3: Tìm hiểu dữ liệu

- 3.1. Khảo sát hiện trạng trang tư vấn Sinh viên
- 3.2 Khảo sát Sổ tay Sinh viên năm 2023

3.3. Kết luận sau khi khảo sát

3.4 Tiết xử lý dữ liệu

3.4.1 Tiết xử lý dữ liệu Sổ tay Sinh viên

3.4.2 Tiết xử lý dữ liệu trang Tư vấn Sinh viên

Chương 4: Thiết kế hệ thống

4.1. Cấu hình phần cứng

4.2 Kiến trúc hệ thống

4.3 Mô tả cơ sở dữ liệu

4.3.1 Vector Database

4.3.2 FAQ

4.3.3 QuestionCount

4.3.4 Feedback

4.3.5. Cache

4.4 Thiết kế giao diện

4.4.1 Trang đăng ký

4.4.2 Trang đăng nhập

4.4.3 Trang bảng tin

4.4.4. Trang đăng bài viết

4.4.5. Trang chi tiết bài viết

4.4.6. Trang nhắn tin với chatbot

4.5 Môi trường học tăng cường cho việc lựa chọn 2 phản hồi FAQ

Chương 5: Chi tiết các tham số và siêu tham số

5.1. Mô hình ngôn ngữ lớn

5.2. Văn bản tài liệu

5.3. Môi trường Deep Q-Learning

5.4 Mạng Deep Q-Learning

5.5 Deep Q-Learning Agent

Chương 6: Kết quả đạt được

6.1 Kết quả

6.1.1 Về mặt lý thuyết

6.1.2 Về mặt sản phẩm

6.2 Kiểm thử mô hình

6.2.1 Phản hồi FAQ

6.2.2 Phản hồi từ RAG

6.2.3 Lịch sử Chat

6.2.4 Cache

Chương 7: Kết luận

7.1 Tổng kết

7.2 Đóng góp của đê tài

7.3 Hạn chế của đê tài

7.4 Hướng phát triển cho tương lai

Tài liệu tham khảo

[1] Mô hình ngôn ngữ lớn: <https://aws.amazon.com/vi/what-is/large-language-model/>

[2] Tìm hiểu về Langchain: <https://aws.amazon.com/vi/what-is/langchain/>

[3] Tìm hiểu PostgreSQL: <https://www.w3schools.com/postgresql/>

[4] Tìm hiểu tổng quan SQL Server: <https://itviec.com/blog/sql-server-la-gi/>

[5] RAG là gì? Ứng dụng của RAG: <https://lacviet.vn/retrieval-augmented-generation/>

[6] Bài báo khoa học về RAG: <https://arxiv.org/pdf/2312.10997>

[7] Deep Q-Learning: <https://vietnix.vn/deep-learning-la-gi/>

[8] LangChain: <https://python.langchain.com/docs/introduction>

[9] Tìm hiểu Q-Learning: <https://www.geeksforgeeks.org/q-learning-in-python/>

[10] Sổ tay SV 2023: <https://sao.hcmute.edu.vn/ArticleId/e8db4387-3f43-4770-a67c-90fa85ce009f/so-tay-sinh-vien-2023>

[11] Django: <https://docs.djangoproject.com/en/5.1/>

[12] Llama cpp: <https://github.com/ggerganov/llama.cpp>

KẾ HOẠCH THỰC HIỆN

STT	Thời gian	Công việc	Ghi chú
1	15/06 – 16/06/2024	Tìm kiếm dữ liệu sổ tay Sinh viên	
2	17/06 – 02/07/2024	Nghiên cứu thiết kế kiến trúc mô hình	
3	03/07 – 22/07/2024	Nghiên cứu xây dựng Vector Database	
4	23/07 – 20/08/2024	Nghiên cứu triển khai mô hình ngôn ngữ lớn (LLMs) và chuỗi xử lý (Chain)	
5	21/08 – 27/09/2024	Nghiên cứu tối ưu hóa quy trình RAG	
6	28/09 – 17/10/2024	Nghiên cứu xây dựng FAQ	
7	18/10 – 15/11/2024	Nghiên cứu thuật toán Deep Q-Learning giúp tăng trải nghiệm người dùng	
8	16/11 – 30/11/2024	Nghiên cứu xây dựng UI/UX tương tác theo thời gian thực	
9	01/12 – 07/12/2024	Kiểm thử và tối ưu	

Ngày tháng năm 20

Người viết đề cương

Ý kiến của giáo viên hướng dẫn

(Ký và ghi rõ họ tên)

MỤC LỤC

LỜI CẢM ƠN.....	4
MỤC LỤC	11
BẢNG PHÂN CÔNG.....	15
TỪ NGỮ VIẾT TẮT	17
DANH MỤC HÌNH.....	18
DANH MỤC BẢNG	20
CHƯƠNG 1: TỔNG QUAN	21
1.1. Tính cấp thiết của đề tài.....	21
1.2. Mục tiêu của đề tài.....	22
1.3. Cách tiếp cận và phương pháp nghiên cứu.....	22
1.3.1. Đối tượng nghiên cứu	22
1.3.2. Phạm vi nghiên cứu	23
1.3.3. Nội dung nghiên cứu	23
1.4. Phương pháp nghiên cứu	23
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT.....	25
2.1. Hiện thực hóa dự án.....	25
2.2. Giới thiệu về mô hình ngôn ngữ lớn	26
2.2.1. Định nghĩa mô hình ngôn ngữ lớn.....	26
2.2.2. Sự quan trọng của mô hình ngôn ngữ lớn	27
2.2.3. Cách thức hoạt động của mô hình ngôn ngữ lớn.....	28
2.2.4. Huấn luyện mô hình ngôn ngữ lớn	28
2.3. Giới thiệu về LangChain	29
2.3.1. Định nghĩa LangChain	29
2.3.2 Các module của LangChain.....	30
2.4. Giới thiệu về PostgreSQL.....	32
2.4.1 Định nghĩa PostgreSQL.....	32
2.4.2 Các tính năng của PostgreSQL.....	33
2.5 Giới thiệu về SQL Server	33

MỤC LỤC

2.5.1 Định nghĩa SQL Server	33
2.5.2 Các tính năng của SQL Server	34
2.6 Giới thiệu về Chroma	34
2.6.1. Cơ sở dữ liệu vector.....	34
2.6.2. Định nghĩa Chroma	35
2.7. Giới thiệu về Python Django	36
2.7.1. Định nghĩa Django	36
2.7.2. Những điểm nổi bật của Django.....	37
2.7.2. Cách hoạt động của Django.....	37
2.8 Giới thiệu về Python Flask	38
2.8.1. Định nghĩa Flask.....	38
2.8.2. Các tính năng của Flask.....	39
2.9. Giới thiệu về SQLAlchemy	40
2.9.1. Định nghĩa SQLAlchemy	40
2.9.2. Các tính năng nổi bật của SQLAlchemy	40
2.10. Giới thiệu về RAG.....	41
2.10.1. Định nghĩa RAG.....	41
2.10.2. Thành phần của RAG	41
2.10.3. Quy trình hoạt động của RAG.....	42
2.11. Giới thiệu về Cache	43
2.11.1. Định nghĩa Cache	43
2.11.2. Đặc điểm của Cache	44
2.11.3. Phân loại Cache	44
2.11.4. Tối ưu hóa phản hồi hệ thống bằng Cache cho câu hỏi lặp lại.....	44
2.11.5. Quy trình truy vấn Cache của Chatbot	45
2.12. Giới thiệu về học tăng cường	45
2.13. Giới thiệu về Q-Learning	47
2.13.1. Định nghĩa Q-Learning.....	47
2.13.2. Các thành phần của Q-Learning	47
2.13.3. Công thức của Q-Learning	48
2.14. Giới thiệu về Deep Q-Learning	49

MỤC LỤC

2.14.1 Nguyên lý hoạt động	49
2.14.2 Kỹ thuật tăng cường cho Deep Q-Learning	49
2.14.3 Công thức cập nhật	50
CHƯƠNG 3: TÌM HIỂU DỮ LIỆU.....	51
3.1. Khảo sát hiện trạng trang tư vấn Sinh viên	51
3.2 Khảo sát Sổ tay Sinh viên năm 2023	52
3.3. Kết luận sau khi khảo sát.....	55
3.4 Tiết xu lý dữ liệu	55
3.4.1 Tiết xu lý dữ liệu Sổ tay Sinh viên	55
3.4.2 Tiết xu lý dữ liệu trang Tư vấn Sinh viên	57
CHƯƠNG 4: THIẾT KẾ HỆ THỐNG	58
4.1. Cấu hình phần cứng	58
4.2 Kiến trúc hệ thống	59
4.3 Mô tả cơ sở dữ liệu	60
4.3.1 Vector Database.....	60
4.3.2 FAQ	61
4.3.3 QuestionCount	62
4.3.4 Feedback	62
4.3.5. Cache	64
4.4 Thiết kế giao diện	65
4.4.1 Trang đăng ký	65
4.4.2 Trang đăng nhập	66
4.4.3 Trang bảng tin	66
4.4.4. Trang đăng bài viết	67
4.4.5. Trang chi tiết bài viết	67
4.4.6. Trang nhắn tin với chatbot.....	69
4.5 Môi trường học tăng cường cho việc lựa chọn 2 phản hồi FAQ	69
CHƯƠNG 5: CHI TIẾT CÁC THAM SỐ VÀ SIÊU THAM SỐ	71
5.1 Mô hình ngôn ngữ lớn	71
5.2 Văn bản tài liệu	72
5.3 Môi trường Deep Q-Learning	72

MỤC LỤC

5.4 Mạng Deep Q-Learning.....	73
5.5 Deep Q-Learning Agent	74
CHƯƠNG 6. KẾT QUẢ ĐẠT ĐƯỢC.....	75
6.1 Kết quả.....	75
6.1.1 Về mặt lý thuyết	75
6.1.2 Về mặt sản phẩm	75
6.2 Kiểm thử mô hình.....	87
6.2.1 Phản hồi FAQ	87
6.2.2 Phản hồi từ RAG	91
6.2.3 Lịch sử Chat.....	94
6.2.4. Cache	94
CHƯƠNG 7. KẾT LUẬN.....	95
7.1 Tổng kết.....	95
7.2 Đóng góp của đề tài	95
7.3 Hạn chế của đề tài.....	96
7.4 Hướng phát triển cho tương lai.....	96
TÀI LIỆU THAM KHẢO	98

BẢNG PHÂN CÔNG

STT	Nhiệm vụ	Mức độ	Thực hiện
1	Tự động chọn n_layer tối ưu cho mỗi lần chạy server	100%	GVHD hỗ trợ
2	Chuyển Sổ tay Sinh viên từ dạng pdf sang txt và làm sạch dữ liệu	100%	GVHD hỗ trợ
3	Kết hợp LLM và Flask Server trả về các chunks	100%	GVHD hỗ trợ
4	Thiết kế, xây dựng Cơ sở dữ liệu cho bảng FAQ, Feedback, QuestionCount	100%	Nguyễn Thị Phú
5	Xây dựng chức năng và thêm dữ liệu FAQ cũng như Cache để truy xuất nhanh	100%	Nguyễn Thị Phú
6	Xử lý ngôn ngữ tiếng Việt bằng thư viện underthesea của Pytorch	100%	Nguyễn Thị Phú
7	Áp dụng học tăng cường vào hệ thống: xây dựng môi trường, mô hình mạng và agent bằng thuật toán Deep Q-Learning	100%	Nguyễn Thị Phú
8	Vẽ biểu đồ mô tả reward theo các episodes	100%	Nguyễn Thị Phú
9	Sử dụng mô hình Gemma từ HuggingFace và lưu vào hub	100%	Bùi Quang Thiện
10	Kết hợp phương pháp RAG với LlamaCPP	100%	Bùi Quang Thiện
11	Triển khai phương pháp RAG	100%	Bùi Quang Thiện
12	Viết API Flask để giao tiếp giữa back-end và front-end	100%	Bùi Quang Thiện

13	Thực hiện streaming chuỗi kết quả trả về từ FAQ và RAG	100%	Bùi Quang Thiện
14	Xây dựng giao diện bằng Python Django	100%	Bùi Quang Thiện

TỪ NGỮ VIẾT TẮT

Từ viết tắt	Nghĩa tiếng Anh	Tạm dịch
AI	Artificial Intelligence	Trí tuệ nhân tạo
RDBMS	Relational Database Management System	Hệ quản trị cơ sở dữ liệu quan hệ
RAG	Retrieval Augmented Generation	Tạo sinh tăng cường truy xuất
LLM(s)	Large Language Model(s)	Mô hình ngôn ngữ lớn
RNN	Recurrent Neural Network	Mạng nơ-ron hồi quy
API	Application Programming Interface	Giao diện lập trình ứng dụng
FAQ	Frequently Asked Questions	Các câu hỏi thường xuyên được hỏi
ORM	Object-Relational Mapping	Ánh xạ Đối tượng-Quan hệ
MA	Moving Average	Đường trung bình động

DANH MỤC HÌNH

Hình 1: Quy mô các mô hình ngôn ngữ cho đến tháng 03/2023.....	26
Hình 2: GPU của máy tính hãng MSI	27
Hình 3: LLM hoàn thành các tác vụ khác nhau.....	27
Hình 4: Mô hình embedding biểu diễn các từ thành các vector.....	28
Hình 5: Logo LangChain	29
Hình 6: Các thành phần của module LangChain.....	30
Hình 7: Cấu trúc chi tiết Model I/O module của LangChain	31
Hình 8: Sơ đồ chi tiết Retrieval module của LangChain.....	31
Hình 9: Logo PostgreSQL	32
Hình 10: Logo Microsoft SQL Server.....	33
Hình 11: Cơ sở dữ liệu vector	35
Hình 12: Logo Chroma.....	35
Hình 13: Quy trình của Chroma	36
Hình 14: Python Django	36
Hình 15: Cách hoạt động của Django.....	38
Hình 16: Logo Flask	38
Hình 17: Logo SQLAlchemy	40
Hình 18: Mô tả quy trình của RAG	41
Hình 19: Cấu trúc và vị trí của các loại bộ nhớ cache trong hệ thống máy tính	43
Hình 20: Quy trình truy vấn Cache của Chatbot	45
Hình 21: Quy trình tương tác giữa Agent và môi trường trong học tăng cường.....	46
Hình 22: Công thức hàm chính sách tối ưu	46
Hình 23: Nguyên lý hoạt động của Deep Q-Learning.....	49
Hình 24: Công thức cập nhật	50
Hình 25: Toàn bộ quá trình áp dụng Deep Q-Learning	50
Hình 26: Sổ tay SINH VIÊN 2023	52
Hình 27: Sổ tay Sinh viên được chuyển sang dạng txt.....	56
Hình 28: Mô tả kiến trúc hệ thống Chatbot	59
Hình 29: Vector Database.....	61
Hình 30: Bảng FAQ từ PostgreSQL.....	61
Hình 31: Bảng QuestionCount từ PostgreSQL	62
Hình 32: Bảng Feedback từ PostgreSQL	62
Hình 33:Trang đăng ký	65
Hình 34: Trang đăng nhập	66
Hình 35: Trang bảng tin.....	66
Hình 36: Trang đăng bài viết.....	67
Hình 37: Trang chi tiết bài viết.....	67
Hình 38: Bình luận bài viết.....	68

Hình 39: Trang nhắn tin với chatbot	69
Hình 40: Công thức reward	70
Hình 41: Công thức reward	72
Hình 42: Sơ đồ mạng nơ-ron nhân tạo thiết kế cho Deep Q-Learning	73
Hình 43: Reward theo từng episode	85
Hình 44: Mô tả các reward qua từng episodes không áp dụng đường trung bình động	86
Hình 45: Mô tả các reward qua từng episodes nhưng áp dụng đường trung bình động	86
Hình 46: Các phản hồi của chatbot từ FAQ	91
Hình 47: Các phản hồi của chatbot từ RAG	93
Hình 48: Tính năng lưu lịch sử chat của chatbot.....	94
Hình 49: Tính năng lưu trong bộ nhớ cache.....	94

DANH MỤC BẢNG

Bảng 1: Hiện thực hóa dự án	25
Bảng 2: Kết luận sau khi khảo sát thực tế	55
Bảng 3: Cấu hình phần cứng	58
Bảng 4: Mô tả bảng FAQ	62
Bảng 5: Mô tả bảng QuestionCount	62
Bảng 6: Mô tả bảng Feedback	63
Bảng 7: Thông số cho mô hình ngôn ngữ lớn	71
Bảng 8: Thông số cho văn bản tài liệu	72
Bảng 9: Thông số của Deep Q-Learning	73
Bảng 10: Mạng nơ-ron thiết kế cho Deep Q-Learning	74
Bảng 11: Thông số cho Agent	74
Bảng 12: Nhận xét tổng quát sản phẩm	76
Bảng 13: Đánh giá hệ thống	85

CHƯƠNG 1: TỔNG QUAN

1.1. Tính cấp thiết của đề tài

Trong thời đại ngày nay, giáo dục đang đứng trước những cơ hội và thách thức chưa từng có; thuận lợi trong việc tiếp nhận nguồn kiến thức kỹ thuật số phát triển, khó khăn trong việc triển khai, áp dụng các công nghệ hiện đại đó hỗ trợ các hoạt động trong trường học sao cho hiệu quả. Sự phát triển nhanh chóng của Trí tuệ nhân tạo đã làm thay đổi hoàn toàn phương pháp tìm kiếm, truy vấn thông tin. Đặc biệt, các chatbot đã nổi lên như một công cụ hữu ích, hỗ trợ Sinh viên tiếp cận thông tin một cách nhanh chóng và chính xác trong việc tư vấn về nội quy, quy chế đào tạo và các quy định liên quan, nó đóng vai trò như một cầu nối hiệu quả giữa nhà trường và Sinh viên, giúp giảm tải khối lượng công việc cho bộ phận quản lý và nâng cao mức độ hài lòng của người học.

Tuy nhiên, có hai vấn đề lớn trong công tác tư vấn, giải đáp thắc mắc Sinh viên. Một là các phương pháp truyền thống như tạo ra diễn đàn, Sinh viên đăng câu hỏi, sau đó các tư vấn viên sẽ phản hồi sau, việc này hoàn toàn không đáp ứng được nhu cầu phản hồi theo thời gian thực – ảnh hưởng rất lớn, tác động sâu sắc tới trải nghiệm người dùng. Hai là các chatbot hiện nay trong môi trường giáo dục gặp phải hạn chế về khả năng cung cấp thông tin đúng ngữ cảnh và được cập nhật thường xuyên. Nhiều hệ thống dựa hoàn toàn vào dữ liệu huấn luyện sẵn, khiến thông tin dễ lỗi thời hoặc không đáp ứng được nhu cầu đặc thù của phòng Tuyển sinh và Công tác Sinh viên. Để giải quyết vấn đề này, phương pháp RAG được áp dụng nhằm cải thiện hiệu quả hoạt động của chatbot. Bằng cách kết hợp khả năng sinh văn bản của các mô hình ngôn ngữ lớn với cơ chế truy xuất thông tin từ nguồn dữ liệu bên ngoài, RAG giúp chatbot không chỉ trả lời các câu hỏi chính xác hơn mà còn đảm bảo thông tin luôn được cập nhật và phù hợp với bối cảnh.

Đó cũng chính là lý do mà nhóm chúng em chọn đề tài “Chatbot tư vấn cho Sinh viên trường Đại Học Sư Phạm Kỹ Thuật TP.HCM”. Đây không chỉ là đề tài mang tính thực tiễn cao mà còn giúp chúng em áp dụng kiến thức và kỹ năng đã học vào bài toán thực tế.

1.2. Mục tiêu của đề tài

Trước hết, đề tài hướng đến việc tối ưu hóa truy cập thông tin, tạo ra một hệ thống tự động và thông minh giúp Sinh viên dễ dàng tra cứu và truy xuất các thông tin liên quan đến quy chế đào tạo và nội quy của trường đại học. Hệ thống này sẽ giúp giảm thiểu thời gian và công sức cần thiết để tìm kiếm thông tin so với phương pháp truyền thống, mang lại sự tiện lợi và nhanh chóng trong việc nắm bắt các quy định cần thiết.

Bên cạnh đó, mục tiêu quan trọng không kém là nâng cao độ chính xác và tin cậy của thông tin. Việc đảm bảo thông tin được cung cấp cho Sinh viên luôn chính xác, nhất quán và cập nhật là một thách thức lớn trong môi trường giáo dục đại học. Phương pháp RAG sẽ giúp tổng hợp và trình bày thông tin một cách rõ ràng và dễ hiểu, từ đó giảm thiểu các hiểu lầm và vi phạm nội quy do thiếu hiểu biết. Điều này không chỉ bảo vệ quyền lợi của Sinh viên mà còn duy trì trật tự và kỷ luật trong trường học.

Cuối cùng là cải thiện trải nghiệm học tập của Sinh viên. Bằng cách tạo điều kiện thuận lợi cho Sinh viên trong việc tiếp cận và hiểu rõ các quy định và nội quy của nhà trường, đề tài này không chỉ giúp Sinh viên tuân thủ tốt hơn mà còn hỗ trợ họ trong việc lên kế hoạch học tập và hoạt động một cách hiệu quả. Khi Sinh viên có thể dễ dàng truy cập thông tin và nhận được sự hỗ trợ nhanh chóng, họ sẽ cảm thấy tự tin và an tâm hơn trong quá trình học tập.

1.3. Cách tiếp cận và phương pháp nghiên cứu

1.3.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài này là ứng dụng chatbot dựa trên phương pháp RAG, để tư vấn Sinh viên về các quy định của Trường Đại học Sư phạm Kỹ thuật TP.HCM. Cụ thể, nghiên cứu tập trung vào:

- Phương pháp RAG và ứng dụng của nó trong việc xây dựng chatbot.
- Dữ liệu từ sổ tay Sinh viên của Trường Đại học Sư phạm Kỹ thuật TP.HCM.
- Quy trình xử lý ngôn ngữ tự nhiên để hiểu và phản hồi các câu hỏi của Sinh viên.

CHƯƠNG 1: TỔNG QUAN

- Các mô hình ngôn ngữ lớn đang thịnh hành, tiêu biểu là Gemma.
- Giao diện tương tác giữa chatbot và Sinh viên.

1.3.2. Phạm vi nghiên cứu

- Nghiên cứu các phương pháp xử lý ngôn ngữ tự nhiên phù hợp cho Chatbot hướng tới.
- Nghiên cứu mô hình học sâu Transformer.
- Thu thập và tổng hợp các tài liệu sổ tay Sinh viên để tạo bộ dữ liệu huấn luyện.
- Sử dụng kiến thức học tăng cường để chọn ra phản hồi tốt nhất góp phần tăng trải nghiệm người dùng.
- Xây dựng giao diện Chatbot để Sinh viên của Trường Đại học Sư Phạm Kỹ Thuật TP.HCM sử dụng.

1.3.3. Nội dung nghiên cứu

- Tìm hiểu về mô hình ngôn ngữ lớn.
- Tìm hiểu về phương pháp RAG.
- Tìm hiểu về framework LangChain để thực thi RAG.
- Tìm hiểu thuật toán Deep Q-Learning trong học tăng cường.
- Tìm hiểu về Python Django, Flask.
- Tìm hiểu cách làm việc với PostgreSQL, SQL Server.
- Tìm hiểu cách làm việc với Database ORM - SQLAlchemy

1.4. Phương pháp nghiên cứu

- Phương pháp phân tích và tổng hợp các tài liệu, sách, báo, các tạp chí khoa học, các bài viết, tài liệu tham khảo, các video liên quan đến mô hình ngôn ngữ lớn, framework LangChain, phương pháp RAG, thuật toán Deep Q-Learning, Python Django, Flask,...
- Phương pháp chuyên gia, nhờ sự góp ý, gợi ý, giúp đỡ từ Giảng viên hướng dẫn, các thầy cô trong Khoa Công nghệ thông tin trường Đại học Sư Phạm Kỹ Thuật TP.HCM.

CHƯƠNG 1: TỔNG QUAN

- Phương pháp thu thập số liệu, dữ liệu từ Sổ tay Sinh viên năm 2023 được công bố bởi phòng Tuyển sinh và Công tác Sinh viên cũng như các câu hỏi thường được hỏi từ trang Tư vấn Sinh viên truyền thống.

Qua việc áp dụng các phương pháp nghiên cứu đa dạng và có hệ thống, kết hợp với việc phát triển các tính năng mở rộng, dự án này hướng đến việc phát triển một ứng dụng chatbot toàn diện, có tính thực tiễn cao, đáp ứng hiệu quả nhu cầu đa dạng của Sinh viên và góp phần nâng cao chất lượng dịch vụ hỗ trợ Sinh viên của Trường Đại học Sư phạm Kỹ thuật TP.HCM.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

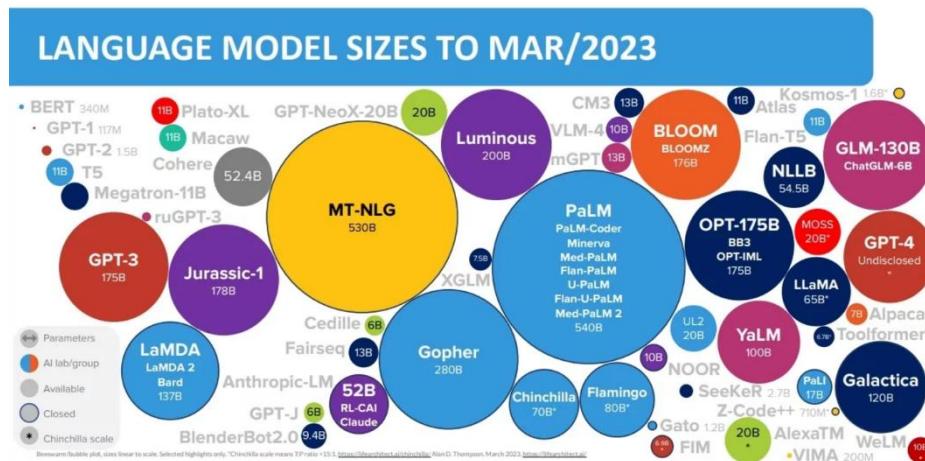
2.1. Hiện thực hóa dự án

Thành phần	Mô tả	Cụ thể
Framework xây dựng RAG	Xây dựng hệ thống RAG để kết hợp truy xuất thông tin và tạo phản hồi.	LangChain
Mô hình ngôn ngữ lớn	Tạo ra các phản hồi tự nhiên và chính xác dựa trên văn bản truy xuất.	gemma-2-9b-it-GGUF
Embedding	Biểu diễn văn bản thành vector để hỗ trợ tìm kiếm và so khớp ngữ nghĩa.	BAAI/bge-base-en-v1.5
Vector Store	Lưu trữ embedding, phục vụ truy vấn dữ liệu ngữ nghĩa.	Chroma
Đồng bộ hóa dữ liệu	Lưu trữ vector embedding lên máy chủ an toàn.	FTP Server
Database cho FAQ, Feedback	Quản lý các thông tin câu hỏi-câu trả lời FAQ, Feedback.	SQLAlchemy kết hợp PostgreSQL
Database cho Cache	Cache được câu hỏi đã hỏi	SQLAlchemy kết hợp PostgreSQL
Database cho users	Quản lý thông tin người dùng, hội thoại với chatbot, bài viết, thông báo,...	SQL Server
Giao diện người dùng	Thiết kế giao diện đơn giản, thân thiện với người dùng.	Django
Giao tiếp API	Xây dựng API để tích hợp giữa các thành phần front-end và back-end của hệ thống.	Flask

Bảng 1: Hiện thực hóa dự án

2.2. Giới thiệu về mô hình ngôn ngữ lớn

2.2.1. Định nghĩa mô hình ngôn ngữ lớn



Hình 1: Quy mô các mô hình ngôn ngữ cho đến tháng 03/2023

Mô hình ngôn ngữ lớn, còn gọi là LLM, là các mô hình học sâu rất lớn, được đào tạo trước dựa trên một lượng dữ liệu khổng lồ. Bộ chuyển hóa cơ bản là tập hợp các mạng nơron có một bộ mã hóa và một bộ giải mã với khả năng tự tập trung. Bộ mã hóa và bộ giải mã trích xuất ý nghĩa từ một chuỗi văn bản và hiểu mối quan hệ giữa các từ và cụm từ trong đó.

Bộ chuyển hóa LLM có khả năng đào tạo không có giám sát, mặc dù lời giải thích chính xác hơn là bộ chuyển hóa thực hiện việc tự học. Thông qua quá trình này, bộ chuyển hóa học cách hiểu ngữ pháp, ngôn ngữ và kiến thức cơ bản.

Khác với các mạng nơ-ron hồi quy RNN trước đó thường xử lý tuần tự dữ liệu đầu vào, bộ chuyển hóa xử lý song song toàn bộ trình tự. Điều này cho phép các nhà khoa học dữ liệu sử dụng GPU để đào tạo các LLM dựa trên bộ chuyển hóa, qua đó giảm đáng kể thời gian đào tạo.

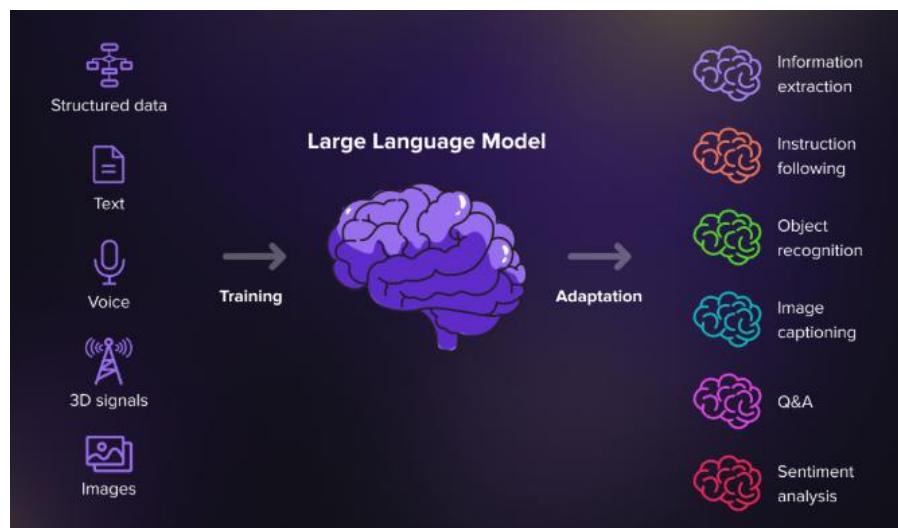


Hình 2: GPU của máy tính hãng MSI

Kiến trúc mạng nơ-ron của bộ chuyển hóa cho phép việc sử dụng các mô hình rất lớn, thường có hàng trăm tỷ tham số. Các mô hình quy mô lớn như vậy có thể thu nạp một lượng dữ liệu khổng lồ, thường là từ Internet, nhưng cũng từ các nguồn, ví dụ như Common Crawl với hơn 50 tỷ trang web, và Wikipedia với khoảng 57 triệu trang.

2.2.2. Sự quan trọng của mô hình ngôn ngữ lớn

Các mô hình ngôn ngữ lớn vô cùng linh hoạt. Một mô hình có thể thực hiện các tác vụ hoàn toàn khác nhau, ví dụ như trả lời câu hỏi, trích xuất thông tin, nhận diện vật thể, chủ thích ảnh, tóm tắt tài liệu, dịch ngôn ngữ và hoàn thành câu,... LLM có khả năng làm gián đoạn việc sáng tạo nội dung và cách thức mọi người sử dụng công cụ tìm kiếm và trợ lý ảo.



Hình 3: LLM hoàn thành các tác vụ khác nhau

Mặc dù không hoàn hảo, nhưng LLM đang thể hiện khả năng đáng kinh ngạc của mình trong việc đưa ra các dự đoán dựa trên một số lượng lời nhắc hoặc dữ liệu đầu vào tương đối nhỏ. LLM có thể được sử dụng cho hướng AI tạo sinh để tạo ra nội dung dựa trên lời nhắc được nhập vào bằng ngôn ngữ của con người.

2.2.3. Cách thức hoạt động của mô hình ngôn ngữ lớn

Một yếu tố quan trọng trong cách thức hoạt động của LLM là cách chúng biểu diễn các từ. Các hình thức máy học trước đây sử dụng một bảng số để biểu diễn từng từ. Tuy nhiên, hình thức biểu diễn này không thể nhận ra mối quan hệ giữa các từ, chẳng hạn như các từ có nghĩa tương tự. Hạn chế này đã được khắc phục bằng cách sử dụng các vector đa chiều, thường được gọi là embedding từ, để biểu diễn các từ sao cho các từ có nghĩa theo ngữ cảnh tương tự nhau hoặc các mối quan hệ khác sẽ gần nhau trong không gian vector.



Hình 4: Mô hình embedding biểu diễn các từ thành các vector

Bằng cách sử dụng embedding từ, bộ chuyển hóa có thể xử lý trước văn bản dưới dạng phần biểu diễn bằng số thông qua bộ mã hóa và hiểu ngữ cảnh của các từ lẩn cùm từ có nghĩa tương tự, cũng như các mối quan hệ khác giữa các từ, chẳng hạn như các phần của lời nói. Sau đó, LLM có thể áp dụng kiến thức về ngôn ngữ này thông qua bộ giải mã để tạo ra một kết quả đầu ra độc đáo.

2.2.4. Huấn luyện mô hình ngôn ngữ lớn

Mạng nơ-ron dựa trên bộ chuyển hóa rất lớn. Các mạng này chứa nhiều nút và lớp. Mỗi nút trong một lớp có kết nối với tất cả các nút trong lớp tiếp theo, mỗi nút có trọng số và độ lệch. Trọng số và độ lệch cùng với phần nhúng được gọi là tham số mô hình. Mạng nơ-ron dựa trên bộ chuyển hóa lớn có thể có hàng tỷ tham số. Kích thước của mô hình

thường được xác định bởi mối quan hệ thực nghiệm giữa kích thước mô hình, số lượng tham số và kích thước của dữ liệu đào tạo.

Quá trình đào tạo được thực hiện bằng cách sử dụng một tập ngữ liệu lớn gồm dữ liệu chất lượng cao. Trong quá trình đào tạo, mô hình điều chỉnh liên tục các giá trị tham số cho đến khi mô hình dự đoán chính xác mã thông báo tiếp theo từ trình tự mã thông báo đầu vào trước đó. Mô hình thực hiện điều này thông qua các kỹ thuật tự học, giúp mô hình học được cách điều chỉnh các tham số để tối đa hóa khả năng của các mã thông báo tiếp theo trong các mẫu đào tạo.

Sau khi được đào tạo, LLM có thể dễ dàng được điều chỉnh để thực hiện nhiều tác vụ bằng cách sử dụng các tập dữ liệu có giám sát tương đối nhỏ, một quy trình được gọi là tinh chỉnh.

2.3. Giới thiệu về LangChain

2.3.1. Định nghĩa LangChain



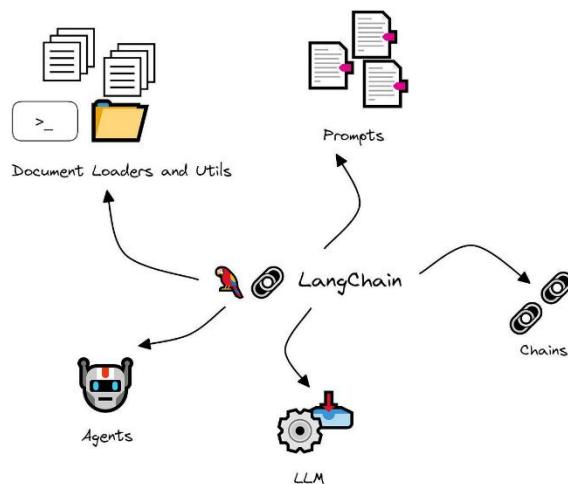
Hình 5: Logo LangChain

LangChain là một framework được viết bằng Python và JavaScript, nó cung cấp các công cụ để thao tác và xây dựng ứng dụng dựa trên LLMs. LangChain được biết đến là một mã nguồn mở, cho phép các nhà phát triển dễ dàng xây dựng các ứng dụng dựa trên các mô hình ngôn ngữ lớn, có thể kể đến như Bard, GPT-3,...

Theo đó, mã nguồn mở này hoạt động như một cầu nối giữa LLM và các ứng dụng. Chúng cung cấp các công cụ để:

- Kết nối LLM với các nguồn dữ liệu bên ngoài: Cho phép người dùng truy cập dữ liệu từ các API, cơ sở dữ liệu, tập tin,... và sử dụng dữ liệu này để cung cấp cho LLM.
- Xử lý đầu ra của LLM: Mã nguồn mở này cung cấp các công cụ để phân tích, hiểu và chuyển đổi đầu ra của LLM thành định dạng phù hợp với ứng dụng của bạn.

2.3.2 Các module của LangChain

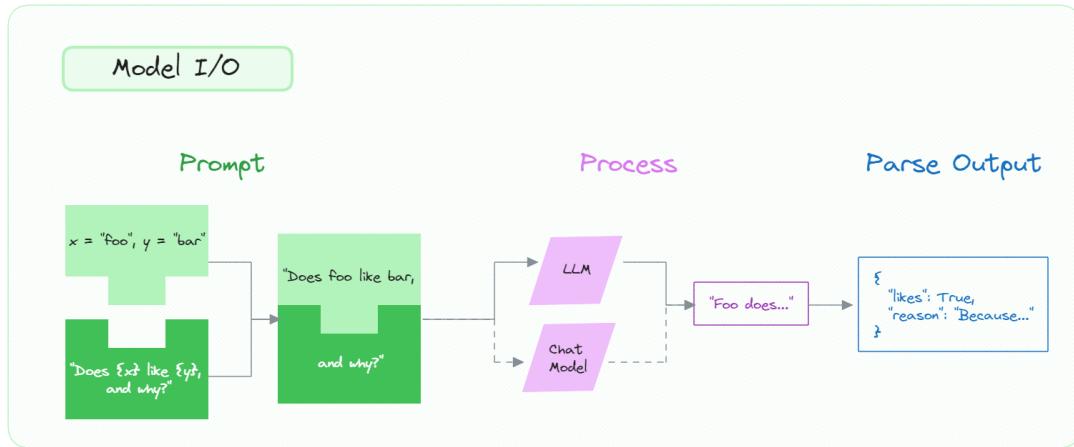


Hình 6: Các thành phần của module LangChain

LangChain framework cung cấp các module cho phép quản lý nhiều khía cạnh cụ thể trong việc tương tác với LLM, bao gồm:

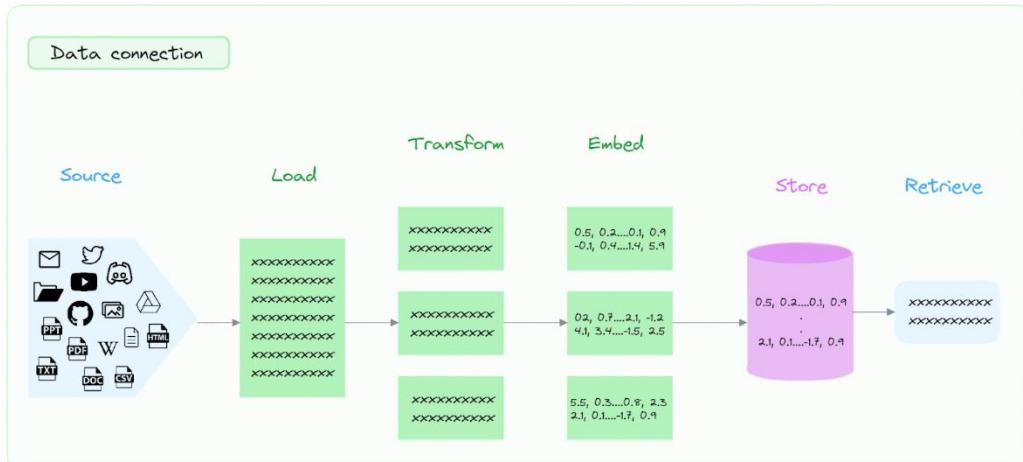
- Model I/O: Là thành phần nền tảng để giao tiếp với các LLM, nó cung cấp giao diện để thực hiện tương tác liền mạch với bất kỳ LLM nào.
- Đầu vào được định nghĩa bằng Prompt - tập hợp các hướng dẫn hay Input giúp LLM hiểu ngữ cảnh và trả lời mạch lạc.
- Phần xử lý là mô hình ngôn ngữ được LangChain tích hợp 2 loại chính là LLM và Chat Model.
- Ngoài ra, module này còn kết hợp Output Parser chuyển đổi kết quả đầu ra từ văn bản thô thành thông tin có cấu trúc.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT



Hình 7: Cấu trúc chi tiết Model I/O module của LangChain

- Retrieval: Là thành phần giúp tạo điều kiện thuận lợi cho việc tích hợp dữ liệu chuyên biệt của người dùng vào giai đoạn Generation của mô hình ngôn ngữ bằng các sử dụng RAG. LangChain hỗ trợ các chức năng như Document Loaders, Document Transformers, Text Embedding Models và nhiều thuật toán truy xuất cũng như cách lưu trữ dữ liệu dưới dạng vector đảm bảo việc sử dụng hiệu quả và phù hợp với ngữ cảnh.



Hình 8: Sơ đồ chi tiết Retrieval module của LangChain

- Chains: Là thành phần giúp thiết kế những xử lý phức tạp bằng cách kết nối các thành phần khác với nhau hoặc LLM với nhau, bao gồm các chuỗi khác. Cách tiếp cận này vừa đơn giản vừa hiệu quả, giúp phát triển các ứng dụng phức tạp, nâng cao khả năng bảo trì.

- Agents: Là thành phần giúp ứng dụng sử dụng mô hình ngôn ngữ làm công cụ suy luận để quyết định thứ tự hành động một cách linh hoạt. LangChain cung cấp nhiều loại Agents kết hợp đa dạng với các Tools giúp chúng tương tác với các tiện ích chung, với Chains hay với các Agents khác, tạo thành framework mạnh mẽ xử lý các tác vụ phức tạp.
- Memory: Là thành phần đóng vai trò then chốt đối với các ứng dụng giao diện đối thoại, cho phép đưa luồng hội thoại trước đó vào LLM, giúp duy trì ngữ cảnh tương tác của người dùng. LangChain cung cấp rất nhiều tiện ích để tích hợp Memory vào hệ thống, hỗ trợ các hoạt động cơ bản như đọc, viết, đảm bảo hệ thống có thể truy cập vào các tin nhắn trong quá khứ hoặc duy trì cập nhật liên tục.
- Callbacks: Là thành phần giúp đăng ký các sự kiện để khai thác các giai đoạn khác nhau của LLM, hữu ích trong các tác vụ như logging, monitoring, streaming,... Các sự kiện này được kích hoạt bởi CallbackHandler tương ứng với từng sự kiện đã đăng ký. Ngoài ra LangChain cũng cung cấp một số built-in-handler như StdOutCallbackHandler giúp ghi lại tất cả sự kiện trong quá trình thực thi.

2.4. Giới thiệu về PostgreSQL

2.4.1 Định nghĩa PostgreSQL



Hình 9: Logo PostgreSQL

PostgreSQL là một hệ thống quản trị cơ sở dữ liệu quan hệ và đối tượng (object-relational database management system) miễn phí và nguồn mở (RDBMS) tiên tiến nhất hiện nay. khả năng mở rộng cao và tuân thủ các tiêu chuẩn kỹ thuật. Nó được thiết kế để xử lý một loạt các khối lượng công việc lớn, từ các máy tính cá nhân đến kho dữ liệu hoặc dịch vụ Web có nhiều người dùng đồng thời.

2.4.2 Các tính năng của PostgreSQL

PostgreSQL tích hợp nhiều tính năng tuyệt vời giúp hỗ trợ nhà phát triển xây dựng app đáp ứng các chức năng phức tạp, truy vấn nhanh chóng và bảo mật duy trì tính toàn vẹn và độ tin cậy. Để đáng tin cậy hơn, Postgresql cung cấp các tùy chọn bảo mật, xác thực và khôi phục thảm họa khác nhau. PostgreSQL được chứng minh là có khả năng mở rộng cao cả về số lượng dữ liệu và số lượng người dùng có thể thao tác cùng lúc.

2.5 Giới thiệu về SQL Server

2.5.1 Định nghĩa SQL Server



Hình 10: Logo Microsoft SQL Server

Microsoft SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ-RDBMS được dùng để lưu trữ và truy xuất dữ liệu theo yêu cầu từ các ứng dụng khác. Thay vì phải viết mã từ đầu để quản lý cơ sở dữ liệu, các nhà phát triển phần mềm thường sử dụng SQL Server để tận dụng các chức năng sẵn có, giúp quá trình phát triển ứng dụng nhanh chóng, tiết kiệm chi phí, đồng thời tăng tính ổn định, bảo mật và khả năng mở rộng.

SQL Server có cả phiên bản mã nguồn mở và mã nguồn đóng, giúp tăng cường tính linh hoạt và khả năng tùy chỉnh trong việc triển khai cơ sở dữ liệu. Tuy nhiên, phiên bản chính của SQL Server chủ yếu là mã nguồn đóng, điều này có nghĩa là mã nguồn đóng này không được công khai và người dùng cần phải tuân theo các điều khoản và điều kiện sử dụng của Microsoft.

Tương tự như các RDBMS khác, trong SQL Server, dữ liệu được lưu trữ dưới dạng các bảng (tables) với các hàng (rows) và cột (columns). SQL Server đảm bảo rằng dữ liệu được lưu trữ hiệu quả và có thể được tìm kiếm, truy xuất nhanh chóng. Ngoài ra nó còn quản lý tốt việc truy cập đồng thời từ nhiều người dùng để tránh xung đột dữ liệu.

2.5.2 Các tính năng của SQL Server

SQL Server nổi bật với nhiều tính năng mạnh mẽ hỗ trợ quản lý và phân tích dữ liệu, đồng thời cung cấp các công cụ về trí tuệ doanh nghiệp (Business Intelligence) để giúp hiểu rõ hơn về doanh nghiệp và khách hàng. Dưới đây là một số tính năng chính của SQL Server:

- SQL Server Data Quality: SQL Server cung cấp các dịch vụ tích hợp để quản lý chất lượng dữ liệu, bao gồm khả năng kiểm tra, sửa lỗi và làm sạch dữ liệu.
- SQL Server Master: Tính năng này cho phép bạn quản lý danh mục dữ liệu và thông tin liên quan, tạo ra một nguồn dữ liệu chung và đáng tin cậy.
- SQL Server Data Tools: Đây là một bộ công cụ dành cho lập trình viên để phát triển cơ sở dữ liệu. SQL Server Data Tools cung cấp các tính năng mạnh mẽ để thiết kế, quản lý và triển khai cơ sở dữ liệu Microsoft SQL Server.
- SQL Server Management Studio (SSMS): Được sử dụng để triển khai, giám sát và quản lý cơ sở dữ liệu. SSMS cung cấp một giao diện đồ họa và dòng lệnh mạnh mẽ cho quản trị viên cơ sở dữ liệu.
- SQL Server Analysis Services (SSAS): Bằng cách sử dụng SSAS, bạn có thể tạo các mô hình phân tích dữ liệu và thực hiện các phân tích phức tạp để hiểu rõ hơn về dữ liệu của bạn.
- SQL Server Reporting Services (SSRS): SSRS cho phép tạo ra các báo cáo dễ dàng hơn. Bạn có thể thiết kế và phát triển các báo cáo tùy chỉnh dựa trên dữ liệu từ cơ sở dữ liệu SQL Server và chia sẻ chúng với người dùng cuối.

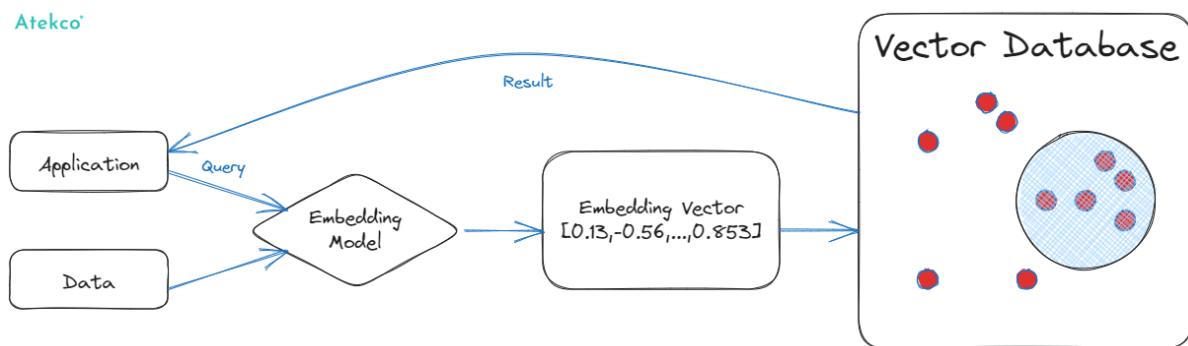
2.6 Giới thiệu về Chroma

2.6.1. Cơ sở dữ liệu vector

Ngày nay, trí tuệ nhân tạo (AI) đang nhanh chóng tạo ra nhiều tác động với sự phát triển của ngành công nghiệp phần mềm. Điều này đã dẫn đến những thay đổi lớn về nhu cầu xử lý và lưu trữ dữ liệu. Những đổi mới trong kiến trúc AI/ML cho phép chúng ta tạo

ra các mô hình nhúng (embedding models). Mô hình nhúng mã hóa tất cả các loại dữ liệu thành dạng vector để biểu diễn ý nghĩa và ngữ cảnh của tài liệu.

Cơ sở dữ liệu vector là cơ sở dữ liệu phi quan hệ (noSQL) chuyên dụng nhằm lưu trữ, xử lý cũng như truy xuất lượng lớn dữ liệu nhiều chiều ở dạng vector. Trong đó, vector là biểu diễn dữ liệu toán học mô tả thông tin của đối tượng dựa trên các đặc điểm hoặc chất lượng khác nhau của chúng. Mỗi vector đại diện cho một điểm dữ liệu, chẳng hạn như hình ảnh, âm thanh, văn bản, hoặc bao gồm một tập hợp các giá trị mô tả nhiều tính chất của chúng. Ví dụ như tính năng chụp ảnh và tìm kiếm hình ảnh tương tự là một điển một cơ sở dữ liệu vector mã nguồn mở dễ sử dụng và tích hợp sâu rộng với hệ sinh thái Python.



Hình 11: Cơ sở dữ liệu vector

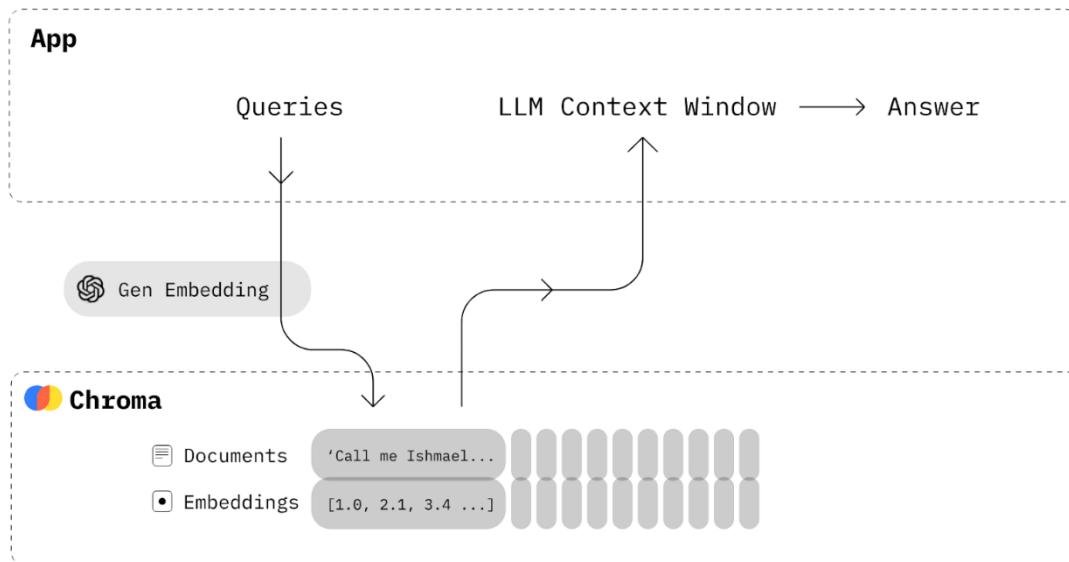
2.6.2. Định nghĩa Chroma



Chroma

Hình 12: Logo Chroma

Chroma là một kho lưu trữ vector nguồn mở được sử dụng để lưu trữ và truy xuất các nhúng vector. Công dụng chính của nó là lưu các nhúng cùng với siêu dữ liệu để các mô hình ngôn ngữ lớn sử dụng sau này. Ngoài ra, nó cũng có thể được sử dụng cho các công cụ tìm kiếm ngữ nghĩa trên dữ liệu văn bản.



Hình 13: Quy trình của Chroma

2.7. Giới thiệu về Python Django

2.7.1. Định nghĩa Django

Là một Web Framework bậc cao miễn phí, sử dụng mã nguồn mở được lập trình bằng Ngôn ngữ Python, Django hiện nay sở hữu một cộng đồng đông đảo người sử dụng và có nhiều tài liệu hỗ trợ cho các developer mới tìm hiểu. Django cho phép người dùng kiểm soát chặt chẽ quá trình phát triển website thông qua một nền tảng duy nhất, có khả năng xử lý các thao tác phức tạp khi lập trình web.



Hình 14: Python Django

2.7.2. Những điểm nổi bật của Django

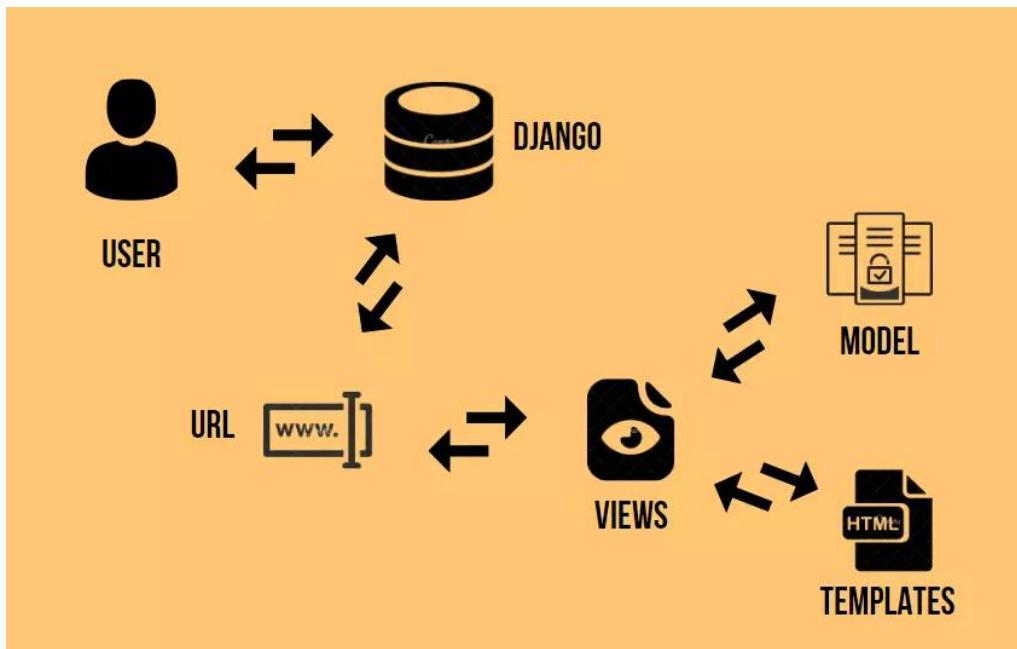
Các nhà phát triển web thường lựa chọn Django hơn so với những khung web khác, bởi vì các lý do sau:

- Tốc độ phát triển: Django được thiết kế với khung web mạnh mẽ để có thể nhanh chóng phát triển bất kỳ trang web nào bằng mã. Nhờ có cấu trúc khung hợp lý và dễ cài đặt vì vậy bạn có thể viết mã web với Django chỉ bằng vài dòng. Một số tác vụ phát triển phổ biến mà Django cung cấp cấu trúc sẵn sàng như: xác thực người dùng, quản trị nội dung, sơ đồ trang web, nguồn cấp dữ liệu RSS.
- Tiết kiệm chi phí: Django Software Foundation là một chức phi lợi nhuận vì vậy bạn không mất phí khi sử dụng phần mềm Django. Ngoài ra còn có một cộng đồng hoạt động tích cực, thường xuyên có các buổi gặp mặt, hội họp nhằm khuyến khích những nhà phát triển đánh giá và bảo trì và phần mềm được tốt hơn.
- Phổ biến: Django được sử dụng phổ biến để phát triển hàng nghìn dự án nguồn mở và các trang web nổi tiếng, có thể kể đến như: Instagram, Mozilla Firefox, Pinterest, National Geographic... Do được sử dụng thông dụng nên khung Django phát triển mạnh mẽ và có cơ sở hạ tầng hỗ trợ vững chắc.

2.7.2. Cách hoạt động của Django

Django hoạt động bằng cách sử dụng kiến trúc web theo mô hình MVT (Model – View – Template), cụ thể như sau:

- Model (Mô hình): Được xem là kho dữ liệu của Django, các cơ sở dữ liệu được quản lý và lưu trữ tại đây.
- View (Giao diện): Có nhiệm vụ xử lý các yêu cầu của người dùng, truy xuất dữ liệu từ Model và cuối cùng là trả kết quả về dạng Template.
- Template (Mẫu): Các tập tin HTML chứa cấu trúc và bộ cục trang web. Dữ liệu được điền vào Template và tạo ra trang web hoàn chỉnh và hiển thị cho người dùng.



Hình 15: Cách hoạt động của Django

2.8 Giới thiệu về Python Flask

2.8.1. Định nghĩa Flask

Flask là loại framework web phổ biến được viết bằng trình lập ngôn ngữ Python. Công nghệ thường được sử dụng để xây dựng trang web từ những ứng dụng đơn giản đến những hệ thống phức tạp hơn.

Flask được thiết kế để hoạt động và mở rộng một cách, đồng thời nó cũng cung cấp các công cụ và thư viện cần thiết để phát triển ứng dụng web hiệu quả. Flask cũng có cộng đồng sáng tạo và hỗ trợ mạnh mẽ từ cộng đồng Python.



Hình 16: Logo Flask

2.8.2. Các tính năng của Flask

Flask Framework sở hữu một số tính năng quan trọng mà nhà phát triển thường sử dụng để xây dựng hiệu ứng web. Dưới đây là một số tính năng chính của Flask:

- Nhẹ và dễ sử dụng: Công nghệ có cấu trúc nhẹ nhàng và mã nguồn dễ đọc, giúp người phát triển dễ dàng tiếp cận và tùy chỉnh theo nhu cầu cụ thể của họ.
- Định tuyến linh hoạt: Flask cung cấp cơ chế hoạt động định tuyến, cho phép người phát triển xác định các mẫu URL và phân bổ chúng cho các hàm xử lý tương ứng. Điều này giúp quản lý và xử lý yêu cầu HTTP một cách hiệu quả.
- Được mở rộng rộng rãi: Mặc dù mang đặc điểm rút gọn nhưng Flask vẫn có khả năng mở rộng mạnh mẽ thông qua việc sử dụng các tiện ích và thư viện của cộng đồng. Người dùng có thể phân tích các tính năng như xác thực, đăng nhập, điều hướng, cơ sở dữ liệu tương tác và nhiều tính năng khác.
- Máy chủ phát triển tích hợp: Flask cung cấp máy chủ phát triển hợp đồng, giúp người phát triển dễ dàng kiểm tra và phát triển ứng dụng mà không cần cấu hình bổ sung.
- Gửi yêu cầu RESTful: Flask hỗ trợ xây dựng API và các ứng dụng RESTful theo cách hoạt động và hiệu quả.
- Cộng đồng lớn và tích cực: Số lượng người dùng Flask rất đông và luôn nhận được hỗ trợ mạnh mẽ từ cộng đồng Python, điều này giúp người phát triển tìm kiếm thông tin và tài liệu một cách dễ dàng.

2.9. Giới thiệu về SQLAlchemy

2.9.1. Định nghĩa SQLAlchemy



Hình 17: Logo SQLAlchemy

SQLAlchemy là loại toolkit và ORM (Object-Relational Mapping) cho ngôn ngữ lập trình Python. Công nghệ được sử dụng để tương tác với cơ sở dữ liệu quan hệ một cách linh hoạt và hiệu quả.

SQLAlchemy cung cấp các công cụ và thư viện cho việc tạo và thực thi các truy vấn SQL, quản lý liên kết giữa đối tượng Python và dữ liệu trong cơ sở dữ liệu. Ngoài ra, nền tảng còn tạo các mô hình đối tượng cho ứng dụng Python.

2.9.2. Các tính năng nổi bật của SQLAlchemy

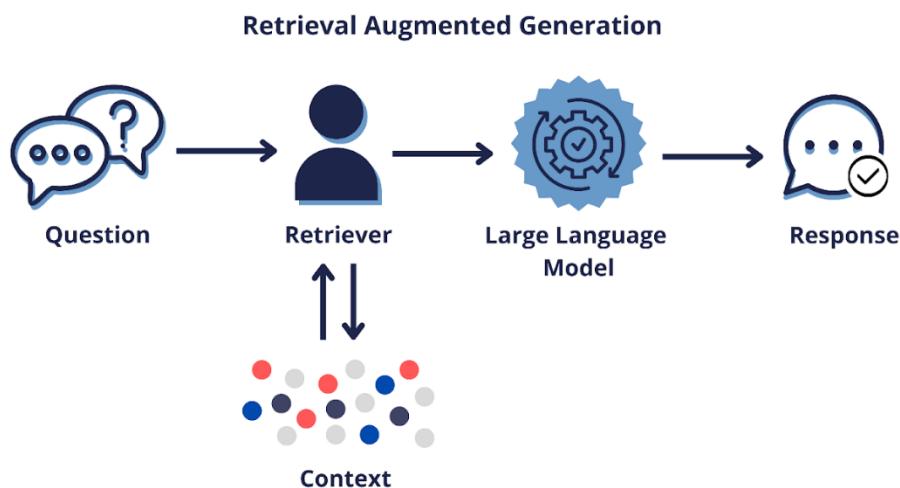
- ORM Layer: SQLAlchemy cung cấp bộ công cụ mạnh mẽ để ánh xạ dữ liệu cơ sở dữ liệu vào các đối tượng Python. Điều này cho phép lập trình viên làm việc với dữ liệu thông qua các đối tượng và phương thức Python thay vì sử dụng ngôn ngữ truy vấn SQL một cách trực tiếp.
- Core Layer: SQLAlchemy cung cấp một API cấp thấp cho việc tạo và thực thi các truy vấn SQL linh hoạt. Người dùng có thể tạo truy vấn một cách tự nhiên và tận dụng các tính năng của cơ sở dữ liệu quan hệ.
- Hỗ trợ nhiều cơ sở dữ liệu: SQLAlchemy hỗ trợ nhiều loại cơ sở dữ liệu quan hệ như SQLite, PostgreSQL, MySQL và Oracle, giúp đơn giản hóa việc chuyển đổi giữa các cơ sở dữ liệu.

- Mô-đun mở rộng: SQLAlchemy có tính linh hoạt cao và hỗ trợ các mô-đun mở rộng để mở rộng tính năng và khả năng tương tác với các cơ sở dữ liệu.

2.10. Giới thiệu về RAG

2.10.1. Định nghĩa RAG

Retrieval-Augmented Generation (RAG) là một kỹ thuật giúp nâng cao khả năng của mô hình sinh (language model generation) kết hợp với tri thức bên ngoài (external knowledge). Phương pháp này thực hiện bằng cách truy xuất thông tin liên quan từ kho tài liệu (tri thức) và sử dụng chúng cho quá trình sinh câu trả lời dựa trên LLMs.



Hình 18: Mô tả quy trình của RAG

2.10.2. Thành phần của RAG

Hai thành phần chính của RAG:

- Retrieval Module (Mô-đun truy xuất): Đây là bộ phận chịu trách nhiệm tìm kiếm, thu thập dữ liệu từ các kho thông tin có sẵn, chẳng hạn như cơ sở dữ liệu nội bộ, tài liệu hướng dẫn, hoặc các nguồn trực tuyến. Mô-đun này đảm bảo rằng các phản hồi được tạo ra luôn dựa trên thông tin chính xác, phù hợp.

- Generation Module (Mô-đun sinh): Sau khi dữ liệu được truy xuất, mô-đun này sẽ xử lý, tạo phản hồi bằng ngôn ngữ tự nhiên thông qua các mô hình ngôn ngữ lớn (Large Language Models – LLM). Mục tiêu của mô-đun sinh là tạo ra các câu trả lời mạch lạc, mang tính cá nhân hóa và thân thiện với người dùng.

2.10.3. Quy trình hoạt động của RAG

- Thu thập dữ liệu: dữ liệu chính là nền tảng để hệ thống hoạt động hiệu quả. Hệ thống lấy dữ liệu từ nhiều nguồn khác nhau, bao gồm: Các tài liệu hướng dẫn, báo cáo kỹ thuật, email, các bài viết chuyên môn, nghiên cứu ngành, ...
- Phân chia dữ liệu: Sau khi thu thập, dữ liệu được phân loại, tổ chức theo các cụm nhỏ để tăng tính hiệu quả trong việc truy xuất.
- Nhúng tài liệu: Nhúng tài liệu là bước quan trọng để biến dữ liệu văn bản thành dạng mà máy tính có thể hiểu và xử lý. Điều này được thực hiện thông qua các thuật toán học sâu (Deep Learning), giúp chuyển đổi văn bản thành biểu diễn số học (vector embeddings).
- Quy trình hoạt động: Các mô hình ngôn ngữ như BERT, RoBERTa, hoặc Sentence Transformers được sử dụng để phân tích ngữ nghĩa của từng đoạn văn bản. Những mô hình này đảm bảo dữ liệu được biểu diễn một cách chính xác, tối ưu.
- Tác dụng: Dữ liệu sau khi nhúng có thể được so khớp ngữ nghĩa với các truy vấn của người dùng, bất kể cách diễn đạt của họ. Ví dụ, dù người dùng hỏi “làm sao để sửa lỗi hệ thống?” hay “khắc phục lỗi IT như thế nào?”, hệ thống đều hiểu đây là cùng một ý định và truy xuất các tài liệu liên quan.
- Xử lý truy vấn người dùng: Khi người dùng gửi truy vấn, hệ thống RAG thực hiện hai nhiệm vụ chính: phân tích ngữ nghĩa và so khớp thông tin.
- Phân tích ngữ nghĩa: Truy vấn được chuyển đổi thành biểu diễn vector tương tự như dữ liệu đã nhúng, nhờ đó hệ thống hiểu được ý định thực sự của người dùng. Ví dụ,

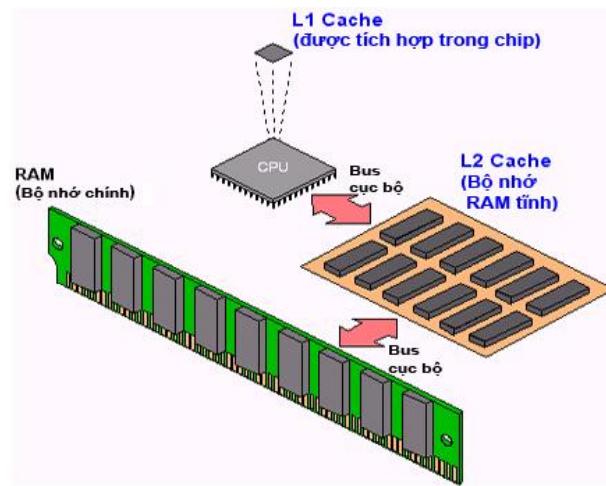
nếu truy vấn là “Làm thế nào để bảo trì máy chủ hiệu quả?”, RAG sẽ nhận diện từ khóa “bảo trì” và “máy chủ” để tìm kiếm các tài liệu phù hợp.

- So khớp thông tin: Dựa trên biểu diễn vector, hệ thống tìm kiếm các dữ liệu tương đồng trong kho và trả về kết quả chính xác nhất.
- Tạo phản hồi với LLM: Mô hình tích hợp thông tin truy xuất với ngữ cảnh truy vấn. Sau đó tạo ra phản hồi tự nhiên, liền mạch và phù hợp với nhu cầu của người dùng.

2.11. Giới thiệu về Cache

2.11.1. Định nghĩa Cache

Cache là một cơ chế lưu trữ tạm thời dữ liệu được truy cập hoặc tính toán trước đó, nhằm tăng tốc độ xử lý và giảm thiểu thời gian cũng như tài nguyên cần thiết để tái sử dụng dữ liệu đó.



Hình 19: Cấu trúc và vị trí của các loại bộ nhớ cache trong hệ thống máy tính

Cache lưu trữ một phiên bản của dữ liệu tại một vị trí nhanh hơn (bộ nhớ nhanh, ổ cứng, hoặc cơ sở dữ liệu) so với nguồn gốc, để khi dữ liệu được yêu cầu lần nữa, nó có thể được truy xuất ngay mà không cần tính toán hoặc tải lại từ nguồn ban đầu.

2.11.2. Đặc điểm của Cache

- Tốc độ cao: Cache thường được lưu trong bộ nhớ nhanh (RAM hoặc bộ nhớ cục bộ) để cải thiện hiệu năng truy xuất dữ liệu.
- Thời gian sống ngắn: Dữ liệu trong cache có thể hết hạn hoặc bị thay thế (evicted) dựa trên các chính sách quản lý.
- Dễ truy xuất: Dữ liệu trong cache thường được tổ chức để truy xuất nhanh, ví dụ: sử dụng khóa (key-value) hoặc chỉ mục.

2.11.3. Phân loại Cache

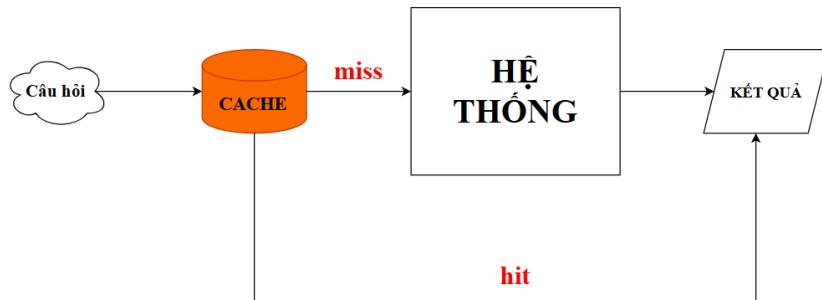
- In-Memory Cache: Lưu trữ trong bộ nhớ (RAM), ví dụ: Redis, Memcached.
- Database Cache: Lưu trữ trong cơ sở dữ liệu, như cơ chế bạn đang sử dụng với PostgreSQL.
- Disk Cache: Lưu trữ trên ổ đĩa, sử dụng trong trường hợp dữ liệu quá lớn để lưu trong RAM.
- Application Cache: Dùng trong các ứng dụng cụ thể.

2.11.4. Tối ưu hóa phản hồi hệ thống bằng Cache cho câu hỏi lặp lại

Khi người dùng đặt lại những câu hỏi trùng lặp hoàn toàn với những câu đã hỏi trước đó, hệ thống sẽ tận dụng cơ chế Cache để tăng tốc độ phản hồi và tối ưu hóa tài nguyên. Bằng cách này, hệ thống không chỉ giảm thiểu thời gian xử lý mà còn cải thiện hiệu suất tổng thể, mang lại trải nghiệm nhanh chóng và liền mạch hơn cho người dùng. Cơ chế này đặc biệt hữu ích trong việc giảm tải cho quy trình RAG-tổn nhiều thời gian, đảm bảo hoạt động hiệu quả ngay cả khi có nhiều truy vấn lặp lại.

2.11.5. Quy trình truy vấn Cache của Chatbot

Khi một yêu cầu prompt được gửi đến LLM, hệ thống sẽ kiểm tra cache trước. Hệ thống sử dụng tìm kiếm toàn văn prompt_tsv để so sánh yêu cầu hiện tại với các yêu cầu đã được lưu trong bảng. Nếu tìm thấy bản ghi phù hợp trong cache, hệ thống sẽ trả về response mà không cần gọi lại LLM.



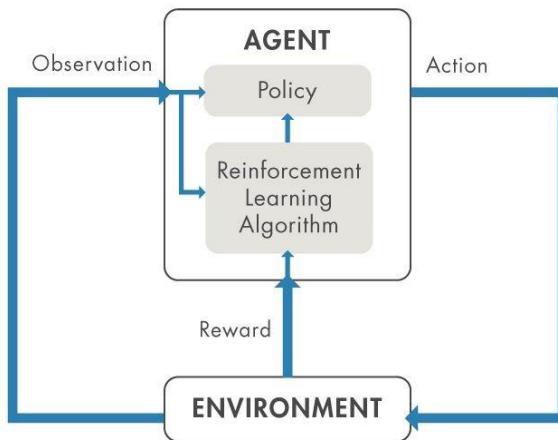
Hình 20: Quy trình truy vấn Cache của Chatbot

2.12. Giới thiệu về học tăng cường

Học tăng cường là một nhánh quan trọng trong Trí tuệ nhân tạo, nơi Agent học cách hành động trong một môi trường để tối đa hóa phần thưởng nhận được. Học tăng cường khác biệt với các phương pháp học máy khác ở chỗ nó tập trung vào việc học thông qua thử và sai (trial-and-error), thay vì học từ một tập dữ liệu cố định.

Một bài toán học tăng cường thường được mô hình hóa bằng Quy trình Quyết định Markov (Markov Decision Process - MDP), bao gồm:

- S: Tập hợp các trạng thái (States) mô tả môi trường.
- A: Tập hợp các hành động (Actions) mà agent có thể thực hiện.
- R: Hàm phần thưởng (Reward) xác định giá trị phần thưởng khi agent thực hiện một hành động ở trạng thái cụ thể.
- P: Hàm chuyển trạng thái (Transition Probability) xác suất di chuyển từ trạng thái hiện tại sang trạng thái tiếp theo.



Hình 21: Quy trình tương tác giữa Agent và môi trường trong học tăng cường

Mục tiêu của học tăng cường là tìm một chính sách tối ưu, thường ký hiệu là π^* , để tối đa hóa tổng phần thưởng kỳ vọng dài hạn:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

Hình 22: Công thức hàm chính sách tối ưu

Trong đó:

- γ : Hệ số chiết khấu, cho biết mức độ ưu tiên giữa phần thưởng hiện tại và phần thưởng tương lai.
- k : Số bước trong tương lai, tính từ thời điểm t .
- R_{t+k+1} : Phần thưởng nhận được tại bước $t+k+1$. Đây là phần thưởng mà agent nhận được khi thực hiện hành động trong môi trường.

2.13. Giới thiệu về Q-Learning

2.13.1. Định nghĩa Q-Learning

Q-Learning là một thuật toán học tăng cường, thuộc loại học không giám sát, được sử dụng để học một chiến lược tối ưu cho một tác nhân trong một môi trường không biết trước đó.



Hình 23: Q-Learning

2.13.2. Các thành phần của Q-Learning

- Giá trị Q (Q-Values) hoặc Giá trị hành động (Action-Values): Giá trị Q được định nghĩa cho các trạng thái và hành động. $Q(S,A)$ là một ước lượng về mức độ tốt khi thực hiện hành động A tại trạng thái S .
- Rewards và Episodes: Trong suốt vòng đời của mình, một tác nhân (agent) bắt đầu từ một trạng thái khởi đầu, thực hiện nhiều lần chuyển tiếp từ trạng thái hiện tại sang trạng thái tiếp theo dựa trên hành động mà nó chọn và môi trường mà nó tương tác. Ở mỗi bước chuyển tiếp, tác nhân từ một trạng thái thực hiện một hành động, quan sát một phần thưởng từ môi trường, và sau đó chuyển sang một trạng thái khác. Nếu tại bất kỳ thời điểm nào, tác nhân rơi vào một trạng thái kết thúc (terminating state),

tức là không còn chuyển tiếp nào có thể thực hiện, tập hợp bước (episode) đó được coi là hoàn thành.

- Lựa chọn Hành động với Chính sách ϵ -greedy: Một phương pháp đơn giản để chọn hành động dựa trên ước lượng hiện tại của giá trị Q (Q-value) là sử dụng chính sách ϵ -greedy
 - Khai thác (Exploitation): Với xác suất $1-\epsilon$, chọn hành động có giá trị Q cao nhất, dựa trên hiểu biết hiện tại.
 - Khám phá (Exploration): Với xác suất ϵ , chọn hành động ngẫu nhiên để khám phá lợi ích tiềm năng của những hành động mới.

2.13.3. Công thức của Q-Learning

$$Q(s, a) = r(s, a) + \gamma \max_a Q(s', a)$$

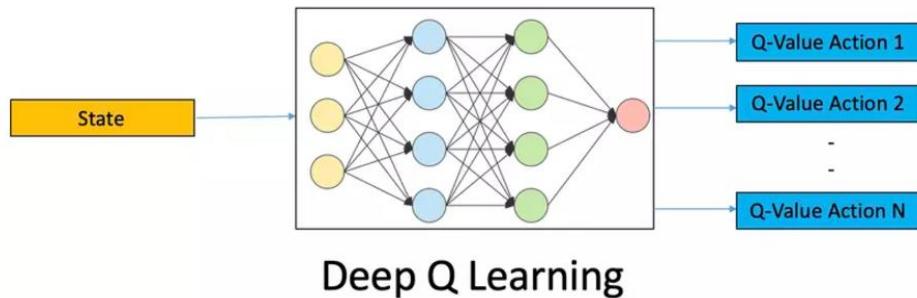
Hình 24: Công thức Q-Learning

Trong đó:

- $Q(s, a)$ là Q-value khi thực hiện action a tại state s
- $r(s, a)$ là reward nhận được
- s' là state kế tiếp.
- γ là hệ số discount, đảm bảo càng "xa" đích Q-value càng nhỏ

2.14. Giới thiệu về Deep Q-Learning

2.14.1 Nguyên lý hoạt động



Hình 25: Nguyên lý hoạt động của Deep Q-Learning

Deep Q-Learning là sự kết hợp giữa học tăng cường và mạng nơ-ron sâu. Chúng được sử dụng để giải quyết vấn đề Q-learning - một kỹ thuật học tăng cường truyền thống. Deep Q-Learning sử dụng một mạng nơ-ron để ước lượng giá trị Q (Q-value), cho phép tác nhân chọn hành động tối ưu trong mỗi trạng thái.

Mạng nơ-ron được sử dụng trong Deep Q-Learning bao gồm:

- Input layer: Nhận đầu vào là trạng thái s.
- Hidden layers: Tích hợp các lớp ẩn để học các đặc trưng từ trạng thái.
- Output layer: Cung cấp giá trị $Q(s,a)$ cho mỗi hành động a.

2.14.2 Kỹ thuật tăng cường cho Deep Q-Learning

Deep Q-Learning có thể kết hợp các kỹ thuật như:

- Replay buffer: Lưu trữ kinh nghiệm từ các lần chơi trước, giúp giảm sự phụ thuộc vào chuỗi dữ liệu gần nhất.
- Target network: Sử dụng một mạng nơ-ron mục tiêu tạm thời để làm giảm sự không ổn định trong quá trình huấn luyện.

2.14.3 Công thức cập nhật

Mạng nơ-ron được tối ưu hóa thông qua hàm mất mát (loss function):

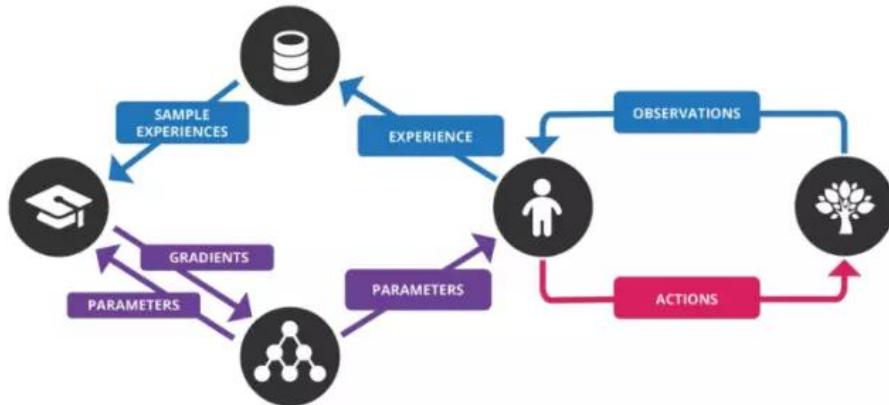
$$L(\theta) = \mathbb{E} \left[\left(R + \gamma \max_{a'} Q_{\text{target}}(s', a'; \theta^-) - Q(s, a; \theta) \right)^2 \right]$$

Hình 26: Công thức cập nhật

Trong đó:

- θ : Trọng số của mạng nơ-ron hiện tại.
- θ^- : Trọng số của mạng nơ-ron mục tiêu.
- Q_{target} : Giá trị Q dự đoán từ mạng mục tiêu.

Nhìn lại toàn bộ quá trình, ta có:



Hình 27: Toàn bộ quá trình áp dụng Deep Q-Learning

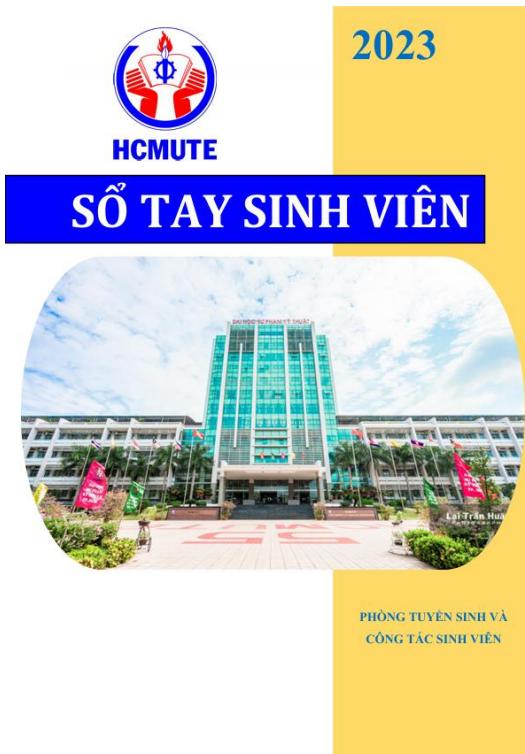
CHƯƠNG 3: TÌM HIỂU DỮ LIỆU

3.1. Khảo sát hiện trạng trang tư vấn Sinh viên

Để có thể thực hiện đề tài, cũng như triển khai giống với thực tế, phù hợp với thực tiễn nhóm tiến hành khảo sát trang web Tư vấn Sinh viên truyền thống của trường Đại học Sư Phạm Kỹ Thuật TP.HCM: <https://tuvansinhvien.hcmute.edu.vn>, trang web là một diễn đàn giải đáp mà Sinh viên sẽ đăng các thắc mắc cá nhân rồi chờ các thầy/cô/cán bộ có thẩm quyền hoặc chịu trách nhiệm về nội dung đó trả lời. Nhóm tập trung phân tích, tổng hợp các câu hỏi thường xuyên được hỏi, cũng như các câu trả lời tương ứng, qua đó xác định được phạm vi các cặp FAQ mà Sinh viên hay thắc mắc và đưa vào hệ thống giúp truy xuất nhanh hơn, không tốn quá nhiều thời gian.

Hình 28: Trang tư vấn sinh viên của trường

3.2 Khảo sát Sổ tay Sinh viên năm 2023



Hình 29: Sổ tay Sinh viên 2023

Sổ tay Sinh viên năm 2023 được đăng tải công khai cho tất cả Sinh viên trường Đại học Sư Phạm Kỹ Thuật TP.HCM áp dụng gồm các nội dung như sau:

1. Hướng dẫn Sinh viên liên hệ với các đơn vị trong trường để giải quyết các yêu cầu.
2. Quyết định về việc ban hành Quy chế đào tạo trình độ Đại học.
3. Quy chế đào tạo trình độ Đại học
 - a. Chương I: Những quy định chung
 - b. Chương II: Lập kế hoạch và tổ chức giảng dạy
 - c. Chương III: Đánh giá kết quả học tập và cấp bằng tốt nghiệp
 - d. Chương IV: Những quy định khác đối với Sinh viên
 - e. Chương V: Tổ chức thực hiện

CHƯƠNG 3: TÌM HIỂU DỮ LIỆU

4. Phụ lục 1: Quy định về thời gian học tập đối với đối tượng ưu tiên.
5. Phụ lục 2: Quy định về chuẩn đầu ra ngoại ngữ.
6. Phụ lục 3: Quy định về tham gia Nghiên cứu Khoa học của các chương trình chính quy Chất lượng cao.
7. Quyết định về việc hướng dẫn thực hiện quy chế Đào tạo Đại học và Cao đẳng hệ Chính quy theo hệ thống tín chỉ.
8. Hướng dẫn thực hiện quy chế
 - a. Chương I: Những quy định chung
 - b. Chương II: Tổ chức đào tạo
 - c. Chương III: Kiểm tra và thi học phần
 - d. Chương IV: Xét và công nhận tốt nghiệp
 - e. Chương V: Tổ chức thực hiện
9. Quyết định về việc tính điểm học phần đối với Sinh viên Đại học Chính quy.
10. Quyết định về việc chuyển đổi điểm của các học phần ngoại ngữ.
11. Quy định chuyển đổi điểm các học phần ngoại ngữ.
12. Quyết định về việc ban hành quy chế quản lý, tổ chức Đào tạo.
13. Quy chế quản lý, tổ chức đào tạo trực tuyến
 - a. Chương I: Những quy định chung
 - b. Chương II: Quy định tổ chức thẩm định học liệu điện tử UTEX MOOC trước khi đăng tải
 - c. Chương III: Quy định quản lý, tổ chức, đào tạo trực tuyến
 - d. Chương IV: Tổ chức thực hiện

CHƯƠNG 3: TÌM HIỂU DỮ LIỆU

14. Quyết định về việc ban hành quy định đánh giá kết quả rèn luyện của Sinh viên Đại học Chính quy.
15. Quy định đánh giá kết quả rèn luyện của Sinh viên Đại học Chính quy
 - a. Chương I: Quy định chung
 - b. Chương II: Các tiêu chí đánh giá và khung điểm
 - c. Chương IV: Tổ chức đánh giá và sử dụng kết quả đánh giá
 - d. Chương V: Tổ chức thực hiện
16. Quyết định về việc ban hành quy định tổ chức các chương Công tác Xã hội cho Sinh viên Đại học Chính quy.
17. Quy định tổ chức các chương Công tác Xã hội cho Sinh viên Đại học Chính quy.
18. Quyết định về việc ban hành thực hiện các chương trình Công tác Xã hội cho Sinh viên hệ Đại học Chính quy.
19. Quy định về công tác Tư vấn đối với Sinh viên hệ Đại học Chính quy.
20. Hướng dẫn về việc xét, cấp học bổng Khuyến khích học tập đối với Sinh viên hệ Đại học Chính quy.
21. Hướng dẫn về việc thực hiện chính sách miễn giảm học phí, trợ cấp xã hội.
22. Quyết định về việc ban hành quy chế thi đua khen thưởng đối với Sinh viên hệ Đại học Chính quy.
23. Quy chế thi đua khen thưởng đối với Sinh viên hệ Đại học Chính quy.
24. Nghị quyết của cuộc họp Hội đồng trường.
25. Quyết định về việc ban hành quy tắc văn hóa ứng xử của Sinh viên Đại học Sư Phạm Kỹ Thuật TP.HCM.

3.3. Kết luận sau khi khảo sát

Phân tích	Sổ tay Sinh viên	Trang Tư vấn Sinh viên
Ưu điểm	Tổng hợp đầy đủ các thông tin Tiện lợi, có thể tra cứu qua Internet hoặc dạng sổ tay giấy	Câu trả lời chính xác, chi tiết hơn Lưu trữ thông tin lâu dài, có thể tra cứu lại hoặc tìm kiếm vấn đề tương tự
Nhược điểm	Quá nhiều thông tin dư thừa, không cần thiết Chỉ cung cấp các thông tin tổng quát Sinh viên tốn nhiều thời gian để tra cứu ra vấn đề mình gặp phải	Thời gian chờ đợi lâu, phải chờ vài tiếng, thậm chí vài ngày Chất lượng câu trả lời không đồng nhất Có nhiều bài viết mà nội dung mơ hồ, hoặc thậm chí là spam

Bảng 2: Kết luận sau khi khảo sát thực tế

Từ những ưu nhược điểm đã phân tích và tổng hợp, nhóm quyết định tiền xử lý dữ liệu, để có chất lượng dữ liệu tốt nhất, đáp ứng được kỳ vọng mà Sinh viên mong đợi, giúp tăng trải nghiệm sử dụng của Sinh viên, đồng thời việc tiền xử lý dữ liệu cũng giúp tăng hiệu quả mô hình ngôn ngữ trong việc truy xuất và tìm ra câu trả lời phù hợp.

3.4 Tiền xử lý dữ liệu

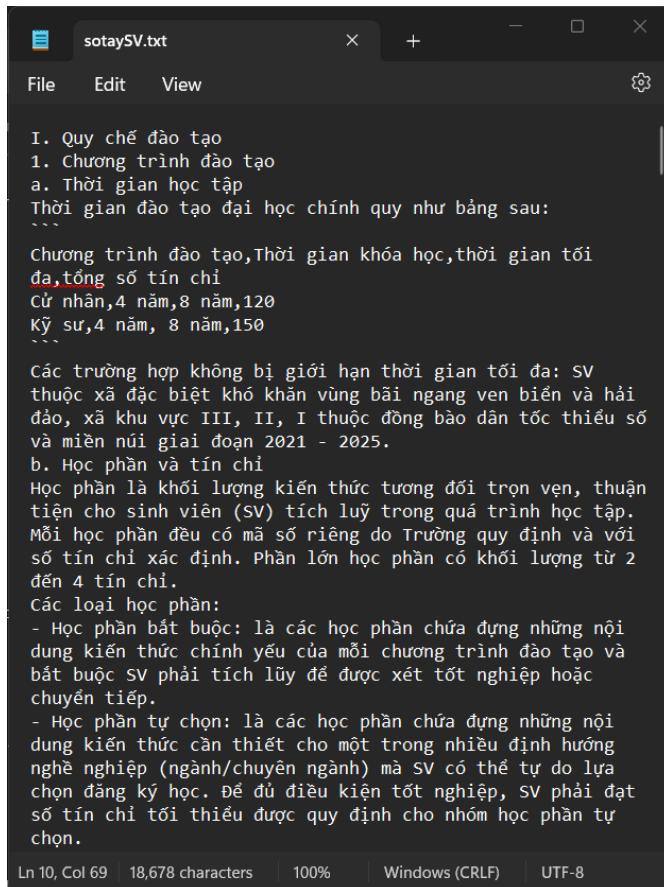
3.4.1 Tiền xử lý dữ liệu Sổ tay Sinh viên

Trước hết, nhóm quyết định chỉ giữ lại các quy định, quyết định, điều lệ dành cho Sinh viên hệ Đại học Chính quy, bỏ qua các quyết định, quy định, điều lệ dành cho Giảng viên, cá nhân, tổ chức mà chỉ hướng chung đến đối tượng là Sinh viên tại trường.

Dữ liệu ban đầu được thu thập từ các tài liệu chính thức của trường, chủ yếu ở định dạng pdf. Để thuận tiện cho việc xử lý, tệp pdf này được chuyển đổi sang định dạng văn

CHƯƠNG 3: TÌM HIẾU DỮ LIỆU

bản thuần txt bằng thủ công. Các bảng trong tệp pdf cũng được chuyển sang dạng text để phù hợp với mô hình ngôn ngữ.



Hình 30: Sổ tay Sinh viên được chuyển sang dạng txt

Sau khi chuyển đổi, dữ liệu được làm sạch, loại bỏ các ký tự đặc biệt và thông tin không cần thiết như số trang, đầu trang, cuối trang.

Cuối cùng, văn bản sau khi làm sạch được tách từ thành các token bằng thư viện *underthesea* của Pytorch, đảm bảo phù hợp với việc phân tích tiếng Việt.

3.4.2 Tiết xử lý dữ liệu trang Tư vấn Sinh viên

Trước hết nhóm vào trang <https://tuvansinhvien.hcmute.edu.vn> chọn ra những cặp câu hỏi, và câu trả lời hay được hỏi nhất, bỏ qua những câu hỏi mang tính chất quá chi tiết, cá nhân,... Chỉ hướng tới các câu hỏi mang tính chung chung mà Sinh viên thường hay gặp phải.

Sau đó, giữ nguyên câu trả lời của Giảng viên/Cán bộ có trách nhiệm phụ trách trả lời thắc mắc đó, nhưng có sửa lại văn phong của câu hỏi Sinh viên, thay thế các từ ngữ vùng miền, từ ngữ khó hiểu, loại bỏ các ký tự đặc biệt.

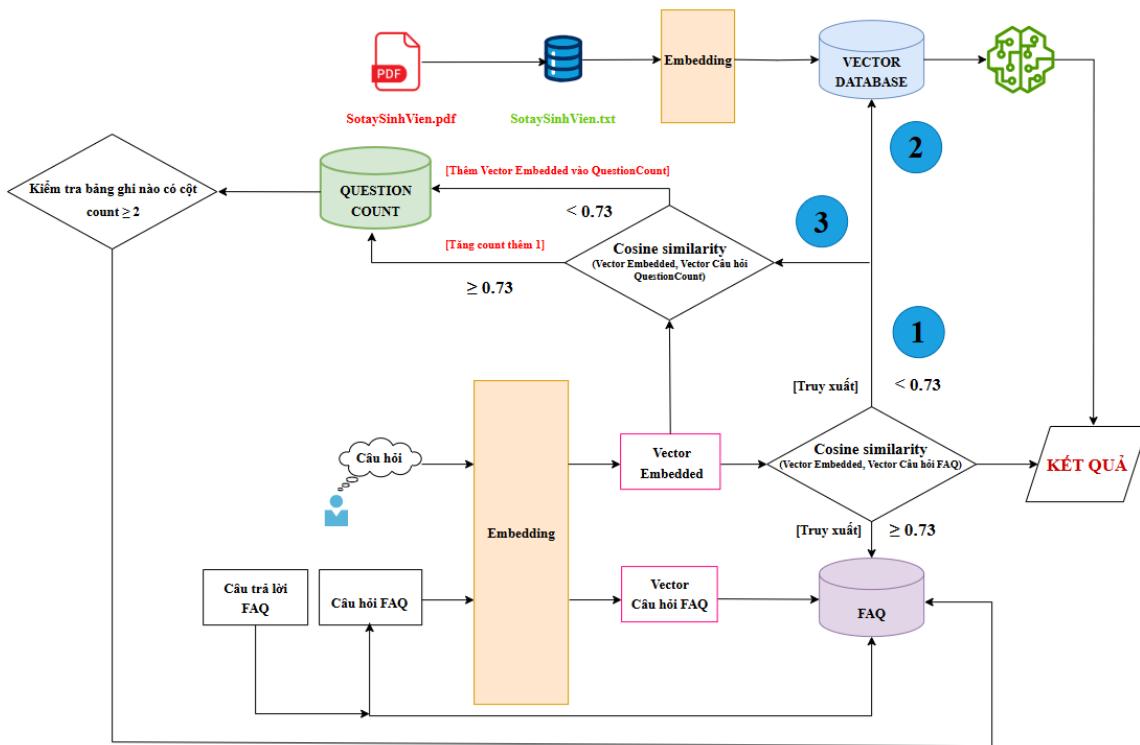
CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

4.1. Cấu hình phần cứng

Thành phần	Cấu hình chi tiết			Chú thích
Máy chủ	Zeppelin Server			Triển khai phương pháp RAG và xây dựng Flask API
GPU	Tesla P100-PCIE-16GB			Tối ưu n_layers cho mô hình ngôn ngữ lớn cũng như tăng tốc quá trình tính toán embedding
CPU	Intel Core i9 10900X			Các tác vụ thông thường
RAM	64GB			Đảm bảo khả năng xử lý các tập dữ liệu lớn trong bộ nhớ.
VRAM	16GB			Hỗ trợ GPU trong việc lưu trữ và xử lý các tensor lớn khi chạy mô hình ngôn ngữ lớn
Laptop	MSI GF63 Thin 11UC-1228VN			Gọi API từ Zeppelin Server, để xây dựng các trang giao diện
	Bộ vi xử lý Intel Core i7-13800H	Card đồ họa GEFORCE GTX NVIDIA	RAM 8GB DDR4 và ổ cứng SSD 512GB	

Bảng 3: Cấu hình phần cứng

4.2 Kiến trúc hệ thống



Hình 31: Mô tả kiến trúc hệ thống Chatbot

Ứng dụng “Chatbot tư vấn Sinh viên trường Đại Học Sư Phạm Kỹ Thuật TP.HCM”. Đầu tiên, dữ liệu từ tài liệu PDF (SotaySinhVien.pdf) chứa nội quy và quy chế được trích xuất và chuyển đổi thành tệp văn bản (SotaySinhVien.txt) chỉ lưu trữ các quy định dành riêng cho Sinh viên hệ Đại học Chính quy. Nội dung văn bản sau đó được xử lý và chuyển đổi thành các vector embedding bằng cách sử dụng mô hình HuggingFaceEmbeddings cụ thể là mô hình embedding BAAI/bge-base-en-v1.5, nhằm biểu diễn dữ liệu dưới dạng số hóa phù hợp cho việc so sánh và truy vấn. Các vector này được lưu trữ trong một Vector Database của Chroma để phục vụ cho việc tìm kiếm và truy xuất thông tin hiệu quả. Bên cạnh đó nhóm cũng chuẩn bị sẵn bộ câu hỏi – câu trả lời thường xuyên được hỏi, sau đó tổ chức bảng FAQ trong hệ quản trị cơ sở dữ liệu PostgreSQL theo cấu trúc (câu hỏi FAQ, vector câu hỏi FAQ đã embedding, câu trả lời FAQ).

Khi người dùng nhập một câu hỏi vào hệ thống chatbot, câu hỏi sẽ được xử lý và chuyển đổi thành một vector embedding bằng mô hình embedding đã sử dụng trước đó.

Vector của câu hỏi này được so sánh với các vector của các câu hỏi trong hệ thống FAQ bằng cách sử dụng độ tương đồng Cosine-Cosine Similarity. Nếu kết quả Cosine Similarity giữa vector của câu hỏi và vector câu hỏi từ cơ sở dữ liệu FAQ lớn hơn hoặc bằng 73%, chatbot sẽ ngay lập tức truy xuất câu hỏi tương tự trong cơ sở dữ liệu FAQ và trả lời dựa trên nội dung của hệ thống này. Ngược lại nếu nhỏ hơn 73%, quy trình RAG sẽ được thực thi, cụ thể sẽ truy xuất vào Vector Database đã được khởi tạo từ đầu lấy ra 7 đoạn văn liên quan nhất từ cơ sở dữ liệu, sau đó kết quả sẽ được gửi cho mô hình ngôn ngữ lớn để sinh phản hồi tạo kết quả. Bên cạnh đó, hệ thống còn tự động xác định câu hỏi thường xuyên được hỏi nếu tần suất lớn hơn hay bằng 2, kết hợp với câu trả lời từ RAG, rồi thêm vào bảng FAQ, thuận tiện cho Sinh viên truy xuất mà không cần trải qua giai đoạn thêm thủ công.

Hệ thống này được xây dựng nhằm đảm bảo tính hiệu quả, chính xác, nhanh chóng trong việc giải đáp thắc mắc của Sinh viên. Việc sử dụng Vector Database và kỹ thuật xử lý ngôn ngữ tự nhiên, giúp chatbot có khả năng tìm kiếm thông tin nhanh chóng, đồng thời đảm bảo khả năng mở rộng khi cập nhật thêm tài liệu hoặc nâng cấp mô hình. Nguồn độ tương đồng 73% được chọn dựa trên thử nghiệm thực tế để tối ưu hóa việc phân loại câu hỏi và nguồn trả lời phù hợp dựa trên tài liệu đã cung cấp.

4.3 Mô tả cơ sở dữ liệu

4.3.1 Vector Database

Nội dung của Sổ tay Sinh viên 2023 gồm 18,678 kí tự, được chia thành 38 chunks mỗi chunk có 768 kí tự-phù hợp với kích thước mô hình embedding, và giữa các chunk có thể chồng lên nhau 300 kí tự, do đó ta có được 38 vector embedding, nhưng do Chroma đã băm nên giá trị của các vector thẻ truy cập hay nhìn thấy.

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

The screenshot shows a database interface with two main windows. On the left, a table named 'embedding_fulltext_search_content' is displayed with columns 'id' and 'content'. The table contains 22 rows of text. On the right, a text editor window titled 'Edit Database Cell' is open, showing a list of numbered items related to education and training programs. The text editor has various options like 'Mode: Text', 'Edit Pragmas', 'Execute SQL', and 'File'.

Hình 32: Vector Database

4.3.2 FAQ

	COLUMN NAME	#	DATA TYPE	IDENTITY	LOCAL	COLLATION	NOT NULL	DEFAULT	GENERATED	COMMENT
<input type="checkbox"/>	123	id	int4		true		true			
<input type="checkbox"/>	ABC	question	varchar		true	default	true			
<input type="checkbox"/>	ABC	answer	varchar		true	default	true			
<input type="checkbox"/>	ABC	embedding	varchar		true	default	true			

Hình 33: Bảng FAQ từ PostgreSQL

Tên trường	Kiểu dữ liệu	Chức năng
id	int4	Trường lưu trữ số thứ tự các bảng ghi
question	varchar	Trường lưu trữ câu hỏi thường được hỏi
answer	varchar	Trường lưu trữ câu trả lời cho câu hỏi thường được hỏi
embedding	varchar	Trường lưu trữ vector embedding của câu hỏi thường được hỏi đã được embedding

Bảng 4: Mô tả bảng FAQ

4.3.3 QuestionCount

	COLUMN NAME	#	DATA TYPE	IDENTITY	LOCAL	COLLATION	NOT NULL	DEFAULT	GENERATED	COMMENT
<input type="checkbox"/>	123 id	1	int4		true		true			
<input type="checkbox"/>	ABC question_embedding	2	varchar		true	default	true			
<input type="checkbox"/>	123 count	3	int4		true		false			

Hình 34: Bảng QuestionCount từ PostgreSQL

Tên trường	Kiểu dữ liệu	Chức năng
id	int4	Trường lưu trữ số thứ tự các bảng ghi
question_embedding	varchar	Trường lưu trữ vector embedding của câu hỏi mà người dùng hỏi.
count	int4	Trường lưu trữ tần suất câu hỏi được hỏi

Bảng 5: Mô tả bảng QuestionCount

4.3.4 Feedback

	COLUMN NAME	#	DATA TYPE	IDENTITY	LOCAL	COLLATION	NOT NULL	DEFAULT	GENERATED	COMMENT
<input type="checkbox"/>	123 id	1	int4		true		true			
<input type="checkbox"/>	ABC question	2	varchar		true	default	true			
<input type="checkbox"/>	ABC embedding_question	3	varchar		true	default	true			
<input type="checkbox"/>	ABC response_1	4	varchar		true	default	true			
<input type="checkbox"/>	ABC response_2	5	varchar		true	default	true			
<input type="checkbox"/>	ABC embedding_1	6	varchar		true	default	true			
<input type="checkbox"/>	ABC embedding_2	7	varchar		true	default	true			
<input type="checkbox"/>	123 likes_1	8	int4		true		false			
<input type="checkbox"/>	123 dislikes_1	9	int4		true		false			
<input type="checkbox"/>	123 likes_2	10	int4		true		false			
<input type="checkbox"/>	123 dislikes_2	11	int4		true		false			
<input type="checkbox"/>	∅ t	12	timestamp		true		false			

Hình 35: Bảng Feedback từ PostgreSQL

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

Tên trường	Kiểu dữ liệu	Chức năng
id	int4	Trường lưu trữ số thứ tự các bảng ghi
question	varchar	Trường lưu trữ câu hỏi của người dùng
embedding_question	varchar	Trường lưu trữ vector embedding câu hỏi của người dùng đã embedding
response_1	varchar	Trường lưu trữ câu trả lời FAQ có độ tương đồng cosine cao nhất
response_2	varchar	Trường lưu trữ câu trả lời FAQ có độ tương đồng cosine cao nhì
embedding_1	varchar	Trường lưu trữ vector embedding của response_1 đã embedding
embedding_2	varchar	Trường lưu trữ vector embedding của response_2 đã embedding
likes_1	int4	Trường lưu trữ số lượt thích response_1 của người dùng
dislikes_1	int4	Trường lưu trữ số lượt không thích response_1 của người dùng
likes_2	int4	Trường lưu trữ số lượt thích response_2 của người dùng
dislikes_2	int4	Trường lưu trữ số lượt không thích response_2 của người dùng
t	timestamp	Trường lưu trữ thời gian hệ thống đưa ra 2 phản hồi cho người dùng

Bảng 6: Mô tả bảng Feedback

4.3.5. Cache

		COLUMN NAME	#	DATA TYPE	IDENTITY	LOCAL	COLLATION	NOT NULL	DEFAULT	GENERATED	COMMENT
□	123	id	1	int4		true		true			
□	RBC	prompt	2	varchar		true	default	true			
□	RBC	llm	3	varchar		true	default	true			
□	123	idx	4	int4		true		false			
□	RBC	response	5	varchar		true	default	false			
□	[+]	prompt_tsv	6	tsvector		true		false			to_tsvector('simple')::regconfig, (llm::text ' '::tex...

Hình 36: Bảng Cache từ PostgreSQL

Tên trường	Kiểu dữ liệu	Chức năng
id	int4	Trường lưu trữ số thứ tự các bảng ghi
prompt	varchar	Trường lưu trữ câu hỏi đầu vào mà người dùng đã gửi đến LLM
llm	varchar	Trường lưu trữ danh sách các cặp key-value (dạng tuple) mô tả các tham số cấu hình được sử dụng khi gọi mô hình ngôn ngữ lớn
response	varchar	Trường lưu trữ phản hồi từ mô hình LLM tương ứng với prompt
prompt_tsv	tsvector	Trường lưu trữ dạng biểu diễn tìm kiếm toàn văn của kết hợp llm và prompt (tối ưu hóa cho truy vấn full-text search)

Bảng 7: Mô tả bảng Cache

4.4 Thiết kế giao diện

4.4.1 Trang đăng ký

First name

Last name

Username*

Required. 150 characters or fewer. Letters, digits and @/./+/-/_ only.

Email*

Password*

• Your password can't be too similar to your other personal information.
• Your password must contain at least 8 characters.
• Your password can't be a commonly used password.
• Your password can't be entirely numeric.

Password confirmation*

Enter the same password as before, for verification.

Đăng ký

Bạn đã có tài khoản? [Đăng nhập](#)

Hình 37:Trang đăng ký

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

4.4.2 Trang đăng nhập

The screenshot shows a blue header bar with the HCMUTE logo on the left and navigation links 'Trang Chủ', 'Đăng Nhập', and 'Đăng Kí' on the right. Below the header is a white form titled 'Đăng Nhập'. It contains two input fields: 'Username*' and 'Password*', both with red asterisks indicating required fields. Below these fields are two buttons: a blue-bordered 'Đăng nhập' button and a smaller 'Quên mật khẩu?' link. At the bottom of the form is a note: 'Bạn chưa có tài khoản? [Đăng ký ngay](#)'.

Hình 38: Trang đăng nhập

4.4.3 Trang bảng tin

The screenshot shows a news feed interface with a blue header bar featuring the HCMUTE logo and navigation links: 'Trang Chủ', 'Thông Báo', 'Chatbot', 'Bài viết', and a user profile icon. The main content area displays two posts:

- Nguyễn Thị Phú** (November 12, 2024-09:41 PM)
Thắc mắc về trợ cấp cho sinh viên khó khăn HK1 2024-2025
Minh tên là: Nguyễn Thị Phú. MSSV: 21110600
Hiện tại đã qua ngày thông báo của trường nhưng em vẫn chưa ...
[\[Xem chi tiết\]](#)
- Bùi Quang Thiện** (November 12, 2024-07:56 PM)
Chuẩn đầu ra ngoại ngữ của trường
Em học lớp CLC thì cần bằng TOEIC bao nhiêu để đủ chuẩn tốt nghiệp a?
[\[Xem chi tiết\]](#)

To the right of the posts is a sidebar titled 'Bài viết gần đây' containing three recent posts:

- Thắc mắc về trợ cấp cho sinh viên khó khăn HK1 2024-2025
- Chuẩn đầu ra ngoại ngữ của trường
- Hello

Hình 39: Trang bảng tin

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

4.4.4. Trang đăng bài viết

 HCMUTE

Trang Chủ Thông Báo Chatbot Bài viết  buithien 

Bài viết mới

Title*

Content

Styles | Format | **B** *I* U ~~S~~ ↪ ↩ |                                         <img alt="list icon" data-bbox="8828 358 8

Hình 40: Trang đăng bài viết

4.4.5. Trang chi tiết bài viết



Trang Chủ Thông Báo Chatbot Bài viết    



thiphu
November 12, 2024



Thắc mắc về trợ cấp cho sinh viên khó khăn HK1 2024-2025

Minh tên là: Nguyễn Thị Phú. MSSV: 21110600

Hiện tại đã qua ngày thông báo của trường nhưng em vẫn chưa thấy công bố kết quả chính thức về tiền trợ cấp và cũng chưa có tiền gửi vào tài khoản.

Minh muốn hỏi mọi người là khi nào mới có danh sách và khi nào em mới nhận được tiền trợ cấp ạ?

 - 2 lượt thích

Thêm bình luận



 Thêm bình luận

Bình luận (17) :

Hình 41: Trang chi tiết bài viết

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

The screenshot shows a social media feed with four identical posts from a user named 'minhtan'. Each post is timestamped at 'Nov. 12, 2024, 9:44 p.m.' and contains the same message: 'Chào em Kết quả họp xét Học bổng trợ cấp khó khăn HK1-2024-2025 vừa được thông qua vào thứ 5 tuần trước. Do đó trong tuần này hoặc tuần sau sẽ có Danh sách chính thức và nhà trường sẽ chuyển Học bổng vào tài khoản cá nhân mà Nhà trường và Ngân hàng đã cung cấp cho các em.' Below each message are two engagement icons: a red heart icon with '0' and a blue speech bubble icon with '0'.

Trang Chủ Thông Báo Chatbot Bài viết ▾ buithien ▾

minhtan
Nov. 12, 2024, 9:44 p.m.

Chào em Kết quả họp xét Học bổng trợ cấp khó khăn HK1-2024-2025 vừa được thông qua vào thứ 5 tuần trước. Do đó trong tuần này hoặc tuần sau sẽ có Danh sách chính thức và nhà trường sẽ chuyển Học bổng vào tài khoản cá nhân mà Nhà trường và Ngân hàng đã cung cấp cho các em.

0 0

minhtan
Nov. 12, 2024, 9:44 p.m.

Chào em Kết quả họp xét Học bổng trợ cấp khó khăn HK1-2024-2025 vừa được thông qua vào thứ 5 tuần trước. Do đó trong tuần này hoặc tuần sau sẽ có Danh sách chính thức và nhà trường sẽ chuyển Học bổng vào tài khoản cá nhân mà Nhà trường và Ngân hàng đã cung cấp cho các em.

0 0

minhtan
Nov. 12, 2024, 9:44 p.m.

Chào em Kết quả họp xét Học bổng trợ cấp khó khăn HK1-2024-2025 vừa được thông qua vào thứ 5 tuần trước. Do đó trong tuần này hoặc tuần sau sẽ có Danh sách chính thức và nhà trường sẽ chuyển Học bổng vào tài khoản cá nhân mà Nhà trường và Ngân hàng đã cung cấp cho các em.

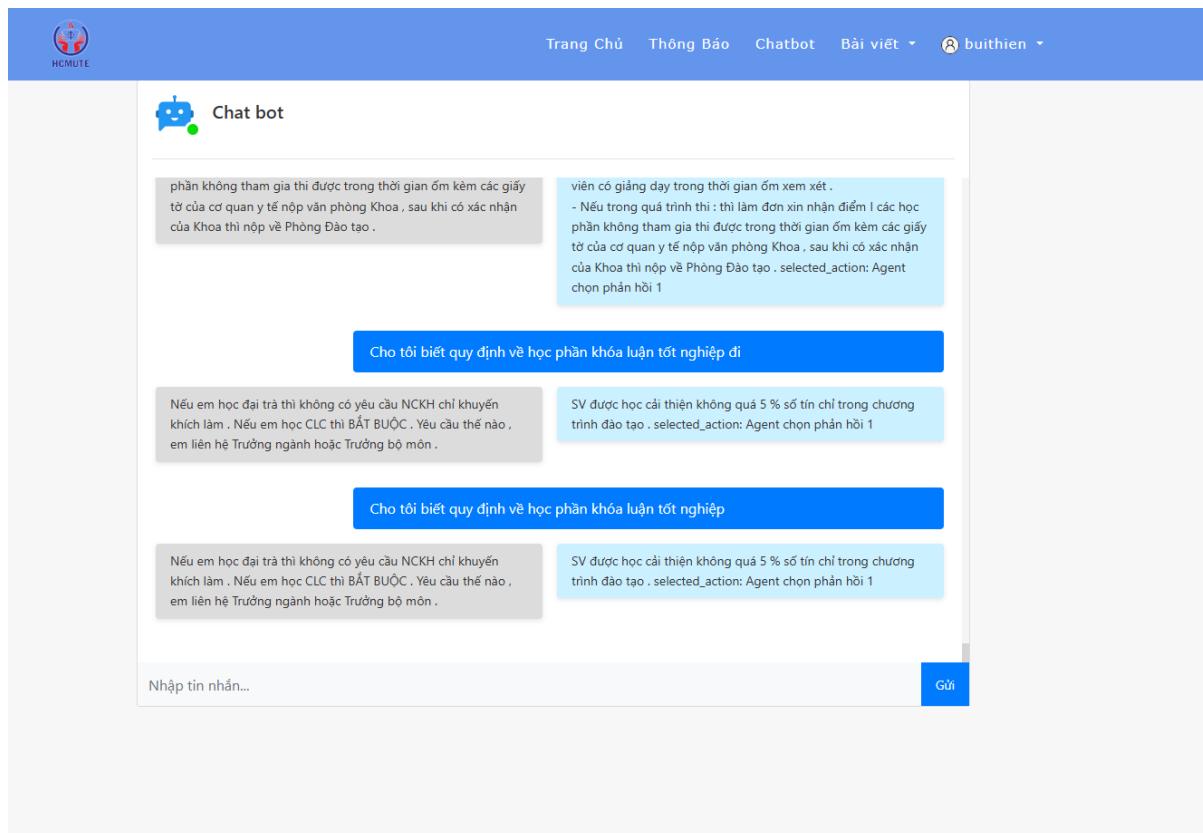
0 0

minhtan
Nov. 12, 2024, 9:44 p.m.

Hình 42: Bình luận bài viết

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

4.4.6. Trang nhắn tin với chatbot



Hình 43: Trang nhắn tin với chatbot

4.5 Môi trường học tăng cường cho việc lựa chọn 2 phản hồi FAQ

Nhóm đã xây dựng một môi trường học tăng cường, cụ thể sử dụng thuật toán Deep Q-Learning nhằm tối ưu hóa việc chọn giữa hai phản hồi FAQ có độ tương đồng cao nhất và cao nhì. Môi trường này được triển khai trong lớp DQLEnv, nơi mỗi trạng thái là một vector embedding được tạo ra bởi phép toán *concatenate* bao gồm embedding của câu hỏi và 2 phản hồi. Không gian hành động bao gồm hai lựa chọn: chọn phản hồi thứ nhất hoặc thứ hai.

Phản thưởng trong môi trường được thiết kế để phản ánh mức độ hài lòng của người dùng thông qua tỷ lệ giữa số lượt likes và dislikes, đồng thời giảm dần theo thời gian để khuyến khích các phản hồi có giá trị lâu dài. Công thức tính phản thưởng kết hợp logarit tự nhiên của tỷ lệ thích trên tổng số tương tác với tích yếu tố decay thời gian. Việc chọn hàm logarit tự nhiên của tỷ lệ thích trên tổng số tương tác là do quá trình thực nghiệm, nếu không sử dụng hàm logarit tự nhiên thì độ biến thiên reward giữa các episode quá lớn làm cho đồ thị giữa 2 biến reward và episodes giao động mạnh mẽ, không tuân theo quy tắc

CHƯƠNG 4: THIẾT KẾ HỆ THỐNG

chung khi tích lũy reward theo episodes - điểm thưởng tăng chậm đều, rồi đến mức bão hòa hoặc tăng rất nhỏ.

$$\log \left(\frac{\text{likes}}{\text{likes} + \text{dislikes} + \varepsilon} \right) \cdot e^{-\lambda \cdot t}$$

Hình 44: Công thức reward

Quy trình hoạt động của môi trường bắt đầu từ việc khởi tạo trạng thái qua hàm reset, nơi trạng thái bao gồm embedding của câu hỏi và 2 phản hồi. Khi thực hiện một hành động, hàm step sẽ trả về phần thưởng dựa trên dữ liệu phản hồi người dùng được lưu trong bảng Feedback, đồng thời giữ trạng thái ban đầu để đánh giá liên tục. Thiết kế này giúp hệ thống chatbot học cách ưu tiên các phản hồi phù hợp nhất với nhu cầu người dùng.

CHƯƠNG 5: CHI TIẾT CÁC THAM SỐ VÀ SIÊU THAM SỐ

5.1 Mô hình ngôn ngữ lớn

Tham số	Tên biến	Mô tả	Giá trị
Temperature	temperature	Điều chỉnh mức độ sáng tạo của mô hình khi sinh đầu ra.	0.3
Number of context tokens	n_ctx	Số lượng token ngữ cảnh mà mô hình có thể xử lý trong một lần chạy.	8192
Maximum number of tokens	max_tokens	Giới hạn số lượng token đầu ra mà mô hình có thể sinh ra.	500
Number of GPU layers	n_gpu_layers	Quy định số lớp đầu tiên của mô hình sẽ được tải lên GPU để tăng tốc xử lý.	43

Bảng 8: Thông số cho mô hình ngôn ngữ lớn

5.2 Văn bản tài liệu

Tham số	Tên biến	Mô tả	Giá trị
Chunk Size	chunk_size	Kích thước tối đa của mỗi đoạn văn bản (chunk) sau khi chia nhỏ.	768
Chunk Overlap	chunk_overlap	Số lượng ký tự bị trùng lặp giữa hai chunk liên tiếp.	300
Length Function	length_function	Hàm được sử dụng để tính độ dài của văn bản.	Sử dụng hàm len
Separator Regex	is_separator_regex	Chỉ định cách xác định các dấu ngắt. Giá trị mặc định là False nghĩa là dấu ngắt sẽ được xác định bởi các ký tự thông thường (mặc định là khoảng trắng hoặc xuống dòng)	False

Bảng 9: Thông số cho văn bản tài liệu

5.3. Môi trường Deep Q-Learning

Cấu trúc phần thưởng được xây dựng theo công thức bên dưới, trong đó có 2 siêu tham số quan trọng đó là ε , λ

$$\log \left(\frac{\text{likes}}{\text{likes} + \text{dislikes} + \varepsilon} \right) \cdot e^{-\lambda \cdot t}$$

Hình 45: Công thức reward

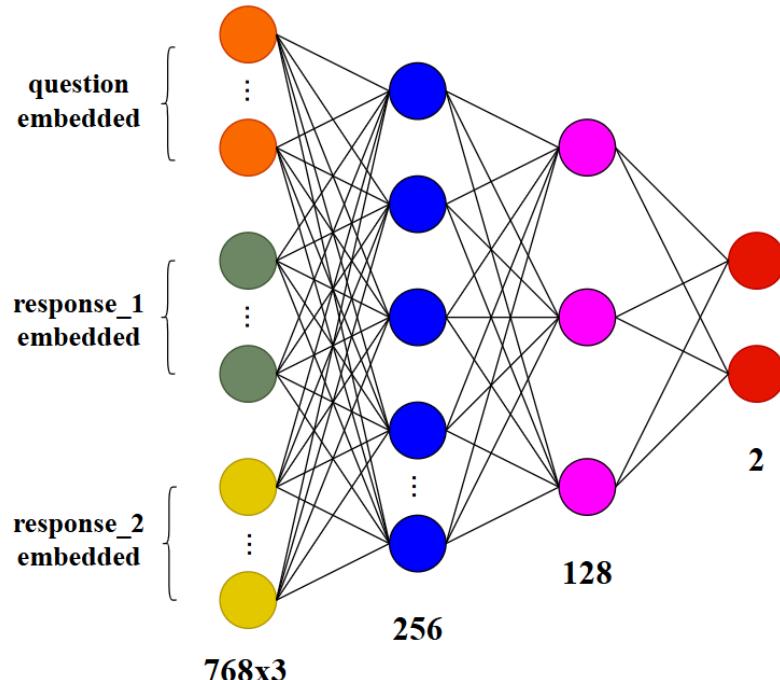
Tham số	Tên biến	Mô tả	Giá trị
Lambda Decay	lambda_decay	Thể hiện mức độ quan trọng của độ biến thiên thời gian, từ lúc tạo phản hồi tới hiện tại.	0.0001
Error	err	Một con số rất nhỏ, tránh trường hợp likes và dislikes bằng 0, khiến cho giá trị phân số không xác định.	0.00001

Bảng 10: Thông số của Deep Q-Learning

5.4 Mạng Deep Q-Learning

Mạng nơ-ron được thiết kế với 3 layers:

- Input layer: nhận trạng thái của môi trường dưới dạng vector 1D.
- Hidden layer: tầng thứ nhất có 256 nodes, tầng sau có 128 nodes và sử dụng hàm kích hoạt ReLU.
- Output layer: trả về giá trị Q tương ứng với các hành động khả thi trong trạng thái hiện tại.



Hình 46: Sơ đồ mạng nơ-ron nhân tạo thiết kế cho Deep Q-Learning

Layer	Input Size	Output Size	Activation
Input Layer	state_dim (768x3)	256	None
Fully Connected (fc1)	256	128	ReLU
Fully Connected (fc2)	128	action_dim (2)	None

Bảng 11: Mạng nơ-ron thiết kế cho Deep Q-Learning

Mục tiêu của mạng là học hàm Q để dự đoán giá trị Q tối ưu cho từng hành động

5.5 Deep Q-Learning Agent

Các siêu tham số được tinh chỉnh thủ công, không qua bất kỳ công cụ tự động nào để được kết quả như kỳ vọng.

Siêu tham số	Tên biến	Giá trị
Buffer Size	buffer_size	20
Batch Size	batch_size	16
Gamma	gamma	0.995
Learning Rate	lr	0.00004
Tau	tau	0.003
Epsilon	epsilon	0.01

Bảng 12: Thông số cho Agent

CHƯƠNG 6. KẾT QUẢ ĐẠT ĐƯỢC

6.1 Kết quả

6.1.1 Về mặt lý thuyết

Về cơ bản, nhóm đã có cái nhìn tổng quan về một trong những giải pháp áp dụng Trí tuệ nhân tạo vào môi trường Sư phạm, qua đó giảm tải khối lượng của các phòng ban và hoàn thành được các mục tiêu đã đề ra từ đầu. Trải qua quá trình này, nhóm đã nắm vững quy trình phát triển ứng dụng mô hình ngôn ngữ lớn, hiểu rõ cách sử dụng các công cụ lập trình, các thư viện của Python, thực hiện tạo lập, phân tích, tương tác với cơ sở dữ liệu và áp dụng đa phương pháp như RAG, hay học tăng cường qua đó giải quyết những vấn đề còn mắc phải của chatbot như: ảo giác, thiếu hiểu biết ngữ cảnh sâu, hạn chế trong việc cập nhật thông tin, tiêu tồn tài nguyên, đặc biệt năm bắt được hành vi người dùng từ đó đưa ra phản hồi thích hợp.

Không những hoàn thành nhiệm vụ về xây dựng hệ thống có áp dụng Trí tuệ nhân tạo, nhóm còn hoàn thành việc trực quan hóa ý tưởng của mình thông qua giao diện web bằng công nghệ Python Django và Flask, qua đó người dùng có thể dễ dàng tiếp cận, sử dụng ứng dụng một cách thuận tiện, hài lòng.

Tóm lại, thông qua dự án “Chatbot hỗ trợ Tư vấn cho Sinh viên trường Đại học Sư” này nhóm không chỉ có được kiến thức thông qua việc chủ động tìm, đọc các tài liệu, sách báo mà còn được nâng cao kỹ năng và làm dày thêm kinh nghiệm lập trình của bản thân. Đây là một bước quan trọng trong quá trình phát triển sau này và hy vọng rằng có thể áp dụng kiến thức và kỹ năng này một cách tự tin hơn nữa vào các dự án trong tương lai.

6.1.2 Về mặt sản phẩm

Các tiêu chí nhóm đưa ra để đánh giá hiệu suất, khả năng trả lời của chatbot là: độ chính xác, thời gian phản hồi, tính bao quát, tính cá nhân hóa, khả năng phán đoán phản hồi FAQ của Agent.

CHƯƠNG 7: KẾT LUẬN

Tiêu chí	Kết quả	
Độ chính xác	16/20 bộ câu hỏi-câu trả lời được chuẩn bị ngẫu nhiên trước. Mô tả cụ thể ở bảng 13	
Thời gian phản hồi cho FAQ	$\sim 0-1$ s	
Thời gian phản hồi cho RAG	Khó, cần suy luận	$\sim 30-45$ s
	Dễ, truy xuất ngay	$\sim 10-15$ s
Tính bao quát	Sinh viên chỉ có thể các câu hỏi về nội quy, quy chế, học bỗng, điểm tích lũy, điểm rèn luyện, điểm Công tác Xã hội, điểm ngoại ngữ.	
	Các câu hỏi quá chi tiết về nợ học phí cá nhân, bảo hiểm Y tế, bảo hiểm Xã hội, thông tin các lớp học phần, thông tin Giảng viên chatbot chưa có thông tin để trả lời.	
Tính cá nhân hóa	Chatbot có thể ghi nhớ và cung cấp thông tin, hiểu ngữ cảnh dựa trên lịch sử tương tác của người dùng.	
Khả năng phán đoán phản hồi FAQ	<p>Sau 2000 episodes huấn luyện, Agent được xây dựng qua thuật toán Deep Q-Learning có cải thiện reward khoảng 480 episodes đầu, sau đó rơi vào trạng thái bão hòa không cải thiện thêm.</p> <ul style="list-style-type: none"> • Hình 40: Mô tả reward tích lũy qua các episodes • Hình 41: Mô tả reward tích lũy qua các episodes nhưng áp dụng kĩ thuật đường trung bình động - xác định xu hướng của dữ liệu theo thời gian. 	

Bảng 13: Nhận xét tổng quát sản phẩm

CHƯƠNG 7: KẾT LUẬN

STT	Câu hỏi	Câu trả lời kỳ vọng	Kết quả của hệ thống
1	Cho tôi biết học phần tương đương là gì?	Học phần tương đương: Là học phần hay một nhóm học phần thuộc chương trình đào tạo một ngành/chuyên ngành đang tổ chức đào tạo tại trường. SINH VIÊN được phép tích lũy để thay cho học phần hay một nhóm học phần trong chương trình đào tạo của một ngành hay chuyên ngành khác.	Học phần tương đương: Là học phần hay một nhóm học phần thuộc chương trình đào tạo một ngành/chuyên ngành đang tổ chức đào tạo tại trường. SINH VIÊN được phép tích lũy để thay cho học phần hay một nhóm học phần trong chương trình đào tạo của một ngành hay chuyên ngành khác.
2	Khi nào Sinh viên bị buộc thôi học?	SINH VIÊN bị buộc thôi học nếu thuộc một trong các trường hợp sau: - Có 3 lần cảnh báo học tập liên tiếp. - Thời gian học tập vượt quá giới hạn theo quy định (Kỹ sư/Cử nhân: tối đa 8 năm, Kiến trúc sư: tối đa 9 năm) - Sinh viên thi hộ hoặc nhờ người thi hộ bị kỷ luật lần thứ 2 - Người học sử dụng hồ sơ, văn bằng, chứng chỉ giả làm điều kiện trúng tuyển hoặc điều kiện tốt nghiệp sẽ bị buộc thôi học.	SINH VIÊN bị buộc thôi học nếu thuộc một trong các trường hợp sau: - Có 3 lần cảnh báo học tập liên tiếp. - Thời gian học tập vượt quá giới hạn theo quy định (Kỹ sư/Cử nhân: tối đa 8 năm, Kiến trúc sư: tối đa 9 năm) - Sinh viên thi hộ hoặc nhờ người thi hộ bị kỷ luật lần thứ 2 - Người học sử dụng hồ sơ, văn bằng, chứng chỉ giả làm điều kiện trúng tuyển hoặc điều kiện tốt nghiệp sẽ bị buộc thôi học.
3	Khi nào thì SINH VIÊN được công nhận đã tham gia nghiên cứu khoa học khi thực hiện một trong những yêu cầu sau:	SINH VIÊN được công nhận đã tham gia nghiên cứu khoa học khi thực hiện một trong những yêu cầu sau:	SINH VIÊN được công nhận đã tham gia nghiên cứu khoa học khi thực hiện một trong những yêu cầu sau:

CHƯƠNG 7: KẾT LUẬN

	nghiên cứu khoa học?	<ul style="list-style-type: none"> - Tham gia nghiên cứu khoa học theo nhóm nghiên cứu do các giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên. - Tham gia đề tài nghiên cứu khoa học SINH VIÊN cấp trường. - Tham gia các cuộc thi nghiên cứu khoa học, sân chơi nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức. - Trình bày báo cáo khoa học tại các hội nghị nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức. - Tham gia các hình thức nghiên cứu khoa học khác, phải có sự chấp thuận của các thầy cô trưởng ngành. 	<ul style="list-style-type: none"> - Tham gia nghiên cứu khoa học theo nhóm nghiên cứu do các giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên. - Tham gia đề tài nghiên cứu khoa học SINH VIÊN cấp trường. - Tham gia các cuộc thi nghiên cứu khoa học, sân chơi nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức. - Trình bày báo cáo khoa học tại các hội nghị nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức. - Tham gia các hình thức nghiên cứu khoa học khác, phải có sự chấp thuận của các thầy cô trưởng ngành.
4	Có mấy khoảng thời gian để SINH VIÊN được rút bớt học phần vậy?	<p>Có hai khoảng thời gian để SINH VIÊN được rút bớt học phần:</p> <ul style="list-style-type: none"> - Trong thời gian điều chỉnh đăng ký học phần, SINH VIÊN rút bớt học phần qua mạng và không phải đóng học phí cho các học phần này. - Sau thời gian trên, SINH VIÊN được phép rút bớt học phần trong tuần 8 với học kỳ chính, tuần 3 với học kỳ phụ theo thông báo của phòng Đào tạo. SINH VIÊN phải đóng học phí cho các học phần đã rút. 	<p>Có hai khoảng thời gian để SINH VIÊN được rút bớt học phần:</p> <ul style="list-style-type: none"> - Trong thời gian điều chỉnh đăng ký học phần, SINH VIÊN rút bớt học phần qua mạng và không phải đóng học phí cho các học phần này. - Sau thời gian trên, SINH VIÊN được phép rút bớt học phần trong tuần 8 với học kỳ chính, tuần 3 với học kỳ phụ theo thông báo của phòng Đào tạo. SINH VIÊN phải đóng học phí cho các học phần đã rút.

CHƯƠNG 7: KẾT LUẬN

		học phí cho các học phần đã rút.	
5	Tôi muốn xin giấy hoãn NVQS thì phải làm sao?	Nếu SINH VIÊN vẫn còn trong thời gian đào tạo (dưới 4 năm) thì SINH VIÊN vào trang online.hcmute.edu.vn chọn phần Đăng ký giấy nhận , sau khi đăng ký SINH VIÊN xem thời gian hẹn trả giấy rồi tới Phòng Tuyển sinh và Công tác Sinh viên theo đúng lịch hẹn .	Nếu SINH VIÊN vẫn còn trong thời gian đào tạo (dưới 4 năm) thì SINH VIÊN vào trang online.hcmute.edu.vn chọn phần Đăng ký giấy nhận , sau khi đăng ký SINH VIÊN xem thời gian hẹn trả giấy rồi tới Phòng Tuyển sinh và Công tác Sinh viên theo đúng lịch hẹn .
6	Không ghi MSSINH VIÊN trong lúc chuyển tiền học phí thì phải làm sao?	Nếu Sinh viên quên ghi MSSINH VIÊN khi chuyển tiền học phí, em vui lòng liên hệ với Phòng Tài chính của trường để được hướng dẫn xử lý.	Nếu bạn quên ghi MSSINH VIÊN khi chuyển tiền học phí, bạn vui lòng liên hệ với Phòng Tài chính của trường để được hướng dẫn xử lý. Thông thường, Phòng Tài chính sẽ yêu cầu bạn cung cấp những thông tin sau: - Họ tên đầy đủ - Lớp - Năm sinh - Số điện thoại liên lạc - Mã số thuế (nếu có) - Ghi chú chuyển khoản Phòng Tài chính sẽ xác minh thông tin và hỗ trợ ghi nhận học phí cho bạn. Hãy nhớ giữ lại các bằng chứng về việc bạn đã chuyển tiền, như biên lai hoặc giao dịch ngân hàng, để phục vụ cho quá trình xác minh.
7	Tôi là SINH VIÊN thuộc đồng bào dân tộc thiểu số, vậy thời gian đào	Không bị giới hạn thời gian đào tạo	Bạn là Sinh viên thuộc đồng bào dân tộc thiểu số nên thời gian đào tạo tối đa của bạn là 2 năm. Điều này được quy định đặc thù cho đối tượng Sinh viên người dân tộc thiểu

CHƯƠNG 7: KẾT LUẬN

	tạo tối đa của tôi là?		số theo đúng như Quyết định của Nhà trường.
8	Hội đồng đánh giá khóa luận tốt nghiệp bao gồm những ai?	Hội đồng đánh giá khóa luận tốt nghiệp bao gồm tối thiểu 3 thành viên và tối đa 5 thành viên trong đó giảng viên hướng dẫn không tham gia hội đồng.	Hội đồng đánh giá khóa luận tốt nghiệp bao gồm tối thiểu 3 và tối đa 5 thành viên. Trong đó có giảng viên hướng dẫn, nhưng không tham gia hội đồng đánh giá.
9	Tôi đã rút học phần vào tuần 8, vậy tôi có phải thanh toán các học phần đã rút không?	SINH VIÊN phải đóng học phí cho các học phần đã rút.	Trường hợp bạn rút học phần trong tuần thứ 8, bạn sẽ phải thanh toán các học phần đã rút. Theo quy định của trường, thời gian rút bớt học phần không tính phí là trong khoảng thời gian điều chỉnh đăng ký học phần. Sau thời điểm này, nếu muốn rút bớt học phần thì bạn cần phải đóng học phí cho các học phần đã rút.
10	Tôi bỏ học nhưng quên rút học phần trong thời gian quy định cho phép rút học phần, vậy thì bị gì	Trường sẽ tổ chức cho SINH VIÊN rút bớt các học phần đã đăng ký. Sau thời gian quy định cho phép rút bớt học phần, các học phần sẽ được giữ nguyên trong kết quả đăng ký học phần và nếu SINH VIÊN không đi học các học phần này sẽ được xem như tự ý bỏ học và phải nhận điểm 0.	Nếu bạn bỏ học nhưng quên rút học phần trong thời gian quy định, tức là sau thời gian điều chỉnh đăng ký học phần và trước khi đến tuần 8 (học kỳ chính) hoặc tuần 3 (học kỳ phụ), thì trường sẽ mặc định tính điểm cho các học phần đó là 0. Điều này có nghĩa là bạn sẽ không được nhận tín chỉ cho các học phần đã bỏ, ảnh hưởng đến kết quả học tập và khả năng tốt nghiệp của bạn. Để tránh trường hợp này xảy ra trong lần sau, bạn nên lưu ý thời gian quy định rút học phần và thực hiện thủ tục rút bớt học phần nếu có nhu cầu.

CHƯƠNG 7: KẾT LUẬN

11	Tôi đăng ký tham gia hoạt động công tác xã hội, nhưng không tham gia hay không hoàn thành nhiệm vụ, nội dung đã đăng ký thì sao?	<p>SINH VIÊN đăng ký tham gia hoạt động công tác xã hội, nhưng không thực hiện nội dung đã đăng ký mà không có lý do hoặc lý do không chính đáng bị trừ 0.5 ngày công tác xã hội.</p> <p>SINH VIÊN đăng ký tham gia nhưng không hoàn thành nhiệm vụ, gây ảnh hưởng xấu đến hình ảnh của hoạt động, tổ chức, nhà trường, tùy theo mức độ vi phạm sẽ bị xử lý kỷ luật từ khiển trách trở lên và bị trừ ngày CTXH từ 1 đến toàn bộ số ngày đã tích lũy được.</p>	<p>Chào bạn! Cảm ơn bạn đã đặt câu hỏi. Theo quy định của Trường Đại học Sư phạm Kỹ thuật TP.HCM, Sinh viên đăng ký tham gia hoạt động công tác xã hội nhưng không hoàn thành nhiệm vụ sẽ bị xử lý kỷ luật từ khiển trách trở lên và bị trừ ngày Công tác xã hội từ 1 đến toàn bộ số ngày đã tích lũy được tùy theo mức độ vi phạm. Bạn nên cân nhắc kỹ trước khi đăng ký tham gia bất kỳ hoạt động nào, đảm bảo bạn có thể hoàn thành nhiệm vụ để không ảnh hưởng đến hình ảnh của mình và các hoạt động đang diễn ra.</p>
12	Các hoạt động nào không mang tính chất công tác xã hội bao gồm?	<p>SINH VIÊN được công nhận đã tham gia nghiên cứu khoa học khi thực hiện một trong những yêu cầu sau:</p> <ul style="list-style-type: none"> - Tham gia nghiên cứu khoa học theo nhóm nghiên cứu do các giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên. - Tham gia đề tài nghiên cứu khoa học SINH VIÊN cấp trường. - Tham gia các cuộc thi nghiên cứu khoa học, sân chơi nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức. - Trình bày báo cáo khoa học tại các hội nghị nghiên cứu 	<p>Các hoạt động không mang tính chất công tác xã hội bao gồm:</p> <ul style="list-style-type: none"> - Xem biểu diễn văn hóa nghệ thuật, chiếu phim; - Xem & cổ vũ thi đấu thể thao; - Tham quan bảo tàng, di tích lịch sử; - Đi du lịch; - Tham dự hội thảo; - Sinh hoạt tại các Câu lạc bộ...

CHƯƠNG 7: KẾT LUẬN

		<p>khoa học do Trường hoặc các Đơn vị đào tạo tổ chức.</p> <ul style="list-style-type: none"> - Tham gia các hình thức nghiên cứu khoa học khác, phải có sự chấp thuận của các thầy cô trưởng ngành. 	
13	Cho tôi biết quy định về học phần khóa luận tốt nghiệp	<p>Quy định học phần khóa luận tốt nghiệp: học phần khóa luận tốt nghiệp chỉ có điểm tổng kết được tính từ các điểm thành phần bao gồm điểm của giảng viên hướng dẫn, giảng viên phản biện và điểm của các thành viên hội đồng đánh giá. Hội đồng đánh giá khóa luận tốt nghiệp bao gồm tối thiểu 3 thành viên và tối đa 5 thành viên trong đó giảng viên hướng dẫn không tham gia hội đồng. Điểm tổng kết được tính từ trung bình cộng của các điểm thành phần. Các điểm thành phần không được lệch nhau quá 2 điểm, trường hợp lệch quá 2 điểm thì Hội đồng phải thống nhất và điều chỉnh phù hợp.</p>	<p>Quy định học phần khóa luận tốt nghiệp như sau:</p> <ul style="list-style-type: none"> - Học phần khóa luận tốt nghiệp chỉ có điểm tổng kết được tính từ các điểm thành phần bao gồm điểm của giảng viên hướng dẫn, giảng viên phản biện và điểm của các thành viên hội đồng đánh giá. - Hội đồng đánh giá khóa luận tốt nghiệp bao gồm tối thiểu 3 thành viên và tối đa 5 thành viên trong đó giảng viên hướng dẫn không tham gia hội đồng. - Điểm tổng kết được tính từ trung bình cộng của các điểm thành phần. Các điểm thành phần không được lệch nhau quá 2 điểm, trường hợp lệch quá 2 điểm thì Hội đồng phải thống nhất và điều chỉnh phù hợp.
14	Cách đăng ký các hoạt động CTXH để tích lũy đủ điểm CTXH?	<p>Sinh viên cần đăng ký tham gia các hoạt động CTXH theo quy định của nhà trường</p>	<p>Để tích lũy đủ điểm CTXH, Sinh viên cần đăng ký tham gia các hoạt động CTXH theo quy định của nhà trường. Thông tin chi tiết về cách đăng ký và danh sách các hoạt động CTXH được cập nhật trên trang web của nhà trường hoặc tại văn phòng Hợp tác xã hội.</p>

CHƯƠNG 7: KẾT LUẬN

15	Đăng ký tham gia hoạt động CTXH nhưng bản thân không thể tham gia thì sao?	Trường hợp vì lý do sức khỏe nên SINH VIÊN rất khó khăn hoặc không thể tham gia các hoạt động công tác xã hội , SINH VIÊN phải làm đơn kèm xác nhận về tình trạng sức khỏe của cơ quan y tế từ cấp huyện trở lên để Nhà trường xem xét việc miễn , giảm tham gia hoạt động công tác xã hội .	Trường hợp vì lý do sức khỏe nên SINH VIÊN rất khó khăn hoặc không thể tham gia các hoạt động công tác xã hội , SINH VIÊN phải làm đơn kèm xác nhận về tình trạng sức khỏe của cơ quan y tế từ cấp huyện trở lên để Nhà trường xem xét việc miễn , giảm tham gia hoạt động công tác xã hội .
16	Học cải thiện điểm có ảnh hưởng đến xếp loại tốt nghiệp không?	SINH VIÊN được học cải thiện không quá 5 % số tín chỉ trong chương trình đào tạo .	SINH VIÊN được học cải thiện không quá 5 % số tín chỉ trong chương trình đào tạo .
17	Tôi bỏ học nhưng quên rút học phần trong thời gian quy định cho phép rút học phần, vậy thì bị gì	Trường sẽ tổ chức cho SINH VIÊN rút bớt các học phần đã đăng ký. Sau thời gian quy định cho phép rút bớt học phần, các học phần sẽ được giữ nguyên trong kết quả đăng ký học phần và nếu SINH VIÊN không đi học các học phần này sẽ được xem như tự ý bỏ học và phải nhận điểm 0 . Điều kiện : Có hai khoảng thời gian để SINH VIÊN được rút bớt học phần :	<p>Trường sẽ tổ chức cho SINH VIÊN rút bớt các học phần đã đăng ký. Sau thời gian quy định cho phép rút bớt học phần, các học phần sẽ được giữ nguyên trong kết quả đăng ký học phần và nếu SINH VIÊN không đi học các học phần này sẽ được xem như tự ý bỏ học và phải nhận điểm 0 . Điều kiện : Có hai khoảng thời gian để SINH VIÊN được rút bớt học phần :</p> <ul style="list-style-type: none"> - Trong thời gian điều chỉnh đăng ký học phần , SINH VIÊN rút bớt học phần qua mạng và không phải đóng học phí cho các học phần này . - Sau thời gian trên , SINH VIÊN được phép rút bớt học phần trong

CHƯƠNG 7: KẾT LUẬN

			tuần 8 với học kỳ chính , tuần 3 với học kỳ phụ theo thông báo của phòng Đào tạo . SINH VIÊN phải đóng học phí cho các học phần đã rút
18	Điều kiện tối thiểu nghiệp tại trường	<p>Tích lũy đủ học phần , số tín chỉ và hoàn thành các nội dung bắt buộc khác theo yêu cầu của chương trình đào tạo , đạt chuẩn đầu ra của chương trình đào tạo .</p> <ul style="list-style-type: none"> - Điểm trung bình tích lũy của toàn khóa học đạt từ trung bình trở lên . - Tại thời điểm xét tốt nghiệp không bị truy cứu trách nhiệm hình sự hoặc không đang trong thời gian bị kỷ luật ở mức đình chỉ học tập . - Điểm rèn luyện toàn khóa đạt từ trung bình trở lên . - Tích lũy đủ số ngày công tác xã hội theo quy định hiện hành của Trường . - Hoàn thành giáo dục quốc phòng và giáo dục thể chất đối với các ngành đào tạo không chuyên về quân sự và thể dục - thể thao . - Đạt yêu cầu chuẩn đầu ra ngoại ngữ . - Đối với chương trình chính quy CLC , SINH VIÊN phải tham gia nghiên cứu khoa học theo nhóm nghiên cứu do giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên . 	<ul style="list-style-type: none"> - Tích lũy đủ học phần , số tín chỉ và hoàn thành các nội dung bắt buộc khác theo yêu cầu của chương trình đào tạo , đạt chuẩn đầu ra của chương trình đào tạo . - Điểm trung bình tích lũy của toàn khóa học đạt từ trung bình trở lên . - Tại thời điểm xét tốt nghiệp không bị truy cứu trách nhiệm hình sự hoặc không đang trong thời gian bị kỷ luật ở mức đình chỉ học tập . - Điểm rèn luyện toàn khóa đạt từ trung bình trở lên . - Tích lũy đủ số ngày công tác xã hội theo quy định hiện hành của Trường . - Hoàn thành giáo dục quốc phòng và giáo dục thể chất đối với các ngành đào tạo không chuyên về quân sự và thể dục - thể thao . - Đạt yêu cầu chuẩn đầu ra ngoại ngữ . - Đối với chương trình chính quy CLC , SINH VIÊN phải tham gia nghiên cứu khoa học theo nhóm nghiên cứu do giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên .

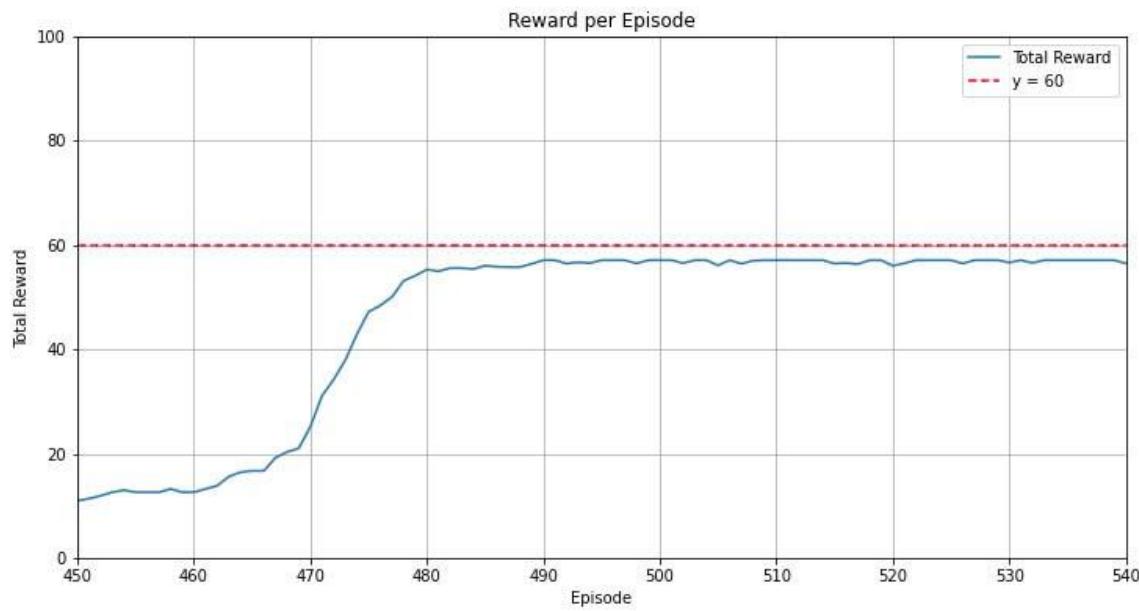
CHƯƠNG 7: KẾT LUẬN

		tham gia đề tài nghiên cứu khoa học với giảng viên .	
19	Học lại bao nhiêu % thì bị hạ 1 bậc xét tốt nghiệp ?	SINH VIÊN học lại / rót quá 5 % tổng số chỉ thi sẽ bị hạ bằng xuất sắc xuống giỏi , giỏi xuống khá từ loại khá thi không hạ dù rót 5 % số chỉ . Chi tiết liên hệ qua mail cô Quỳnh : quynhbt@hcmute.edu.vn để được kiểm tra	SINH VIÊN học lại / rót quá 5 % tổng số chỉ thi sẽ bị hạ bằng xuất sắc xuống giỏi , giỏi xuống khá từ loại khá thi không hạ dù rót 5 % số chỉ . Chi tiết liên hệ qua mail cô Quỳnh : quynhbt@hcmute.edu.vn để được kiểm tra
20	Chào bạn	Chào bạn, rất vui được hỗ trợ bạn. Bạn có bất cứ thắc mắc gì về việc đăng ký học phần không?	Chào bạn ! Rất vui được chào đón bạn đến với chatbot tư vấn Sinh viên của trường Đại học Sư phạm Kỹ thuật TP.HCM. Bạn có bất kỳ thắc mắc gì về tuyển sinh , đào tạo , đời sống Sinh viên hay các chính sách của trường , đừng ngần ngại hỏi tôi nhé ! Tôi sẽ cố gắng hết sức để giúp bạn .

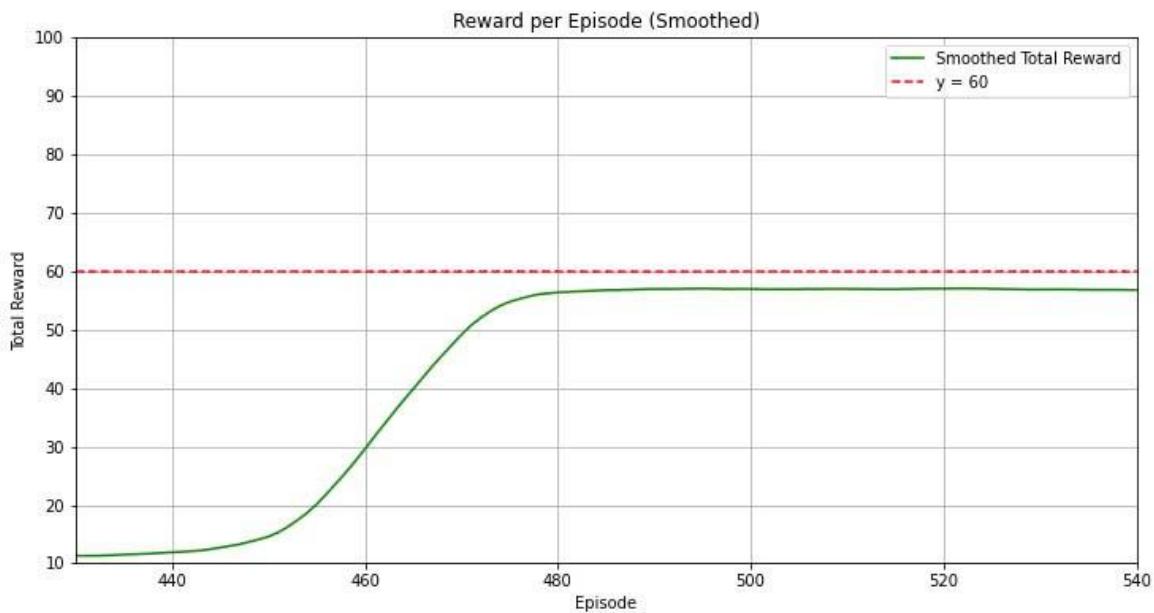
Bảng 14: Đánh giá hệ thống

Episode 1/2000 - Total Reward: 11.03	Episode 481/2000 - Total Reward: 56.62
Episode 2/2000 - Total Reward: 11.54	Episode 482/2000 - Total Reward: 56.62
Episode 3/2000 - Total Reward: 11.57	Episode 483/2000 - Total Reward: 56.62
Episode 4/2000 - Total Reward: 11.57	Episode 484/2000 - Total Reward: 57.15
Episode 5/2000 - Total Reward: 12.34	Episode 485/2000 - Total Reward: 56.62
Episode 6/2000 - Total Reward: 11.71	Episode 486/2000 - Total Reward: 55.33
Episode 7/2000 - Total Reward: 12.27	Episode 487/2000 - Total Reward: 56.62
Episode 8/2000 - Total Reward: 11.03	Episode 488/2000 - Total Reward: 56.62
Episode 9/2000 - Total Reward: 11.03	Episode 489/2000 - Total Reward: 56.62
Episode 10/2000 - Total Reward: 11.69	Episode 490/2000 - Total Reward: 56.62
Episode 11/2000 - Total Reward: 11.03	Episode 491/2000 - Total Reward: 56.62
Episode 12/2000 - Total Reward: 10.38	Episode 492/2000 - Total Reward: 56.62
Episode 13/2000 - Total Reward: 11.48	Episode 493/2000 - Total Reward: 56.62
Episode 14/2000 - Total Reward: 11.03	Episode 494/2000 - Total Reward: 55.98
Episode 15/2000 - Total Reward: 11.03	Episode 495/2000 - Total Reward: 55.26
Episode 16/2000 - Total Reward: 11.03	Episode 496/2000 - Total Reward: 56.62
Episode 17/2000 - Total Reward: 11.03	Episode 497/2000 - Total Reward: 56.62

Hình 47: Reward theo từng episode



Hình 48: Mô tả các reward qua từng episodes không áp dụng đường trung bình động



Hình 49: Mô tả các reward qua từng episodes nhưng áp dụng đường trung bình động

CHƯƠNG 7: KẾT LUẬN

6.2 Kiểm thử mô hình

6.2.1 Phản hồi FAQ

Cho tôi biết học phần tương đương là gì?

Học phần tương đương : Là học phần hay một nhóm học phần thuộc chương trình đào tạo một ngành / chuyên ngành đang tổ chức đào tạo tại Trường SV được phép tích lũy để thay cho học phần hay một nhóm học phần trong chương trình đào tạo của một ngành / chuyên ngành khác .



Học phần tự chọn : là các học phần chưa đựng những nội dung kiến thức cần thiết cho một trong nhiều định hướng nghề nghiệp (ngành / chuyên ngành) mà SV có thể tự do lựa chọn đăng ký học . Để đủ điều kiện tốt nghiệp , SV phải đạt số tín chỉ tối thiểu được quy định cho nhóm học phần tự chọn .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

Khi nào thì sinh viên sẽ bị buộc thôi học?

SV bị buộc thôi học nếu thuộc một trong các trường hợp sau :

- Có 3 lần cảnh báo học tập liên tiếp .
- Thời gian học tập vượt quá giới hạn theo quy định (Kí túc / Cử nhân : tối đa 8 năm , Kiến trúc sư : tối đa 9 năm) - Sinh viên thi hộ hoặc nhờ người thi hộ bị kỉ luật lần thứ 2 - Người học sử dụng hồ sơ , văn bằng , chứng chỉ giả làm điều kiện trúng tuyển hoặc điều kiện tốt nghiệp sẽ bị buộc thôi học .



Cuối mỗi học kỳ chính , SV bị cảnh báo học tập khi thuộc một trong các điều kiện sau :

- Tổng số tín chỉ không đạt trong học kỳ vượt quá 50 % số tín chỉ đã đăng ký trong học kỳ , hoặc tổng số tín chỉ nợ đọng từ đầu khóa học vượt quá 24 .
- ĐTBHK đạt dưới 0.8 đối với học kỳ đầu khóa học , dưới 1.0 đối với các học kỳ tiếp theo .
- ĐTBTL đạt dưới 1.2 đối với SV trình độ năm thứ nhất , dưới 1.4 đối với SV trình độ năm thứ hai , dưới 1.6 đối với SV trình độ năm thứ ba , dưới 1.8 đối với SV các năm tiếp theo .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

CHƯƠNG 7: KẾT LUẬN

Khi tốt nghiệp thì bằng tốt nghiệp được xếp loại dựa trên thang điểm 4 hay thang điểm 10 vậy?

Từ khóa 2021 trở về sau điểm xếp loại bằng tốt nghiệp sẽ theo thang 4 , SV đọc chi tiết Quy chế Đào tạo số 1727 / QĐ-ĐHSPKT ngày 06/9/2021 trên trang web phòng Đào tạo .



- Tích lũy đủ học phần , sổ tin chỉ và hoàn thành các nội dung bắt buộc khác theo yêu cầu của chương trình đào tạo , đạt chuẩn đầu ra của chương trình đào tạo .
- Điểm trung bình tích lũy của toàn khóa học đạt từ trung bình trở lên .
- Tại thời điểm xét tốt nghiệp không bị truy cứu trách nhiệm hình sự hoặc không đang trong thời gian bị kỷ luật ở mức định chỉ học tập .
- Điểm rèn luyện toàn khóa đạt từ trung bình trở lên .
- Tích lũy đủ sổ ngày công tác xã hội theo quy định hiện hành của Trường .
- Hoàn thành giáo dục quốc phòng và giáo dục thể chất đối với các ngành đào tạo không chuyên về quân sự và thể dục - thể thao .
- Đạt yêu cầu chuẩn đầu ra ngoại ngữ .
- Đối với chương trình chính quy CLC , SV phải tham gia nghiên cứu khoa học theo nhóm nghiên cứu do giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

Cho tôi biết học phần song hành là gì?

Học phần song hành : Học phần A là học phần song hành của học phần B thì điều kiện bắt buộc để đăng ký học học phần B là đã đăng ký học học phần A. SV được phép đăng ký học học phần B vào cùng học kỳ đã đăng ký học học phần A hoặc vào các học kỳ tiếp sau đó .



Học phần tự chọn : là các học phần chứa đựng những nội dung kiến thức cần thiết cho một trong nhiều định hướng nghề nghiệp (ngành / chuyên ngành) mà SV có thể tự do lựa chọn đăng ký học . Để đủ điều kiện tốt nghiệp , SV phải đạt số tín chỉ tối thiểu được quy định cho nhóm học phần tự chọn .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

CHƯƠNG 7: KẾT LUẬN

Đăng ký tham gia hoạt động CTXH nhưng bản thân không thể tham gia thì sao?

Trường hợp vì lý do sức khỏe nên SV rất khó khăn hoặc không thể tham gia các hoạt động công tác xã hội , SV phải làm đơn kèm xác nhận về tình trạng sức khỏe của cơ quan y tế từ cấp huyện trở lên để Nhà trường xem xét miễn , giảm tham gia hoạt động công tác xã hội .



SV đăng ký tham gia hoạt động công tác xã hội , nhưng không thực hiện nội dung đã đăng ký mà không có lý do hoặc lý do không chính đáng bị trừ 0.5 ngày công tác xã hội . SV đăng ký tham gia nhưng không hoàn thành nhiệm vụ , gây ảnh hưởng xấu đến hình ảnh của hoạt động , tổ chức , nhà trường , tùy theo mức độ vi phạm sẽ bị xử lý kỷ luật từ khiển trách trở lên và bị trừ ngày CTXH từ 1 đến toàn bộ số ngày đã tích lũy được .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

Học cải thiện điểm có ảnh hưởng đến xếp loại tốt nghiệp không?

SV được học cải thiện không quá 5 % số tín chỉ trong chương trình đào tạo .



- Tích lũy đủ học phần , số tín chỉ và hoàn thành các nội dung bắt buộc khác theo yêu cầu của chương trình đào tạo , đạt chuẩn đầu ra của chương trình đào tạo .

- Điểm trung bình tích lũy của toàn khóa học đạt từ trung bình trở lên .

- Tại thời điểm xét tốt nghiệp không bị truy cứu trách nhiệm hình sự hoặc không đang trong thời gian bị kỷ luật ở mức định chỉ học tập .

- Điểm rèn luyện toàn khóa đạt từ trung bình trở lên .

- Tích lũy đủ số ngày công tác xã hội theo quy định hiện hành của Trường .

- Hoàn thành giáo dục quốc phòng và giáo dục thể chất đối với các ngành đào tạo không chuyên về quân sự và thể dục - thể thao .

- Đạt yêu cầu chuẩn đầu ra ngoại ngữ .

- Đối với chương trình chính quy CLC , SV phải tham gia nghiên cứu khoa học theo nhóm nghiên cứu do giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

CHƯƠNG 7: KẾT LUẬN

Nguyên tắc xét học bổng của trường là gì?

Căn cứ Quyết định số 10/CT-ĐT về việc xác định số lượng suất học bổng cho sinh viên năm học 2019-2020. Nếu số lượng sinh viên đủ tiêu chuẩn xét học bổng nhiều hơn số suất học bổng đã được xác định thì sẽ tiến hành xét, cấp theo thứ tự: thành tích từ cao trào xuống cho đến khi hết học bổng. Lưu ý: Thứ tự ưu tiên khi xét HBKKHT như sau : + Xếp loại học bổng : Xuất sắc , Giỏi , Khá . + Trong cùng xếp loại học bổng , thứ tự ưu tiên căn cứ lần lượt vào điểm trung bình chung học tập , điểm rèn luyện , số tín chỉ . + Trường hợp có nhiều sinh viên có cùng các tiêu chí thành tích mà không đủ số tiền cấp sẽ do Hội đồng xem xét , quyết định .
- Khi cấp học bổng , trường hợp số tiền dư ở suất lẻ cuối cùng không đủ một suất học bổng thì sẽ được cấp thành nguyên một suất khi số tiền dư lớn hơn hoặc bằng 70 % giá trị suất học bổng . Trường hợp nhỏ hơn 70 % giá trị suất học bổng sẽ do Hội đồng xem xét quyết định .



Học phần trước : Học phần A là học phần trước của học phần B thì điều kiện bắt buộc để đăng ký học học phần B là đã học và có điểm tổng kết của học phần A khác không .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

Tôi muốn xin giấy hoãn NVQS thì phải làm sao?

Nếu SV vẫn còn trong thời gian đào tạo (dưới 4 năm) thì SV vào trang online.hcmute.edu.vn chọn phần Đăng ký giấy nhận , sau khi đăng ký SV xem thời gian hẹn trả giấy rồi tới Phòng Tuyển sinh và Công tác Sinh viên theo đúng lịch hẹn .



SV vào trang online.hcmute.edu.vn chọn phần Đăng ký giấy nhận , sau khi đăng ký SV xem thời gian hẹn trả giấy rồi tới Phòng Tuyển sinh và Công tác Sinh viên theo đúng lịch hẹn .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

CHƯƠNG 7: KẾT LUẬN

Học lại bao nhiêu % thì bị hạ 1 bậc xét tốt nghiệp ?

SV học lại / rớt quá 5 % tổng số chỉ thì sẽ bị hạ bằng xuất sắc xuống giỏi , giỏi xuống khá từ loại khá thì không hạ dù rớt 5 % số chỉ . Chi tiết liên hệ qua mail có Quỳnh : quynhbt@hcmute.edu.vn để được kiểm tra .



SV được học cải thiện không quá 5 % số tín chỉ trong chương trình đào tạo .



Agent chọn phản hồi 1

Nhập tin nhắn...

Gửi

Hình 50: Các phản hồi của chatbot từ FAQ

6.2.2 Phản hồi từ RAG

Chuẩn đầu ra ngoại ngữ của trường

Để đạt được tiêu chuẩn tốt nghiệp, sinh viên phải chứng minh khả năng ngoại ngữ bằng chứng chỉ quốc tế. Cụ thể:

- **Đối với các chương trình giảng dạy bằng tiếng Việt:**
 - Đại học chính quy đại trà: TOEIC 550 điểm hoặc N4 (JLPT) / J4 (Japanese Language Proficiency Test).
 - Đại học chính quy CLC tiếng Việt: TOEIC 600 điểm hoặc N4 (JLPT) / J4 (Japanese Language Proficiency Test).
 - Đại học chính quy CLC tiếng Nhật: N3 (JLPT) / J3 (Japanese Language Proficiency Test).
- **Đối với các chương trình giảng dạy bằng tiếng Anh:**
 - Cần chứng minh khả năng tiếng Anh đạt chuẩn theo tiêu chí của trường. Hãy liên hệ với phòng Đào tạo để biết thêm thông tin chi tiết về chính sách thi và cấp chứng chỉ ngoại ngữ nhé!

Khi nào thì SV được công nhận đã tham gia nghiên cứu khoa học?

SV được công nhận đã tham gia nghiên cứu khoa học khi thực hiện một trong những yêu cầu sau:

- Tham gia nghiên cứu khoa học theo nhóm nghiên cứu do các giảng viên hướng dẫn hoặc tham gia đề tài nghiên cứu khoa học với giảng viên.
- Tham gia đề tài nghiên cứu khoa học SV cấp trường.
- Tham gia các cuộc thi nghiên cứu khoa học, sân chơi nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức.
- Trình bày báo cáo khoa học tại các hội nghị nghiên cứu khoa học do Trường hoặc các Đơn vị đào tạo tổ chức.

91

CHƯƠNG 7: KẾT LUẬN

Có mấy khoảng thời gian để SV được rút bớt học phần vậy?

Có hai khoảng thời gian để sinh viên được rút bớt học phần:

- Trong thời gian điều chỉnh đăng ký học phần, sinh viên rút bớt học phần qua mạng và không phải đóng học phí cho các học phần này.
- Sau thời gian trên, sinh viên được phép rút bớt học phần trong tuần 8 với học kỳ chính, tuần 3 với học kỳ phụ theo thông báo của phòng Đào tạo. Sinh viên phải đóng học phí cho các học phần đã rút.

Nhập tin nhắn...

Gửi

Hội đồng đánh giá khóa luận tốt nghiệp bao gồm những ai?

Hội đồng đánh giá khóa luận tốt nghiệp bao gồm tối thiểu 3 và tối đa 5 thành viên. Trong đó có giảng viên hướng dẫn, nhưng không tham gia hội đồng đánh giá.

Tôi đã rút học phần vào tuần 8, vậy tôi có phải thanh toán các học phần đã rút không?

Trường hợp bạn rút học phần trong tuần thứ 8, bạn sẽ **phải thanh toán** các học phần đã rút. Theo quy định của trường, thời gian rút bớt học phần không tính phí là trong khoảng thời gian điều chỉnh đăng ký học phần. Sau thời điểm này, nếu muốn rút bớt học phần thì bạn cần phải đóng học phí cho các học phần đã rút.

Tôi đăng ký tham gia hoạt động công tác xã hội, nhưng không tham gia hay không hoàn thành nhiệm vụ, nội dung đã đăng ký thì sao?

Chào bạn! Cảm ơn bạn đã đặt câu hỏi. Theo quy định của Trường Đại học Sư phạm Kỹ thuật TP.HCM, sinh viên đăng ký tham gia hoạt động công tác xã hội nhưng không hoàn thành nhiệm vụ sẽ bị xử lý kỷ luật từ khiển trách trở lên và bị trừ ngày Công tác xã hội từ 1 đến toàn bộ số ngày đã tích lũy được tùy theo mức độ vi phạm. Bạn nên cân nhắc kỹ trước khi đăng ký tham gia bất kỳ hoạt động nào, đảm bảo bạn có thể hoàn thành nhiệm vụ để không ảnh hưởng đến hình ảnh của mình và các hoạt động đang diễn ra.

Tôi bỏ học nhưng quên rút học phần trong thời gian quy định cho phép rút học phần, vậy thì bị gì?

Nếu bạn bỏ học nhưng quên rút học phần trong thời gian quy định, tức là sau thời gian điều chỉnh đăng ký học phần và trước khi đến tuần 8 (học kỳ chính) hoặc tuần 3 (học kỳ phụ), thì trường sẽ mặc định tính điểm cho các học phần đó là 0. Điều này có nghĩa là bạn sẽ không được nhận tín chỉ cho các học phần đã bỏ, ảnh hưởng đến kết quả học tập và khả năng tốt nghiệp của bạn. Để tránh trường hợp này xảy ra trong lần sau, bạn nên lưu ý thời gian quy định rút học phần và thực hiện thủ tục rút bớt học phần nếu có nhu cầu.

CHƯƠNG 7: KẾT LUẬN

Không ghi MSSV trong lúc chuyển tiền học phí thì phải làm sao?

Nếu bạn quên ghi MSSV khi chuyển tiền học phí, bạn vui lòng liên hệ với Phòng Tài chính của trường để được hướng dẫn xử lý. Thông thường, Phòng Tài chính sẽ yêu cầu bạn cung cấp những thông tin sau:

- Họ tên đầy đủ - Lớp - Năm sinh - Số điện thoại liên lạc - Mã số thuế (nếu có) - Ghi chú chuyển khoản Phòng Tài chính sẽ xác minh thông tin và hỗ trợ ghi nhận học phí cho bạn. Hãy nhớ giữ lại các bằng chứng về việc bạn đã chuyển tiền, như biên lai hoặc giao dịch ngân hàng, để phục vụ cho quá trình xác minh.

Tôi có thể đăng ký môn học của hệ chất lượng cao không nếu tôi học hệ đại trà?

Chào bạn! Cảm ơn bạn đã lựa chọn trường Đại học Sư phạm Kỹ thuật TP.HCM. Theo như quy định của nhà trường, sinh viên hệ đại trà **không thể đăng ký** các môn học thuộc hệ chất lượng cao. Hệ đại trà và hệ chất lượng cao là hai chương trình đào tạo riêng biệt với những tiêu chuẩn và nội dung khác nhau. Hệ chất lượng cao thường có mức độ khó khăn cao hơn và yêu cầu điểm tốt nghiệp cao hơn. Let me know if you have any other questions!

Tôi là SV thuộc thuộc đồng bào dân tộc thiểu số, vậy thời gian đào tạo tối đa của tôi là

Bạn là sinh viên thuộc đồng bào dân tộc thiểu số nên thời gian đào tạo tối đa của bạn là **2 năm**. Điều này được quy định đặc thù cho đối tượng sinh viên người dân tộc thiểu số theo đúng như Quyết định của Nhà trường.

Các hoạt động nào không mang tính chất công tác xã hội bao gồm?

Các hoạt động **không** mang tính chất công tác xã hội bao gồm:

- Xem biểu diễn văn hóa nghệ thuật, chiếu phim; - Xem & cỗ vũ thi đấu thể thao; - Tham quan bảo tàng, di tích lịch sử; - Đi du lịch; - Tham dự hội thảo; - Sinh hoạt tại các Câu lạc bộ...

Cho tôi biết quy định về học phần khóa luận tốt nghiệp

Quy định học phần khóa luận tốt nghiệp như sau:

- Học phần khóa luận tốt nghiệp chỉ có điểm tổng kết được tính từ các điểm thành phần bao gồm điểm của giảng viên hướng dẫn, giảng viên phản biện và điểm của các thành viên hội đồng đánh giá.
- Hội đồng đánh giá khóa luận tốt nghiệp bao gồm tối thiểu 3 thành viên và tối đa 5 thành viên trong đó giảng viên hướng dẫn không tham gia hội đồng.
- Điểm tổng kết được tính từ trung bình cộng của các điểm thành phần. Các điểm thành phần không được lệch nhau quá 2 điểm, trường hợp lệch quá 2 điểm thì Hội đồng phải thống nhất và điều chỉnh phù hợp.

Hình 51: Các phản hồi của chatbot từ RAG

CHƯƠNG 7: KẾT LUẬN

6.2.3 Lịch sử Chat

Cho tôi biết phổ điểm từ thang điểm 10 sang thang điểm 4

Để tiện theo dõi, phổ điểm từ thang điểm 10 sang thang điểm 4 được chuyển đổi như sau:

- **9 - 10 điểm:** Học lực xuất sắc (A)
- **7 - 8.9 điểm:** Học lực giỏi (B)
- **5 - 6.9 điểm:** Học lực khá (C)
- **3 - 4.9 điểm:** Học lực trung bình (D)
- **0 - 2.9 điểm:** Học lực yếu (F)

tôi thấy không đúng

Tôi hiểu bạn thấy thông tin này không đầy đủ. Để tôi có thể giúp bạn tốt hơn, vui lòng cung cấp thêm thông tin về những gì bạn muốn biết hoặc thắc mắc gì liên quan đến việc học tập tại trường Đại học Sư phạm Kỹ thuật TP.HCM? Hãy hỏi rõ ràng câu hỏi của bạn để tôi có thể trả lời chính xác nhất!

ví dụ phổ điểm từ 9-10 là A+, 8.5-9 là A. Tôi muốn xem đầy đủ bảng phổ điểm

Bảng phổ điểm của trường đại học sư phạm kỹ thuật TP.HCM:

- 9 - 10: A+, 4 - 8.5 - 8.9; A, 3.7 - 8.0 - 8.4; B+, 3.5 - 7.0 - 7.9; B, 3.0 - 6.5 - 6.9; C+, 2.5 - 5.5 - 6.4; C, 2.0 - 5.0 - 5.4; D+, 1.5 - 4.0 - 4.9; D, 1.0 - <4.0; F, 0
- Học phần chỉ được tính tích lũy khi đạt từ điểm D trở lên. Một số trường hợp đặc biệt sử dụng các điểm chữ thay vì điểm số:
- P: áp dụng cho học phần giáo dục quốc phòng và giáo dục thể chất không tính vào điểm trung bình học tập, không phân mức.
- I: Điểm chưa hoàn thiện do được phép hoãn thi, kiểm tra.

Hình 52: Tính năng lưu lịch sử chat của chatbot

6.2.4. Cache

35s

Tôi bỏ học nhưng quên rút học phần trong thời gian quy định cho phép rút học phần, vậy thì bị gì

Nếu bạn bỏ học nhưng quên rút học phần trong thời gian quy định, tức là sau thời gian điều chỉnh đăng ký học phần và trước khi đến tuần 8 (học kỳ chính) hoặc tuần 3 (học kỳ phụ), thì trường sẽ mặc định tính điểm cho các học phần đó là

0. Điều này có nghĩa là bạn sẽ không được nhận tín chỉ cho các học phần đã bỏ, ảnh hưởng đến kết quả học tập và khả năng tốt nghiệp của bạn. Để tránh trường hợp này xảy ra trong lần sau, bạn nên lưu ý thời gian quy định rút học phần và thực hiện thủ tục rút bỏ học phần nếu có nhu cầu.

0s

Tôi bỏ học nhưng quên rút học phần trong thời gian quy định cho phép rút học phần, vậy thì bị gì

Bạn ơi, tôi hiểu bạn đang băn khoăn về việc bỏ học mà quên rút học phần trong thời gian quy định. Nếu bạn đã bỏ học nhưng không rút học phần trước khi thời hạn kết thúc thì hệ thống sẽ ghi nhận điểm "0" cho các môn bạn đã đăng ký. Điều này có thể ảnh hưởng đến kết quả học tập của bạn trong kỳ thi sắp tới. Để khắc phục tình trạng này, bạn nên liên hệ ngay với phòng Đào tạo để được tư vấn và giải quyết thắc mắc cụ thể. Bạn có thể tìm thêm thông tin chi tiết trên trang web hoặc bộ phận hỗ trợ sinh viên của trường đại học Sư phạm Kỹ thuật TP. HCM nhé! Chúc bạn sớm được giải quyết bài toán này.

Hình 53: Tính năng lưu trong bộ nhớ cache

CHƯƠNG 7. KẾT LUẬN

7.1 Tổng kết

Nhóm đã đạt được cơ bản những mục tiêu đã đề ra, nắm bắt, hiểu rõ các lý thuyết liên quan, từ đó áp dụng thực tế vào dự án “Chatbot hỗ trợ tư vấn Sinh viên trường Đại học Sư Phạm Kỹ Thuật TP.HCM”. Đò án tương đối đáp ứng những yêu cầu cơ bản của Chatbot, nhưng cũng còn những hạn chế nhất định cần khắc phục. Qua dự án này, nhóm học được nhiều bài học thực tế cách làm việc ở môi trường doanh nghiệp vào trong quá trình phát triển đề tài, cũng như tôi rèn được khả năng tự học, tự nghiên cứu tài liệu nước ngoài, qua đó tiếp cận gần hơn với nguồn tri thức nhân loại.

7.2 Đóng góp của đề tài

Đề tài “Chatbot hỗ trợ tư vấn cho sinh viên trường Đại học Sư phạm Kỹ thuật TP.HCM” mang lại nhiều đóng góp thiết thực cho nhà trường và sinh viên. Trước hết, ứng dụng giúp tự động hóa quy trình giải đáp thắc mắc, giảm tải áp lực cho đội ngũ cán bộ tư vấn và các phòng ban, từ đó nâng cao hiệu quả quản lý và hỗ trợ sinh viên. Chatbot cung cấp thông tin nhanh chóng, chính xác và hoạt động 24/7, giúp sinh viên dễ dàng tiếp cận thông tin về lịch học, chính sách học tập, yêu cầu tốt nghiệp và các hoạt động của trường, bất kể thời gian hay địa điểm. Bên cạnh đó, sản phẩm đã đạt được nhiều mục tiêu quan trọng, như triển khai thành công trên nền tảng web với giải pháp công nghệ tiên tiến như RAG, đảm bảo khả năng tìm kiếm thông tin hiệu quả từ các tài liệu liên quan. Độ chính xác của mô hình trả lời đạt mức cao nhờ áp dụng mô hình ngôn ngữ Gemma và phương pháp tạo vector embedding bằng HuggingFaceBgeEmbeddings. Hơn nữa, cơ chế Cache câu hỏi thường gặp giúp cải thiện tốc độ phản hồi, trong khi hệ thống học tăng cường hỗ trợ tối ưu hóa sự hài lòng của người dùng. Nhóm thực hiện rất hài lòng với hiệu quả và sự thân thiện của sản phẩm web, từ giao diện trực quan đến khả năng tương tác dễ dàng, độ chính xác cao, tốc độ phản hồi nhanh của chatbot, cũng như hệ thống tương tác likes, dislikes và góp ý từ người dùng, giúp chatbot ngày càng hoàn thiện. Sản phẩm không chỉ nâng cao trải nghiệm học tập của sinh viên mà còn góp phần tạo nên hình ảnh một trường đại học hiện đại, ứng dụng công nghệ tiên tiến trong quản lý và hỗ trợ giáo dục.

7.3 Hạn chế của đề tài

Mặc dù đề tài "Chatbot hỗ trợ tư vấn cho Sinh viên trường Đại học Sư Phạm Kỹ Thuật TP.HCM" mang lại nhiều lợi ích, nhưng vẫn tồn tại một số hạn chế nhất định khi:

- Phạm vi dữ liệu còn hạn chế: Nhóm chỉ tiếp cận được dữ liệu từ trang Tư vấn Sinh viên và Sổ tay Sinh viên năm 2023. Các dữ liệu thông báo từ các Khoa, các cấp không được thống nhất ở một trang chủ chung, mà phân bố riêng lẻ, không công khai toàn thể Sinh viên nên rất khó tiếp cận.
- Khả năng hiểu ngữ cảnh phức tạp: Chatbot chỉ có thể giải đáp, tư vấn những thắc mắc chung, không cụ thể chi tiết, qua đây cần có giải pháp liên kết thắc mắc với thầy cô/cán bộ phụ trách để có hướng giải quyết, giúp tăng trải nghiệm cho các bạn Sinh viên.
- Số lượng dữ liệu còn ít: Dữ liệu Sổ tay Sinh viên, bảng Feedback còn ít gây hạn chế trong việc đào tạo RAG cũng như huấn luyện Agent chọn ra phản hồi còn chưa tốt.
- Giới hạn khả năng xử lý ngôn ngữ tự nhiên: Mặc dù đã sử dụng thư viện underthesea của Pythorch để tiền xử lý tiếng Việt, nhưng hệ thống vẫn gặp khó khăn trong việc hiểu các câu hỏi không rõ ràng, sử dụng từ ngữ địa phương hoặc cách diễn đạt không chuẩn.
- Thời gian phản hồi chưa đạt đến thời gian thực: Mặc dù phản hồi cho những câu hỏi FAQ gần như là thời gian thực, tuy nhiên nếu câu hỏi này do RAG đảm nhiệm, thì khoảng thời gian kéo dài gần 1 phút, điều này cần thêm GPU để xử lý tốt hơn.

7.4 Hướng phát triển cho tương lai

Nhóm sẽ khắc phục những hạn chế đã nêu ở trên, cập nhật lại dữ liệu cho phù hợp với thời điểm để đảm bảo độ chính xác về thông tin, nhất là các thông tin về điểm ngoại ngữ, điểm học phần,... Đồng thời, cũng như thiết kế lại giao diện một cách tỉ mỉ, đẹp mắt, thân thiện với người dùng hơn nữa.

Trong tương lai, ứng dụng “Chatbot tư vấn Sinh viên trường Đại Học Sư Phạm Kỹ Thuật TP.HCM” sẽ được phát triển để nâng cao hiệu quả và tính chính xác, đáp ứng nhiều

CHƯƠNG 7: KẾT LUẬN

yêu cầu phức tạp hơn và mang lại trải nghiệm toàn diện hơn. Hệ thống sẽ được nâng cấp với khả năng phân tích ngữ cảnh chuyên sâu, giúp chatbot hiểu rõ hơn ý định và nhu cầu cụ thể của người dùng, từ đó cung cấp câu trả lời chi tiết và chính xác hơn. Dữ liệu từ nhiều nguồn khác nhau như lịch học, sự kiện của trường, thông báo quan trọng từ trường và các Khoa Đào tạo, bên cạnh đó các chính sách mới sẽ được tích hợp để đảm bảo thông tin luôn được cập nhật và phong phú.

Hơn thế nữa, “Chatbot tư vấn Sinh viên trường Đại Học Sư Phạm Kỹ Thuật TP.HCM” sẽ được cải tiến để các bạn Sinh viên có thể nộp hồ sơ, quy đổi điểm ngoại ngữ trực tuyến mà không cần mất thời gian đến các Phòng ban tiếp nhận hồ sơ như truyền thống, điều này cũng đòi hỏi việc triển khai tính bảo mật của hệ thống để đảm bảo thông tin của Sinh viên được giữ bí mật không bị đánh cắp.

Cuối cùng, với những hướng phát triển trên, nhóm tin rằng trong tương lai ứng dụng sẽ trở nên hoàn thiện hơn với đầy đủ các tính năng hấp dẫn, góp phần cải thiện trải nghiệm của người dùng tốt hơn, tiết kiệm chi phí, thời gian từ đó nâng cao được chất lượng trải nghiệm của các bạn Sinh viên, các thầy cô/cán bộ phụ trách quản lý Sinh viên, đang học tập và làm việc tại trường Đại học Sư Phạm Kỹ Thuật TP.HCM.

TÀI LIỆU THAM KHẢO

- [1] Mô hình ngôn ngữ lớn: <https://aws.amazon.com/vi/what-is/large-language-model/>
- [2] Tìm hiểu về Langchain: <https://aws.amazon.com/vi/what-is/langchain/>
- [3] Tìm hiểu PostgreSQL: <https://www.w3schools.com/postgresql/>
- [4] Tìm hiểu tổng quan SQL Server: <https://itviec.com/blog/sql-server-la-gi/>
- [5] RAG là gì? Ứng dụng của RAG: <https://lacviet.vn/retrieval-augmented-generation/>
- [6] Bài báo khoa học về RAG: <https://arxiv.org/pdf/2312.10997.pdf>
- [7] Deep Q-Learning: <https://vietnix.vn/deep-learning-la-gi/>
- [8] LangChain: <https://python.langchain.com/docs/introduction>
- [9] Tìm hiểu Q-Learning: <https://www.geeksforgeeks.org/q-learning-in-python/>
- [10] Sô tay SV 2023: <https://sao.hcmute.edu.vn/ArticleId/e8db4387-3f43-4770-a67c-90fa85ce009f/so-tay-sinh-vien-2023>
- [11] Django: <https://docs.djangoproject.com/en/5.1/>
- [12] Llama cpp: <https://github.com/ggerganov/llama.cpp>