

Phát triển trợ lý ảo học tập tích hợp LLM hỗ trợ học môn Cấu trúc dữ liệu và Giải thuật

Hoàng Ngọc Long*, Võ Hải Đăng, Lê Hiếu Nghiêm

Trường Đại học Quốc tế Hồng Bàng

TÓM TẮT

Trợ lý ảo học tập là một trong những ứng dụng tiềm năng của trí tuệ nhân tạo trong giáo dục đại học, đặc biệt khi kết hợp với các mô hình ngôn ngữ lớn (LLM). Bài báo này trình bày một cách tiếp cận phát triển hệ thống trợ lý ảo hỗ trợ học tập cho môn học Cấu trúc dữ liệu và Giải thuật (CTDL>) - một trong những học phần nền tảng và bắt buộc trong chương trình đào tạo ngành Công nghệ thông tin, giữ vai trò quan trọng trong việc rèn luyện tư duy thuật toán và năng lực giải quyết vấn đề cho sinh viên. Hệ thống được xây dựng dựa trên kiến trúc tích hợp giữa mô hình ngôn ngữ lớn và kỹ thuật truy hồi kết hợp sinh (Retrieval-Augmented Generation - RAG). Nghiên cứu cũng đề xuất cách tiếp cận đánh giá trợ lý ảo học tập dựa trên ba tiêu chí: Cấu trúc câu trả lời, chất lượng nội dung và độ liên quan đến truy vấn người học. Kết quả đánh giá cho thấy hệ thống giúp sinh viên tiếp cận và ôn tập kiến thức một cách trực quan, linh hoạt và hiệu quả. Nghiên cứu mở ra hướng phát triển khả thi cho việc ứng dụng LLM trong xây dựng công cụ học tập thông minh và cá nhân hóa trong môi trường giáo dục đại học.

Từ khóa: cá nhân hóa học tập, cấu trúc dữ liệu và giải thuật, kỹ thuật RAG, mô hình ngôn ngữ lớn, trợ lý ảo học tập

1. ĐẶT VẤN ĐỀ

Trí tuệ nhân tạo (AI) đang phát triển mạnh mẽ, từ các thuật toán đơn giản ban đầu như hệ chuyên gia (Expert Systems), AI đã tiến tới các mô hình học sâu (Deep Learning) hiện đại với khả năng tự học và tối ưu hóa dựa trên dữ liệu đầu vào. AI đang chứng minh vai trò ngày càng quan trọng trong đời sống, khi góp phần nâng cao chất lượng cuộc sống và tối ưu hóa nhiều hoạt động thường nhật. Trợ lý ảo, hay còn gọi chatbot, là một trong những ứng dụng tiêu biểu của AI hiện nay trong việc xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP), cho phép máy tính giao tiếp với con người thông qua ngôn ngữ tự nhiên một cách tự động và thông minh. Công nghệ này có khả năng tối ưu hóa quy trình làm việc trong nhiều lĩnh vực như giáo dục, y tế, thương mại điện tử và chăm sóc khách hàng.

Trong lĩnh vực giáo dục, chatbot đã và đang chứng minh tiềm năng to lớn khi có thể đóng vai trò như một trợ lý ảo học tập thông minh. Thực tế cho thấy, trong giáo dục hình thức giảng dạy truyền thống gặp khó khăn trong việc theo sát từng sinh viên và giải đáp đầy đủ mọi thắc mắc. Điều này khiến nhiều người học chủ động tìm đến các công cụ học tập hỗ trợ. Trong bối cảnh đó, chatbot trở thành “trợ lý ảo học tập” đắc lực, có thể giải đáp nhanh chóng các câu hỏi về lý thuyết, bài tập, tài liệu..., giúp người

học tiết kiệm thời gian và nâng cao tính chủ động. Đồng thời, giáo viên cũng có điều kiện tập trung vào chuyên môn giảng dạy và phát triển kỹ năng mềm cho sinh viên. Đối với các môn học có độ phức tạp cao như CTDL>, chatbot còn được đánh giá là công cụ hỗ trợ hiệu quả, nhờ khả năng phản hồi tức thì các câu hỏi về thuật toán, độ phức tạp, cấu trúc dữ liệu phù hợp, thậm chí đưa ra ví dụ minh họa và lời giải cụ thể. Với tính linh hoạt trong tương tác và khả năng thích ứng với từng trình độ người học, chatbot hỗ trợ quá trình học trở nên trực quan, dễ hiểu và dễ tiếp cận hơn.

Theo như nghiên cứu của tác giả Lê Thủy Hiền [1] tại Trường Đại học Kinh tế Thành phố Hồ Chí Minh, việc ứng dụng chatbot trong học tập có tác động tích cực đến nhận thức và hành vi học tập của sinh viên, cho thấy tiềm năng của công nghệ này trong việc hỗ trợ quá trình tự học. Wollny và cộng sự [2] cũng đã nêu ra rằng chatbot trong giáo dục có thể giúp cải thiện khả năng tiếp thu kiến thức, hỗ trợ sinh viên hiểu rõ hơn về nội dung học tập và nâng cao hiệu suất học tập. Một bước tiến đáng chú ý trong lĩnh vực AI gần đây là sự ra đời và phát triển của các mô hình ngôn ngữ lớn (Large Language Models - LLM) như GPT (Generative Pre-trained Transformer). Những mô hình này đã cải thiện rõ rệt khả năng hiểu và tạo sinh

Tác giả liên hệ: Hoàng Ngọc Long

Email: longhn@hiu.vn

ngôn ngữ tự nhiên, giúp chatbot trở nên mượt mà và chính xác hơn trong phản hồi. LLM không chỉ hiểu sâu sắc về ngữ nghĩa, mà còn đủ sức xử lý những tình huống học tập phức tạp, mang lại trải nghiệm học cá nhân hóa và hiệu quả. Nhờ khả năng học từ lượng dữ liệu khổng lồ, LLM có thể đề xuất giải pháp phù hợp, tài liệu hữu ích và giải thích rõ ràng các khái niệm chuyên sâu trong các môn học như CTDL>.

Nghiên cứu này đề xuất một hướng tiếp cận tích hợp mô hình LLM vào ứng dụng hỗ trợ học tập chuyên biệt cho môn học CTDL>. Các phần tiếp theo của bài báo sẽ trình bày cơ sở lý thuyết, thiết kế hệ thống, kết quả thử nghiệm và hướng phát triển tương lai của hệ thống.

2. CƠ SỞ LÝ THUYẾT VÀ NGHIÊN CỨU LIÊN QUAN

2.1. Tổng quan về trợ lý ảo và mô hình ngôn ngữ lớn

Chatbot là các chương trình máy tính đối thoại thông minh mô phỏng cuộc trò chuyện của con người một cách tự nhiên. Một chatbot có thể xử lý đầu vào từ người dùng và tạo ra đầu ra tương ứng [3]. Thông thường, chatbot nhận văn bản ngôn ngữ tự nhiên làm đầu vào, và đầu ra cần là câu trả lời phù hợp nhất với câu đầu vào của người dùng. Chatbot cũng có thể được định nghĩa là “hệ thống đối thoại giữa người và máy tính trực tuyến sử dụng ngôn ngữ tự nhiên” [4]. Như vậy, chatbot là một hệ thống đối thoại tự động có khả năng phục vụ hàng nghìn người dùng cùng lúc. Chatbot có thể được phân loại dựa trên nhiều tiêu chí khác nhau, bao gồm: Loại dịch vụ cung cấp, miền tri thức, phương pháp tạo phản hồi, mục tiêu, quyền truy cập hệ thống và mức độ hỗ trợ từ con người [4].

Trong bối cảnh giáo dục, chatbot đang được phát triển thành một hình thái cao cấp hơn gọi là trợ lý ảo học tập (virtual learning assistant). Đây là các hệ thống đối thoại thông minh không chỉ tương tác ngôn ngữ mà còn hỗ trợ quá trình học tập một cách chủ động, có định hướng và cá nhân hóa theo trình độ, mục tiêu và hành vi của người học. Trợ lý ảo học tập có khả năng trả lời câu hỏi học thuật, giải thích khái niệm, cung cấp ví dụ minh họa. Ngoài ra, hệ thống còn đưa ra gợi ý ôn tập và theo dõi tiến trình học của người dùng. Không giống các chatbot truyền thống, trợ lý ảo học tập thường được tích hợp với các nguồn tri thức chuyên ngành. Các hệ thống này có khả năng thích ứng theo từng đối tượng người học, từ đó giúp tăng cường hiệu quả và tương tác trong môi trường học tập số.

Để xây dựng một hệ thống chatbot hoặc trợ lý ảo học tập cơ bản, hiện nay có hai phương pháp phổ

biến [3, 5]:

- Phương pháp dựa trên luật (Rule-based) là phương pháp truyền thống trong đó chatbot hoạt động dựa trên các tập hợp quy tắc (rule-based) được định nghĩa sẵn như từ khóa, mẫu câu và cây điều kiện (if-else). Giả sử, nếu người dùng hỏi “ngăn xếp là gì?”, chatbot sẽ nhận diện từ “ngăn xếp” và đưa ra phản hồi phù hợp. Phương pháp này dễ phát triển, kiểm soát nội dung phản hồi tốt. Tuy nhiên, nó lại thiếu khả năng mở rộng và không thích ứng tốt với các truy vấn phức tạp.
- Phương pháp dựa trên AI (AI-based) sử dụng các kỹ thuật trí tuệ nhân tạo như học máy (Machine Learning - ML) và học sâu (Deep Learning - DL) là để học từ dữ liệu và cung cấp các phản hồi thông minh, tự động phù hợp với ngữ cảnh. Những chatbot này giúp cải thiện hiệu quả giao tiếp, tăng tính linh hoạt và mang lại trải nghiệm người dùng tốt hơn so với các phương pháp xây dựng chatbot truyền thống. Học sâu (DL) là một bước tiến quan trọng trong lĩnh vực AI, cho phép mô hình hóa các cấu trúc dữ liệu phức tạp bằng cách sử dụng mạng nơ-ron nhiều tầng. Trong xử lý ngôn ngữ tự nhiên (NLP), học sâu đã tạo tiền đề cho sự phát triển của các kiến trúc hiện đại như Transformer [6 - 8]. Phương pháp này nổi bật nhờ độ chính xác cao, khả năng sinh ngôn ngữ tự nhiên tốt và khả năng thích ứng với dữ liệu phức tạp. Tuy vậy, nó cũng đòi hỏi tài nguyên tính toán lớn, thời gian huấn luyện dài và có thể khó kiểm soát do có nhiều tham số ẩn.

Mô hình ngôn ngữ lớn (Large Language Models - LLMs) là một ứng dụng tiêu biểu của học sâu, được huấn luyện trên tập dữ liệu văn bản khổng lồ với mục tiêu hiểu và sinh ngôn ngữ ở mức độ tự nhiên cao. Về mặt kỹ thuật, LLM thường được xây dựng dựa trên kiến trúc Transformer và sử dụng hàng tỷ tham số, cho phép mô hình học các biểu diễn ngôn ngữ phức tạp. Các mô hình nổi bật như GPT (OpenAI) [6], BART (Google) [7] và LLaMa (Meta) [8] đã chứng minh khả năng vượt trội trong nhiều nhiệm vụ như trả lời câu hỏi, hội thoại, dịch máy, và tóm tắt văn bản. Nhờ khối lượng dữ liệu huấn luyện đa dạng, LLM có khả năng thu nhận tri thức tổng quát và biểu hiện ngôn ngữ linh hoạt. Tuy nhiên, LLM cũng đối mặt với hiện tượng “ảo giác” (hallucination), tức là tạo ra nội dung không chính xác hoặc không có cơ sở thực tế - một hạn chế phổ biến do thiếu khả năng truy cập thông tin cập nhật và tri thức chuyên biệt. Để khắc phục điều này, các kỹ thuật tăng cường như Retrieval-Augmented Generation (RAG) [9] đã được đề xuất. RAG kết hợp giữa khả năng sinh ngôn ngữ của mô hình với hệ

thống truy hồi tài liệu ngữ cảnh từ kho tri thức ngoài, cho phép phản hồi chính xác hơn và có thể kiểm chứng. Đây là một bước tiến quan trọng trong việc phát triển trợ lý ảo học tập dựa trên LLM, vì nó cho phép mô hình vừa hiểu câu hỏi theo ngữ cảnh, vừa truy cập các tài liệu học thuật phù hợp - từ đó đưa ra lời giải thích chính xác, cập nhật và minh bạch.

2.2. Các nghiên cứu liên quan

Để xây dựng một ứng dụng trợ lý ảo học tập có thể đáp ứng tốt các nhu cầu và hỗ trợ toàn diện cho người dùng, nhóm đã tiến hành khảo sát một số nghiên cứu liên quan. Chatbot tư vấn tuyển sinh ở công trình [10] tập trung vào việc phát triển một chatbot tư vấn tuyển sinh cho trường đại học. Chatbot được thiết kế để trả lời các câu hỏi thường gặp của sinh viên về quy trình tuyển sinh, các ngành học, học phí và các thông tin liên quan khác. Nghiên cứu này sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên và học máy để cải thiện khả năng hiểu và phản hồi của chatbot. Vấn đề lớn của hệ thống này là không kiểm soát được dữ liệu và thiếu chuyên môn trong một số lĩnh vực, dẫn đến khả năng cung cấp thông tin không chính xác hoặc không đầy đủ. Công trình [11] xây dựng một chatbot tư vấn người dùng trong lĩnh vực y tế da liễu. Chatbot này sử dụng nền tảng Rasa để phát triển, với các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) như phân loại ý định, trích xuất thực thể và quản lý hội thoại. Mô hình PhoBERT được áp dụng để cải thiện hiệu quả nhận dạng ý định và tên riêng. Kết quả thử nghiệm cho thấy chatbot có độ chính xác cao trong việc tư vấn các triệu chứng và phương pháp điều trị bệnh da liễu. Tuy nhiên, chatbot này dựa trên quy tắc (rule-based chatbot), làm hạn chế sự thoải mái cho người dùng và làm giảm tính linh hoạt của hệ thống. Nghiên cứu [12] đã trình bày phương pháp xây dựng ứng dụng hỗ trợ học tập dựa trên ChatGPT API nhằm giúp sinh viên học tập hiệu quả hơn trong môn CTD>. Các kỹ thuật nổi bật được áp dụng trong nghiên cứu [12] như prompting, pre-prompting, function call và fine-tuning dùng để tạo ra phản hồi phù hợp với nội dung môn học CTD>. Dù vậy, nghiên cứu [12] vẫn đang còn phụ thuộc nhiều vào API của bên thứ ba, dẫn đến các mặt hạn chế như chi phí cao, khó kiểm soát dữ liệu và hệ thống được xây dựng vẫn chưa có cơ chế đánh giá tự động sự liên quan của nội dung phản hồi.

Nghiên cứu này đề xuất một số hướng khắc phục trong quá trình xây dựng ứng dụng trợ lý ảo, cụ thể như:

(1) Tích hợp trực tiếp mô hình ngôn ngữ lớn (LLM)

vào hệ thống thay vì phụ thuộc vào các API bên ngoài. Cách tiếp cận này giúp giảm chi phí vận hành, kiểm soát tốt hơn dữ liệu đầu vào/đầu ra của chatbot, đồng thời giảm độ trễ trong xử lý truy vấn do không phải truyền tải dữ liệu qua mạng đến máy chủ bên ngoài.

(2) Phát triển cơ chế đánh giá phản hồi nhằm cung cấp cho người dùng cơ sở để theo dõi cấu trúc, tính chính xác và mức độ liên quan của dữ liệu trả về.

2.3. Các mô hình ngôn ngữ lớn nguồn mở

Nhóm nghiên cứu đã tiến hành khảo sát các mô hình ngôn ngữ lớn nguồn mở sau nhằm tìm ra mô hình phù hợp cho mục tiêu của nghiên cứu:

- Mô hình aya-23-8b là mô hình LLM do Cohere phát triển nổi bật nhờ khả năng xử lý linh hoạt các ngôn ngữ ít phổ biến và phù hợp với các tác vụ phân tích, sinh văn bản phức tạp. Theo bài báo so sánh giữa các mô hình LLM phổ biến [13] cho thấy so với llama3, aya-23-8b tránh được lỗi hiểu sai chỉ dẫn hoặc sinh văn bản thừa bằng tiếng Anh. Mô hình này cũng đạt điểm cao trên các thước đo BLEU, ROUGE [14] và độ trung thực ngữ nghĩa (semantic fidelity), đặc biệt hiệu quả trong các tác vụ đa ngôn ngữ.
- Mô hình qwen2-vl-7b-instruct [15] là dòng mô hình ngôn ngữ lớn (LLM) do Alibaba Cloud phát triển, có khả năng xử lý đồng thời hình ảnh và văn bản, giúp nâng cao độ chính xác và chất lượng đầu ra. Nhờ khả năng tương tác theo hướng dẫn, mô hình đặc biệt phù hợp với các tác vụ cần linh hoạt và tùy biến theo yêu cầu.
- Mô hình hermes-3-llama-3.2-3b [16] là dòng mô hình ngôn ngữ lớn (LLM) do Nous Research phát triển, được tinh chỉnh từ Llama-3.2-3B, nổi bật với khả năng đối thoại đa lượt, lập luận, vai diễn và tạo mã.

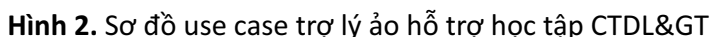
3. THIẾT KẾ HỆ THỐNG VÀ PHƯƠNG PHÁP

3.1. Khảo sát hiện trạng và phân tích chức năng

Nhằm tìm hiểu nhu cầu và mong muốn sử dụng hệ thống trợ lý ảo học tập của sinh viên, nhóm nghiên cứu đã tiến hành một cuộc khảo sát với mục tiêu thu thập thông tin về nhu cầu sử dụng trợ lý ảo trong quá trình học tập môn học CTD>. Kết quả khảo sát được sử dụng để xây dựng các chức năng phù hợp cho hệ thống trợ lý ảo học tập môn CTD>. Đối tượng khảo sát là sinh viên ngành Công nghệ thông tin (CNTT) tại Trường Đại học Quốc tế Hồng Bàng. Hình thức khảo sát là trực tuyến bằng Google Forms với các câu hỏi tập trung vào tìm hiểu nhu cầu và mong muốn của sinh viên khi sử dụng trợ lý ảo học tập môn CTD>.



mục tiêu cung cấp môi trường trao đổi, học tập và tra cứu thông tin về môn học CTDL> trong chương trình đào tạo ngành CNTT tại Trường Đại học Quốc tế Hồng Bàng một cách trực quan và tương tác.



tập, giúp tạo ra một chuỗi truy vấn logic, có tính liên kết giữa các chủ đề. Về mặt kiến trúc, mô-đun này đảm bảo rằng các phiên giao tiếp được duy trì liên tục và có tính liên kết cao, ngay cả khi truy vấn của người dùng thay đổi theo thời gian.

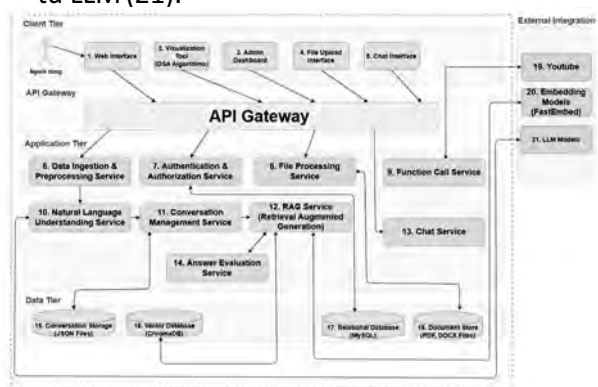
- Mô-đun Hỏi đáp & Sinh câu trả lời là thành phần chính của hệ thống giao tiếp thông tin, được thiết kế nhằm tăng cường khả năng tương tác giữa người dùng và hệ thống. Mô-đun này đảm nhiệm nhiệm vụ nhận các câu hỏi do người dùng nhập vào, phân tích nội dung, truy xuất dữ liệu liên quan và đưa ra câu trả lời phù hợp cùng với các đề xuất thông tin bổ trợ nếu cần thiết. Các thành phần chính và quy trình trong mô-đun:

- Mô-đun RAG (12) có nhiệm vụ truy xuất tri thức từ các kho dữ liệu có sẵn (16, 17, 18) và dùng mô hình LLM (21) để sinh câu trả lời.
- + Mô-đun Đánh giá câu trả lời (14) có nhiệm vụ đánh giá chất lượng và mức độ phù hợp của câu trả lời trước khi trả về người dùng.
- + Mô-đun Hỏi đáp (13) kết nối giữa backend và giao diện người dùng để hiển thị phản hồi.

Toàn bộ chuỗi xử lý hỏi-đáp diễn ra theo thứ tự: Truy vấn → Tiền xử lý (6) → Hiểu ngôn ngữ (10) → Quản lý hội thoại (11) → Truy xuất/sinh câu trả lời (12) → Đánh giá (14) → Gửi lại qua giao diện người dùng (13).

- Mô-đun Tích hợp và giao tiếp với hệ thống bên ngoài được thiết kế nhằm tạo cầu nối cho trợ lý ảo với các ứng dụng khác, từ đó mở rộng các tính năng của hệ thống, gồm các mô-đun sau:

- + Mô-đun Xác thực (7) có nhiệm vụ xác thực người dùng khi đăng nhập từ giao diện.
- + Mô-đun Xử lý tập tin (8) hỗ trợ xử lý tài liệu (pdf, docx), lưu trữ vào mô-đun (18).
- + Mô-đun Gọi hàm (9) cho phép gọi hàm tùy chỉnh từ LLM (21).



Hình 3. Kiến trúc tổng quan hệ thống trợ lý ảo học tập môn CTDL>

3.3. Cách thức đánh giá phản hồi từ trợ lý ảo học tập

Về tiêu chí đánh giá, hệ thống đánh giá dựa trên 3 khía cạnh chính:

- Cấu trúc: Đánh giá tính đầy đủ và cấu trúc phù hợp của câu trả lời.
- Nội dung: Đánh giá chất lượng nội dung thực tế.
- Độ liên quan: Đánh giá mức độ phù hợp với câu hỏi.

3.3.1. Đánh giá cấu trúc

Hệ thống đánh giá sự hiện diện và nội dung của các thẻ bắt buộc trong câu trả lời:

<CONCEPT></CONCEPT>: Giải thích khái niệm.

<EXAMPLE></EXAMPLE>: Ví dụ liên quan.

<IMPLEMENTATION></IMPLEMENTATION>: Triển khai mã nguồn (nếu có).

<EXPLANATION></EXPLANATION>: Giải thích chi tiết vấn đề.

<DESCRIPTION></DESCRIPTION>: Minh họa thuật toán hoặc cấu trúc.

<COMPLEXITY></COMPLEXITY>: Phân tích độ phức tạp (nếu có).

<LAB></LAB>: Phần thực hành (nếu có).

<VIDEO></VIDEO>: Phần video liên quan (nếu có)

Hệ thống đánh giá sự hiện diện và nội dung của các thẻ bắt buộc trong câu trả lời:

- Thẻ có nội dung đầy đủ: +10 → 15 điểm.
- Thẻ tồn tại nhưng trống: +5 điểm.
- Thẻ bị thiếu hoặc lỗi: 0 điểm.
- Cấu trúc tổng thể: Tối đa 100 điểm.

3.3.2. Đánh giá nội dung

Để đánh giá chất lượng nội dung phản hồi của hệ thống trợ lý ảo học tập AlgoBot trong môn CTDL>, nghiên cứu đề xuất cơ chế chấm điểm dựa trên bốn yếu tố chính: độ chi tiết và rõ ràng của giải thích khái niệm, tính minh họa và liên quan của ví dụ, chất lượng mã nguồn, và độ chính xác của phân tích độ phức tạp thuật toán. Các yếu tố này được gán trọng số tương ứng (30% cho giải thích khái niệm, 20% cho ví dụ, 20% cho mã nguồn, 30% cho phân tích độ phức tạp) và điều chỉnh dựa trên độ dài phản hồi, nhằm đảm bảo tính khách quan và phù hợp với yêu cầu học thuật.

- Chất lượng giải thích khái niệm: yếu tố này đánh giá mức độ rõ ràng, đầy đủ và chính xác của phần giải thích khái niệm trong phản hồi ở thẻ <CONCEPT>. Các tiêu chí bao gồm: (1) độ dài văn bản, sử dụng số từ (word_count) trong phần giải thích, với ngưỡng tối thiểu 100 từ để đảm bảo tính chi tiết; số lượng thuật ngữ chuyên ngành có trong phản hồi; và đánh giá cấu trúc phản hồi bằng cách so sánh với cấu trúc của từng loại câu hỏi.

- Chất lượng ví dụ: Yếu tố này đánh giá khả năng minh họa khái niệm thông qua nội dung phản hồi ở thẻ <EXAMPLE>, được chấm dựa trên tỷ lệ từ khóa chung giữa câu hỏi và ví dụ.

- Chất lượng mã nguồn: Yếu tố này đánh giá phần mã nguồn trong phản hồi ở thẻ <IMPLEMENTATION> dựa trên tính đầy đủ của mã nguồn và tỷ lệ chú thích comment_lines / max(1, code_lines).

- Phân tích độ phức tạp: Yếu tố này đánh giá phần phân tích độ phức tạp ở thẻ <COMPLEXITY> và được chấm theo các tiêu chí như +25 điểm nếu có phân tích đầy đủ độ phức tạp thuật toán; +15 điểm nếu có phân tích một phần; +5 điểm nếu có phân tích cơ bản.

- Điều chỉnh theo độ dài phản hồi: Nếu word_count ≥ 500 thì điểm được điều chỉnh tăng 10%; nếu word_count < 500 thì điểm được điều chỉnh giảm 20%.

3.3.3. Đánh giá độ liên quan

Đánh giá mức độ phù hợp của câu trả lời với câu hỏi:

- Tỷ lệ từ khóa chung (45% trọng số):

$$\text{keyword_ratio} = \frac{\text{len}(\text{common_keywords})}{\max(1, \text{len}(\text{query_keywords}))}$$

$$\text{keyword_score} = \min(100, \text{keyword_ratio} * 100)$$

- Tương đồng ngữ nghĩa (45% trọng số):

$$\text{semantic_score} = \cosine_similarity(\text{query_embedding}, \text{response_embedding}) * 100$$

3.3.4. Tính điểm tổng hợp và phân loại đánh giá

Dựa trên điểm tổng hợp có trọng số, câu trả lời được phân loại thành ba mức:

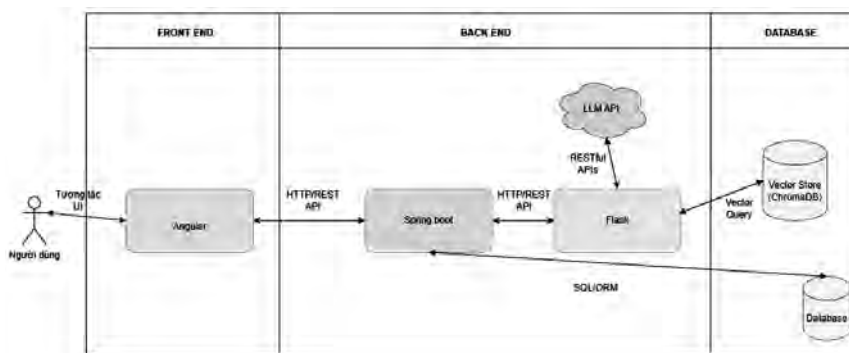
- Tốt (≥ 70 điểm): Câu trả lời rõ ràng, đúng định dạng và đầy đủ thông tin.
- Trung bình (50 - 69 điểm): Câu trả lời có cấu trúc nhưng thiếu nhất quán hoặc chưa đầy đủ.
- Tệ (< 50 điểm): Thiếu cấu trúc, sai định dạng, thông tin không chính xác.

4. KẾT QUẢ NGHIÊN CỨU VÀ THỬ NGHIỆM

4.1. Cài đặt hệ thống

Để phát triển trợ lý ảo học tập tích hợp LLM hỗ trợ học CTD> hệ thống được nhóm phát triển theo mô hình 3 lớp như ở Hình 4 với Angular cho giao diện người dùng, Spring Boot cho xử lý nghiệp vụ và kết nối cơ sở dữ liệu, cùng Flask đảm nhiệm các chức năng kết nối với ChromaDB và LangChain. Ứng dụng gồm các trang sau:

- Trang Home: Giới thiệu về hệ thống, chuyển tiếp người dùng đến trang chat.
- Trang Đăng ký/ Đăng nhập: Người dùng sẽ xác thực thông tin trước khi sử dụng ứng dụng.
- Trang Chat: Là trang người dùng tương tác với trợ lý ảo như ở Hình 5.
- Trang Minh họa thuật toán: Người dùng có thể thấy rõ hoạt ảnh minh họa chi tiết về cách thức vận hành của cấu trúc dữ liệu với các thông số được nhập từ người dùng như ở Hình 6.



Hình 4. Triển khai hệ thống theo mô hình 3 lớp



Hình 5. Giao diện trang Chat



Hình 6. Giao diện trang minh họa thuật toán

4.2. Xây dựng dữ liệu

Trong quá trình kiểm thử, để nâng cao tính chính xác câu trả lời của trợ lý ảo học tập, hệ thống được nạp tài liệu giáo trình và bài thực hành sử dụng trong môn học CTD> ngành CNTT tại Trường Đại học Quốc tế Hồng Bàng [17]. Dữ liệu là các tập tin *.pdf hoặc *.docx sẽ được nạp vào được ChromaDB xử lý thông qua quy trình RAG. Quy trình này bao gồm:

- (1) Phân đoạn nội dung: Tài liệu được chia thành các đoạn nhỏ để dễ dàng quản lý và truy xuất sau này.
- (2) Tạo vector nhúng cho mỗi đoạn nhỏ: Chuyển đổi từng đoạn văn bản thành vector nhúng có kích thước cố định.
- (3) Lưu trữ trong ChromaDB: Các vector nhúng cùng với metadata tương ứng sẽ được lưu trữ trong ChromaDB để phục vụ cho việc truy xuất nhanh chóng khi có yêu cầu từ người dùng.

#	ID	Tên file	Thời gian tải lên	Thao tác
1	4	Đồ Vẽ Nhon. Giáo trình Cấu trúc Dữ Liệu & Giải thuật.pdf	19/02/2025 16:44:06	
2	6	CTDL>_Lab01.pdf	31/03/2025 23:09:20	
3	7	CTDL>_Lab02.pdf	31/03/2025 23:09:25	
4	8	CTDL>_Lab03.pdf	31/03/2025 23:09:53	
5	9	CTDL>_Lab04.pdf	31/03/2025 23:10:13	
6	10	CTDL>_Lab05.pdf	31/03/2025 23:10:24	
7	11	CTDL>_Lab06.pdf	31/03/2025 23:10:35	
8	12	CTDL>_Lab07.pdf	31/03/2025 23:10:49	
9	13	CTDL>_Lab08.pdf	31/03/2025 23:11:01	
10	14	CTDL>_Lab09.pdf	31/03/2025 23:11:16	

Hình 7. Các tài liệu được nạp cho hệ thống

4.3. Thử nghiệm truy vấn

4.3.1. Phương thức thử nghiệm

Nhằm thử nghiệm hệ thống trợ lý ảo học tập, nhóm sử dụng 68 câu hỏi để đánh giá hệ thống và được chia thành 4 nhóm chính:

Bảng 1. Các bộ câu hỏi thử nghiệm

Bộ câu hỏi giao tiếp cơ bản	Bộ câu hỏi lý thuyết
1) Chào bạn, hãy giới thiệu bản thân và chức năng chính của bạn. 2) Chào bạn, bạn có thể giải thích một cách đơn giản cấu trúc dữ liệu là gì không? 3) Nếu tôi không hiểu bài tập về cây nhị phân, bạn sẽ giải thích lại như thế nào? 4) Bạn nghĩ gì về tầm quan trọng của giải thuật trong việc phát triển phần mềm? ...	1) Định nghĩa cấu trúc dữ liệu là gì? 2) Giải thích Quick sort. 3) So sánh ưu nhược điểm giữa mảng (array) và danh sách liên kết (linked list). 4) Mô tả đặc điểm của cây nhị phân (binary tree) và các ứng dụng của nó. 5) Giải thích khái niệm đệ quy và nêu ví dụ cụ thể. ...
Bộ câu hỏi về Code	Bộ câu hỏi về bài tập và tài liệu
1) Viết hàm đảo ngược một danh sách liên kết (Linked list). 2) Viết code cài đặt thuật toán sắp xếp nổi bọt (bubble sort). 3) Cách cài đặt thuật toán Quick Sort trong C/C++ (hoặc ngôn ngữ khác). 4) Viết hàm thực hiện tìm kiếm nhị phân trong mảng đã được sắp xếp. ...	1) Đề bài: Cho một danh sách số nguyên, viết hàm C/C++ tìm cặp số có tổng lớn nhất. 2) Đề bài: Cho đồ thị có trọng số, viết code tìm chu trình Euler (nếu tồn tại). 3) Tham khảo: Liệt kê các bài tập về cấu trúc dữ liệu và giải thuật. 4) Tham khảo: 10 bài tập về Binary Search. ...

Mỗi bộ câu hỏi được thiết kế nhằm kiểm tra các khía cạnh cụ thể trong lĩnh vực CTDL>. Cấu trúc ứng với mỗi bộ câu hỏi cụ thể được mô tả như ở Bảng 2.

Bảng 2. Cấu trúc phản hồi của từng loại câu hỏi thử nghiệm

Bộ câu hỏi giao tiếp cơ bản	Bộ câu hỏi lý thuyết
<CONCEPT></CONCEPT> <EXAMPLE></EXAMPLE> <DESCRIPTION></DESCRIPTION>	<CONCEPT></CONCEPT> <EXAMPLE></EXAMPLE> <DESCRIPTION></DESCRIPTION> <VIDEOS></VIDEOS>
Bộ câu hỏi về Code	Bộ câu hỏi về bài tập và tài liệu
<CONCEPT></CONCEPT> <EXAMPLE></EXAMPLE> <DESCRIPTION></DESCRIPTION> <IMPLEMENTATION></IMPLEMENTATION> <EXPLANATION></EXPLANATION> <VIDEOS></VIDEOS> <LAB></LAB> <COMPLEXITY></COMPLEXITY>	<CONCEPT></CONCEPT> <EXAMPLE></EXAMPLE> <DESCRIPTION></DESCRIPTION> <IMPLEMENTATION></IMPLEMENTATION> <EXPLANATION></EXPLANATION> <VIDEOS></VIDEOS>

Trong đó, bộ câu hỏi về code chứa tương đối hoàn chỉnh về mặt cấu trúc phản hồi của ứng dụng nhất

và sau khi thử nghiệm thu được kết quả như ở Hình 8:

Khái niệm

Quick Sort là một thuật toán sắp xếp phân chia hiệu quả, có độ phức tạp thời gian trung bình là $O(n \log n)$. Nó hoạt động bằng cách chọn một phần tử làm chốt và sắp xếp các phần tử khác sao cho nhỏ hơn và lớn hơn phần tử chốt.

Ví dụ

Hãy xem ví dụ với danh sách các số nguyên: [12, 5, 8, 3, 9, 1].

- Chọn phần tử chốt là 12.
- Tách danh sách thành hai phần: [5, 8, 3] và [9, 1].
- Tiếp tục áp dụng Quick Sort cho mỗi phần:
 - Đối với [5, 8, 3]:
 - Chọn 3 làm chốt.
 - Tách: [5] và [8, 3].
 - Áp dụng lại:

Thực hành

```
#include
#include

void QuickSort(int a[], int left, int right) {
    if(left >= right){
        return;
    }
    int i = left, j = right, x = a[(left + right)/2];

    while(i < j){
        while(a[i] < x){i++;}
        while(a[j] > x){j--;}

        if(i < j){
            int temp = a[i]; a[i] = a[j]; a[j] = temp;
            i++;
            j--;
        }
    }

    QuickSort(a, left, j);
    QuickSort(a, i+1, right);
}
```

Độ phức tạp

Thuật toán QuickSort có độ phức tạp thời gian $O(n \log n)$ với một pivot được chọn ngẫu nhiên mỗi lần. Độ phức tạp không phụ thuộc vào pivot.

Video:

DSA - Quick Sort, Selection Sort, Heap Sort

Tài liệu liên quan

Đồ Văn Nhơn. Giáo trình Cấu Trúc Dữ Liệu & Giải thuật.pdf (độ liên quan: 27.0%)

Hình 8. Kết quả phản hồi của bộ câu hỏi về Code (1)

Theo Hình 8, nội dung phản hồi từ trợ lý ảo được trình bày theo bố cục rõ ràng, dựa trên cấu trúc mẫu tương ứng với từng dạng câu hỏi khác nhau (trừ những phần không bắt buộc). Thêm vào đó,

trợ lý ảo có đề xuất các tài liệu, trích dẫn nội dung liên quan và hiển thị độ liên quan tương ứng để hỗ trợ người dùng dễ dàng hơn trong quá trình tìm hiểu.

4.3.2. Đánh giá các chỉ số về cấu trúc, nội dung và độ liên quan

Hệ thống cung cấp đánh giá cho từng câu hỏi dựa trên các tiêu chí: cấu trúc, nội dung, và độ liên quan theo các tiêu chuẩn đánh giá đã trình bày ở mục 3.3. Bên cạnh đó, hệ thống cũng đưa ra các tóm tắt đánh giá tổng quan nhằm hỗ trợ người dùng dễ dàng nhận định chất lượng phản hồi. Hình 9 cho thấy trợ lý ảo đưa ra kết quả với độ liên quan cao, cấu trúc phản hồi tương đồng với cấu trúc mẫu đã định nghĩa. Tuy nhiên đánh giá về nội dung vẫn còn hạn chế và cần tiến hành cải thiện thêm.

Trong thực nghiệm, nhóm đã tiến hành truy vấn 68 câu hỏi đã xây dựng với lần lượt 3 mô hình ngôn ngữ lớn nguồn mở là aya-23-8b, hermes-3-llama-3.2-3b, qwen2-vl-7b-instruct. Sau đó, nhóm tính

điểm trung bình của 68 câu truy vấn này theo các tiêu chí cấu trúc, nội dung, và độ liên quan để so sánh như ở Bảng 3.



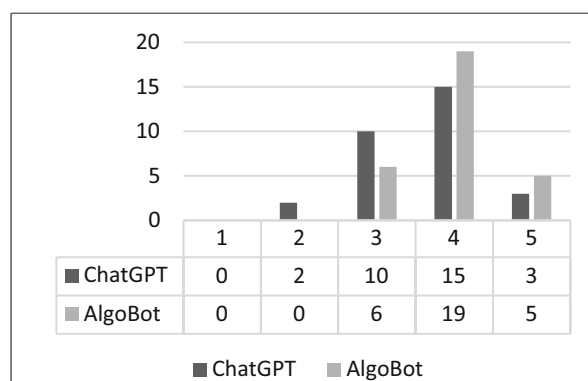
Hình 9. Kết quả phản hồi của bộ câu hỏi về Code (2)

Bảng 3. Bảng điểm so sánh ba mô hình LLM

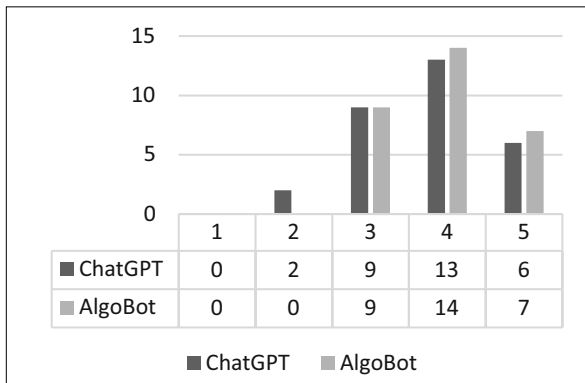
Mô hình	Cấu trúc	Nội dung	Độ liên quan	Ưu điểm	Nhược điểm
aya-23-8b	90.0	36.4	60.0	- Phản hồi đa phần có cấu trúc rõ ràng với đầy đủ thẻ đã định nghĩa. - Luôn bao gồm các phần khái niệm, ví dụ, minh họa. - Định dạng mã nguồn rõ ràng kèm giải thích.	- Chất lượng nội dung đôi khi thiếu chi tiết. - Ví dụ chưa đủ toàn diện. - Một số giải thích quá ngắn gọn.
hermes-3-llama-3.2-3b	13.75	12.0	74.0	- Độ liên quan tới câu hỏi khá tốt khi hoạt động đúng. - Bao gồm phần videos với các tài nguyên tham khảo. - Ví dụ mã nguồn thực tế.	- System prompt bị lộ trong phản hồi. - Thường xuyên bị "hallucinate". - Thiếu cấu trúc và định dạng phù hợp (không tuân thủ system prompt). - Phản hồi không nhất quán (đôi khi trống rỗng).
qwen2-vl-7b-instruct	12.85	11.5	54.0	Phản hồi trực tiếp, ngắn gọn.	- Thiếu định dạng cấu trúc, phản hồi quá ngắn gọn. - Chất lượng nội dung thấp. - Không tuân theo cấu trúc mong đợi (system prompt).

4.3.3. Đánh giá khách quan từ người dùng

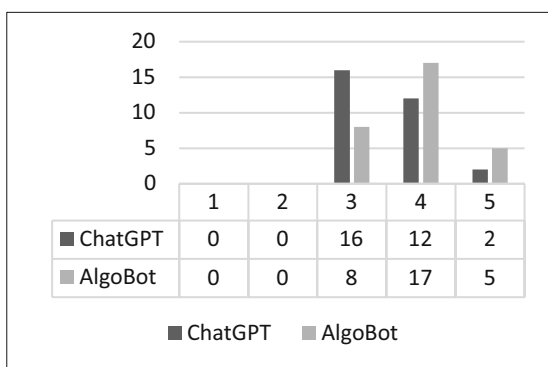
Sau khi hoàn thành ứng dụng, nhóm đã tiến hành khảo sát hiệu quả của ứng dụng trên ý kiến của người dùng để có cái nhìn khách quan về tính năng và trải nghiệm của người dùng về trợ lý ảo học tập đã xây dựng. Đối tượng khảo sát là 30 sinh viên ngành CNTT Trường Đại học Quốc tế Hồng Bàng, 10 sinh viên K24 (10.33%), 16 sinh viên K23 (16.53%), 2 sinh viên K22 (2.7%) và 2 sinh viên K21 (2.7%). Phương thức khảo sát được tiến hành thông qua GoogleForm và đánh giá theo thang điểm 5.



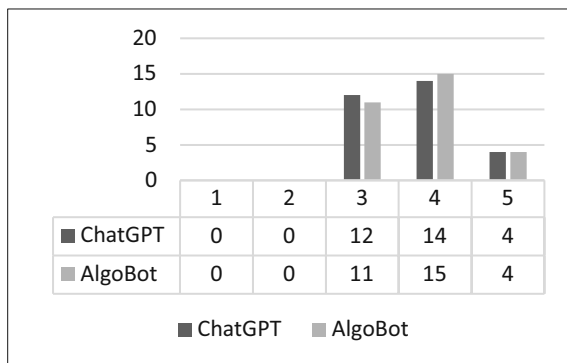
Hình 10. Đánh giá cấu trúc phản hồi giữa ChatGPT và Algotbot về kiến thức môn CTD>



Hình 11. Đánh giá về độ phù hợp với nội dung môn học CTDL> đã học trong chương trình đào tạo giữa ChatGPT và Algotbot



Hình 12. Đánh giá độ dễ hiểu của nội dung phản hồi giữa ChatGPT và Algotbot về kiến thức môn CTDL>



Hình 13. Đánh giá độ ưa thích các chức năng giữa ChatGPT và Algotbot về kiến thức môn CTDL>

Dựa trên kết quả khảo sát ở Hình 10 - 13, hệ thống trợ lý ảo học tập được xây dựng, AlgoBot, nhìn chung được đánh giá cao và có phần nhỉnh hơn ChatGPT ở một số khía cạnh đặc biệt là về cấu trúc phản hồi, độ dễ hiểu, còn về độ ưa thích thì cả hai đều khá tương đương nhau và trong đó AlgoBot giữ ổn định hơn với đa số đánh giá tích cực.

4.4. Bàn luận kết quả

Kết quả thử nghiệm cho thấy hệ thống trợ lý ảo học tập AlgoBot mang lại hiệu quả đáng kể trong việc

hỗ trợ sinh viên tiếp cận và nắm bắt kiến thức môn CTDL>. Khi so sánh với ChatGPT, AlgoBot vượt trội về cấu trúc phản hồi, độ dễ hiểu và mức độ liên quan đến nội dung giảng dạy, cho thấy việc kết hợp LLM với cơ chế đánh giá phản hồi có thể đảm bảo chất lượng câu trả lời.

Dựa trên Bảng 3, mô hình aya-23-8b đạt điểm cao nhất về cấu trúc phản hồi (90.0), phản ánh khả năng tổ chức thông tin tốt. Tuy nhiên, chất lượng nội dung còn hạn chế (36.4), có thể xuất phát từ việc kho tri thức chưa bao phủ đầy đủ các chủ đề nâng cao hoặc kỹ thuật RAG chưa được tối ưu hóa trong quá trình truy xuất tài liệu liên quan. Trong khi đó, mô hình hermes-3-llama-3.2-3b có điểm độ liên quan cao nhất (74.0), phản ánh khả năng xử lý ngữ cảnh mạnh mẽ. Tuy nhiên, mô hình này đôi khi gặp hiện tượng “ảo giác”, dẫn đến các phản hồi không chính xác, làm giảm độ tin cậy trong kết quả. Ngược lại, mô hình qwen2-vl-7b-instruct có phản hồi ngắn gọn nhưng thiếu định dạng cấu trúc và chi tiết, khiến điểm số về nội dung thấp nhất.

Kết quả khảo sát người dùng (Hình 10 - 13) củng cố những phát hiện này, cho thấy AlgoBot được đánh giá cao hơn ChatGPT về cấu trúc phản hồi, độ phù hợp với nội dung môn học và độ dễ hiểu. Đáng chú ý, sinh viên năm nhất (K24) đánh giá cao các giải thích khái niệm cơ bản, trong khi sinh viên năm cuối (K21, K22) lại ưu tiên tính năng minh họa thuật toán (Hình 6). Điều này cho thấy hệ thống cần tích hợp cơ chế học thích nghi, cho phép điều chỉnh phản hồi dựa trên trình độ và nhu cầu học tập của người dùng, nhằm tối ưu hóa trải nghiệm học tập.

Nhìn chung, hệ thống trợ lý ảo học tập tích hợp LLM đã chứng minh tiềm năng trong việc nâng cao khả năng tiếp thu kiến thức CTDL>, đặc biệt là về tính cấu trúc và độ liên quan của phản hồi. Tuy nhiên, để phát triển một công cụ học tập hoàn thiện hơn, nhóm nghiên cứu cần tiếp tục cải tiến mô hình, mở rộng bộ dữ liệu thử nghiệm, đồng thời tích hợp thêm các phương pháp đánh giá nội dung một cách toàn diện, hướng đến việc xây dựng một hệ thống học tập thông minh và thích ứng tốt hơn với từng nhóm đối tượng sinh viên.

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Nghiên cứu này trình bày việc xây dựng hệ thống trợ lý ảo học tập tích hợp LLM nhằm hỗ trợ sinh viên trong quá trình học tập và ôn luyện môn học CTDL> tại Trường Đại học Quốc tế Hồng Bàng. Hệ thống được phát triển dựa trên kiến trúc Truy hồi kết hợp sinh (Retrieval-Augmented Generation

- RAG) cùng với cơ chế đánh giá phản hồi theo ba tiêu chí chính gồm cấu trúc, nội dung và độ liên quan. Hệ thống hoạt động độc lập, không phụ thuộc vào các dịch vụ API bên ngoài, qua đó giúp giảm chi phí vận hành, nâng cao khả năng kiểm soát dữ liệu đầu vào và đầu ra, đồng thời giảm thiểu độ trễ khi xử lý các truy vấn từ người dùng. Bên cạnh chức năng chính là trả lời câu hỏi, hệ thống còn tích hợp cơ chế đánh giá phản hồi trực quan, giúp người dùng dễ dàng đánh giá chất lượng của các câu trả lời. Kết quả thử nghiệm trên tập gồm 68 câu hỏi được chia thành các nhóm câu hỏi về kiến thức cơ bản, khái niệm, lập trình, bài tập và tài liệu cho thấy hệ thống trợ lý ảo AlgoBot cung cấp phản hồi phù hợp hơn với nội dung môn học so với ChatGPT, đặc biệt thể hiện ưu điểm rõ rệt về tính trực quan, cấu trúc và mức độ dễ hiểu. Tuy nhiên, nghiên cứu cũng ghi nhận một số hạn chế còn tồn tại như sự thiếu đồng đều về chất lượng nội dung phản hồi, hiện tượng “ảo giác” (hallucination) trong một số mô hình LLM và khả năng thích nghi theo trình độ người học chưa đạt mức tối ưu. Để khắc phục những hạn chế này, nghiên cứu đề xuất các hướng phát triển cụ thể: (1) nâng cao chất lượng nội dung phản hồi thông qua việc bổ sung và cập nhật định kỳ các tài liệu chuyên sâu về CTDL> trong cơ sở dữ liệu ChromaDB, qua đó cải thiện khả năng truy xuất và giảm thiểu phản hồi thiếu chính xác; (2) kiểm soát hiện tượng “ảo giác” bằng kỹ thuật kiểm tra chéo giữa kết quả phản hồi của LLM với các tài liệu tham chiếu có sẵn trong ChromaDB sử dụng

chỉ số tương đồng ngữ nghĩa (cosine similarity), đồng thời tiến hành tinh chỉnh các mô hình nguồn mở như aya-23-8b trên tập dữ liệu chuyên biệt; (3) thực hiện khảo sát đánh giá hệ thống trên quy mô lớn hơn, tối thiểu 100 sinh viên thuộc các khóa học khác nhau trong khoảng thời gian ba tháng nhằm đánh giá hiệu quả sử dụng lâu dài và thu thập phản hồi định tính nhằm cải tiến trải nghiệm người dùng và cá nhân hóa; (4) phân tích lịch sử truy vấn, trình độ của người dùng để phân loại người dùng và điều chỉnh phản hồi (ví dụ: Giải thích chi tiết hơn cho sinh viên năm nhất, hoặc tập trung vào minh họa thuật toán cho sinh viên năm cuối); (5) áp dụng các thước đo định lượng như BLEU và ROUGE [14] để đánh giá khách quan chất lượng phản hồi, kết hợp thêm tính năng cho phép người dùng đánh giá trực tiếp độ chính xác của câu trả lời; (6) phát triển một mô hình ngôn ngữ chuyên biệt dành riêng cho môn CTDL> bằng cách huấn luyện hoặc tinh chỉnh trên tập dữ liệu học thuật chuyên sâu, qua đó nâng cao độ chính xác và chiều sâu chuyên môn của các phản hồi. Các đề xuất này sẽ được triển khai trong các giai đoạn nghiên cứu tiếp theo, nhằm xây dựng một hệ thống trợ lý ảo học tập thông minh, hiệu quả và đáp ứng tốt hơn nhu cầu học tập của sinh viên ngành Công nghệ thông tin.

LỜI CẢM ƠN

Nghiên cứu này được Trường Đại học Quốc tế Hồng Bàng cấp kinh phí thực hiện dưới mã số đề tài SVTC18.30 và đề tài GVTC18.02.

TÀI LIỆU THAM KHẢO

- [1] L. T. Hiền, “Tác động của Chatbot AI đến việc học tập của sinh viên,” 2024 [Trực tuyến]. Available: <https://digital.lib.ueh.edu.vn/handle/UEH/72395>.
- [2] S. Wollny, J. Schneider, D. Di Mitri, J. Weidlich, M. Rittberger and H. Drachsler, "Are We There Yet? - A Systematic Literature Review on Chatbots in Education," *Frontiers in Artificial Intelligence*, vol. Volume 4, July 2021.
- [3] G. Caldarini, S. Jaf and K. McGarry, "A Literature Survey of Recent Advances in Chatbots," *Information*, vol. 13, 2022.
- [4] X. Huang, "CHATBOT: Architecture, Design, & Development," in *Ph.D. Thesis*, Pennsylvania, University of Pennsylvania, School of Engineering and Applied Science, 2017.
- [5] B. A. Alazzam, M. Alkhatib and K. Shaanlan, "Artificial Intelligence Chatbots: A Survey of

- Classical versus Deep Machine Learning Techniques," 28 April 2023. [Online]. Available: <https://digitalcommons.aaru.edu.jo/isl/vol12/iss4/37>.
- [6] A. Radford and K. Narasimhan, "Improving Language Understanding by Generative Pre-Training," 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:49313245>.
- [7] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," 2019.
- [8] Meta, "Llama 2: Open Foundation and Fine-Tuned Chat Models," 2023. [Online]. Available: <https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/>.

[Accessed 30 April 2025].

[9] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," 2021.

[10] Q. T. Thợ, T. N. Oanh, B. C. Tuấn, N. V. Phương, H. N. N. Bảo, N. S. T. Long and B. H. Thắng, "Phát triển trợ lý ảo thông minh bằng mô hình ngôn ngữ lớn hỗ trợ giảng dạy," 24 May 2024. [Online]. Available: <https://tapchikhoahochongbang.vn/js/article/view/819>.

[11] B. N. Quang, "Nghiên cứu và xây dựng chatbot tư vấn người dùng trong y tế da liễu," March 2022. [Online]. [Accessed March 2025].

[12] H. N. Long, N. Đ. Thịnh, L. D. Thưởng, V. H. Đăng and T. H. Việt, "Phương pháp xây dựng ứng dụng hỗ trợ học tập dựa trên ChatGPT API," 24 May 2024. [Online]. Available: <https://tapchikhoahochongbang.vn/js/article/view/820>. [Accessed 01 May 2025].

[13] L. Richter, X. He, P. Minervini and M. J. Kusner,

"An Auditing Test To Detect Behavioral Shift in Language Models," October 2024. [Online]. Available: <https://arxiv.org/abs/2410.19406>. [Accessed 30 April 2025].

[14] Microsoft, "A list of metrics for evaluating LLM-generated content," Microsoft, 25 June 2024. [Online]. Available: <https://learn.microsoft.com/en-us/ai/playbook/technology-guidance/generative-ai/working-with-llms/evaluation/list-of-eval-metrics>.

[15] P. Wang, S. Bai, S. Tan, S. Wang, Z. Fan and J. Bai, "Qwen2-VL: Enhancing Vision-Language Model's Perception," Alibaba, 03 October 2024. [Online]. Available: <https://arxiv.org/pdf/2409.12191>.

[16] HuggingFace, "Hermes 3 - Llama-3.2 3B," Teknium, [Online]. Available: <https://huggingface.co/NousResearch/Hermes-3-Llama-3.2-3B>. [Accessed 30 April 2025].

[17] Đ. V. Nhơn và T. Q. Sơn, Giáo trình Cấu trúc dữ liệu & Giải thuật, NXB ĐHQG, TP. HCM, 2015.

Developing an LLM-Integrated virtual learning assistant for the course of Data Structures and Algorithms

Hoang Ngoc Long, Vo Hai Dang, Le Hieu Nghiem

ABSTRACT

Virtual learning assistants represent a promising application of artificial intelligence in higher education, especially when integrated with large language models (LLMs). This paper presents a development approach for an AI-powered assistant designed to support student learning in the course Data Structures and Algorithms - a foundational and compulsory subject in computer science education, essential for building algorithmic thinking and problem-solving skills. The system architecture integrates advanced LLMs with the Retrieval-Augmented Generation (RAG) technique. The study also proposes an evaluation method for the assistant based on three criteria: Structural quality, content accuracy, and contextual relevance of the responses. Experimental results indicate that the system facilitates more intuitive, flexible, and effective access to course knowledge compared to traditional study methods. This research demonstrates a feasible and practical approach for applying LLMs to the development of intelligent, personalized learning tools in higher education environments.

Keywords: data structures and algorithms, large language models, personalized learning, retrieval-augmented generation (RAG), virtual learning assistant

Received: 07/5/2025

Revised: 05/7/2025

Accepted for publication: 07/7/2025