



**ỨNG DỤNG MÔ HÌNH NGÔN NGỮ LỚN KẾT HỢP
RETRIEVAL-AUGMENTED GENERATION TRONG PHÁT TRIỂN
AI CHATBOT TƯ VẤN TUYỂN SINH: NGHIÊN CỨU TRƯỜNG HỢP TẠI
TRƯỜNG CAO ĐẲNG Y TẾ BẠC LIÊU
APPLICATION OF LARGE LANGUAGE MODELS COMBINED WITH
RETRIEVAL-AUGMENTED GENERATION IN DEVELOPING AN
AI CHATBOT FOR ADMISSION SUPPORT: A CASE STUDY AT BAC
LIEU MEDICAL COLLEGE**

Võ Mĩ Duyên*, Nguyễn Quốc Sĩ

Trường Cao đẳng Y tế Bạc Liêu

*vmduyen@cdytbaclieu.edu.vn

Ngày nhận bài:

28/5/2026

Ngày chấp nhận đăng:

07/9/2026

Keywords: AI Chatbot;
Large Language Model
(LLM); Retrieval-
Augmented Generation
(RAG); Admission
Support; Educational
Digital Transformation.

ABSTRACT

In the context of educational digital transformation, the application of artificial intelligence in admission support can help reduce the workload of admission staff and improve learners' access to information. This study develops and evaluates an AI chatbot for admission information retrieval at Bac Lieu Medical College. The system was developed on the Dify platform by integrating a Large Language Model (LLM) with Retrieval-Augmented Generation (RAG). The knowledge base was constructed from the College's official admission data, organized into information categories, and integrated with semantic retrieval to support response generation. Experimental results showed that the system operated stably, with a response time of less than two seconds and an average of 6.37 interactions per session. A post-deployment survey involving 59 users and 9 admission staff members indicated a relatively high level of acceptance regarding information accuracy, user interface, and 24/7 support capability. The findings provide empirical evidence for the feasibility of integrating LLMs, RAG, and a domain-specific admission knowledge base in developing AI chatbots for admission support at vocational education institutions.

TÓM TẮT

Trong bối cảnh chuyển đổi số giáo dục, ứng dụng trí tuệ nhân tạo trong hỗ trợ tuyển sinh góp phần giảm tải cho đội ngũ tư vấn và nâng cao khả năng tiếp cận thông tin của người học. Nghiên cứu này xây dựng và đánh giá hệ thống AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh tại Trường Cao đẳng Y tế Bạc Liêu. Hệ thống được phát triển trên nền tảng Dify, ứng dụng mô hình ngôn ngữ lớn (Large Language Model – LLM) kết hợp Retrieval-Augmented Generation (RAG). Cơ sở tri thức được xây dựng từ dữ liệu tuyển sinh chính thức của nhà trường, chuẩn hóa theo các nhóm thông tin và tích hợp cơ chế truy xuất ngữ nghĩa phục vụ

Từ khóa: AI Chatbot; Large Language Model (LLM); Retrieval-Augmented Generation (RAG); hỗ trợ tuyển sinh; chuyển đổi số giáo dục.

tạo phản hồi. Kết quả thực nghiệm cho thấy hệ thống hoạt động ổn định, thời gian phản hồi dưới 2 giây và mức độ tương tác trung bình đạt 6,37 lượt/phần. Khảo sát sau triển khai thử nghiệm trên 59 người dùng và 9 cán bộ tuyển sinh ghi nhận mức độ chấp nhận tương đối cao về độ chính xác thông tin, giao diện và khả năng hỗ trợ 24/7. Kết quả cung cấp minh chứng thực nghiệm về tính khả thi của việc kết hợp LLM, RAG và cơ sở tri thức tuyển sinh chuyên biệt trong phát triển AI Chatbot hỗ trợ tuyển sinh tại cơ sở giáo dục nghề nghiệp.

1. Giới thiệu

1.1 Đặt vấn đề

Trong bối cảnh chuyển đổi số giáo dục, nhu cầu ứng dụng trí tuệ nhân tạo (Artificial Intelligence – AI) trong hỗ trợ tuyển sinh ngày càng gia tăng. Hoạt động tư vấn tuyển sinh hiện chủ yếu được thực hiện thông qua website, fanpage hoặc đội ngũ cán bộ tuyển sinh, nhưng còn phụ thuộc vào thời gian làm việc của nhân sự và khó đáp ứng đồng thời số lượng lớn câu hỏi trong giai đoạn cao điểm. Trong khi đó, người học ngày càng có nhu cầu tiếp cận thông tin trực tuyến nhanh chóng, chính xác và liên tục.

Sự phát triển của các mô hình ngôn ngữ lớn (Large Language Model – LLM) mở ra nhiều tiềm năng trong xây dựng AI Chatbot có khả năng xử lý ngôn ngữ tự nhiên và tương tác linh hoạt với người dùng (Brown et al., 2020). Tuy nhiên, LLM còn hạn chế trong truy xuất dữ liệu chuyên biệt và có nguy cơ tạo phản hồi thiếu chính xác. Retrieval-Augmented Generation (RAG) cho phép kết hợp khả năng sinh ngôn ngữ của LLM với truy xuất dữ liệu từ cơ sở tri thức chuyên biệt, góp phần nâng cao độ chính xác và tính cập nhật của phản hồi (Gao et al., 2023; Lewis et al., 2020).

Mặc dù chatbot đã được ứng dụng trong giáo dục, các nghiên cứu triển khai AI Chatbot kết hợp LLM và RAG trong hỗ trợ tuyển sinh tại Việt Nam còn tương đối hạn chế, đặc biệt tại các cơ sở giáo dục nghề nghiệp và y tế (Lê Trung Thanh et al., 2023; Phan Thị Thanh Nga et al., 2022). Đồng thời, chưa có nhiều nghiên cứu đánh giá hiệu quả triển khai thực tế thông qua các tiêu chí như tốc độ phản hồi, khả năng duy trì tương tác và mức độ chấp nhận của người dùng. Đây là khoảng trống nghiên cứu cần được tiếp tục làm rõ.

1.2 Mục tiêu và câu hỏi nghiên cứu

Nghiên cứu này nhằm phát triển và đánh giá hệ thống AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh tại Trường Cao đẳng Y tế Bạc Liêu thông qua việc ứng dụng mô hình LLM kết hợp kỹ thuật RAG trên nền tảng Dify. Trên cơ sở khảo sát nhu cầu thực tế của người dùng và cán bộ tuyển sinh, nghiên cứu tập trung xây dựng cơ sở tri thức tuyển sinh số hóa, triển khai hệ thống AI Chatbot và đánh giá hiệu quả hoạt động của hệ thống trong môi trường thực nghiệm.

Để đạt được mục tiêu nghiên cứu, bài báo tập trung trả lời các câu hỏi nghiên cứu sau:

1. Những khó khăn và nhu cầu thực tế nào đặt ra đối với hoạt động hỗ trợ tra cứu thông tin tuyển sinh trực tuyến tại nhà trường?

2. Hệ thống AI Chatbot ứng dụng mô hình LLM kết hợp RAG có vận hành ổn định và đáp ứng tốt các yêu cầu kỹ thuật trong môi trường thực tế?

3. Mức độ chấp nhận của người dùng và cán bộ tuyển sinh đối với hệ thống AI Chatbot sau triển khai thử nghiệm được ghi nhận như thế nào?

Đối tượng nghiên cứu của đề tài là hệ thống AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh và khả năng ứng dụng mô hình LLM kết hợp RAG trong hoạt động hỗ trợ tra cứu thông tin tuyển sinh. Phạm vi nghiên cứu được triển khai với dữ liệu tuyển sinh thực tế tại Trường Cao đẳng Y tế Bạc Liêu, tập trung vào các nhóm thông tin như ngành đào tạo, phương thức xét tuyển, học phí và hồ sơ tuyển sinh.

1.3 Ý nghĩa và đóng góp nghiên cứu

Về mặt khoa học, nghiên cứu bổ sung minh chứng thực nghiệm về khả năng ứng dụng mô hình LLM kết hợp RAG trong phát triển AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh tại cơ sở giáo dục nghề nghiệp. Khác với các chatbot tuyển sinh chủ yếu dựa trên kịch bản hoặc nhận diện ý định, nghiên cứu xây dựng cơ sở tri thức từ dữ liệu tuyển sinh chính thức của nhà trường, chuẩn hóa thành các nhóm tri thức chuyên biệt và tích hợp với cơ chế truy xuất ngữ nghĩa trên nền tảng Dify. Trên cơ sở đó, hệ thống kết hợp khả năng xử lý ngôn ngữ tự nhiên của mô hình Gemini 2.5 Flash với dữ liệu tuyển sinh được truy xuất từ cơ sở tri thức để tạo phản hồi phù hợp với câu hỏi của người dùng. Nghiên cứu đồng thời đánh giá hệ thống thông qua các chỉ số vận hành và khảo sát người dùng sau giai đoạn triển khai thử nghiệm, qua đó cung cấp cơ sở thực nghiệm để xem xét tính khả thi của mô hình LLM kết hợp RAG trong môi trường tuyển sinh thực tế.

Về mặt thực tiễn, nghiên cứu đã xây dựng và triển khai được hệ thống AI Chatbot hỗ trợ tra cứu các nhóm thông tin tuyển sinh như ngành đào tạo, học phí, phương thức xét tuyển và hồ sơ đăng ký. Hệ thống có khả năng hỗ trợ tra cứu thông tin liên tục 24/7, phản hồi nhanh và xử lý các cách diễn đạt câu hỏi khác nhau, góp phần giảm tải các câu hỏi lặp lại cho đội ngũ tuyển sinh và nâng cao khả năng tiếp cận thông tin của người học.

2. Tổng quan nghiên cứu và cơ sở lý thuyết

2.1 AI Chatbot và Large Language Model (LLM) trong giáo dục

AI Chatbot đã được ứng dụng trong giáo dục nhằm hỗ trợ học tập, tư vấn và cung cấp thông tin cho người học. Trong hoạt động tuyển sinh tại Việt Nam, một số nghiên cứu đã phát triển chatbot dựa trên xử lý ngôn ngữ tự nhiên, học sâu hoặc cơ chế nhận diện ý định nhằm nâng cao khả năng tương tác và tự động hóa hoạt động tư vấn (Lê Trung Thanh et al., 2023; Phan Thị Thanh Nga et al., 2022). Tuy nhiên, các hệ thống dựa trên intent hoặc kịch bản hội thoại được xây dựng trước có thể gặp hạn chế trong việc mở rộng và cập nhật tri thức khi thông tin thay đổi.

Sự phát triển của các mô hình LLM đã mở rộng khả năng xử lý ngôn ngữ tự nhiên và tạo phản hồi linh hoạt của chatbot (Brown et al., 2020). Tuy nhiên, việc bảo đảm phản hồi dựa trên nguồn dữ liệu chuyên biệt và có thể kiểm soát vẫn là vấn đề cần được xem xét, từ đó thúc đẩy các hướng tiếp cận kết hợp LLM với cơ chế truy xuất tri thức trong các ứng dụng giáo dục (Swacha & Gracel, 2025).

2.2 Retrieval-Augmented Generation (RAG)

RAG kết hợp khả năng sinh ngôn ngữ của LLM với cơ chế truy xuất thông tin từ nguồn dữ liệu bên ngoài (Lewis et al., 2020; Gao et al., 2023). Trong kiến trúc này, dữ liệu liên quan đến truy vấn được tìm kiếm từ cơ sở tri thức và cung cấp làm ngữ cảnh cho mô hình tạo phản hồi, qua đó tăng khả năng khai thác thông tin chuyên biệt và giảm sự phụ thuộc vào tri thức có sẵn của mô hình (Izacard & Grave, 2021).

Các nghiên cứu gần đây cho thấy RAG đang được ứng dụng trong nhiều bối cảnh giáo dục như trợ lý học tập, chatbot khai thác tài liệu chuyên biệt và hệ thống hỗ trợ truy xuất thông tin tại các cơ sở giáo dục (Li et al., 2025; Swacha & Gracel, 2025). Đáng chú ý, Chen et al. (2024) đã triển khai chatbot kết hợp LLM và RAG trong hỗ trợ tuyển sinh đại học, cho thấy khả năng sử dụng cơ sở tri thức tuyển sinh để cung cấp thông tin phù hợp với truy vấn của người dùng. Bên cạnh đó, nền tảng Dify hỗ trợ tích hợp mô hình LLM, cơ sở tri thức và workflow RAG, tạo điều kiện thuận lợi cho việc triển khai hệ thống trong môi trường thực tế (Dify, 2024).

2.3 Khoảng trống nghiên cứu

Các nghiên cứu chatbot tuyển sinh tại Việt Nam đã cho thấy khả năng ứng dụng xử lý ngôn ngữ tự nhiên và học sâu trong tự động hóa hoạt động tư vấn, nhưng chủ yếu tập trung vào các hệ thống dựa trên intent hoặc mô hình hội thoại được xây dựng trước (Lê Trung Thanh et al., 2023; Phan Thị Thanh Nga et al., 2022). Trong khi đó, các nghiên cứu quốc tế gần đây đã mở rộng sang chatbot kết hợp LLM, RAG và cơ sở tri thức chuyên biệt trong giáo dục (Li et al., 2025; Swacha &

Gracel, 2025), trong đó đã có nghiên cứu ứng dụng trực tiếp vào hỗ trợ tuyển sinh đại học (Chen et al., 2024).

Tuy nhiên, trong phạm vi các tài liệu được tổng quan, còn ít minh chứng thực nghiệm về việc triển khai LLM kết hợp RAG trong hoạt động hỗ trợ tuyển sinh tại các cơ sở giáo dục nghề nghiệp ở Việt Nam, đặc biệt là các nghiên cứu đồng thời xem xét việc xây dựng cơ sở tri thức tuyển sinh chuyên biệt, khả năng vận hành của hệ thống và mức độ chấp nhận của người dùng sau triển khai. Trên cơ sở đó, nghiên cứu này xây dựng và đánh giá AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh tại Trường Cao đẳng Y tế Bạc Liêu bằng cách tích hợp cơ sở tri thức tuyển sinh với mô hình LLM theo kiến trúc RAG trên nền tảng Dify.

3. Phương pháp nghiên cứu

3.1 Thiết kế nghiên cứu

Nghiên cứu được thực hiện theo hướng nghiên cứu ứng dụng, kết hợp khảo sát thực tiễn và triển khai hệ thống thực nghiệm. Quy trình gồm ba giai đoạn: (1) khảo sát nhu cầu và khó khăn trong tra cứu thông tin tuyển sinh; (2) xây dựng hệ thống AI Chatbot ứng dụng LLM kết hợp RAG; và (3) đánh giá hiệu quả hệ thống thông qua các chỉ số vận hành và khảo sát người dùng sau triển khai thử nghiệm.

Phương pháp khảo sát bằng bảng hỏi được sử dụng nhằm thu thập dữ liệu về nhu cầu tra cứu thông tin tuyển sinh, các vấn đề thường gặp trong quá trình tư vấn và mức độ kỳ vọng đối với hệ thống AI Chatbot. Bên cạnh đó, nghiên cứu triển khai mô hình thực nghiệm trên nền tảng Dify nhằm đánh giá khả năng ứng dụng thực tế của hệ thống trong môi trường tuyển sinh.

3.2 Đối tượng và dữ liệu nghiên cứu

Đối tượng nghiên cứu bao gồm hệ thống AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh và các nhóm người tham gia khảo sát, gồm học sinh lớp 12 tại các trường THPT, sinh viên và cán bộ tuyển sinh tại Trường Cao đẳng Y tế Bạc Liêu.

Dữ liệu khảo sát được thu thập qua hai giai đoạn. Giai đoạn khảo sát nhu cầu sử dụng phương pháp chọn mẫu tự nguyện đối với nhóm người dùng, thu được 86 phiếu khảo sát, và chọn mẫu toàn bộ đối với 9 cán bộ tuyển sinh nhằm xác định nhu cầu tra cứu thông tin, những khó khăn trong hoạt động tuyển sinh và yêu cầu đối với hệ thống AI Chatbot. Giai đoạn đánh giá sau triển khai thử nghiệm áp dụng phương pháp chọn mẫu tương tự, thu được 59 phiếu từ người dùng và 9 phiếu từ cán bộ tuyển sinh nhằm đánh giá mức độ hài lòng và khả năng chấp nhận đối với hệ thống.

Nguồn dữ liệu phục vụ xây dựng hệ thống gồm các tài liệu tuyển sinh chính thức của nhà trường về ngành đào tạo, phương thức xét tuyển, học phí, hồ sơ tuyển sinh và các quy định liên quan. Các tài liệu được chuẩn hóa và chuyển đổi thành cơ sở tri thức số hóa phục vụ quá trình truy xuất thông tin của AI Chatbot.

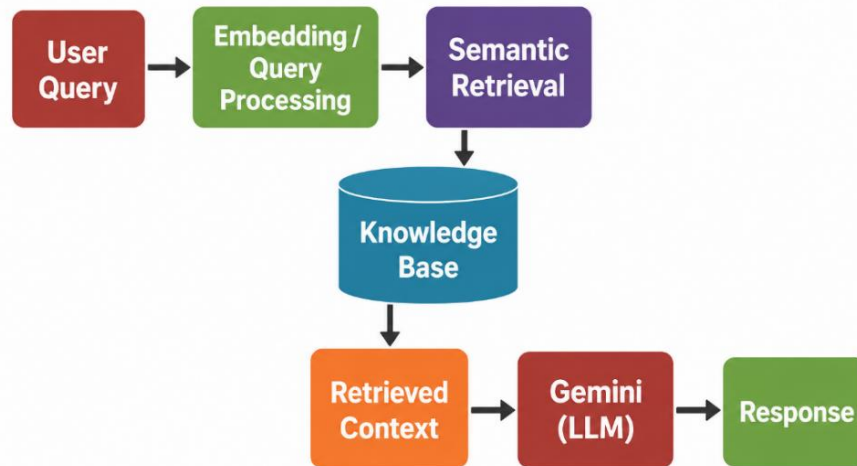
3.3 Kiến trúc và quy trình triển khai hệ thống AI Chatbot

Hệ thống AI Chatbot được triển khai trên Dify theo kiến trúc kết hợp LLM và RAG (Gao et al., 2023; Lewis et al., 2020), gồm giao diện người dùng, workflow xử lý, cơ sở tri thức tuyển sinh và mô hình Gemini 2.5 Flash. Mô hình được kết nối thông qua API từ Google AI Studio và cấu hình trên Dify (Google, 2025).

Cơ sở tri thức được xây dựng từ tài liệu tuyển sinh chính thức của Trường Cao đẳng Y tế Bạc Liêu, chuẩn hóa thành 10 tệp định dạng Markdown (.md) theo các nhóm thông tin về hệ đào tạo, ngành đào tạo và hỗ trợ tuyển sinh. Cách tổ chức này nhằm giảm trùng lặp và hỗ trợ truy xuất thông tin phù hợp với câu hỏi của người dùng.

Trong quá trình xây dựng hệ thống, dữ liệu văn bản được biểu diễn dưới dạng vector embedding để phục vụ truy xuất dựa trên độ tương đồng ngữ nghĩa. Nghiên cứu tiến hành thử nghiệm các phương thức truy xuất gồm keyword search, vector search và hybrid search; đồng thời điều chỉnh việc phân đoạn dữ liệu (chunking), truy vấn tái cấu trúc (reconstructed query) và số lượng kết quả truy xuất (Top-K) nhằm cải thiện mức độ phù hợp của dữ liệu được truy xuất.

Khi người dùng gửi câu hỏi, workflow xử lý truy vấn và truy xuất dữ liệu liên quan từ cơ sở tri thức. Nội dung truy xuất được cung cấp làm ngữ cảnh cho Gemini 2.5 Flash tạo phản hồi và trả kết quả về giao diện chatbot. Quy trình này kết hợp khả năng xử lý ngôn ngữ tự nhiên của LLM với dữ liệu tuyển sinh chính thức của nhà trường.



Hình 1. Kiến trúc hệ thống AI Chatbot ứng dụng LLM kết hợp RAG

Hình 1 mô tả luồng xử lý truy vấn từ người dùng đến quá trình truy xuất dữ liệu trong cơ sở tri thức và sinh phản hồi bằng mô hình LLM. Kiến trúc này cho phép nội dung phản hồi được tạo trên cơ sở ngữ cảnh truy xuất từ nguồn dữ liệu tuyển sinh của nhà trường thay vì chỉ phụ thuộc vào tri thức có sẵn của mô hình.

3.4 Phương pháp đánh giá hệ thống

Hiệu quả hoạt động của hệ thống được đánh giá theo hai nhóm: (1) các chỉ số vận hành kỹ thuật và (2) mức độ chấp nhận của người dùng. Các chỉ số vận hành được ghi nhận từ chức năng Monitoring và dữ liệu log của hệ thống, tập trung vào thời gian phản hồi, trạng thái xử lý và số lượt tương tác trung bình trên mỗi phiên (Average Session Interactions). Trong đó, thời gian phản hồi được xác định dựa trên khoảng chênh lệch giữa thời điểm tạo truy vấn (Created Time) và thời điểm hoàn tất xử lý (Updated Time); trạng thái xử lý được sử dụng để theo dõi tính ổn định trong quá trình vận hành; và số lượt tương tác trung bình trên mỗi phiên phản ánh khả năng duy trì hội thoại của chatbot.

Bên cạnh đánh giá kỹ thuật, nghiên cứu thực hiện khảo sát định lượng sau triển khai thử nghiệm đối với nhóm người dùng và cán bộ tuyển sinh. Các tiêu chí đánh giá gồm độ chính xác và phù hợp của thông tin, khả năng hiểu câu hỏi, tốc độ phản hồi, giao diện tương tác và khả năng hỗ trợ liên tục 24/7. Các tiêu chí được đánh giá theo thang đo Likert và tổng hợp theo phân bố các mức đánh giá nhằm xem xét mức độ hài lòng và khả năng chấp nhận hệ thống trong môi trường tuyển sinh thực tế.

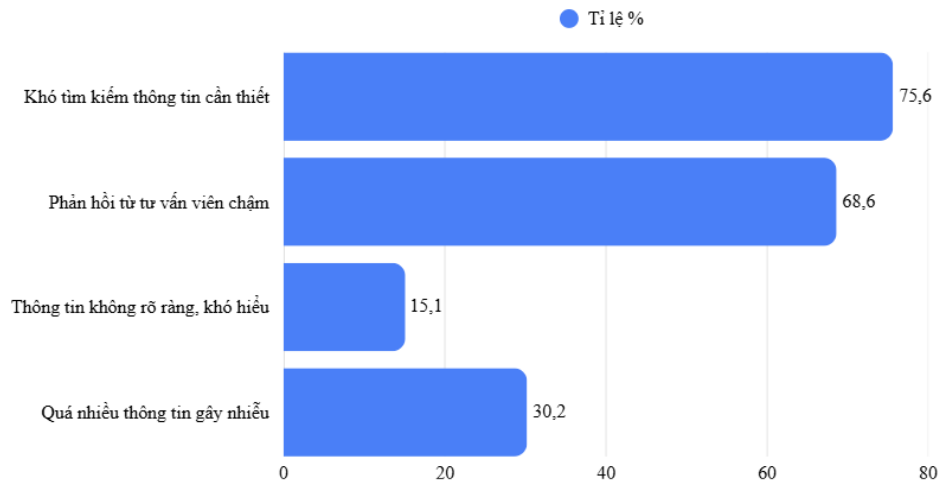
4. Kết quả nghiên cứu và thảo luận

4.1 Kết quả khảo sát nhu cầu và khó khăn trong hoạt động tuyển sinh

Kết quả khảo sát cho thấy người học có nhu cầu tra cứu thông tin tuyển sinh trực tuyến, chủ yếu thông qua website, fanpage và mạng xã hội để tìm kiếm thông tin về ngành đào tạo, phương thức xét tuyển, học phí và hồ sơ tuyển sinh. Kết quả này phản ánh nhu cầu tiếp cận thông tin tuyển sinh qua các kênh trực tuyến.

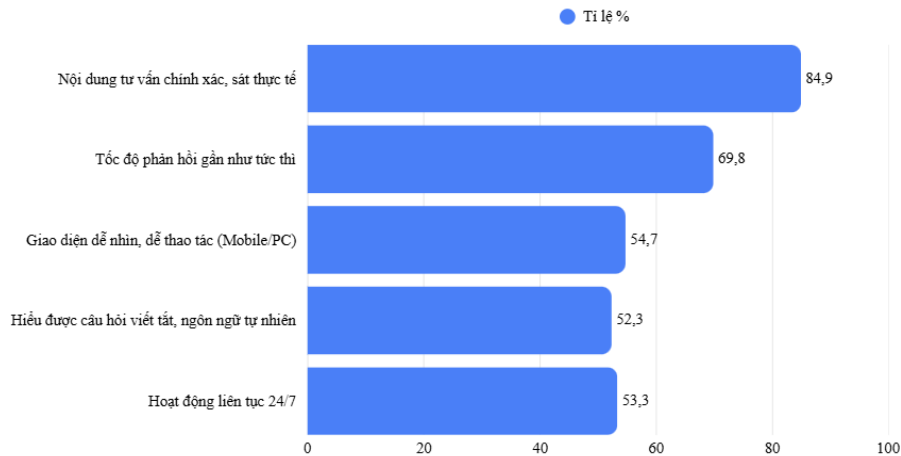
Kết quả ở Biểu đồ 1 cho thấy những khó khăn nổi bật trong quá trình tra cứu thông tin tuyển sinh là khó tìm kiếm thông tin cần thiết (75,6%) và phản hồi từ tư vấn viên còn chậm (68,6%). Đối với cán bộ tuyển sinh, áp lực tư vấn tăng trong giai đoạn cao điểm, đặc biệt với các câu hỏi lặp lại về ngành đào tạo, điều kiện xét tuyển, học phí và hồ sơ đăng ký. Phần lớn cán bộ

được khảo sát đánh giá việc ứng dụng AI Chatbot là cần thiết để hỗ trợ tra cứu thông tin và giảm tải công tác tư vấn.



Biểu đồ 1. Những khó khăn trong hoạt động tra cứu thông tin tuyển sinh

Ngoài những khó khăn trong quá trình tra cứu, người dùng cũng đặt ra các yêu cầu đối với hệ thống AI Chatbot. Kết quả ở Biểu đồ 2 cho thấy người dùng quan tâm nhiều nhất đến nội dung tư vấn chính xác, sát thực tế (84,9%) và tốc độ phản hồi gần như tức thì (69,8%); đồng thời quan tâm đến giao diện dễ sử dụng, khả năng hỗ trợ 24/7 và hiểu câu hỏi viết tắt, ngôn ngữ tự nhiên.



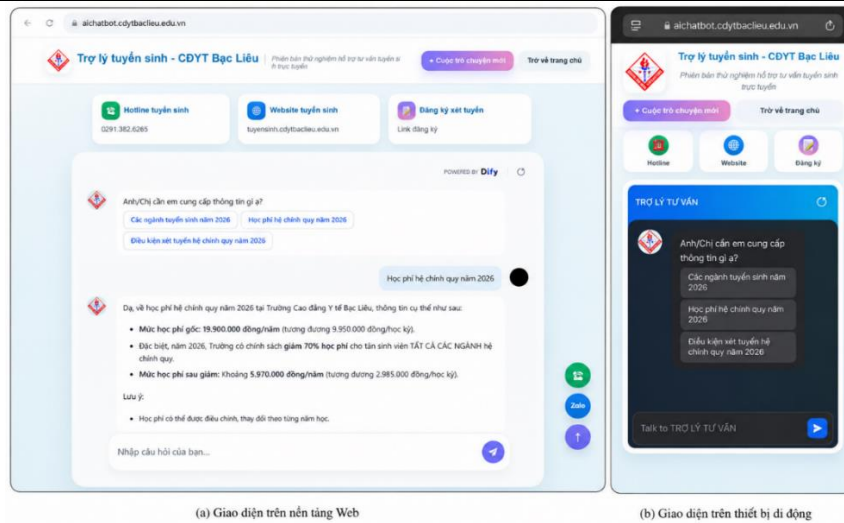
Biểu đồ 2. Nhu cầu đối với chức năng của hệ thống AI Chatbot

Các kết quả trên là cơ sở xác định yêu cầu đối với hệ thống và góp phần trả lời câu hỏi nghiên cứu thứ nhất.

4.2 Kết quả triển khai hệ thống AI Chatbot ứng dụng LLM kết hợp RAG

Hệ thống AI Chatbot được triển khai trên nền tảng Dify với kiến trúc kết hợp LLM và RAG, hỗ trợ tra cứu các nhóm thông tin chính gồm ngành đào tạo, phương thức xét tuyển, học phí, hồ sơ đăng ký và các quy định tuyển sinh liên quan.

Cơ sở tri thức được xây dựng từ tài liệu tuyển sinh chính thức của nhà trường, chuẩn hóa thành 10 tệp chuyên biệt định dạng Markdown (.md) và tích hợp vào hệ thống. Kết quả triển khai cho thấy chatbot có khả năng khai thác cơ sở tri thức để xử lý các câu hỏi tuyển sinh phổ biến và một số cách diễn đạt khác nhau, kể cả câu hỏi viết tắt hoặc chưa đầy đủ.



Hình 2. Giao diện hệ thống AI Chatbot hỗ trợ tuyển sinh trên các thiết bị

Kết quả triển khai cho thấy LLM kết hợp RAG có thể tích hợp với cơ sở tri thức tuyển sinh chuyên biệt và vận hành trong môi trường thực tế của nhà trường. Việc sử dụng dữ liệu tuyển sinh chính thức làm nguồn truy xuất giúp phản hồi được tạo dựa trên ngữ cảnh từ cơ sở tri thức, góp phần kiểm soát nguồn thông tin trong quá trình sinh phản hồi. Hiệu năng vận hành và mức độ chấp nhận của người dùng được trình bày ở mục tiếp theo.

4.3 Đánh giá hiệu quả hoạt động và mức độ chấp nhận của người dùng

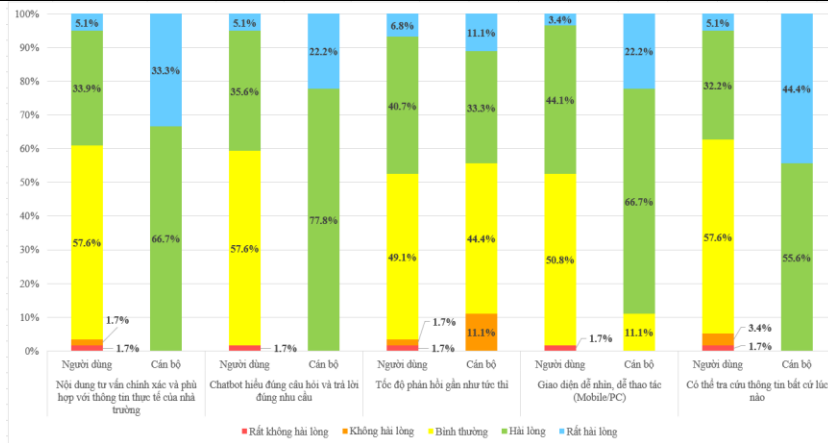
Kết quả theo dõi trong giai đoạn triển khai thử nghiệm cho thấy hệ thống AI Chatbot duy trì hoạt động ổn định đối với các truy vấn được ghi nhận. Dữ liệu log cho thấy thời gian phản hồi của hệ thống đối với các nhóm câu hỏi tuyển sinh phổ biến ở mức dưới 2 giây. Các phiên hội thoại được theo dõi có trạng thái xử lý thành công (Success), không ghi nhận lỗi xử lý đáng kể trong giai đoạn thử nghiệm. Đồng thời, số lượt tương tác trung bình trên mỗi phiên (Average Session Interactions) đạt 6,37 lượt/phần, cho thấy người dùng có xu hướng tiếp tục trao đổi với chatbot trong cùng một phiên hội thoại.

Bảng 1. Kết quả đánh giá hiệu năng hệ thống AI Chatbot

Tiêu chí đánh giá	Kết quả
Thời gian phản hồi của hệ thống	< 2 giây
Trạng thái xử lý	Success
Average Session Interactions	6,37 lượt/phần

Nguồn: Kết quả phân tích dữ liệu của tác giả

Bên cạnh các chỉ số vận hành, khảo sát định lượng sau triển khai thử nghiệm được thực hiện với 59 người dùng theo phương pháp chọn mẫu tự nguyện và 9 cán bộ tuyển sinh theo phương pháp chọn mẫu toàn bộ nhằm đánh giá mức độ chấp nhận đối với hệ thống. Kết quả khảo sát được trình bày tại Biểu đồ 3.



Biểu đồ 3. Kết quả đánh giá mức độ chấp nhận hệ thống AI Chatbot của người dùng và cán bộ tuyển sinh

Kết quả ở Biểu đồ 3 cho thấy tỷ lệ đánh giá từ mức Bình thường trở lên tương đối cao ở cả hai nhóm khảo sát. Đối với tiêu chí “Nội dung tư vấn chính xác và phù hợp với thông tin thực tế của nhà trường”, tỷ lệ này đạt 96,6% ở nhóm người dùng và 100% ở nhóm cán bộ tuyển sinh. Ở nhóm người dùng, các tiêu chí về tốc độ phản hồi và khả năng hỗ trợ 24/7 lần lượt đạt 96,6% và 94,9%.

Kết quả từ các chỉ số vận hành và khảo sát sau triển khai cho thấy hệ thống đáp ứng các yêu cầu về tốc độ phản hồi, khả năng duy trì tương tác và hỗ trợ tra cứu thông tin tuyển sinh trong môi trường thử nghiệm. Kết quả khảo sát đồng thời ghi nhận mức độ chấp nhận tương đối cao của người dùng và cán bộ tuyển sinh đối với hệ thống. Các kết quả này cung cấp bằng chứng thực nghiệm ban đầu về tính khả thi của AI Chatbot trong hoạt động hỗ trợ tuyển sinh, đồng thời góp phần trả lời câu hỏi nghiên cứu thứ hai về khả năng vận hành và đáp ứng yêu cầu kỹ thuật, cũng như câu hỏi nghiên cứu thứ ba về mức độ chấp nhận của người dùng sau triển khai thử nghiệm.

4.4 Thảo luận

Kết quả nghiên cứu cho thấy việc tích hợp LLM với RAG và cơ sở tri thức tuyển sinh chuyên biệt có tính khả thi trong môi trường tuyển sinh của một cơ sở giáo dục nghề nghiệp. Kết quả này phù hợp với xu hướng ứng dụng RAG trong giáo dục được ghi nhận bởi Li et al. (2025) và Swacha và Gracel (2025), trong đó cơ sở tri thức chuyên biệt và cơ chế truy xuất đóng vai trò quan trọng trong việc cung cấp ngữ cảnh cho mô hình tạo phản hồi. Đặc biệt, kết quả có sự tương đồng với Chen et al. (2024) về khả năng ứng dụng LLM kết hợp RAG trong hỗ trợ tuyển sinh. Tuy nhiên, nghiên cứu này được triển khai trong bối cảnh một cơ sở giáo dục nghề nghiệp tại Việt Nam, với cơ sở tri thức được xây dựng từ dữ liệu tuyển sinh thực tế của nhà trường và hệ thống được đánh giá đồng thời thông qua các chỉ số vận hành và khảo sát người dùng sau triển khai thử nghiệm.

Các chỉ số vận hành ghi nhận thời gian phản hồi dưới 2 giây, trạng thái xử lý thành công và trung bình 6,37 lượt tương tác mỗi phiên; đồng thời, khảo sát cho thấy hệ thống nhận được mức độ chấp nhận tương đối cao từ người dùng và cán bộ tuyển sinh. Những kết quả này bổ sung minh chứng thực nghiệm về khả năng triển khai LLM kết hợp RAG trong hỗ trợ tuyển sinh tại cơ sở giáo dục nghề nghiệp. Tuy nhiên, kết quả cần được xem xét trong phạm vi nghiên cứu thử nghiệm tại một cơ sở; việc đánh giá trên quy mô lớn hơn và tại nhiều cơ sở giáo dục là cần thiết để kiểm chứng thêm tính ổn định và khả năng khái quát hóa của mô hình.

5. Kết luận

Nghiên cứu đã xây dựng và đánh giá hệ thống AI Chatbot hỗ trợ tra cứu thông tin tuyển sinh tại Trường Cao đẳng Y tế Bạc Liêu trên nền tảng Dify, ứng dụng LLM kết hợp RAG và cơ sở tri thức tuyển sinh chuyên biệt. Hệ thống có khả năng khai thác dữ liệu tuyển sinh chính thức

của nhà trường để hỗ trợ các câu hỏi phổ biến về ngành đào tạo, phương thức xét tuyển, học phí và hồ sơ đăng ký.

Trong giai đoạn triển khai thử nghiệm, hệ thống ghi nhận thời gian phản hồi dưới 2 giây, trạng thái xử lý thành công (Success) và trung bình 6,37 lượt tương tác/phần. Khảo sát sau triển khai ghi nhận tỷ lệ đánh giá từ mức Bình thường trở lên tương đối cao về độ chính xác và phù hợp của thông tin, tốc độ phản hồi và khả năng hỗ trợ 24/7. Các kết quả này cung cấp bằng chứng thực nghiệm ban đầu về tính khả thi và mức độ chấp nhận của AI Chatbot ứng dụng LLM kết hợp RAG trong hỗ trợ tuyển sinh tại cơ sở giáo dục nghề nghiệp.

Về mặt khoa học, nghiên cứu bổ sung minh chứng thực nghiệm về việc triển khai LLM kết hợp RAG trên cơ sở tri thức tuyển sinh chuyên biệt trong bối cảnh giáo dục nghề nghiệp. Về mặt thực tiễn, hệ thống cho thấy khả năng hỗ trợ tra cứu thông tin tuyển sinh trực tuyến, cung cấp thông tin liên tục và hỗ trợ giảm tải các câu hỏi lặp lại cho đội ngũ tuyển sinh.

Nghiên cứu vẫn còn một số hạn chế. Giai đoạn đánh giá sau triển khai thử nghiệm được thực hiện với 59 người dùng theo phương pháp chọn mẫu tự nguyện và 9 cán bộ tuyển sinh theo phương pháp chọn mẫu toàn bộ. Do quy mô mẫu còn hạn chế và nghiên cứu được thực hiện tại một cơ sở giáo dục nghề nghiệp, kết quả chủ yếu phản ánh tính khả thi của hệ thống trong bối cảnh triển khai tại Trường Cao đẳng Y tế Bạc Liêu và chưa đủ cơ sở để khái quát cho các cơ sở giáo dục khác. Bên cạnh đó, hệ thống hiện tập trung vào các nhóm thông tin tuyển sinh cơ bản và hiệu quả phản hồi phụ thuộc vào chất lượng cũng như mức độ cập nhật của cơ sở tri thức. Các nghiên cứu tiếp theo cần mở rộng quy mô và phạm vi mẫu, triển khai tại nhiều cơ sở giáo dục và bổ sung các phương pháp đánh giá kỹ thuật chuyên sâu hơn để kiểm chứng tính ổn định, độ chính xác và khả năng khái quát hóa của mô hình.

TÀI LIỆU THAM KHẢO

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Chen, Z., Zou, D., Xie, H., Lou, H., & Pang, Z. (2024). Facilitating university admission using a chatbot based on large language models with retrieval-augmented generation. *Educational Technology & Society*, 27(4), 454–470. [https://doi.org/10.30191/ETS.202410_27\(4\).TP02](https://doi.org/10.30191/ETS.202410_27(4).TP02)
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). *Retrieval-augmented generation for large language models: A survey*. arXiv. <https://doi.org/10.48550/arXiv.2312.10997>
- Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open domain question answering. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 874–880. <https://doi.org/10.18653/v1/2021.eacl-main.74>
- Lê Trung Thanh, Nguyễn Văn Hòa, & Trần Minh Tuấn. (2023). Xây dựng chatbot tư vấn tuyển sinh tiếng Việt ứng dụng học sâu trên nền tảng Rasa. *Tạp chí Khoa học và Công nghệ*, 61(4), 55–63.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-augmented generation for educational application: A systematic survey. *Computers and Education: Artificial Intelligence*, 8, 100417. <https://doi.org/10.1016/j.caeai.2025.100417>

- Phan Thị Thanh Nga, Nguyễn Văn Tâm, & Lê Hoàng Phúc. (2022). Ứng dụng mô hình BERT trong xây dựng chatbot tư vấn tuyển sinh. *Tạp chí Công nghệ Thông tin và Truyền thông*, 18(2), 34–42.
- Swacha, J., & Gracel, M. (2025). Retrieval-Augmented Generation (RAG) chatbots for education: A survey of applications. *Applied Sciences*, 15(8), 4234. <https://doi.org/10.3390/app15084234>
- Dify. (2024). *Dify automation workflow and knowledge base documentation*. Dify Open Source Project. <https://docs.dify.ai>
- Google. (2025). *Gemini 2.5 Flash technical overview and API reference*. Google AI Studio. <https://ai.google.dev/gemini-api/docs>