

RESEARCH ON BUILDING AN INTELLIGENT CHATBOT TO SUPPORT LEARNING FOR HIGH SCHOOL STUDENTS

NGHIÊN CỨU XÂY DỰNG CHATBOT THÔNG MINH HỖ TRỢ HỌC TẬP CHO HỌC SINH TRUNG HỌC PHỔ THÔNG

Hoàng Đình Tuyền

Trường Đại học Quảng Bình

ABSTRACT: In the context of rapid digital transformation in education, intelligent learning support systems have become an inevitable trend to enhance learners' self-study capabilities, personalization, and interaction levels. This paper proposes an intelligent learning chatbot model that applies natural language processing, deep learning, and knowledge retrieval techniques to help learners access educational content quickly and in a manner tailored to their individual needs. The system architecture includes four main components: data preprocessing, a language model, a knowledge base, and an adaptive response generation module. In addition to providing intelligent conversational functions, the chatbot is designed for flexible integration into social platforms such as Zalo, thereby expanding accessibility, increasing interaction, and supporting learning anytime, anywhere. Experimental results on the high school Chemistry curriculum under the 2018 General Education Program demonstrate that the model achieves high performance in response accuracy and user satisfaction, confirming its strong potential for widespread application in the current digital education landscape.

Keywords: Educational Chatbot, Intelligent Chatbot, AI-Assisted Learning, Chatbot Zalo.

TÓM TẮT: Trong bối cảnh chuyển đổi số giáo dục diễn ra mạnh mẽ, các hệ thống hỗ trợ học tập thông minh trở thành xu hướng tất yếu nhằm nâng cao khả năng tự học, cá nhân hoá và mức độ tương tác của người học. Bài báo đề xuất một mô hình chatbot học tập thông minh ứng dụng các kỹ thuật xử lý ngôn ngữ tự nhiên, học sâu và truy vấn tri thức, giúp người học tiếp cận nội dung giáo dục nhanh chóng và phù hợp với nhu cầu cá nhân. Kiến trúc hệ thống bao gồm các thành phần chính: tiền xử lý dữ liệu, mô hình ngôn ngữ, cơ sở tri thức và bộ sinh phản hồi thích ứng. Bên cạnh chức năng hội thoại thông minh, chatbot còn được thiết kế để tích hợp linh hoạt vào các nền tảng mạng xã hội như Zalo, từ đó mở rộng khả năng tiếp cận, tăng cường tương tác và hỗ trợ học tập mọi lúc, mọi nơi. Kết quả thực nghiệm trên môn Hóa học THPT theo chương trình GDPT 2018 cho thấy mô hình đạt hiệu quả cao về độ chính xác phản hồi và mức độ hài lòng của người dùng, khẳng định tiềm năng ứng dụng rộng rãi trong bối cảnh giáo dục số hiện nay.

Từ khóa: Chatbot giáo dục, Chatbot thông minh, AI hỗ trợ học tập, Chatbot Zalo.

1. ĐẶT VẤN ĐỀ

Trong kỷ nguyên chuyển đổi số mạnh mẽ hiện nay, giáo dục phổ thông đang chứng kiến sự thay đổi sâu rộng cả về chương trình, phương pháp giảng dạy và môi trường học tập. Sự phát triển nhanh chóng của các công nghệ như trí tuệ nhân tạo (AI), học máy (Machine Learning), xử

lý ngôn ngữ tự nhiên (NLP) và điện toán đám mây đã mở ra nhiều cơ hội để cải thiện chất lượng dạy và học [7]. Đặc biệt, đối tượng học sinh trung học phổ thông (THPT) - nhóm đang trong giai đoạn hoàn thiện tư duy logic, năng lực tự học và định hướng nghề nghiệp - rất cần các công cụ học tập linh hoạt, cá nhân hóa và sẵn sàng

hỗ trợ mọi lúc, mọi nơi.

Trong thực tế, nhiều học sinh THPT gặp khó khăn khi tiếp cận kiến thức ở các môn học đòi hỏi tư duy như Toán, Vật lý, Hóa học hoặc trong việc chuẩn bị cho các kỳ thi quan trọng. Các thách thức phổ biến bao gồm: (1) thiếu nguồn tài liệu phù hợp, (2) không có người hướng dẫn kịp thời khi gặp bài khó, (3) áp lực học tập trong môi trường cạnh tranh và (4) hạn chế trong kỹ năng tự học. Các nền tảng học trực tuyến và mạng xã hội tuy mang lại lượng thông tin phong phú nhưng lại thiếu tính chọn lọc, hệ thống và khả năng tương tác sâu.

Trong bối cảnh đó, chatbot thông minh nổi lên như một giải pháp tiềm năng giúp hỗ trợ cá nhân hóa quá trình học tập [1], [8], [9]. Chatbot không chỉ có khả năng trả lời câu hỏi tự động, mà còn có thể gợi ý bài học phù hợp, hỗ trợ giải bài tập, nhắc nhở lịch học, tạo môi trường giao tiếp thân thiện và cung cấp phản hồi theo thời gian thực [5]. Tuy nhiên, để chatbot thực sự có hiệu quả trong môi trường giáo dục THPT Việt Nam, cần xây dựng hệ thống dựa trên cơ sở dữ liệu giáo dục chính xác, phù hợp chương trình GDPT 2018, đồng thời đảm bảo ngôn ngữ tiếng Việt được xử lý hiệu quả.

Mặc dù đã có nhiều nghiên cứu về chatbot giáo dục trên thế giới, số lượng các hệ thống tập trung hỗ trợ riêng cho học sinh THPT Việt Nam còn hạn chế [9], [2]. Phần lớn chatbot hiện nay chỉ dừng lại ở mức trả lời câu hỏi cơ bản, chưa khai thác đầy đủ sức mạnh của các mô hình NLP hiện đại như Retrieval-Augmented Generation (RAG), GPT hoặc các mô hình sinh ngôn ngữ đa tầng. Chính vì vậy, bài báo này tập trung đề xuất một mô hình chatbot thông minh dựa trên kiến trúc kết hợp (Hybrid Model), nhằm tối ưu khả năng hiểu ngôn

ngữ tự nhiên của học sinh, tăng độ chính xác trong trả lời câu hỏi và cải thiện hiệu quả học tập [8].

Nghiên cứu hướng đến ba mục tiêu chính: (1) phân tích nhu cầu thực tế của học sinh THPT trong quá trình học tập, (2) xây dựng mô hình chatbot thông minh sử dụng dữ liệu và kiến trúc phù hợp và (3) đánh giá hiệu quả hệ thống thông qua thực nghiệm với người dùng thực tế. Đây sẽ là cơ sở quan trọng để phát triển các nền tảng đào tạo trực tuyến thông minh hơn trong tương lai.

2. CÁC NGHIÊN CỨU LIÊN QUAN

Chatbot xuất hiện từ những năm 1960 với các hệ thống rule-based điển hình như ELIZA [7]. Qua thời gian, kiến trúc chatbot phát triển theo ba hướng chính:

Rule-based: Dựa trên luật cố định (pattern matching). Ưu điểm là điều khiển được ngữ cảnh và trả lời chính xác trong phạm vi hẹp; nhược điểm là thiếu khả năng mở rộng và khó xử lý ngôn ngữ tự nhiên phong phú.

Retrieval-based: Tìm kiếm và chọn câu trả lời phù hợp từ kho câu trả lời có sẵn bằng các kỹ thuật so khớp tri thức (similarity matching, BM25, dense retrieval). Phù hợp cho các ứng dụng cần độ chính xác cao và kiểm soát nội dung.

Generative-based: Dùng mô hình sinh (seq2seq, Transformer, GPT) để tạo câu trả lời mới. Ưu điểm là linh hoạt, có thể giải thích và mở rộng; nhược điểm là dễ sinh thông tin sai lệch nếu không có cơ chế kiểm soát.

Hybrid (Kết hợp): Kết hợp retrieval và generative, hoặc kết hợp rule với mô hình học sâu để vừa đảm bảo tính chính xác vừa có khả năng sinh ngôn ngữ tự nhiên, điển hình nhất là các phương pháp Retrieval-Augmented Generation (RAG).

Nhiều nghiên cứu trong và ngoài nước đã khảo sát vai trò của chatbot trong các nhiệm vụ giáo dục [1], [3], [8]:

Trợ giảng và hỗ trợ giải bài tập: Chatbot có thể cung cấp lời giải từng bước, gợi ý hướng giải và kiểm tra hiểu biết của học sinh thông qua câu hỏi kiểm tra. Các nghiên cứu chỉ ra rằng tương tác tức thời giúp cải thiện động lực và kết quả học tập [9].

Học tập cá nhân hóa (personalized tutoring): Bằng cách theo dõi lịch sử tương tác và điểm yếu của học sinh, chatbot có thể gợi ý bài tập phù hợp và lộ trình học tập thích ứng.

Đánh giá và phản hồi tức thì: Chatbot có thể chấm bài tự động các dạng bài trắc nghiệm và một số bài tự luận ngắn, cung cấp phản hồi phản xạ nhanh.

Hỗ trợ hành vi và động lực học tập: Thông qua nhắc nhở lịch học, mục tiêu và ghi nhận tiến trình, chatbot giúp học sinh duy trì thói quen học tập.

Các nghiên cứu gần đây đã cho thấy sự hiệu quả của Chatbot [4], [5], [6], [10]. Tuy nhiên, vẫn còn một số điểm cần nghiên cứu sau thêm:

+ Thiếu hệ thống tối ưu cho tiếng Việt và chương trình THPT Việt Nam: Hầu hết mô hình gốc được huấn luyện trên dữ liệu tiếng Anh; nghiên cứu chuyên sâu cho ngôn ngữ Việt còn hạn chế.

+ Kiểm soát chất lượng thông tin (fact-checking): Mô hình sinh vẫn có nguy cơ trả lời sai thông tin chuyên môn nếu không có cơ chế truy xuất tri thức đáng tin cậy.

+ Tương tác đa phương thức chưa phổ biến: Ít hệ thống hỗ trợ giải thích bằng đồ thị, hình vẽ hoặc video cho bài toán vật lý, hoá học.

+ Vấn đề đạo đức và quyền riêng tư:

Thu thập dữ liệu học sinh, lưu trữ và phân tích lịch sử học tập đặt ra thách thức về bảo mật và chính sách.

3. ĐỀ XUẤT MÔ HÌNH GIẢI PHÁP

3.1. Kiến trúc tổng thể hệ thống

Hệ thống chatbot AI hỗ trợ học tập được thiết kế theo kiến trúc nhiều tầng, gồm bốn thành phần chính:

(1) Lớp giao tiếp người dùng tích hợp với nền tảng mạng xã hội;

(2) Lớp xử lý ngôn ngữ tự nhiên (NLP) dùng để phân tích và hiểu yêu cầu học sinh;

(3) Lớp mô hình ngôn ngữ chịu trách nhiệm sinh câu trả lời và gợi ý học tập;

(4) Lớp quản lý tri thức và dịch vụ học tập cung cấp kho dữ liệu môn học, ngân hàng câu hỏi và hệ thống đánh giá.

Kiến trúc này đảm bảo khả năng mở rộng, tái sử dụng dữ liệu và tích hợp linh hoạt với nhiều nền tảng số.

3.2. Mô hình toán học tổng quát

Bài toán chatbot được mô hình hoá như một hệ thống ánh xạ từ câu đầu vào x (phát biểu của học sinh) sang câu trả lời y .

Hệ thống gồm hai hàm chính:

(1) Hàm hiểu ngôn ngữ

$$z = f_{NLU}(x) \quad (1)$$

Trong đó z là biểu diễn ngữ nghĩa của câu truy vấn sau khi xử lý tokenization, embedding và phân loại ý định. $f_{NLU}(x)$ hàm ánh xạ từ câu tự nhiên sang không gian ngữ nghĩa.

(2) Hàm sinh phản hồi

$$y = f_{LLM}(z, K) \quad (2)$$

Trong đó K là tập tri thức môn học (Knowledge Base). $f_{LLM}(z, K)$ là mô hình ngôn ngữ lớn tạo ra câu trả lời dựa trên z và tri thức K .

Hệ thống sử dụng cơ chế retrieval-

augmented generation (RAG) để trích xuất tri thức liên quan:

$$K' = \text{Retrieval}(z, K) \quad (3)$$

$$y = f_{LLM}(z, K') \quad (4)$$

Mô hình toán học này giúp chatbot vừa khai thác tri thức chuẩn hoá vừa bảo đảm khả năng sinh ngôn ngữ linh hoạt.

3.3. Thuật toán xử lý truy vấn

Quy trình xử lý được mô tả ngắn gọn như sau:

Thuật toán 1. Xử lý và phản hồi truy vấn học tập

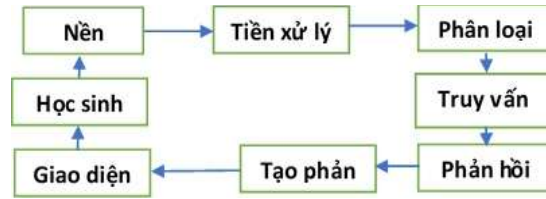
1. *Input: câu hỏi x từ học sinh.*
2. *Thực hiện tiền xử lý: chuẩn hoá văn bản, tách từ, tạo embedding.*
3. *Xác định ý định và loại câu hỏi bằng mô hình phân loại f_{NLU} .*
4. *Truy xuất tri thức liên quan: $K' = \text{Retrieval}(z, K)$.*
5. *Sinh phản hồi tối ưu bằng mô hình ngôn ngữ: $y = f_{LLM}(z, K')$.*
6. *Xuất kết quả sang giao diện mạng xã hội.*
7. *Output: câu trả lời y.*

Thuật toán được tối ưu để đáp ứng nhanh trong môi trường mạng xã hội, đảm bảo độ trễ thấp và độ chính xác cao.

3.4. Workflow của hệ thống

Quy trình làm việc được thể hiện rõ ở **Hình 1** bao gồm các bước:

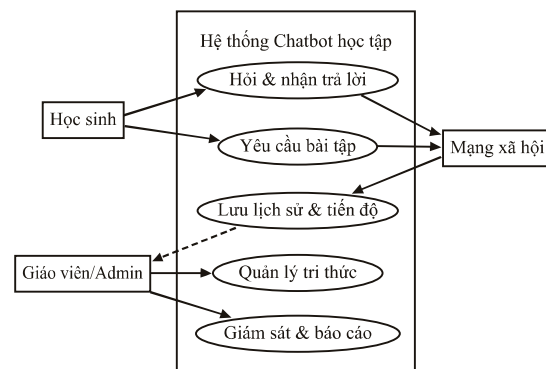
1. Học sinh gửi câu hỏi trên nền tảng mạng xã hội.
2. Hệ thống nhận dạng và phân tích câu hỏi.
3. Truy xuất dữ liệu môn học phù hợp.
4. Mô hình ngôn ngữ sinh câu trả lời.
5. Hệ thống phản hồi lại cho học sinh.
6. Lưu lịch sử để cá nhân hoá học tập.



Hình 1. Sơ đồ Workflow của hệ thống

Hình 2 minh họa mô hình tương tác giữa các thành phần chính trong hệ thống chatbot học tập. Học sinh tương tác với chatbot thông qua nền tảng mạng xã hội (như Zalo) để đặt câu hỏi, nhận câu trả lời và yêu cầu bài tập. Mọi hoạt động học tập đều được hệ thống ghi nhận nhằm theo dõi lịch sử và tiến độ học tập.

Ở phía quản trị, giáo viên hoặc admin có thể cập nhật và quản lý cơ sở tri thức, đồng thời truy cập các chức năng giám sát và báo cáo để theo dõi tình hình học tập của học sinh. Các luồng tương tác hai chiều giữa học sinh, giáo viên và hệ thống đảm bảo hoạt động học tập diễn ra liên tục, chính xác và có tính cá nhân hóa cao.



Hình 2. Sơ đồ usecase của hệ thống

4. THỰC NGHIỆM VÀ KẾT QUẢ ĐẠT ĐƯỢC

4.1. Cài đặt hệ thống

Quá trình triển khai chatbot trên Zalo được thực hiện theo 4 bước chính ở **Bảng 1**:

Bảng 1. Quy trình cài đặt hệ thống

Bước	Nội dung thực hiện	Kết quả đạt được
1	Tạo OA Zalo và kết nối API	Khởi tạo kênh giao tiếp chính thức
2	Cấu hình backend (Python/FastAPI)	Xử lý yêu cầu, định tuyến hội thoại
3	Tích hợp mô hình ChatGPT-tinh chỉnh	Chatbot hiểu đúng ngữ cảnh môn Hóa
4	Thiết kế giao diện hội thoại	Tối ưu trải nghiệm người dùng

4.2. Quy trình thử nghiệm

Kiểm thử chức năng: đo độ chính xác, độ trễ, tính ổn định.

Kiểm thử người dùng: 50 học sinh tham gia, thực hiện 320 câu hỏi ở 4 mức độ (nhận biết, thông hiểu, vận dụng, vận dụng cao) mà các giáo viên chuyên môn đã thẩm định.

Thu thập phản hồi: đánh giá chất lượng câu trả lời và tần suất lỗi.

Cấu trúc bộ dữ liệu thử nghiệm được thể hiện ở **Bảng 2**.

Bảng 2. Cấu trúc bộ dữ liệu thử nghiệm

Mức độ kiến thức	Số lượng	Tỷ lệ
Nhận biết	110	34%
Thông hiểu	95	30%
Vận dụng	70	22%
Vận dụng cao	45	14%

4.3. Kết quả thử nghiệm

Bảng 3 dưới đây tổng hợp kết quả

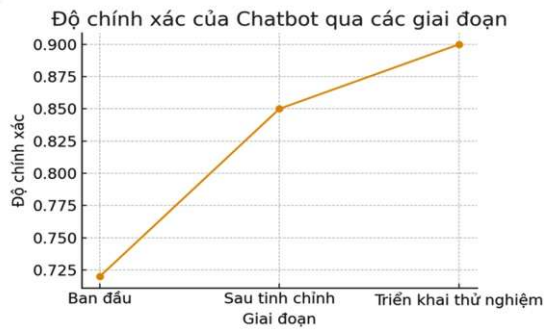
đánh giá theo ba giai đoạn vận hành. Để tính độ chính xác, chúng tôi mời hội đồng gồm 3 giáo viên Hóa chấm theo tiêu chí đúng, đủ, phù hợp mức độ nhận thức. Độ chính xác = số câu trả lời đúng / tổng số câu hỏi.

Bảng 3. Tổng hợp kết quả đánh giá theo ba giai đoạn

Giai đoạn	Độ chính xác	Tốc độ phản hồi (giây)
Ban đầu	0.72	2.1
Sau tinh chỉnh	0.85	1.6
Thử nghiệm thực tế	0.90	1.3

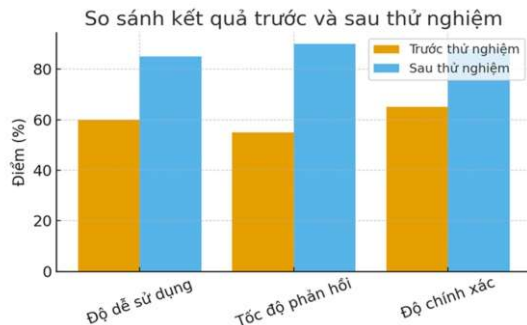
Kết quả thử nghiệm ở **Hình 3** cho thấy độ chính xác của chatbot cải thiện rõ rệt theo từng giai đoạn triển khai. Ở giai đoạn ban đầu, hệ thống đạt 72%, phản ánh hiệu quả xử lý còn phụ thuộc nhiều vào dữ liệu và khả năng khái quát của mô hình. Sau quá trình tinh chỉnh trên bộ dữ liệu chuyên biệt môn Hóa học THPT, độ chính xác tăng lên 85%, cho thấy sự cải thiện đáng kể trong nhận diện ý định và truy xuất tri thức.

Giai đoạn thử nghiệm thực tế với học sinh ghi nhận mức độ chính xác 90%, là mức cao nhất và thể hiện khả năng thích ứng tốt của chatbot với ngôn ngữ tự nhiên trong bối cảnh sử dụng thật. Xu hướng tăng đều qua ba giai đoạn cho thấy quy trình tinh chỉnh và cải thiện tri thức nền có tác động trực tiếp và hiệu quả đến hiệu suất của hệ thống.



Hình 3. Độ chính xác của chatbot qua các giai đoạn.

Hình 4 thể hiện sự cải thiện rõ rệt về độ chính xác của chatbot qua ba giai đoạn phát triển. Ở giai đoạn ban đầu, hệ thống đạt mức chính xác khoảng 72%. Sau khi tinh chỉnh mô hình và bổ sung tri thức, độ chính xác tăng lên 85%. Trong giai đoạn thử nghiệm thực tế với học sinh, hệ thống tiếp tục đạt mức cao nhất, khoảng 90%. Kết quả cho thấy hiệu quả của việc tối ưu mô hình và rà soát dữ liệu trong quá trình triển khai.

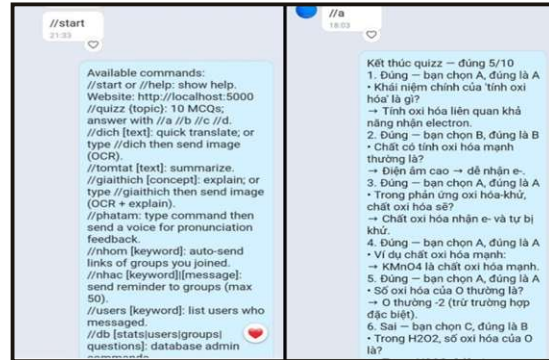


Hình 4. So sánh kết quả trước và sau thử nghiệm

4.4. Minh họa giao diện

Hình 5 dưới đây thể hiện giao diện tương tác của chatbot trên nền tảng Zalo OA. Giao diện được thiết kế tối giản, tập trung vào vùng nhập liệu và khung phản hồi, giúp học sinh đặt câu hỏi nhanh và nhận câu trả lời ngay lập tức. Các nút chọn nhanh (menu, chủ đề, dạng bài) được bố trí trực quan nhằm hỗ trợ quá trình tra cứu kiến thức và luyện tập. Cách trình bày rõ ràng

giúp người dùng dễ theo dõi lời giải, phương trình hóa học và các bước lập luận, qua đó nâng cao trải nghiệm học tập và giảm thời gian thao tác.



Hình 5. Minh họa giao diện chatbot

5. KẾT LUẬN

Nghiên cứu đã xây dựng và thử nghiệm mô hình chatbot hỗ trợ học tập môn Hóa học THPT dựa trên mô hình ngôn ngữ lớn kết hợp cơ chế truy xuất tri thức (RAG). Hệ thống đáp ứng tốt các yêu cầu cốt lõi của một trợ lý học tập số: hiểu ngữ cảnh câu hỏi, truy xuất thông tin chính xác và tạo phản hồi phù hợp với nội dung chương trình môn học.

Kết quả thử nghiệm cho thấy chatbot đạt độ chính xác cao hơn rõ rệt sau giai đoạn tinh chỉnh, đồng thời tốc độ phản hồi ổn định trong môi trường mạng xã hội. Việc triển khai thử nghiệm với học sinh cho thấy mức độ chấp nhận và hiệu quả sử dụng khả quan, đặc biệt ở các dạng câu hỏi lý thuyết, bài tập cơ bản và gợi ý phương pháp giải.

Mô hình đề xuất có tính khả thi cao và có thể mở rộng theo nhiều hướng, bao gồm bổ sung dữ liệu theo chuyên đề, tích hợp đánh giá tự động, cá nhân hóa lộ trình học tập và tối ưu hiệu suất trên các nền tảng giáo dục số. Trong tương lai, việc mở rộng phạm vi thử nghiệm và phát triển bộ chỉ số đánh giá toàn diện sẽ giúp hoàn thiện hơn vai trò của chatbot như một công cụ hỗ trợ học tập thông minh trong giáo dục phổ thông.

TÀI LIỆU THAM KHẢO

Tiếng Việt:

- [1] Đỗ, Ngọc Thanh, “Ứng dụng AI xây dựng Chatbot hỗ trợ sinh viên trong công tác học tập và sinh hoạt tại VKU”, Kỷ yếu Hội thảo Nghiên cứu khoa học của Sinh viên năm học 2021-2021; từ trang 39-44.
- [2] Nguyễn, Thành Thủy (2018), “Ứng dụng thuật toán học có giám sát multi-class svm trong xây dựng hệ thống chatbot hỏi đáp tiếng việt”, *The 7th conference on information technology and its applications*. 177-184.
- [3] Hồ, Đỗ Đạt, Nhựt Tân Nguyễn, and Minh Hiếu Võ (2025), “Ảnh hưởng của sử dụng trí tuệ nhân tạo đến học tập của sinh viên: Trường hợp khảo sát ở hai cơ sở giáo dục tại thành phố Hồ Chí Minh và thành phố Cần Thơ”, *Tạp chí Nghiên cứu khoa học và Phát triển kinh tế Trường Đại học Tây Đô* 22: 11-25.
- [4] Phạm, Trinh Hong Phi, and Mai Huong Tran (2024), “Ứng dụng trí tuệ nhân tạo trong giao tiếp giữa cơ sở dịch vụ công và người dân Việt Nam thông qua ứng dụng dựa vào Chatbot tại Việt Nam”. *Tạp chí Khoa học và kỹ thuật trường Đại học Bình Dương* (2024), 7.1.
- [5] Anh, Nguyễn Kim, Hồ Thị Hương Thơm, and Nguyễn Trung Đức (2025), “ỨNG DỤNG DIALOGFLOW PHÁT TRIỂN CHATBOT VỀ AN TOÀN GIAO THÔNG”, *Tạp chí Khoa học Công nghệ*

Hàng hải 83 (2025): 168-174

- [6] Khanh, Nguyễn Thị Phương (2025), “Tác động của chatbot AI đến sự hài lòng của khách hàng: trường hợp hãng dịch vụ chăm sóc xe ECOSPACE tại Thành phố Hồ Chí Minh”, Luận văn Thạc sỹ Đại học Kinh tế Thành phố Hồ Chí Minh.

Tiếng Anh:

- [7] Adamopoulou, Eleni, and Lefteris Moussiades (2020), “An overview of chatbot technology”, *IFIP international conference on artificial intelligence applications and innovations*. Cham: Springer International Publishing.
- [8] Trang, Nguyen Ngoc, Nguyen Anh Tuan and Nguyen Lan Phuong (2023), “Integrating Artificial Intelligence with Social Network Supporting New Student Onboarding: Factors Create Digital Learning Ecosystem” *Proceedings of the 14th ICSIT*: 123-128.
- [9] My-Thanh Nguyen, Thi, et al (2021), “Design of online learning platform with vietnamese virtual assistant” *Proceedings of the 2021 6th International Conference on Intelligent Information Technology*.
- [10] Ranoliya, Bhavika R. (2017), Nidhi Raghuwanshi, and Sanjay Singh. “Chatbot for university related FAQs”. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE.

Liên hệ:

TS. Hoàng Đình Tuyền

Khoa Công nghệ thông tin, Trường Đại học Quảng Bình
Địa chỉ: 18 Nguyễn Văn Linh, Phường Đồng Thuận, Tỉnh Quảng Trị
Email:hoangdinh TUYEN@gmail.com

Ngày nhận bài: 17/11/2025

Ngày gửi phản biện: 17/11/2025

Ngày duyệt đăng: 15/12/2025