

DOI: <https://doi.org/10.59294/HIUJS.KHTT.2026.047>

XÂY DỰNG HỆ THỐNG ĐA TÁC TỬ HỖ TRỢ TÓM TẮT BÀI ĐỌC TIẾNG VIỆT BẬC TIỂU HỌC

Hoàng Ngọc Long*, Nguyễn Minh Vũ, Phạm Chí Khanh,
Lê Hiếu Nghiêm, Bùi Hồ Chí Cường, Huỳnh Thanh Phú
Trường Đại học Quốc tế Hồng Bàng

TÓM TẮT

Sự bùng nổ của trí tuệ nhân tạo (AI) và xử lý ngôn ngữ tự nhiên (NLP) đang mở ra nhiều cơ hội ứng dụng trong lĩnh vực giáo dục, đặc biệt trong việc hỗ trợ phát triển năng lực đọc hiểu cho học sinh tiểu học. Tuy vậy, đối với tiếng Việt, các giải pháp hiện nay vẫn chưa đáp ứng đầy đủ yêu cầu về nội dung và đối tượng người học. Bài báo này đề xuất một hệ thống hỗ trợ học tập dựa trên AI, giúp học sinh tiếp cận văn bản và nắm bắt nội dung cốt lõi một cách hiệu quả hơn, đồng thời phù hợp với định hướng chuyển đổi số trong giáo dục. Hệ thống được thiết kế theo kiến trúc đa tác tử (MAS), cho phép phân rã bài toán tóm tắt thành các nhiệm vụ thành phần và xử lý thông qua các tác tử chuyên biệt. Cụ thể, mô hình PhoBERT được sử dụng cho tóm tắt trích xuất, trong khi ViT5 đảm nhiệm tóm tắt diễn giải, với các tham số được tinh chỉnh nhằm tối ưu hiệu quả. Ngoài ra, đầu ra của hệ thống có thể được điều chỉnh linh hoạt theo độ dài, mức độ từ vựng và bộ từ điển tương ứng với từng cấp lớp. Kết quả thực nghiệm cho thấy phương pháp đề xuất đạt hiệu quả cao trên các thước đo như ROUGE và BERTScore. Nghiên cứu không chỉ đáp ứng nhu cầu thực tiễn trong dạy và học mà còn góp phần định hướng phát triển các hệ thống tóm tắt văn bản tiếng Việt phù hợp với học sinh tiểu học.

Từ khóa: tóm tắt văn bản, đa tác tử, trí tuệ nhân tạo, xử lý ngôn ngữ tự nhiên, giáo dục tiểu học

DEVELOPING A MULTI-AGENT SYSTEM FOR VIETNAMESE READING SUMMARIZATION IN PRIMARY EDUCATION

Hoang Ngoc Long, Nguyen Minh Vu, Pham Chi Khanh,
Le Hieu Nghiem, Bui Ho Chi Cuong, Huynh Thanh Phu

ABSTRACT

The rapid advancement of artificial intelligence (AI) and natural language processing (NLP) has created significant opportunities for educational applications, particularly in enhancing reading comprehension skills for primary school students. However, for the Vietnamese language, existing technological solutions remain limited in effectively addressing content requirements and learner-specific needs. This paper proposes an AI-based learning support system designed to help students better access textual materials and grasp key ideas, while aligning with ongoing digital transformation initiatives in education. The proposed system is built upon a multi-agent system (MAS) architecture, which decomposes the text summarization task into smaller, specialized sub-tasks handled by dedicated agents. Specifically, the PhoBERT model is employed for extractive summarization, whereas ViT5 is utilized for abstractive summarization, with both models fine-tuned to optimize performance. In addition, the system allows flexible control over the generated summaries in terms of length, vocabulary complexity, and grade-level-specific lexicons. Experimental results demonstrate that the proposed approach achieves strong performance on evaluation metrics such as ROUGE and BERTScore. This study not only addresses practical needs in educational contexts but

* Tác giả liên hệ: Hoàng Ngọc Long, Email: longhn@hiu.vn

(Ngày nhận bài: 05/3/2026; Ngày nhận bản sửa: 09/4/2026; Ngày duyệt đăng: 15/4/2026)

also contributes to the development of AI-driven Vietnamese text summarization systems tailored for primary education.

Keywords: *text summarization, multi-agent system, artificial intelligence, natural language processing, primary education*

1. ĐẶT VẤN ĐỀ

Tóm tắt văn bản là quá trình bày ngắn gọn, chính xác và đầy đủ các ý chính, nội dung cốt lõi của tài liệu gốc. Quá trình này không chỉ đơn thuần là rút ngắn độ dài mà còn đòi hỏi sự chính xác về ngữ pháp, chính tả, tính liên kết trong diễn đạt, cũng như sự phù hợp về ngữ nghĩa và phong cách. Trong bối cảnh giáo dục, đặc biệt ở cấp tiểu học, tóm tắt văn bản đóng vai trò thiết yếu trong việc hỗ trợ phát triển kỹ năng đọc hiểu, giúp học sinh tiếp cận và nắm bắt kiến thức hiệu quả hơn, nhất là khi các em phải xử lý những văn bản dài, trong khi vốn từ và khả năng tập trung còn hạn chế [1 - 3].

Tại Việt Nam, các nghiên cứu về NLP phục vụ giảng dạy và học tập Tiếng Việt đã ghi nhận nhiều kết quả tích cực, đặc biệt trong các bài toán liên quan đến đọc hiểu và tóm tắt. Một số công trình tiêu biểu đã xây dựng hệ thống AI hỗ trợ tóm tắt cho học sinh tiểu học (lớp 1 - 5), đạt hiệu quả cao trên các chỉ số đánh giá như ROUGE, BLEU và BERTScore, đồng thời nhận được phản hồi tích cực từ giáo viên, học sinh và phụ huynh [4]. Những kết quả này cho thấy tiềm năng lớn của AI trong việc hỗ trợ học sinh nắm bắt nội dung chính và nâng cao hứng thú học tập. Tuy nhiên, việc triển khai rộng rãi các giải pháp này trong thực tiễn giáo dục tiểu học vẫn còn gặp nhiều hạn chế. Phần lớn các hệ thống hiện nay được phát triển cho mục đích chung hoặc người dùng phổ thông, chưa thực sự tính đến đặc điểm nhận thức, vốn từ và khả năng tiếp thu của học sinh nhỏ tuổi. Mặc dù các mô hình như PhoBERT và ViT5 đã chứng minh hiệu quả trong nhiều bài toán tiếng Việt, chúng chủ yếu được huấn luyện trên dữ liệu tổng quát và chưa được điều chỉnh theo từng cấp lớp. Vì vậy, dù nền tảng công nghệ đã sẵn sàng, các công cụ tóm tắt văn bản chuyên biệt cho học sinh tiểu học vẫn còn thiếu và cần được nghiên cứu, phát triển thêm.

Khoảng trống này đặt ra nhu cầu xây dựng một ứng dụng AI được thiết kế riêng cho ngữ cảnh giáo dục tiểu học, thay vì điều chỉnh lại các mô hình tổng quát hỗ trợ tóm tắt bài đọc tiếng Việt giúp trẻ vượt qua rào cản đọc hiểu. Nghiên cứu không chỉ giúp học sinh rèn luyện kỹ năng đọc hiểu mà còn hỗ trợ giáo viên và phụ huynh trong việc kiểm tra, đánh giá khả năng tiếp thu của học sinh. Hệ thống được xây dựng theo kiến trúc MAS, kết hợp hai mô hình tóm tắt đã được tinh chỉnh với tham số tối ưu, cùng bộ dữ liệu đủ lớn có chọn lọc và kho từ điển phân theo cấp lớp. Cách tiếp cận này hướng đến việc nâng cao tính thích ứng theo từng trình độ, đồng thời cải thiện chất lượng đầu ra so với mô hình đơn lẻ. Nghiên cứu vì vậy không chỉ đề xuất một công cụ ứng dụng, mà còn gợi mở một hướng tiếp cận có tính hệ thống cho việc phát triển các giải pháp AI phù hợp với giáo dục tiểu học trong bối cảnh chuyển đổi số. Các phần tiếp theo của bài báo sẽ trình bày cơ sở lý thuyết, phương pháp thiết kế, 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 tóm tắt văn bản trong các nghiên cứu

Tại Việt Nam, học sinh tiểu học thường gặp khó khăn khi đọc các văn bản truyện, bài đọc dài, do vốn từ vựng hạn chế và khả năng tập trung chưa phát triển dẫn đến hiệu quả học tập giảm sút [3]. Một nghiên cứu thực nghiệm của Ophuis-Cox & L. Rozendal cho học sinh lớp 3 và lớp 4 cho thấy việc áp dụng chiến lược tóm tắt văn bản giúp học sinh cải thiện đáng kể khả năng hiểu và ghi nhớ thông tin so với các phương pháp học khác [5]. Hay trong nhiều khảo sát nghiên cứu khác, việc sử dụng kỹ thuật tóm tắt để trích xuất từ khóa và kết hợp thành bản tóm tắt đã được áp dụng trong giáo dục sơ cấp, minh chứng rằng các phương pháp tóm tắt có thể được sử dụng để hỗ trợ học tập qua các bài giảng hoặc tài liệu dài. Thông qua đó, phản ánh tích cực thành tích học tập, khả năng đọc hiểu, năng lực ngôn ngữ và cải thiện kỹ năng, nhận thức của học sinh tiểu học khi đã áp dụng kỹ năng tóm tắt [1 - 3]. Tuy nhiên, việc triển khai một hệ thống AI tóm tắt văn bản trong giáo dục tiểu học tiếng Việt

đổi mặt với nhiều thách thức. Các nghiên cứu hiện tại chủ yếu tập trung vào ngôn ngữ tiếng Anh hoặc các ứng dụng thương mại, thiếu sự chú trọng đến đặc thù ngôn ngữ học và nhu cầu sư phạm tại Việt Nam [6]. Các bộ dữ liệu tiếng Việt thường được thiết kế cho các tác vụ tổng quát, chẳng hạn như tóm tắt tin tức, không đáp ứng được yêu cầu về ngôn ngữ đơn giản và nội dung phù hợp với trình độ nhận thức của học sinh tiểu học [7]. Những hạn chế này tạo ra nhu cầu cấp thiết về một hệ thống hỗ trợ tóm tắt chuyên biệt cho giáo dục tiểu học tại Việt Nam.

2.2. Phương pháp tóm tắt và mô hình ngôn ngữ lớn

Nghiên cứu sử dụng hai hướng tóm tắt chính: Trích xuất và diễn giải. Phương pháp trích xuất chọn các câu quan trọng từ văn bản gốc nên giữ nguyên nội dung và hạn chế sai lệch, nhưng có thể thiếu tính mạch lạc. Ngược lại, phương pháp diễn giải tạo lại nội dung theo cách diễn đạt mới, giúp bản tóm tắt tự nhiên và liên kết hơn, song có nguy cơ phát sinh thông tin không chính xác nếu không kiểm soát tốt [8]. Việc áp dụng hai phương pháp này cho tiếng Việt vẫn gặp thách thức do đặc thù ngôn ngữ và thiếu dữ liệu chuyên biệt.

Sự phát triển của kiến trúc Transformer đã thúc đẩy các mô hình như PhoBERT và ViT5 đạt hiệu quả cao trong NLP tiếng Việt. PhoBERT phù hợp cho tóm tắt trích xuất, trong khi ViT5 hỗ trợ tốt tóm tắt diễn giải [7, 9]. Tuy nhiên, các nghiên cứu trước đây [6] chủ yếu huấn luyện các mô hình này trên dữ liệu tin tức hoặc các tác vụ tổng quát, không phù hợp với ngữ cảnh giáo dục tiểu học, vốn yêu cầu ngôn ngữ đơn giản và nội dung phù hợp với trẻ em. Để khắc phục hạn chế này, nghiên cứu hiện tại tinh chỉnh PhoBERT và ViT5 trên bộ dữ liệu chuyên biệt với hơn 10,000 bài đọc trích xuất và hơn 20.000 bài đọc diễn giải đã qua kiểm soát bằng bộ từ điển theo cấp lớp, nhằm tạo ra hai mô hình tóm tắt chính xác, phù hợp và đạt hiệu quả cao với trình độ nhận thức của học sinh tiểu học.

2.3. Hệ thống đa tác tử (Multi-Agent System)

Hệ đa tác tử (MAS) là tập hợp nhiều tác tử cùng hoạt động trong một môi trường để đạt mục tiêu chung. Mỗi tác tử có tính tự chủ, có khả năng nhận thức và đưa ra quyết định dựa trên thông tin sẵn có. MAS nổi bật với tính phân tán, khả năng phối hợp và dễ mở rộng, cho phép chia nhỏ bài toán thành các nhiệm vụ chuyên biệt thay vì phụ thuộc vào một mô hình trung tâm [10]. Trong bài toán tóm tắt văn bản cho học sinh tiểu học, kiến trúc này giúp tách các chức năng như phân tích nội dung, sinh tóm tắt, kiểm soát theo cấp độ lớp và đánh giá kết quả. Một hệ MAS gồm bốn thành phần chính: tác tử, môi trường, cơ chế giao tiếp và cơ chế phối hợp. Các tác tử được thiết kế theo hướng mô-đun, có thể triển khai bằng các framework như LangChain và LangGraph để đảm bảo linh hoạt và khả năng mở rộng. Môi trường cung cấp dữ liệu và ngữ cảnh xử lý, trong khi cơ chế giao tiếp và phối hợp giúp các tác tử trao đổi thông tin và làm việc hiệu quả nhằm đạt mục tiêu chung. Trong lĩnh vực giáo dục, nhiều nghiên cứu đã cho thấy tiềm năng của cách tiếp cận này. Một số hệ thống đa tác tử được sử dụng để cung cấp phản hồi tự động cho bài kiểm tra học sinh đã cải thiện độ chính xác và tính sư phạm so với mô hình đơn tác tử [8]. Gần đây, khung Mixture-of-Agents (MoA) cũng được áp dụng cho bài toán tóm tắt văn bản, trong đó mỗi tác tử đảm nhiệm một chức năng riêng và kết hợp đầu ra nhằm nâng cao độ chính xác và tính mạch lạc. Kết quả thực nghiệm cho thấy các mô hình đa tác tử có thể đạt hiệu quả cao hơn so với các mô hình đơn lẻ trên nhiều tiêu chí đánh giá [11].

2.4. Đánh giá tóm tắt và khoảng trống nghiên cứu

Chất lượng bản tóm tắt được đánh giá bằng các thước đo phổ biến trong xử lý ngôn ngữ tự nhiên, bao gồm ROUGE - phản ánh mức độ bao phủ nội dung so với văn bản nguồn [12] và BERTScore, dùng để đo sự tương đồng ngữ nghĩa [13]. Trong tiếng Việt, các nghiên cứu về tóm tắt trích xuất đã ghi nhận hiệu quả của những thuật toán như TextRank trong việc xác định và lựa chọn câu trọng tâm. Tuy vậy, dù các chỉ số này được sử dụng rộng rãi, việc áp dụng vào văn bản giáo dục tiểu học đòi hỏi sự điều chỉnh nhất định, do tiêu chí đánh giá không chỉ dừng ở mức độ tương đồng nội dung mà còn liên quan đến độ rõ ràng và tính phù hợp về ngôn ngữ [14]. Bên cạnh đó, nhiều nghiên cứu khẳng định tóm tắt có thể hỗ trợ giáo viên trong giảng dạy, với điều kiện hệ thống được thiết kế sát với yêu

cầu của từng cấp lớp [6, 14]. Trên cơ sở đó, nghiên cứu này sử dụng các thang đo ROUGE và BERTScore để đánh giá, đồng thời định hướng điều chỉnh để đảm bảo bản tóm tắt phù hợp với học sinh từ lớp 1 đến lớp 5.

Khoảng trống trong nghiên cứu tóm tắt tiếng Việt dành riêng cho học sinh tiểu học là cơ sở hình thành hướng tiếp cận của đề tài này. Phần lớn các công trình hiện nay [1, 8, 9] tập trung vào tóm tắt tin tức hoặc các bài toán tổng quát, chủ yếu trong bối cảnh tiếng Anh, mà chưa xem xét đầy đủ đặc điểm ngôn ngữ và yêu cầu sư phạm của trẻ em Việt Nam. Chẳng hạn, nhiều nghiên cứu về tóm tắt diễn giải tiếng Việt sử dụng nguồn dữ liệu dành cho người trưởng thành, dẫn đến đầu ra khó phù hợp với trình độ tiếp nhận của học sinh nhỏ tuổi [15]. Bên cạnh đó, các bộ dữ liệu tiếng Việt đang được sử dụng rộng rãi phần lớn phục vụ tóm tắt tin tức hoặc văn bản thông tin, chưa được phân chia theo độ tuổi, cấp lớp hoặc mục tiêu giáo dục cụ thể [7, 15]. Thêm vào đó, một số nền tảng như Canva chỉ cung cấp tính năng tóm tắt nâng cao khi trả phí, trong khi QuillBot lại chưa hỗ trợ tiếng Việt. Các chatbot AI như ChatGPT hay Gemini theo mô hình freemium có thể tóm tắt văn bản, nhưng thường bị giới hạn về độ dài đầu vào. Không chỉ vậy, sự thiếu hụt về dữ liệu, chất lượng mô hình, thiếu cơ chế phối hợp kiểm soát và khả năng mở rộng là điểm hạn chế lớn của nghiên cứu [4]. Nghiên cứu này với mục đích khắc phục hạn chế về đối tượng người dùng, cải thiện chất lượng dữ liệu và hiệu quả mô hình huấn luyện để tạo ra bản tóm tắt đảm bảo độ tin cậy.

Mặt khác, với hệ thống được thiết kế dựa trên kiến trúc MAS, mỗi tác tử đảm nhiệm một vai trò riêng biệt như hiểu ý định, làm rõ yêu cầu, lập kế hoạch, tóm tắt và đánh giá. So với pipeline tuần tự truyền thống hoặc mô hình đơn lẻ, MAS cho phép phân tách rõ ràng từng bước xử lý, thay vì dồn toàn bộ nhiệm vụ vào một mô-đun duy nhất. Trong pipeline tuần tự cơ bản, dữ liệu đi qua các bước cố định và khó điều chỉnh linh hoạt khi có lỗi phát sinh; trong khi đó, MAS cho phép các tác tử tương tác hai chiều, bổ sung thông tin và kích hoạt lại các bước cần thiết (ví dụ cơ chế clarification hoặc retry), từ đó nâng cao khả năng kiểm soát toàn bộ quá trình. Về mặt lợi ích, MAS giúp tăng tính minh bạch (có thể xác định lỗi ở từng bước cụ thể), cải thiện độ chính xác thông qua cơ chế kiểm tra chéo giữa các tác tử, và đặc biệt là tăng khả năng mở rộng khi có thể bổ sung hoặc thay thế từng tác tử mà không ảnh hưởng đến toàn hệ thống. So với các hệ thống cơ bản như bài báo [4], MAS không chỉ tập trung vào việc sinh kết quả đầu ra mà còn kiểm soát chặt chẽ quá trình xử lý từ đầu vào đến đầu ra. Nhờ đó, hệ thống đạt được độ tin cậy cao hơn, đặc biệt trong các bài toán có nhiều ràng buộc như tóm tắt theo cấp lớp trong giáo dục. Trên cơ sở này, nghiên cứu hướng đến xây dựng một ứng dụng AI đa tác tử hỗ trợ tóm tắt bài đọc tiếng Việt cho học sinh tiểu học, với các đóng góp chính được trình bày sau đây:

- Xây dựng bộ từ điển cấp lớp và bộ dữ liệu chuyên biệt cho tóm tắt trích xuất và diễn giải tổng hơn 30,000 văn bản bài đọc được phân loại theo khối lớp 1 - 5.
- Tinh chỉnh các mô hình PhoBERT, ViT5 dựa trên các tham số tối ưu và bộ dữ liệu đã xây dựng để tạo ra các mô hình tóm tắt đạt hiệu suất và chất lượng tốt không chỉ ở các thang đo.
- Xây dựng hệ thống đa tác tử hỗ trợ tóm tắt văn bản và bài đọc tiếng Việt, sẵn sàng triển khai thành một ứng dụng hoàn chỉnh.

3. THIẾT KẾ HỆ THỐNG VÀ PHƯƠNG PHÁP

3.1. Quy trình thu thập và chuẩn bị dữ liệu

Bộ dữ liệu được xây dựng từ các nguồn chính thống như sách giáo khoa của Bộ Giáo dục và Đào tạo, sách lịch sử Việt Nam, truyện thiếu nhi và các tài liệu được khuyến nghị trong chương trình học. Dữ liệu với yêu cầu phù hợp về nội dung, bố cục rõ ràng, nội dung mạch lạc được sao chép, trích xuất từ hình ảnh. Sau đó, dữ liệu được làm sạch và chuẩn hóa nhằm loại bỏ những ký tự, lỗi chính tả sai. Các văn bản thu được sẽ được lưu dưới dạng tệp (.txt) với cấu trúc rõ ràng và được tổ chức bằng Pandas, trong đó mỗi mẫu dữ liệu gồm các thuộc tính như nội dung, cấp độ và nhà xuất bản. Từ nguồn dữ liệu này, nhóm nghiên cứu xây dựng hai tập dữ liệu cho tóm tắt trích xuất và tóm tắt diễn giải. Với tóm tắt trích xuất, văn bản được tách câu và chuẩn hóa bằng VnCoreNLP, sau đó áp dụng TextRank

và LexRank để tính độ tương đồng cosine. Tiếp theo, hệ thống lọc 70% số câu có điểm cao nhất, sau đó sử dụng PhoBERT để đánh giá lại mức độ tương đồng ngữ nghĩa giữa câu và toàn văn bản. Từ tập này, một tập con khoảng 50% số câu được lựa chọn bằng thuật toán Maximal Marginal Relevance (MMR) nhằm cân bằng giữa mức độ quan trọng và tính đa dạng nội dung. Cuối cùng, các câu được sắp xếp lại theo thứ tự ban đầu và ghép thành bản tóm tắt hoàn chỉnh. Với tóm tắt điển giải, văn bản quá dài sẽ được tóm tắt trích xuất bởi các phương pháp vừa rồi, trước khi đưa vào GPT-4o để sinh bản tóm tắt. Kết quả đầu ra được kiểm soát bằng bộ từ điển theo cấp lớp nhằm đảm bảo mức độ phù hợp với năng lực đọc hiểu của học sinh. Toàn bộ dữ liệu được đánh giá tự động bằng ROUGE và BERTScore với các ngưỡng tối thiểu nhằm đảm bảo độ bao phủ nội dung và tương đồng ngữ nghĩa trước khi sử dụng cho huấn luyện.

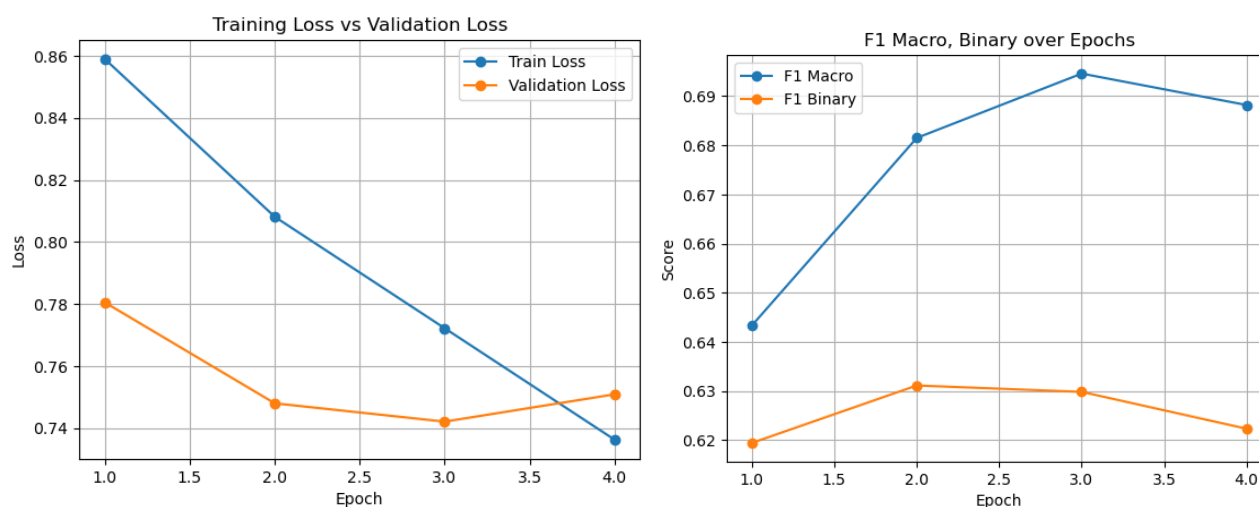
3.2. Mô hình và quy trình huấn luyện

Trong giai đoạn tối ưu hóa mô hình, nghiên cứu sử dụng Optuna để tìm kiếm bộ siêu tham số phù hợp cho cả hai hướng tiếp cận. Cụ thể, với mô hình tóm tắt trích xuất, 35 siêu tham số được khảo sát; trong khi đó, mô hình tóm tắt điển giải được tối ưu với 20 siêu tham số. Quá trình này nhằm quan sát ảnh hưởng của từng cấu hình đến chất lượng đầu ra và lựa chọn bộ tham số tối ưu trước khi tiến hành huấn luyện chính thức.

Bảng 1. Các tham số tối ưu của 2 mô hình sau quá trình Optuna

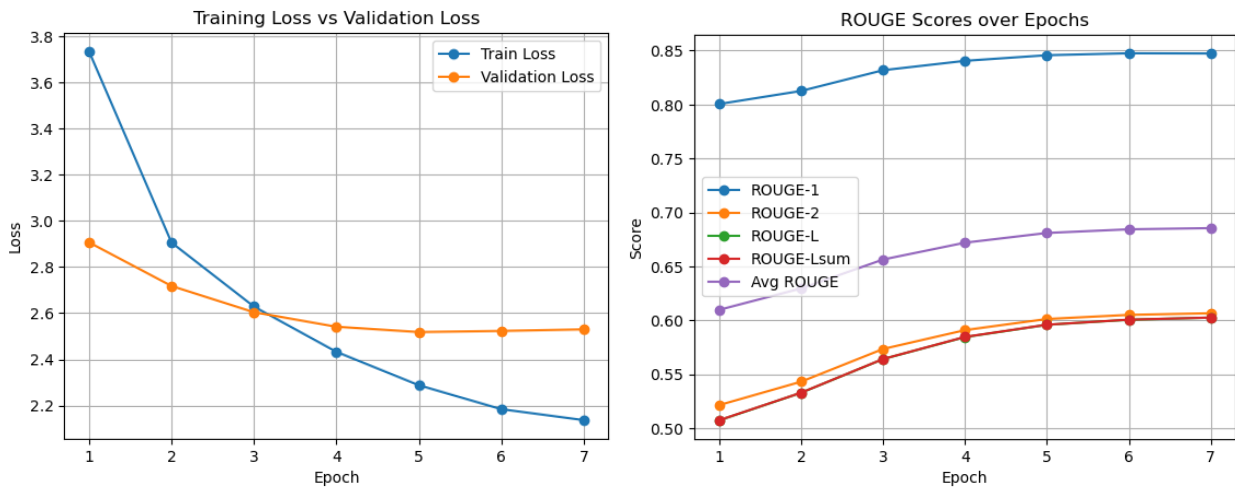
Siêu tham số	PhoBERT (Extractive)	ViT5 (Abstractive)
Learning Rate	2.2414e-05	1.485e-04
Batch Size	16	8
Max Length	256	768 (In) / 256 (Out)
Epochs	6 (Early stopping 4)	5 (Early stopping 7)
Weight Decay	0.0435	0.00165
Warmup Ratio	0.076	0.1
Label smoothing	0.145	0.1

Tinh chỉnh PhoBERT: Mô hình PhoBERT được tinh chỉnh trên tập dữ liệu gồm 240,000 mẫu câu cho tác vụ tóm tắt trích xuất. Dữ liệu được phân chia theo tỷ lệ 80:10:10 lần lượt cho các tập huấn luyện, kiểm định và kiểm tra. Kết quả thực nghiệm tốt nhất đạt được tại epoch thứ 4 với điểm F1 Macro là 0.6882 và F1 Binary là 0.6223.



Hình 1. Các chỉ số đánh giá mô hình tóm tắt trích xuất

Tinh chỉnh ViT5: Đối với tác vụ tóm tắt diễn giải, chúng tôi sử dụng mô hình ViT5 được huấn luyện trên 22.000 cặp văn bản input-output. Mô hình đạt hiệu suất cao với các chỉ số ROUGE-1, ROUGE-2 và ROUGE-L lần lượt là 0.8473, 0.6066 và 0.6026.



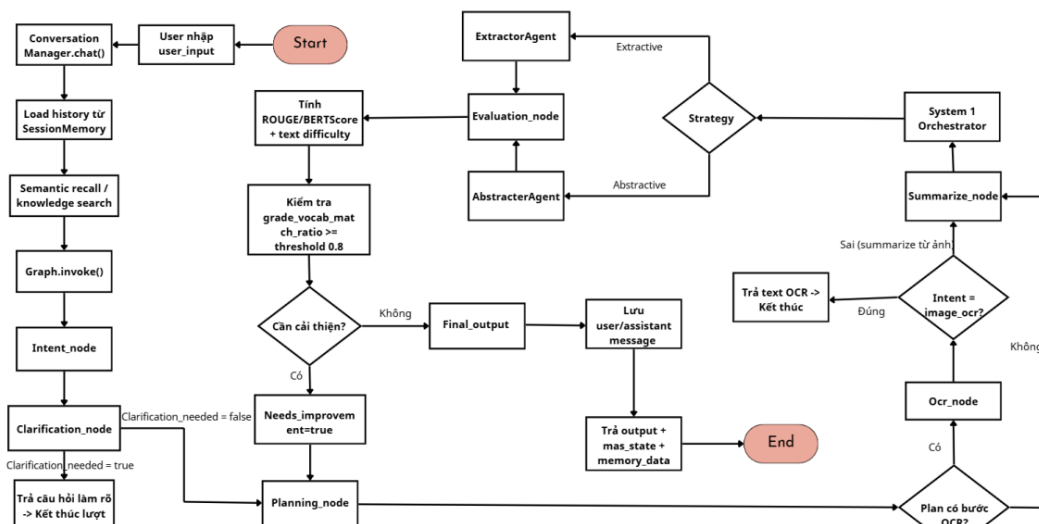
Hình 2. Các chỉ số đánh giá mô hình tóm tắt diễn giải

Quy trình huấn luyện: Quy trình huấn luyện được triển khai theo các bước tuần tự: (1) tiền xử lý dữ liệu, bao gồm chuẩn hóa văn bản, tách từ, gán nhãn và phân chia theo cấp lớp; (2) xây dựng tập huấn luyện, tập xác thực và tập kiểm thử; (3) tinh chỉnh các mô hình ngôn ngữ trên GPU, trong đó mô hình trích xuất sử dụng PhoBERT và mô hình diễn giải sử dụng ViT5; (4) đánh giá hiệu suất trên tập kiểm thử thông qua các chỉ số ROUGE và BERTScore nhằm đo lường mức độ bao phủ nội dung và tương đồng ngữ nghĩa. Sau khi huấn luyện, mô hình đạt yêu cầu được đóng gói và tích hợp vào hệ thống backend, kết hợp với cơ chế giới hạn từ vựng theo lớp và điều chỉnh độ dài văn bản. Cách tiếp cận này giúp hệ thống tạo ra các bản tóm tắt phù hợp với yêu cầu năng lực đọc hiểu và mức độ phức tạp của từng khối lớp trong chương trình giáo dục tiêu học.

3.3. Thiết kế các tác tử

Hệ thống đa tác tử được tổ chức theo hai tầng chức năng gồm Hệ thống 1 và Hệ thống 2. Hệ thống 1 gồm hai tác tử có nhiệm vụ tóm tắt trích xuất và diễn giải, thông qua bộ điều phối hệ thống (*Orchestrator*) quyết định luồng xử lý, lựa chọn tác tử phù hợp. Cách tổ chức này giúp cô lập phần tính toán nặng, đồng thời đảm bảo tính nhất quán trong quá trình sinh văn bản. Hệ thống 2 là tầng tương tác và điều phối mở rộng, chịu trách nhiệm giao tiếp tự nhiên với người dùng, phân loại ý định (*Intent*), chuẩn hóa dữ liệu đầu vào, kiểm tra tính hợp lệ và kích hoạt Hệ thống 1 khi cần thiết. Ngoài ra, tầng này còn thực hiện tính toán các chỉ số đánh giá như ROUGE và BERTScore, cũng như xử lý các yêu cầu bổ sung liên quan đến hệ thống.

- *Intent Agent*: Xác định mục đích yêu cầu (tóm tắt, đánh giá, chuyển đổi định dạng...).
- *Planning Agent*: Lập kế hoạch xử lý và xác định luồng tác vụ.
- *Clarification Agent*: Yêu cầu làm rõ khi đầu vào chưa đủ thông tin hoặc mơ hồ.
- *Agent Memory*: Quản lý ngữ cảnh hội thoại và lưu trữ tạm thời thông tin cần thiết.
- *Evaluation Agent*: Tính toán và phân tích các chỉ số đánh giá chất lượng tóm tắt.
- *Image2Text Agent*: Chuyển đổi nội dung hình ảnh sang văn bản khi đầu vào ở dạng ảnh.



Hình 3. Kiến trúc tổng thể hệ thống đa tác tử tóm tắt văn bản dành cho học sinh tiểu học

3.4. Cơ chế kiểm soát đầu ra tóm tắt theo khối lớp

Để đảm bảo bản tóm tắt phù hợp với từng khối lớp, hệ thống áp dụng cơ chế kiểm soát đa tầng dựa trên đánh giá tự động và ràng buộc từ vựng. Trước hết, mỗi bản tóm tắt sau khi được sinh ra sẽ được kiểm tra bởi Evaluation Agent, chịu trách nhiệm tính toán các chỉ số ROUGE và BERTScore nhằm đo lường mức độ bao phủ nội dung và tương đồng ngữ nghĩa so với văn bản gốc. Các giá trị này được so sánh với ngưỡng yêu cầu tương ứng theo cấu hình của hệ thống. Song song, bản tóm tắt còn được đối chiếu với bộ từ điển phân theo cấp lớp đã được xây dựng trước đó.

Cơ chế này giúp kiểm soát việc sử dụng từ vựng vượt quá phạm vi tiếp nhận của học sinh ở từng khối lớp. Nếu phát hiện tỷ lệ từ ngoài danh mục cho phép vượt ngưỡng thiết lập, hệ thống sẽ ghi nhận và chuyển sang bước điều chỉnh. Trong trường hợp bản tóm tắt không đạt yêu cầu về chỉ số đánh giá hoặc vi phạm ràng buộc từ vựng, cơ chế retry có kiểm soát sẽ được kích hoạt. Hệ thống thực hiện sinh lại bản tóm tắt với tham số điều chỉnh (ví dụ: giảm độ dài, đơn giản hóa diễn đạt hoặc thay đổi chiến lược sinh), sau đó tiếp tục đánh giá cho đến khi đạt tiêu chí hoặc chạm ngưỡng lặp tối đa. Cách tiếp cận này giúp đảm bảo đầu ra không chỉ chính xác về nội dung mà còn phù hợp với năng lực đọc hiểu của lớp, đồng thời duy trì tính nhất quán và khả năng kiểm soát trong toàn bộ quy trình xử lý.

3.5. Phương pháp đánh giá hệ thống

Hiệu quả của hệ thống được đánh giá thông qua các thang đo định lượng phổ biến trong bài toán tóm tắt bao gồm ROUGE và BERTScore. ROUGE được sử dụng để đo mức độ trùng khớp n-gram và mức độ bao phủ nội dung giữa bản tóm tắt sinh ra và bản tham chiếu, trong khi BERTScore đánh giá mức độ tương đồng ngữ nghĩa dựa trên biểu diễn ngữ cảnh của các mô hình ngôn ngữ tiền huấn luyện. Việc lựa chọn các chỉ số này nhằm đảm bảo hai tiêu chí cốt lõi: (i) tính đầy đủ thông tin và (ii) tính nhất quán ngữ nghĩa. Các giá trị trung bình trên tập kiểm thử được sử dụng làm cơ sở so sánh giữa các cấu hình mô hình khác nhau trong quá trình thực nghiệm. Bên cạnh đó, kết quả đánh giá còn được phân tích theo từng nhóm lớp để quan sát mức độ ổn định của hệ thống trong các mức độ ngôn ngữ khác nhau. Cách tiếp cận này giúp phản ánh khách quan hiệu suất tổng thể của hệ thống, đồng thời cung cấp căn cứ để điều chỉnh tham số và cơ chế kiểm soát đầu ra khi cần thiết.

4. KẾT QUẢ NGHIÊN CỨU VÀ THỬ NGHIỆM

4.1. Môi trường và công cụ triển khai

Hệ thống được phát triển trên máy tính cá nhân trang bị chip core i5 11th, GPU NVIDIA GTX 1650 (4GB VRAM), 24GB RAM và sử dụng hệ điều hành Windows 11. Quá trình huấn luyện mô hình sử dụng Kaggle là môi trường để xây dựng, PyTorch, Transformers cùng các thư viện khác. Ứng dụng

hiện tại đang được triển khai với backend là Java Spring Boot và hệ thống AI với Python. Dữ liệu và mô hình được lưu trữ trên ổ SSD tốc độ cao để đảm bảo hiệu năng xử lý.

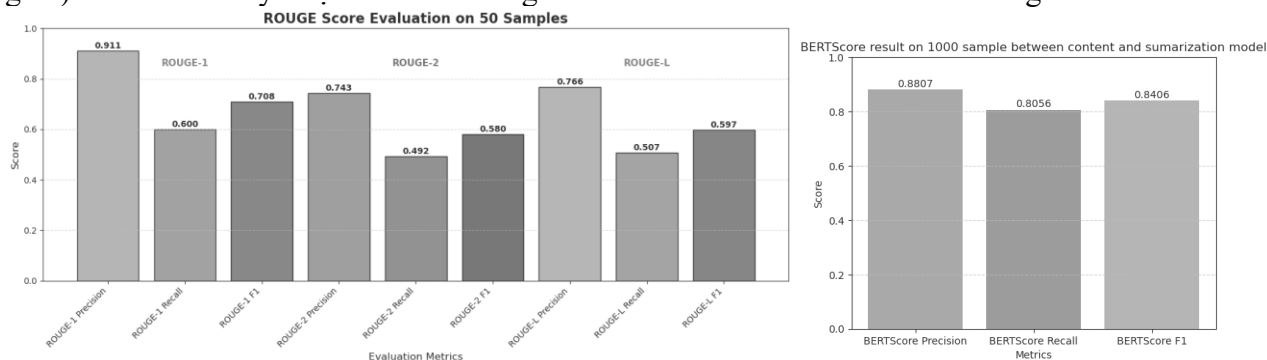
4.2. Triển khai hệ thống

Hệ thống được triển khai theo kiến trúc phân tầng, tách biệt giữa giao diện người dùng, tầng xử lý nghiệp vụ và tầng trí tuệ nhân tạo nhằm đảm bảo tính mở rộng và khả năng bảo trì lâu dài. Tầng giao diện người dùng (frontend) được xây dựng bằng React, cung cấp môi trường tương tác trực quan cho học sinh và giáo viên. Người dùng có thể nhập văn bản, chọn ảnh, lựa chọn khối lớp, yêu cầu tóm tắt hoặc thực hiện các chức năng hỗ trợ khác. Tầng backend được phát triển bằng Java Spring Boot, đóng vai trò điều phối các yêu cầu từ phía người dùng, quản lý phiên làm việc, xử lý xác thực, lưu trữ dữ liệu và giao tiếp với hệ thống đa tác tử. Backend cũng chịu trách nhiệm kết nối với cơ sở dữ liệu MySQL để lưu trữ thông tin người dùng, lịch sử tóm tắt và các cấu hình hệ thống. Tầng trí tuệ nhân tạo được triển khai như một AI Service độc lập (Flask-based), tích hợp hệ thống đa tác tử gồm Hệ thống 1 và Hệ thống 2. Trong đó, Hệ thống 1 chứa các mô hình đã được huấn luyện cho tóm tắt trích xuất và diễn giải, cùng cơ chế điều phối trung tâm. Hệ thống 2 đảm nhiệm xử lý ngôn ngữ tự nhiên, phân loại ý định, chuẩn hóa đầu vào, đánh giá đầu ra và quản lý bộ nhớ ngữ cảnh. Các mô hình được kết nối với cơ sở dữ liệu vector Chroma để hỗ trợ truy xuất ngữ nghĩa khi cần.

Các thành phần được triển khai theo mô hình dịch vụ tách biệt, cho phép mở rộng độc lập từng tầng (ví dụ: tăng tài nguyên GPU cho AI Service mà không ảnh hưởng đến backend). Sau khi huấn luyện, mô hình được đóng gói và tích hợp vào AI Service, sẵn sàng phục vụ thông qua API. Cách tổ chức này giúp hệ thống vận hành ổn định, linh hoạt và phù hợp với yêu cầu triển khai thực tế.

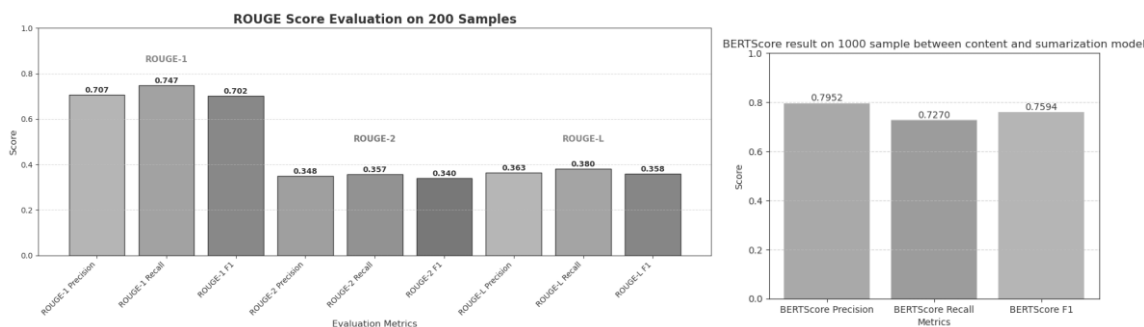
4.3. Đánh giá thực nghiệm

Để đánh giá khách quan chất lượng các bản tóm tắt được tạo ra bởi hệ thống, nghiên cứu sử dụng ba chỉ số để đánh giá: ROUGE-1 (đo mức độ bao phủ nội dung), ROUGE-2 (đo độ mạch lạc và trôi chảy), ROUGE-L (mức độ bảo toàn thứ tự và cấu trúc nội dung) và BERTScore (đo tương đồng ngữ nghĩa). Các chỉ số này được tính toán bằng cách so sánh bản tóm tắt với văn bản gốc.



Hình 4. Kết quả đánh giá trích xuất qua thang đo ROUGE (50 bài) và BERTScore (1,000 bài)

ROUGE-1 đạt giá trị tốt (~0.70), chứng tỏ mô hình có khả năng xác định đúng các nội dung chính trong văn bản. Đây là một chỉ số quan trọng trong bài toán tóm tắt, đặc biệt với tiếng Việt. ROUGE-2 và ROUGE-L ở mức khá cao và ổn (~0.58 - 0.6), phản ánh rằng: Mô hình không chỉ chọn đúng câu mà còn giữ được các cụm từ quan trọng, phản ánh tính liên kết nội dung tốt. Mô hình duy trì được mức độ mạch lạc tương đối trong chuỗi thông tin được trích xuất. BERTScore F1 0.8406 rất cao, cho thấy bản tóm tắt của mô hình bảo toàn tốt nội dung ngữ nghĩa của văn bản gốc, ngay cả khi không trùng hoàn toàn về mặt từ vựng. Giá trị Precision cao hơn Recall cho thấy mô hình có xu hướng lựa chọn các thông tin có độ liên quan cao và chính xác, ưu tiên các nội dung quan trọng và có độ tin cậy cao, dẫn đến việc giảm thiểu nhiễu nhưng đồng thời có thể bỏ sót một số thông tin thứ cấp.



Hình 5. Kết quả đánh giá diễn giải qua thang đo ROUGE (200 bài) và BERTScore (1,000 bài)

ROUGE-1 đạt giá trị cao (~0.70), chứng tỏ mô hình giữ được phần lớn nội dung quan trọng của văn bản gốc (mức độ trùng khớp unigram tốt). Đây là một chỉ số quan trọng trong bài toán tóm tắt, đặc biệt với tiếng Việt. ROUGE-2 và ROUGE-L ở mức trung bình (~0.33 - 0.36), phản ánh rằng mô hình có khả năng tái hiện cụm từ và cấu trúc câu ở mức khá. Tuy nhiên, với đặc thù của tóm tắt diễn giải, việc không đạt ROUGE-2 quá cao là hợp lý, do mô hình có xu hướng diễn đạt lại thay vì sao chép nguyên văn. BERTScore F1 ~0.76 là khá cao, phản ánh mức độ tương đồng ngữ nghĩa tốt giữa bản tóm tắt và văn bản gốc, ngay cả khi cách diễn đạt khác nhau. Chỉ số Precision cao hơn Recall, cho thấy nội dung sinh ra có độ chính xác cao, ít thông tin nhiễu. Tuy nhiên, có thể chưa bao phủ toàn bộ thông tin quan trọng.

Kết quả thực nghiệm được phân tích theo từng cấp lớp cho thấy hệ thống có khả năng kiểm soát tốt nội dung tóm tắt theo trình độ người học. Cụ thể, các bản tóm tắt được sinh ra đều đảm bảo từ vựng nằm trong bộ từ điển tương ứng với từng cấp lớp, góp phần duy trì tính dễ hiểu và phù hợp với năng lực ngôn ngữ của học sinh. Bên cạnh đó, hệ thống hỗ trợ điều chỉnh độ dài đầu ra theo các mức ngắn, trung bình và dài, giúp linh hoạt đáp ứng nhu cầu sử dụng khác nhau. Ngoài ra, độ khó của văn bản tóm tắt được đánh giá dựa trên các tiêu chí như số lượng từ, số câu và tần suất xuất hiện của từ hiếm và nhiều thông số khác, từ đó quy đổi về thang đo cấp lớp tương ứng. Cách tiếp cận này cho phép người dùng dễ dàng đối chiếu và lựa chọn bản tóm tắt phù hợp với trình độ, đồng thời góp phần nâng cao tính ứng dụng của hệ thống trong bối cảnh giáo dục tiêu học.

Để đánh giá tính thực tiễn của hệ thống, chúng tôi tiến hành thu thập ý kiến từ giáo viên Đoàn Văn Lượng, công tác hơn 25 năm tại trường Tiểu học Đình Văn Phú (Tân Thạnh, Tây Ninh). Nhận xét của thầy: “Tôi thấy ứng dụng có ý tưởng tốt, giao diện đẹp, các bước thao tác dễ dùng để cho học sinh nhỏ tuổi dễ làm quen. Nội dung tóm tắt thì bám sát được mục tiêu bài học. Tôi thấy cũng rất có hiệu quả cho học sinh. Còn nói về độ khó theo lớp thì tôi thấy đôi khi nó dùng từ chưa sát với trình độ của học sinh nhỏ và còn lặp ý. Một vài chỗ cần giáo viên chỉnh lại, nhưng nhìn chung tôi cũng đánh giá tốt về ứng dụng.” Bảng 2 dưới đây là đánh giá các bản tóm tắt từng bài đọc theo khối lớp theo như nhận xét của thầy Đoàn Văn Lượng.

Bảng 2. Tổng hợp nhận xét của giáo viên theo từng bài đọc và phương pháp tóm tắt

Lớp	Bài đọc	Phương pháp	Nhận xét của giáo viên
1	Sáng tạo vì cuộc sống	Diễn giải	<i>Phần mở đầu nói “người lớn cần sáng tạo để giúp đời” tôi thấy dễ hiểu cho HS lớp 1, vì câu ngắn, ít học thuật. Nhưng đoạn tóm lại ý “ứng dụng ý tưởng trong khoa học kỹ thuật” nghe hơi xa với các em</i>
2	Kho báu	Diễn giải	<i>Nhìn chung tóm tắt tốt, nhưng phần kết học sinh lớp 2 sẽ không hiểu vì thiếu chi tiết (làm ruộng, bán hàng, mấy lần người cha đi xa...)</i>
3	Chuyện nhỏ trong lớp học	Diễn giải	<i>Phù hợp với độ khó học sinh lớp 3, phù hợp 100% về độ rõ nhân vật/sự kiện</i>

Lớp	Bài đọc	Phương pháp	Nhận xét của giáo viên
4	Chiếc đèn lồng	Trích xuất	<i>Sẽ phù hợp với học sinh lớp 4 nếu có giáo viên kèm theo, nếu các em tự đọc sẽ hơi khó hiểu vì nó thiếu câu nối</i>
4	Chiếc đèn lồng	Diễn giải	<i>Phù hợp với độ khó học sinh lớp 4</i>
5	Giai thoại về bản xônát Ánh trăng	Trích xuất	<i>Giữ được câu nói về “sáng tác nổi tiếng của Beethoven” và “cảm hứng dưới ánh trăng” phù hợp với học sinh lớp 5</i>
5	Giai thoại về bản xônát Ánh trăng	Diễn giải	<i>Nội dung diễn giải khá phù hợp cho học sinh lớp 5 luyện đọc với nêu thông điệp</i>

4.4. Bàn luận kết quả

Nghiên cứu này đã đề xuất và triển khai một hệ thống tóm tắt văn bản tiếng Việt cho học sinh tiểu học dựa trên kiến trúc đa tác tử kết hợp giữa mô hình học sâu và cơ chế kiểm soát đầu ra theo cấp lớp. Về mặt công nghệ, hệ thống tích hợp các mô hình ngôn ngữ đã được tinh chỉnh cho hai hướng tiếp cận trích xuất và diễn giải, cơ chế tối ưu siêu tham số bằng Optuna, bộ từ điển phân tầng theo khối lớp, các chỉ số đánh giá độ khó, cùng các thang đo đánh giá tự động như ROUGE và BERTScore. Kiến trúc triển khai theo hướng phân tách frontend (React), backend (Spring Boot) và AI Service độc lập (Flask-based MAS) giúp đảm bảo tính mô-đun và khả năng mở rộng. Về lợi ích, mô hình MAS cho phép phân tách rõ vai trò giữa các tác tử: sinh tóm tắt, đánh giá, phân loại ý định, chuẩn hóa đầu vào và điều phối xử lý. Cách tiếp cận này giúp tăng tính linh hoạt, dễ bảo trì và có thể mở rộng thêm các chức năng như phản hồi học tập hoặc cá nhân hóa nội dung trong tương lai. Ngoài ra, việc xây dựng bộ dữ liệu chuyên biệt cho bậc tiểu học đóng góp một nguồn tài nguyên có giá trị cho nghiên cứu xử lý ngôn ngữ tự nhiên tiếng Việt trong lĩnh vực giáo dục.

Mặc dù hệ thống đạt được một số kết quả khả quan, vẫn còn tồn tại một số hạn chế cần được cải thiện. Trước hết, do được triển khai trên môi trường cục bộ, thời gian xử lý còn tương đối cao, đặc biệt với các văn bản dài trên 700 từ. Khi toàn bộ pipeline MAS hoạt động, bao gồm tóm tắt, đánh giá, sinh hình ảnh và cơ chế retry khi kết quả chưa đạt yêu cầu, tổng thời gian xử lý có thể dao động từ 2 đến 5 phút tùy theo từng trường hợp. Bên cạnh đó, ở phương pháp tóm tắt diễn giải, mô hình đôi khi vẫn sinh ra các câu có sai lệch nhẹ về nội dung so với văn bản gốc (hallucination), ảnh hưởng đến độ chính xác của bản tóm tắt. Về mặt dữ liệu, mặc dù đã được thu thập và làm sạch tương đối tốt, nguồn dữ liệu chủ yếu từ sách giáo khoa vẫn còn hạn chế về quy mô và độ đa dạng, có thể ảnh hưởng đến khả năng tổng quát hóa của mô hình trong các ngữ cảnh khác nhau. Những hạn chế này cho thấy hệ thống cần tiếp tục được tối ưu về hiệu năng, cải thiện chất lượng mô hình và mở rộng dữ liệu trong các nghiên cứu tiếp theo.

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Nghiên cứu đã xây dựng một hệ thống hỗ trợ tóm tắt văn bản tiếng Việt cho học sinh tiểu học dựa trên kiến trúc đa tác tử, kết hợp hai phương pháp trích xuất và diễn giải. Hệ thống tích hợp các mô hình học sâu đã tinh chỉnh, tối ưu siêu tham số bằng Optuna và đánh giá bằng ROUGE và BERTScore nhằm đảm bảo độ chính xác nội dung và tương đồng ngữ nghĩa. Bộ dữ liệu được xây dựng với cơ chế kiểm soát từ vựng theo cấp lớp, giúp bản tóm tắt phù hợp với năng lực đọc hiểu của học sinh lớp 1 - 5. Kết quả thực nghiệm cho thấy hệ thống đạt hiệu quả ổn định, cho thấy tiềm năng của kiến trúc đa tác tử trong bài toán tóm tắt văn bản giáo dục tiếng Việt.

Trong tương lai, hệ thống có thể được mở rộng theo nhiều hướng nhằm nâng cao khả năng ứng dụng thực tế. Trước hết, cần tiếp tục hoàn thiện hệ thống thông qua việc tinh chỉnh mô hình, sửa lỗi và kiểm nghiệm nhiều lần để đảm bảo độ ổn định trước khi triển khai rộng rãi. Một trong những hạn chế hiện tại là thời gian xử lý còn tương đối cao khi triển khai trên môi trường cục bộ; tuy nhiên, vấn đề này có thể được cải thiện đáng kể khi chuyển sang hạ tầng điện toán đám mây với GPU mạnh hơn, giúp giảm thời gian suy luận và tăng khả năng phục vụ đồng thời nhiều người dùng. Về khả năng mở rộng, kiến trúc đa tác tử cho phép dễ dàng bổ sung các thành phần mới như tăng cường cơ chế tương

tác và đánh giá chéo giữa các tác tử, xây dựng bộ nhớ tác tử nhằm lưu trữ ngữ cảnh và hỗ trợ cá nhân hóa. Ngoài ra, việc mở rộng bộ dữ liệu với dữ liệu gán nhãn từ giáo viên và phản hồi người dùng, cũng như tích hợp các chức năng như tạo câu hỏi đọc hiểu, phân tích độ khó văn bản hoặc sinh nội dung học tập sẽ góp phần hoàn thiện hệ thống. Trong bối cảnh giáo dục, hệ thống có tiềm năng tích hợp với các nền tảng học tập trực tuyến, hỗ trợ giáo viên và học sinh trong quá trình dạy và học. Đồng thời, đây cũng là tiền đề để phát triển theo hướng thương mại hóa, hướng tới xây dựng một nền tảng hỗ trợ học tập thông minh, phù hợp với xu hướng chuyển đổi số trong giáo dục.

LỜI CẢM ƠN

Nghiên cứu này được Trường Đại học Quốc tế Hồng Bàng cấp kinh phí thực hiện dưới mã số đề tài SVTC19.77.

TÀI LIỆU THAM KHẢO

- [1] A. A. Adepoju, "Summarization techniques used by primary school pupils in Kuje, Abuja, Nigeria," *International Journal of Research and Innovation in Social Science*, tập IX, số IIS, pp. 1839-1847, 2025.
- [2] Avivah, F. Hilmiyati và Khaeroni, "The Utilization of Text Summary Techniques to Improve Students' Reading Comprehension Skills," *Journal of Integrated Elementary Education*, tập 2, pp. 95-104, 2022.
- [3] P. Bulut và H. Akyol, "Primary school students and teachers' opinions about summary writing: A qualitative analysis," *Universal Journal of Educational Research*, tập 7, số 6, p. 1424-1434, 2019.
- [4] H. N. Long và H. T. Phú, "Xây dựng ứng dụng AI hỗ trợ tóm tắt bài đọc tiếng Việt cho học sinh tiểu học," *Tạp chí Khoa học Trường Đại học Quốc tế Hồng Bàng*, tập 38, số 7, pp. 173-182, 2025.
- [5] Ophuis-Cox và Rozendal, "The effects of summarization and factual retrieval practice on text comprehension and text retention in elementary education," *Journal of Experimental Psychology*, p. 258-267, 2024.
- [6] T. Shaik, X. Tao, Y. Li, C. Dann, J. McDonald, P. Redmond và L. Galligan, "A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis," tập 10, p. 56720-56739, 2022.
- [7] N. D. Quang, N. T. Thu và P. M. Hieu, "PhoBERT: Pre-trained language models for Vietnamese," *Findings of the Association for Computational Linguistics: EMNLP 2020*, p. 1037-1042, 2020.
- [8] S. Guo và E. Latif, "Using generative AI and multi-agents to provide automatic feedback," 2024.
- [9] L. Phan, H. Tran, H. Nguyen và T. Trinh, "ViT5: Pretrained text-to-text transformer for Vietnamese language generation," 2022.
- [10] S. V. Albrecht, F. Christianos và L. Schäfer, *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*, USA: Massachusetts Institute of Technology, 2024.
- [11] H. Y. Durak, F. Egin và A. Onan, "Enhancing text summarization with AI: A multi-agent system and human comparison in educational contexts," *Thinking Skills and Creativity*, tập 59, 2024.
- [12] S. Kumar và A. Solanki, "ROUGE-SS: A New ROUGE Variant for Evaluation of Text Summarization," *International Conference on Computing, Communication and Automation (ICCCA) 2020*, p. 1-5, 2020.
- [13] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger và Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," trong *International Conference on Learning Representations (ICLR 2020)*, 2020.
- [14] I. Khoshnevis và S. Parvinnejad, "The effect of text summarization as a cognitive strategy on the achievement of male and female language learners' reading comprehension," *International Journal of Learning and Development*, p. 57-69, 2015.
- [15] Q. A. Nguyen, D. C. Can, H. Q. Le và M. V. Tran, "VLSP 2022 Abmusu Task Dataset: A resource for Vietnamese abstractive multi-document summarization," *International Journal of Asian Language Processing*, p. 1-18, 2023.