TRƯỜNG ĐẠI HỌC SƯ PHẠM THÀNH PHỐ HỒ CHÍ MINH

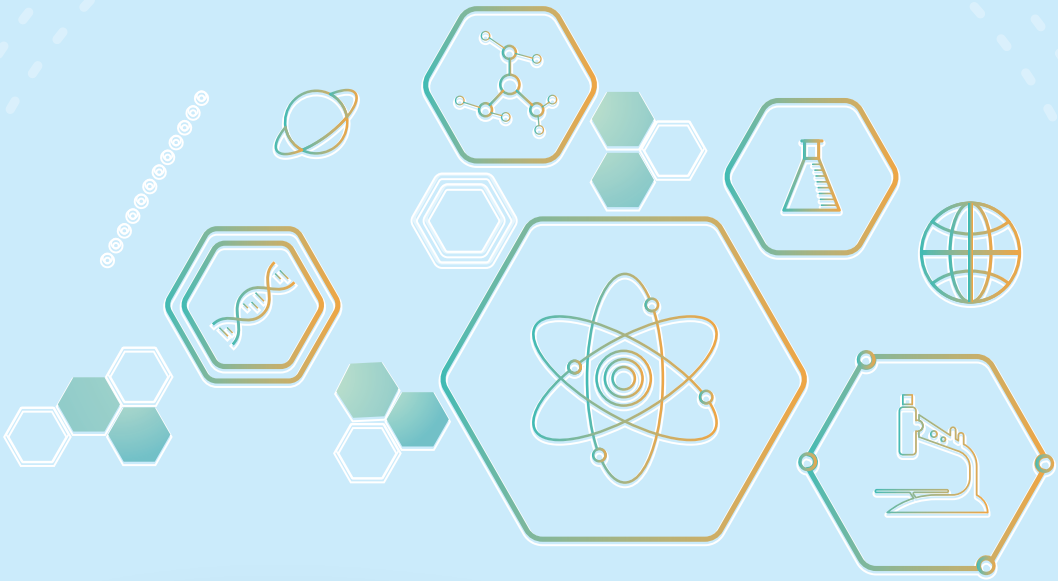
HỘI THẢO KHOA HỌC

CÁN BỘ TRẺ CÁC TRƯỜNG ĐẠI HỌC SƯ PHẠM

LẦN THỨ XI • NĂM 2026

VỚI CHỦ ĐỀ

NGHIÊN CỨU KHOA HỌC VÀ GIÁO DỤC TRONG BỐI CẢNH CHUYỂN ĐỔI SỐ VÀ HỘI NHẬP QUỐC TẾ



NHÀ XUẤT BẢN
ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH

XÂY DỰNG HỆ THỐNG TRÍ TUỆ NHÂN TẠO ĐA PHƯƠNG THỨC NHẰM HỖ TRỢ VIỆC RÈN LUYỆN VÀ ĐÁNH GIÁ NĂNG LỰC NÓI VÀ NGHE CẤP TIỂU HỌC

Hoàng Hữu Phước[1], Phan Thanh Lâm[1], Đinh Thị Phương Loan[1], Nguyễn Thị Hồng Ánh (TU3Q)[1], Nguyễn Thị Hồng Ánh (TU4E)[1]*

TÓM TẮT

Rèn luyện nói và nghe ở tiểu học đòi hỏi học sinh được thực hành nhiều lần, nhận phản hồi cụ thể và có cơ hội điều chỉnh sản phẩm. Trong điều kiện lớp học đông, giáo viên khó theo dõi đầy đủ từng lượt nói, từng biểu hiện nghe hiểu và tiến bộ của mỗi học sinh. Nghiên cứu phát triển hệ thống này trình bày quá trình xây dựng và triển khai VietSpeak AI, một hệ thống trí tuệ nhân tạo đa phương thức được thiết kế cho các nhiệm vụ nói và nghe tiếng Việt ở cấp tiểu học. Nghiên cứu kết hợp phân tích yêu cầu cần đạt của chương trình, thiết kế nhiệm vụ học tập, xây dựng rubric, phát triển phần mềm và phân tích dữ liệu triển khai. Đóng góp trọng tâm của nghiên cứu nằm ở việc kết nối trí tuệ nhân tạo đa phương thức với nhiệm vụ học tập cụ thể, rubric theo chương trình và chu trình luyện tập nhiều lần trong bối cảnh giáo dục tiểu học Việt Nam.

Từ khoá: đánh giá năng lực; nói và nghe; phản hồi; tiểu học; trí tuệ nhân tạo đa phương thức

DEVELOPING A MULTIMODAL ARTIFICIAL INTELLIGENCE SYSTEM TO SUPPORT SPEAKING AND LISTENING PRACTICE AND ASSESSMENT IN PRIMARY EDUCATION

ABSTRACT

Speaking and listening development in primary education requires repeated practice, specific feedback and opportunities for pupils to revise their performance. In large classes, teachers have limited time to monitor each pupil's speaking performance, listening comprehension and progress across multiple attempts. This system development study presents the development and implementation of VietSpeak AI, a multimodal artificial intelligence system designed for Vietnamese speaking and listening tasks in primary education. The study combines analysis of curriculum

[1] Trường Đại học Sư phạm, Đại học Huế; *Email: hhphuoc@hueuni.edu.vn

requirements, learning-task design, rubric development, software development and analysis of implementation data. The main contribution of the study is the integration of multimodal artificial intelligence with curriculum-aligned learning tasks, grade-specific rubrics and repeated practice in the context of Vietnamese primary education.

Keywords: competency assessment; speaking and listening; feedback; primary education; multimodal artificial intelligence

1. Giới thiệu

Năng lực nói và nghe hình thành qua thực hành. Một học sinh muốn kể được câu chuyện rõ ràng cần có cơ hội kể, nhận phản hồi, điều chỉnh và kể lại. Một học sinh muốn nghe hiểu cần nhận diện thông tin, nắm được ý chính, phản hồi đúng nội dung và tham gia phù hợp vào lượt trao đổi. Vì vậy, dạy học nói và nghe ở tiểu học cần được tổ chức như một chu trình liên tục gồm thực hiện, nhận xét, sửa chữa và thực hiện lại. Định hướng này phù hợp với yêu cầu phát triển năng lực giao tiếp trong Chương trình giáo dục phổ thông 2018 (Bộ Giáo dục và Đào tạo, 2018). Ở cấp tiểu học, các nhiệm vụ nói và nghe được triển khai qua nhiều dạng hoạt động như kể chuyện, trình bày, giới thiệu, nghe và trả lời, trao đổi, hỏi đáp. Mỗi dạng nhiệm vụ đòi hỏi những tiêu chí đánh giá khác nhau. Khi kể chuyện, học sinh cần bảo đảm nội dung, trình tự và cách diễn đạt. Khi giới thiệu một sự vật hoặc địa phương, học sinh cần xác định đúng chủ đề, lựa chọn ý, sắp xếp thông tin và trình bày dễ theo dõi. Khi nghe và trả lời, học sinh phải thể hiện mức độ hiểu thông tin qua câu trả lời hoặc lượt đáp phù hợp. Trong bài báo, năng lực nói và nghe được hiểu là khả năng tạo lập lời nói, hiểu thông tin nghe được và phản hồi phù hợp với nhiệm vụ giao tiếp; điểm của một video phản ánh những biểu hiện quan sát được trong nhiệm vụ đó.

Thực tế lớp học đặt ra một trở ngại lớn cho quá trình này. Trong một lớp có nhiều học sinh, thời lượng trên lớp thường chưa đủ để mỗi em thực hiện nhiều lượt nói và nhận phản hồi riêng. Giao nhiệm vụ bằng video giúp mở rộng thời gian luyện tập tại nhà, song số lượng video rất lớn tạo áp lực cho giáo viên. Giáo viên phải xem bài, nghe nội dung, kiểm tra mức độ hoàn thành, ghi nhận xét, lưu kết quả và theo dõi sự thay đổi qua các lần thực hiện. Nếu học sinh luyện nhiều lượt, khối lượng công việc tăng theo số lượng sản phẩm. Nguyễn Thị Quế (2024) cũng lưu ý những bất cập trong dạy học và đánh giá nói, nghe ở tiểu học. Tương tự, khảo sát 30 giáo viên tại 5 trường tiểu học ở Huế tháng 9/2025 cho thấy 24/30 người (80%) gặp khó khăn khi phản hồi riêng cho từng học sinh và 22/30 người (73,3%) có nhu cầu hỗ trợ xử lí bài nói. Đó là động lực để triển khai VietSpeak AI.

Các nghiên cứu về rubric đã cung cấp cơ sở để cụ thể hoá yêu cầu học tập và xây dựng phản hồi theo tiêu chí. Brookhart (2013) nhấn mạnh vai trò của rubric trong đánh giá thường xuyên và định hướng cải thiện sản phẩm. Ở Việt Nam, các nghiên cứu về năng lực nói và nghe cũng đang góp phần mô tả rõ hơn yêu cầu đối với học sinh tiểu học, trong đó có khung mô tả năng lực nói và nghe tiếng Việt của học sinh lớp 3 (Nguyễn Thị Quế, 2024). Tuy vậy, rubric chỉ phát huy hiệu quả khi được gắn với nhiệm vụ cụ thể, được sử dụng nhất quán và được chuyển hoá thành hướng dẫn mà học sinh có thể thực hiện ở lần luyện tiếp theo.

Ở bình diện công nghệ, các hệ thống như SpeechRater đã đặt nền móng cho việc chấm điểm tự động đối với lời nói, đặc biệt trong học ngoại ngữ (Zechner và cộng sựvà cộng sự, 2009). MACH cho thấy khả năng kết hợp giọng nói, hình ảnh và các tín hiệu giao tiếp để hỗ trợ người học trong hoạt động hội thoại (Hoque và cộng sự, 2013). Những hướng tiếp cận này mở ra tiềm năng sử dụng trí tuệ nhân tạo trong luyện nói và giao tiếp. Tuy nhiên, bối cảnh nói và nghe tiếng Việt ở tiểu học có những yêu cầu riêng. Hệ thống cần hiểu nhiệm vụ theo từng khối lớp, đối chiếu sản phẩm với tiêu chí phù hợp, xử lí sự đa dạng của giọng địa phương và tạo ra phản hồi gần với ngôn ngữ học sinh có thể hiểu, có thể làm theo. Các kết quả của SpeechRater và MACH thuộc bối cảnh tiếng Anh và người học khác với học sinh tiểu học Việt Nam, nên chỉ được dùng để tham khảo hướng thiết kế. Huu và cộng sự (2023) nghiên cứu nhận diện lỗi phát âm tiếng Việt trên dữ liệu trẻ em và tách người nói giữa các tập huấn luyện, phát triển, kiểm thử. Nghiên cứu này gợi ý cách kiểm chứng công nghệ lời nói để phát triển độ chính xác của VietSpeak AI.

Khoảng trống quan trọng nằm ở mối liên hệ giữa phân tích tự động và hoạt động dạy học hằng ngày. Một công cụ nhận dạng lời nói hoặc chấm điểm riêng lẻ chưa giải quyết được toàn bộ bài toán. Giá trị sư phạm của hệ thống được quyết định bởi khả năng trả lời bốn câu hỏi:

Học sinh được giao nhiệm vụ gì?

Sản phẩm cần đạt những yêu cầu nào?

Hệ thống dựa vào những bằng chứng nào để nhận xét?

Học sinh cần làm gì ở lần luyện tiếp theo?

VietSpeak AI được xây dựng để giải quyết đồng thời bốn yêu cầu này. Hệ thống phân tích video học sinh gửi lên, đối chiếu dữ liệu với rubric theo lớp và dạng bài, tạo phản hồi sơ bộ, lưu các lượt thực hiện và đưa giáo viên vào vị trí

kiểm tra, điều chỉnh, duyệt kết quả. Cách thiết kế này đặt trí tuệ nhân tạo vào trong quy trình đánh giá thường xuyên, thay vì tách công nghệ khỏi hoạt động dạy học.

Nghiên cứu này có bốn đóng góp chính. Thứ nhất, nghiên cứu xây dựng mô hình nhiệm vụ và rubric dành riêng cho rèn luyện nói và nghe tiếng Việt ở tiểu học, gồm 15 dạng nhiệm vụ và 45 bộ tiêu chí. Thứ hai, nghiên cứu phát triển quy trình phân tích đa phương thức, kết hợp nội dung lời nói, âm thanh và hình ảnh để tạo phản hồi có dẫn chứng. Thứ ba, nghiên cứu thiết kế chu trình người học thực hiện, hệ thống tạo phản hồi sơ bộ, giáo viên kiểm duyệt và người học luyện lại. Thứ tư, nghiên cứu báo cáo dữ liệu vận hành ở quy mô 267 trường và 1.042 giáo viên, qua đó mô tả quy mô sử dụng và những kết quả vận hành cần tiếp tục kiểm chứng.

Mục tiêu của bài báo là trình bày quá trình xây dựng VietSpeak AI, mô tả cách hệ thống được sử dụng trong dạy học nói và nghe, phân tích kết quả triển khai và xác định những hướng phát triển tiếp theo đối với trí tuệ nhân tạo trong giáo dục tiểu học.

2. Đối tượng và phương pháp nghiên cứu

2.1. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu là VietSpeak AI, một hệ thống hỗ trợ rèn luyện và đánh giá nói và nghe tiếng Việt cho học sinh lớp 1 đến lớp 5. Hệ thống được xây dựng và triển khai trong giai đoạn từ tháng 01/2023 đến tháng 06/2026. Quá trình này kết hợp bốn hoạt động: phân tích yêu cầu dạy học, thiết kế rubric, phát triển phần mềm và tổ chức sử dụng tại trường tiểu học. Đây là nghiên cứu phát triển hệ thống kết hợp mô tả dữ liệu vận hành ban đầu.

Mỗi sản phẩm học tập được gắn với học sinh, lớp, nhiệm vụ, khối lớp và bộ tiêu chí tương ứng. Cách tổ chức dữ liệu này cho phép theo dõi một học sinh qua nhiều lần thực hiện, so sánh sản phẩm trước và sau phản hồi, đồng thời hỗ trợ giáo viên xem lại quá trình thay đổi thay vì chỉ xem một sản phẩm đơn lẻ.

VietSpeak AI được phát triển theo nguyên tắc liên kết ba thành phần:

Về nhiệm vụ học tập: xác định học sinh cần nói, nghe, kể hoặc trao đổi về nội dung gì.

Về rubric đánh giá: xác định sản phẩm đạt yêu cầu ở những phương diện nào.

Về phản hồi và luyện lại: chuyển kết quả phân tích thành hướng dẫn cụ thể cho lượt thực hiện tiếp theo.

Quá trình xây dựng bắt đầu từ yêu cầu cần đạt của chương trình. Nhóm nghiên cứu xác định biểu hiện cần quan sát ở từng dạng nhiệm vụ, mô tả các mức độ hoàn thành và lựa chọn dữ liệu cần trích xuất từ video. Từ đó, nhóm thiết kế quy trình phân tích, giao diện phản hồi, chức năng lưu bài và quy trình giáo viên kiểm duyệt.

2.2. Thiết kế nhiệm vụ và rubric

Hệ thống gồm hai nhóm nhiệm vụ chính: nói và nghe, kể chuyện. Các nhiệm vụ được cụ thể hoá thành 15 dạng bài theo khối lớp và yêu cầu cần đạt. Học liệu tích hợp 12 chủ đề văn hoá Huế, tạo bối cảnh gần gũi để học sinh luyện trình bày về cảnh vật, con người, món ăn, sinh hoạt, văn hoá và tình cảm đối với quê hương.

Rubric (bảng tiêu chí kèm mô tả các mức độ thực hiện) được xây dựng theo nguyên tắc mỗi tiêu chí phải gắn với một biểu hiện có thể quan sát hoặc đối chiếu trong sản phẩm. Chẳng hạn, với nhiệm vụ kể chuyện, hệ thống cần xem xét các sự việc được kể, thứ tự xuất hiện và mức độ rõ ràng của lời kể. Với nhiệm vụ giới thiệu hoặc trình bày, hệ thống cần xem nội dung có bám chủ đề hay không, các ý có được sắp xếp hợp lí hay không, tốc độ và âm lượng có phù hợp với người nghe hay không. Với nhiệm vụ nghe và trả lời, câu trả lời cần được đối chiếu với thông tin hoặc câu hỏi mà học sinh đã nghe. Ví dụ, ở lớp 3, nhiệm vụ kể lại truyện đã nghe:

Tiêu chí trình tự:

Mức 2 = kể đúng thứ tự ba sự việc chính;

Mức 1 = đảo vị trí một sự việc nhưng còn nhận ra diễn biến;

Mức 0 = không xác định được trình tự.

Tiêu chí nghe hiểu:

Mức 2 = trả lời đúng hai câu hỏi về thông tin nghe được;

Mức 1 = đúng một câu;

Mức 0 = chưa trả lời đúng câu nào.

Việc đánh giá nghe hiểu cần xác định rõ học liệu đầu vào, câu hỏi và đáp án đối chiếu. Ví dụ, bản ghi 20 giây “Sáng chủ nhật, Lan cùng bố đi chợ Đông Ba mua hoa tặng bà”; hai câu hỏi “Lan đi với ai?” và “Lan mua gì?”. Đáp án tương ứng “với bố” và “hoa”; hệ thống AI chấp nhận cách diễn đạt tương đương, không chấm độ lưu loát thay cho nghe hiểu.

Các nghiên cứu trên học sinh tiểu học cho thấy điều kiện âm thanh ảnh hưởng đến nghe hiểu (Rudner và cộng sự, 2018), trong khi công cụ đo cần phù hợp với bối cảnh lớp học (Mealings và cộng sự, 2025).

Bảng 1. *Yêu cầu đánh giá ở một số nhiệm vụ nói và nghe*

Nhiệm vụ	Yêu cầu trọng tâm	Dữ liệu và bằng chứng cần đối chiếu
Kể chuyện	Kể đúng nội dung, bảo đảm trình tự, diễn đạt rõ ràng	Các sự việc được kể, thứ tự xuất hiện, từ nối, giọng kể
Giới thiệu, trình bày	Nói đúng chủ đề, lựa chọn và sắp xếp ý, diễn đạt dễ hiểu	Nội dung, quan hệ giữa các ý, tốc độ, âm lượng, cách ngắt
Nghe và trả lời	Hiểu thông tin, trả lời đúng điều được hỏi	Nội dung hoặc câu hỏi đã nghe, câu trả lời, mức độ phù hợp
Trao đổi, hỏi đáp	Đáp đúng ý, biết hỏi lại, chờ lượt và duy trì chủ đề	Các lượt lời, quan hệ giữa câu hỏi và câu đáp, thời điểm phản hồi

Việc phân biệt rubric theo nhiệm vụ giúp hệ thống tập trung vào đúng mục tiêu học tập. Một học sinh kể lưu loát nhưng đảo thứ tự sự việc cần nhận phản hồi về tổ chức nội dung. Một học sinh nói chậm nhưng trình bày đủ ý và rõ ràng cần được ghi nhận về nội dung trước khi được hướng dẫn điều chỉnh tốc độ. Với giọng địa phương, hệ thống ưu tiên khả năng truyền đạt và nghe hiểu. Giáo viên kiểm tra những đoạn nhận dạng chưa rõ trước khi xác định đó là lỗi của học sinh.

2.3. Kiến trúc phân tích đa phương thức

VietSpeak AI sử dụng ba nguồn thông tin từ video: nội dung lời nói, âm thanh và hình ảnh. Mỗi nguồn đảm nhận một nhóm câu hỏi riêng:

Về nội dung lời nói, hệ thống làm rõ: học sinh nói gì, có bám nhiệm vụ hay không, các ý được phát triển và liên kết như thế nào.

Về âm thanh, người nghe có thể nghe rõ hay không, tốc độ, âm lượng và khoảng ngắt có hỗ trợ việc truyền đạt hay không.

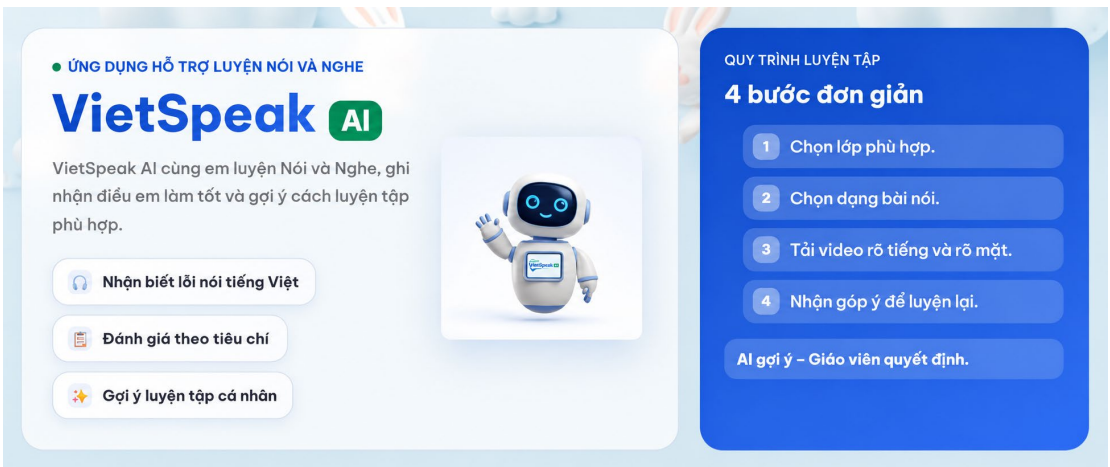
Về hình ảnh, học sinh trình bày với tư thế, hướng nhìn và cử chỉ như thế nào.

Bốn nhóm công nghệ được kết hợp là nhận dạng tiếng nói tự động (Automatic Speech Recognition, ASR), xử lí tín hiệu âm thanh số (Digital Signal Processing, DSP), xử lí ngôn ngữ tự nhiên (Natural Language Processing, NLP) và thị giác máy tính (Computer Vision, CV).Trong đó, ASR dùng Whisper large-v3; xử lí âm thanh dùng librosa 0.10.2; theo dõi điểm cơ thể dùng MediaPipe 0.10.14; nhóm tự viết mô-đun ánh xạ kết quả vào rubric và quản lí lượt luyện.

Trước hết, công cụ nhận dạng tiếng nói chuyển phần lời nói trong video thành văn bản. Bản chép lời được sử dụng để phân tích nội dung, từ khoá, các ý chính, trình tự sự việc và quan hệ giữa các ý. Với bài kể chuyện, hệ thống có thể tìm dấu hiệu về sự xuất hiện của các sự việc và từ chỉ trình tự. Với bài giới thiệu, hệ thống tập trung vào sự phù hợp giữa nội dung lời nói và chủ đề được giao. Trong quá trình đối chiếu, lỗi chép lời tự động cần được phân biệt với lỗi phát âm hoặc diễn đạt của học sinh.

Phần xử lí âm thanh phân tích các đặc điểm liên quan đến khả năng nghe rõ và theo dõi lời nói. Một bài nói có thể đủ ý nhưng vẫn cần luyện thêm nếu âm lượng quá nhỏ, tốc độ quá nhanh hoặc khoảng ngắt khiến người nghe khó theo dõi. Ngược lại, một học sinh nói chậm nhưng phát âm và diễn đạt rõ vẫn cần được ghi nhận đúng mức. Cách phân tích này giúp phản hồi đi vào nguyên nhân cụ thể thay vì quy tất cả khó khăn về một điểm số chung.

Phần xử lí hình ảnh nhận diện một số dấu hiệu có thể quan sát như hướng nhìn, tư thế và cử chỉ. Những dấu hiệu này cung cấp thêm thông tin về cách học sinh trình bày. Hướng nhìn phù hợp có thể hỗ trợ giao tiếp, tư thế ổn định có thể giúp bài trình bày rõ ràng hơn, cử chỉ phù hợp có thể làm nổi bật nội dung. Tuy nhiên, các dấu hiệu hình ảnh chỉ được sử dụng như bằng chứng hỗ trợ. Hướng nhìn về phía người nói chưa đủ để kết luận học sinh đã nghe hiểu. Khả năng nghe hiểu cần được xác định qua việc học sinh trả lời, nhắc lại, hỏi lại hoặc đáp ứng đúng nội dung đã nghe.



Hình 1. *Giao diện giới thiệu và các thao tác chính của VietSpeak AI*

Cấu trúc này tạo ra mô hình người và trí tuệ nhân tạo cùng tham gia đánh giá. Hệ thống đảm nhận các thao tác lặp lại như chuyển lời nói thành văn bản,

tập hợp dẫn chứng và tạo dự thảo nhận xét. Giáo viên đưa ra phán đoán chuyên môn ở những trường hợp cần đọc ngữ cảnh, kiểm tra lỗi nhận dạng hoặc xem xét biểu hiện của học sinh trong mối liên hệ với nhiệm vụ.

2.4. Quy trình luyện tập và phản hồi

Để kiểm soát phản hồi trước khi học sinh tiếp nhận, quy trình luyện tập cần tuân thủ sáu bước:

1. Giáo viên giao nhiệm vụ và lựa chọn rubric.
2. Học sinh chuẩn bị, ghi hình và gửi bài.
3. Hệ thống kiểm tra chất lượng tệp.
4. Hệ thống phân tích video và tạo phản hồi sơ bộ.
5. Giáo viên xem dẫn chứng, điều chỉnh và duyệt phản hồi trước khi gửi cho học sinh.
6. Học sinh đọc hoặc nghe phản hồi đã được duyệt, luyện lại và gửi sản phẩm để giáo viên đánh giá.

Các bài nộp và phản hồi được lưu theo từng học sinh. Nhờ đó, giáo viên có thể xem một sản phẩm tại một thời điểm cụ thể hoặc theo dõi quá trình thay đổi qua nhiều lượt luyện. Khi phát hiện nhận xét sai, giáo viên cần thu hồi hoặc sửa phản hồi; học sinh và phụ huynh cần có kênh yêu cầu đánh giá lại. Mức độ thực hiện cơ chế kiểm duyệt này cần được xác nhận bằng nhật kí xử lí.

Phản hồi trong hệ thống được thiết kế theo hai thành phần. Thành phần thứ nhất ghi nhận điểm học sinh đã thực hiện tốt. Thành phần thứ hai chỉ ra một việc cụ thể cần điều chỉnh. Cách thiết kế này phù hợp với quan điểm của Hattie và Timperley (2007) về phản hồi định hướng cho quá trình học tập. Học sinh cần biết mình đang ở đâu, cần đạt điều gì và cần thực hiện bước tiếp theo ra sao.

Ví dụ, với một bài kể chuyện chưa bảo đảm trình tự, phản hồi có thể viết: “Em đã kể được các sự việc chính. Ở lần kể tiếp theo, em hãy kể việc xảy ra trước, sau đó kể việc xảy ra sau; dùng từ ‘sau đó’ khi chuyển sang sự việc tiếp theo.” Phản hồi này cung cấp cả điểm mạnh, vấn đề cần sửa và cách sửa. Học sinh có thể áp dụng ngay trong lượt luyện sau.

Với bài nói có nội dung đầy đủ nhưng âm lượng thấp, phản hồi cần tập trung vào việc nói đủ nghe, giữ khoảng cách phù hợp với thiết bị và kiểm tra âm thanh trước khi gửi. Với bài trình bày có nhiều ý nhưng liên kết lỏng, phản hồi cần

hướng học sinh nhóm các ý theo chủ đề, sử dụng từ nối và xác định phần mở đầu, phần phát triển, phần kết thúc.

Quy trình phản hồi cũng phân bổ lại công việc giữa học sinh, hệ thống và giáo viên. Học sinh sử dụng phản hồi đã được giáo viên duyệt để tự sửa trong các lượt luyện trung gian. Hệ thống thực hiện các thao tác phân tích và tạo bản nhận xét ban đầu. Giáo viên kiểm tra phản hồi trước khi gửi, chú ý những trường hợp hệ thống nhận diện chưa sát và những đoạn âm thanh chưa rõ; sản phẩm cuối được đánh giá theo rubric. Phụ huynh chỉ hỗ trợ thao tác thiết bị và điều kiện ghi hình, không gợi ý thay nội dung hoặc làm thay bài của học sinh.

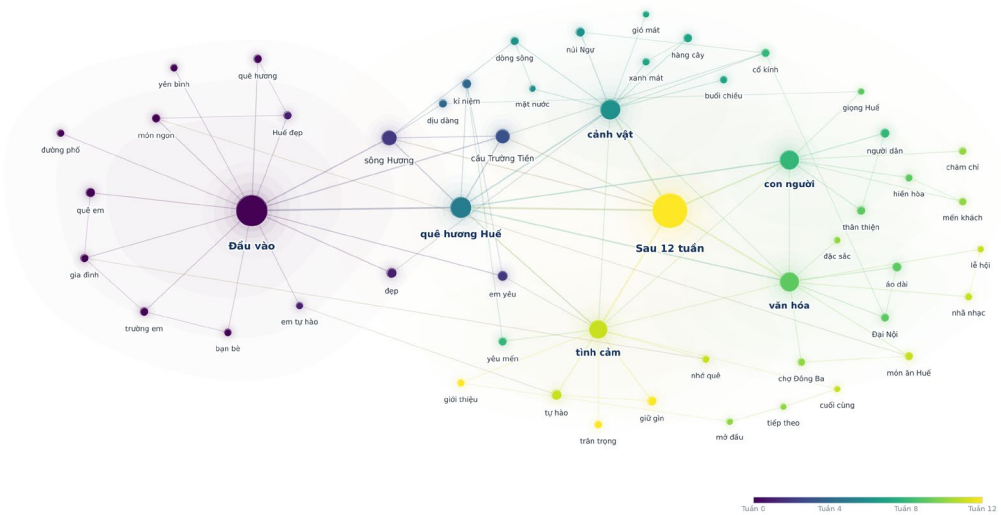
2.5. Dữ liệu và phương pháp đánh giá hệ thống

Dữ liệu nghiên cứu gồm nhiệm vụ học tập, bộ rubric, hơn 3.500 video được báo cáo trong hồ sơ chấm đối chiếu giữa hệ thống và Hội đồng chuyên môn, cùng số liệu tổng hợp trong quá trình triển khai. Phạm vi triển khai gồm 267 trường và 1.042 giáo viên. Tại 5 trường ở Huế, từ 9/2025 đến 5/2026, nhóm thu 3.600 video của 600 học sinh; mỗi khối 1–5 có 120 học sinh và 720 video. Mẫu được chọn thuận tiện từ các lớp đồng ý tham gia. Tập phát triển gồm 400 học sinh/2.400 video; tập điều chỉnh 100 học sinh/600 video; tập kiểm thử 100 học sinh/600 video. Không học sinh nào xuất hiện ở hai tập; không điều chỉnh mô hình hoặc câu lệnh dựa trên tập kiểm thử. Mỗi video do hai giáo viên chấm độc lập, chưa xem kết quả AI; bất đồng được người thứ ba phân xử. Hệ số nhất quán được tính trước phân xử.

Về cách diễn giải, "video đủ điều kiện" cần được xác định bằng tiêu chí đầu vào công bố trước, gồm tệp mở được, âm thanh nghe rõ và có đủ phần thực hiện nhiệm vụ; tiêu chí hình ảnh chỉ áp dụng khi nhiệm vụ cần quan sát hành vi trình bày. Tỉ lệ phân tích cần tính bằng số video nhận được kết quả phân tích chia cho tổng số video đủ điều kiện, nhân 100%. Cụ thể, trong 3.600 video đáp ứng tiêu chí đầu vào, 3.438 video nhận đủ đầu ra phân tích, nên tỉ lệ hoàn thành là 3.438/3.600 × 100 = 95,5%. Trên 600 video kiểm thử với 5 tiêu chí/video, AI và mức chấm tham chiếu trùng hoàn toàn ở 2.781/3.000 cặp video–tiêu chí, tương đương 92,7%.

Nghiên cứu cũng sử dụng một bài nói cùng chủ đề "Nói về quê hương" của học sinh lớp 3 ở thời điểm đầu vào và sau 12 tuần để minh hoạ sự thay đổi trong phát triển và tổ chức ý. HS03, lớp 3, thực hiện cùng nhiệm vụ vào tuần 0 và tuần 12, thời lượng tối đa 2 phút, không có người lớn gợi ý; bản chép lời được kiểm

tra với video trước khi dựng mạng ý. Đây là ví dụ minh hoạ cho chức năng theo dõi quá trình, chưa phải thiết kế thực nghiệm để xác định tác động nhân quả của hệ thống.



Hình 2. *Mạng liên kết ý trong bài nói "Nói về quê hương" trước và sau 12 tuần*

2.6. Nguyên tắc sử dụng dữ liệu học sinh

VietSpeak AI xử lí hình ảnh, tiếng nói và sản phẩm học tập của học sinh trong phạm vi nhiệm vụ giáo dục. Việc triển khai cần có sự đồng thuận của phụ huynh, phân quyền truy cập theo vai trò và quy định thời hạn lưu trữ. Dữ liệu cá nhân cần được quản lí phục vụ hoạt động học tập, không chuyển giao cho mục đích quảng cáo. Phụ huynh kí phiếu đồng thuận riêng cho học tập và nghiên cứu; học sinh được giải thích theo lứa tuổi. Video lưu trên máy chủ của trường tại Việt Nam trong 6 tháng; giáo viên chỉ truy cập lớp được phân công. Phụ huynh gửi yêu cầu xoá qua đầu mối nghiên cứu và được xử lí trong 7 ngày. Không gửi video hoặc bản chép lời cho nhà cung cấp bên ngoài để huấn luyện mô hình. Mục đích học tập cần được tách khỏi mục đích huấn luyện mô hình và nghiên cứu; mỗi mục đích phải được giải thích rõ khi xin đồng thuận. Cần công bố nơi và thời hạn lưu trữ, người được truy cập, cách yêu cầu xoá dữ liệu, phạm vi chuyển dữ liệu cho nhà cung cấp AI và tình trạng xem xét đạo đức nghiên cứu. Đây là các điều kiện cần kiểm chứng bằng hồ sơ thực hiện, phù hợp với định hướng bảo vệ dữ liệu, minh bạch và trách nhiệm giải trình của UNICEF (2025).

Nhà trường cần thống nhất quy trình sử dụng trong tổ chuyên môn, từ cách giao bài, cách kiểm tra chất lượng video, cách đọc phản hồi đến cách giáo viên

điều chỉnh kết quả. Việc sử dụng công nghệ cần đi cùng quy định sử dụng rõ ràng và năng lực chuyên môn của người dạy.

3. Kết quả và thảo luận

3.1. Mô hình nhiệm vụ và rubric theo chương trình

Kết quả đầu tiên của nghiên cứu là mô hình hoá yêu cầu nói và nghe thành cấu trúc nhiệm vụ có thể triển khai trên hệ thống. VietSpeak AI hiện gồm 15 dạng nhiệm vụ và 45 bộ rubric theo lớp và dạng bài. Đây là thành phần nền tảng của hệ thống, bởi AI chỉ có thể tạo phản hồi có ý nghĩa khi biết bài nộp đang được thực hiện cho nhiệm vụ nào và cần đạt những tiêu chí nào.

Việc tích hợp 12 chủ đề văn hoá Huế tạo ra ngữ cảnh học tập gần với đời sống của học sinh. Các chủ đề về cảnh vật, con người, món ăn, sinh hoạt và văn hoá địa phương cung cấp chất liệu để trẻ kể, giới thiệu, trình bày và trao đổi. Nội dung địa phương vì vậy được sử dụng như một môi trường phát triển ngôn ngữ, thay vì chỉ xuất hiện như phần minh hoạ bên ngoài chương trình.

Mối liên hệ giữa nhiệm vụ, rubric và dữ liệu phân tích được thể hiện trong Bảng 2.

Bảng 2. *Cấu trúc liên kết giữa nhiệm vụ học tập, rubric và dữ liệu phân tích*

Thành phần	Câu hỏi định hướng	Ví dụ dữ liệu sử dụng
Nhiệm vụ	Học sinh cần thực hiện hoạt động gì?	Kể chuyện, giới thiệu quê hương, nghe và trả lời
Rubric	Sản phẩm cần đạt yêu cầu nào?	Đúng nội dung, đúng trình tự, rõ lời, phù hợp lượt đáp
Nội dung lời nói	Học sinh đã nói những gì?	Bản chép lời, ý chính, từ nối, sự việc
Âm thanh	Người nghe có thể theo dõi lời nói hay không?	Tốc độ, âm lượng, khoảng ngắt
Hình ảnh	Cách trình bày có hỗ trợ giao tiếp hay không?	Hướng nhìn, tư thế, cử chỉ
Phản hồi	Học sinh cần sửa gì ở lượt sau?	Hướng dẫn cụ thể gắn với lỗi và cách sửa

Cấu trúc này giúp phân biệt lỗi về nội dung với lỗi về cách trình bày. Học sinh có thể nói đủ ý nhưng cần sắp xếp lại. Học sinh có thể sắp xếp ý tốt nhưng cần điều chỉnh âm lượng. Học sinh có thể trả lời đúng nhưng cần chờ lượt hoặc hỏi lại để duy trì trao đổi. Khi mỗi biểu hiện được gắn với một tiêu chí và một loại dữ liệu, phản hồi trở nên cụ thể hơn.

3.2. Phân tích đa phương thức và vai trò của giáo viên

VietSpeak AI được thiết kế để kết hợp các nguồn dữ liệu thay vì sử dụng một tín hiệu duy nhất. Bản chép lời giúp xem xét nội dung. Âm thanh giúp xem xét khả năng truyền đạt. Hình ảnh bổ sung thông tin về cách trình bày. Ba nguồn dữ liệu này tạo thành hồ sơ phong phú hơn về sản phẩm nói. Tuy nhiên, nghiên cứu chưa có phép so sánh trên cùng tập kiểm thử giữa cấu hình đa phương thức với cấu hình chỉ dùng âm thanh hoặc bản chép lời; vì vậy, chưa thể lượng hoá giá trị tăng thêm của hình ảnh. Tính mới được xác định ở thiết kế tích hợp với nhiệm vụ và rubric, chưa phải ở thuật toán AI mới.

Tuy nhiên, hệ thống được thiết kế theo mô hình hỗ trợ giáo viên, không thay thế phán đoán sư phạm. Đây là lựa chọn quan trọng đối với các nhiệm vụ ngôn ngữ ở tiểu học. Cùng một biểu hiện âm thanh có thể cần được giải thích khác nhau tuỳ bối cảnh. Một đoạn nói có âm lượng nhỏ có thể xuất phát từ cách đặt thiết bị, điều kiện phòng hoặc đặc điểm giọng nói. Một câu trả lời ngắn có thể đúng với câu hỏi đóng nhưng chưa đủ với câu hỏi yêu cầu giải thích. Giáo viên cần xem dẫn chứng và ngữ cảnh trước khi duyệt kết quả.

Mô hình giáo viên kiểm duyệt cũng tạo cơ chế kiểm soát chất lượng. Khi hệ thống nhận dạng sai từ, bỏ sót nội dung hoặc phân tích nhầm đoạn âm thanh, giáo viên cần sửa nhận xét trước khi học sinh tiếp nhận. Những điều chỉnh này giúp kết quả cuối cùng gắn với phán đoán chuyên môn và bảo vệ học sinh trước những kết luận tự động thiếu chính xác.

3.3. Phản hồi cá nhân hoá và luyện tập nhiều lần

VietSpeak AI biến mỗi lần nộp bài thành một điểm trong quá trình học tập. Học sinh có thể xem mình đã hoàn thành tiêu chí nào, nhận một hướng dẫn ngắn cho lần sau và gửi sản phẩm mới. Giáo viên có thể xem lại chuỗi sản phẩm để nhận diện sự thay đổi. Cách tổ chức này phù hợp với đánh giá thường xuyên vì sự tiến bộ của học sinh theo Thông tư 27/2020/TT-BGDĐT. Trọng tâm của đánh giá chuyển từ việc ghi nhận một sản phẩm cuối sang việc cung cấp thông tin để học sinh cải thiện sản phẩm kế tiếp.

Giá trị của phản hồi nằm ở mức độ cụ thể. Nhận xét "em cần nói tốt hơn" ít có giá trị hành động. Nhận xét "em đã nêu đủ ba địa điểm; ở lần sau, em hãy nhóm các địa điểm theo cảnh vật và hoạt động, sau đó dùng 'tiếp theo' để chuyển ý" cung cấp một chiến lược rõ ràng. Học sinh có thể hiểu, thực hiện và kiểm tra lại. Phản hồi cá nhân hoá cũng giúp giáo viên phân hoá hỗ trợ. Một học sinh cần

bổ sung ý sẽ nhận hướng dẫn về phát triển nội dung. Một học sinh cần nói rõ hơn sẽ nhận hướng dẫn về âm lượng và tốc độ. Một học sinh cần cải thiện tương tác sẽ nhận hướng dẫn về chờ lượt, hỏi lại và đáp đúng câu hỏi. Cùng một nhiệm vụ có thể tạo ra các đường luyện tập khác nhau theo nhu cầu của từng em. "Phản hồi cá nhân hoá" ở đây là hướng dẫn được điều chỉnh theo nhiệm vụ, tiêu chí và biểu hiện của từng bài nộp, không đồng nghĩa với chẩn đoán toàn diện đặc điểm của học sinh.

3.4. Kết quả triển khai

Bảng 3 tổng hợp các chỉ số về xử lí video và công việc nhận xét bài. Các số liệu được giữ ở mức báo cáo vận hành ban đầu; cách tính, mẫu số và quy trình đo còn cần xác nhận trước khi sử dụng để kết luận về độ chính xác hoặc mức giảm công việc của giáo viên.

Bảng 3. *Một số số liệu vận hành ban đầu được báo cáo của VietSpeak AI*

Chỉ số	Số liệu báo cáo	Phạm vi phản ánh
Video nhận được phân tích	95,5%	Chưa công bố số video ở tử số, mẫu số và tiêu chí sàng lọc đã áp dụng
Mức tương đồng với Hội đồng chuyên môn	92,7%	Chưa xác định rõ đơn vị so sánh, ngưỡng tương đồng và tập kiểm thử độc lập
Thời gian xử lí trung bình được báo cáo	6 phút xuống 1,5 phút/video	Cần xác nhận cỡ mẫu, cách đo và việc tính đủ thời gian giáo viên kiểm duyệt
Phạm vi triển khai	267 trường; 1.042 giáo viên	Quy mô triển khai; không phải bằng chứng về hiệu quả phát triển năng lực

Theo số liệu tổng hợp ban đầu, thời gian xử lí được báo cáo là 6 phút khi nhận xét thủ công và 1,5 phút khi dùng hệ thống cho mỗi video; chênh lệch số học tương đương 75%. Với một lớp giả định có 35 học sinh, thời gian xử lí giảm từ 210 phút xuống 52,5 phút cho một lượt bài. Chênh lệch là 157,5 phút, tương đương hơn hai giờ rưỡi (2,5 giờ) cho mỗi lượt giao bài. Đây là phép quy đổi theo hai mức thời gian đã báo cáo; mức tiết kiệm thực tế chỉ có thể xác nhận khi hai điều kiện được đo với quy trình tương đương và tính đủ thời gian kiểm tra, sửa, duyệt của giáo viên.

Phép quy đổi này minh hoạ khả năng hỗ trợ tổ chức công việc. Nếu mỗi học sinh thực hiện năm lượt cho một nhiệm vụ, lớp 35 học sinh sẽ tạo ra 175 video. Khi giáo viên phải nhận xét đầy đủ tất cả sản phẩm, số lượng video tạo ra áp lực lớn. Khi hệ thống hỗ trợ các lượt luyện trung gian, giáo viên vẫn cần kiểm duyệt phản

hồi cho những lượt đó trước khi học sinh tiếp nhận, đồng thời đánh giá 35 sản phẩm cuối. Vì vậy, khối lượng kiểm duyệt phải được tính trên toàn bộ quy trình.

Tỉ lệ 95,5% được báo cáo phản ánh mức hoàn thành phân tích trong phạm vi đầu vào đã sàng lọc, chưa đủ để xác nhận hệ thống vận hành ổn định ở các điều kiện khác. Chỉ số 92,7% chưa đủ để kết luận mức phù hợp với đánh giá chuyên môn khi chưa làm rõ cách tính và tính độc lập của dữ liệu đối chiếu. Hai chỉ số này cần được hiểu cùng với quy trình giáo viên kiểm tra và duyệt kết quả. Mục đích của hệ thống là cung cấp bản phân tích ban đầu để giáo viên kiểm tra và sử dụng khi phù hợp.

3.5. Theo dõi sự phát triển nội dung qua thời gian

Trong trường hợp minh hoạ ở Hình 2, ở thời điểm đầu, các ý trong bài nói chủ yếu tập trung vào những chi tiết quen thuộc như gia đình, trường học, món ăn và cảnh đẹp. Sau 12 tuần, nội dung được mở rộng thành các nhóm rõ hơn về cảnh vật, con người, văn hoá và tình cảm. Các từ chỉ trình tự như "mở đầu", "tiếp theo", "cuối cùng" cũng xuất hiện.

Sự thay đổi này được thể hiện ở hai phương diện. Thứ nhất, nội dung bài nói phong phú hơn. Học sinh bổ sung thêm chi tiết và mở rộng phạm vi liên tưởng. Thứ hai, quan hệ giữa các ý rõ hơn. Các ý được nhóm theo chủ đề và nối với nhau bằng dấu hiệu trình tự.

Mạng liên kết ý giúp giáo viên quan sát quá trình phát triển bài nói theo cách trực quan. Giáo viên có thể nhận ra nhóm ý nào đã được mở rộng, nhóm ý nào còn ít chi tiết, phần nào thiếu liên kết và chỗ nào cần hướng dẫn thêm. Đây là một hướng sử dụng AI có thể hỗ trợ đánh giá quá trình, bởi hệ thống cung cấp dữ liệu để thảo luận về cách phát triển ý, thay vì chỉ trả về một điểm số. Ví dụ này chưa đủ để kết luận VietSpeak AI tạo ra sự tiến bộ về năng lực nói. Tuy nhiên, nó cho thấy hệ thống có khả năng lưu trữ và biểu diễn sự thay đổi của sản phẩm qua thời gian. Đây là điều kiện cần để các nghiên cứu tiếp theo theo dõi tiến bộ của nhiều học sinh bằng thiết kế có nhóm so sánh và các mốc đo rõ ràng.

3.6. Đóng góp đối với AI trong giáo dục

VietSpeak AI có ba đóng góp ở mức thiết kế ứng dụng đối với nghiên cứu và triển khai trí tuệ nhân tạo trong giáo dục.

Thứ nhất, hệ thống đưa AI vào một nhiệm vụ học tập cụ thể. Thay vì phân tích lời nói một cách khái quát, VietSpeak AI bắt đầu từ yêu cầu của từng dạng bài và từng khối lớp. Cùng một biểu hiện ngôn ngữ có thể được diễn giải khác

nhau trong nhiệm vụ kể chuyện, giới thiệu hoặc trao đổi. Việc gắn phân tích với nhiệm vụ giúp kết quả gần hơn với hoạt động dạy học thực tế.

Thứ hai, hệ thống kết hợp tự động hoá với kiểm soát chuyên môn. Hệ thống thực hiện các thao tác xử lí dữ liệu có tính lặp lại, tạo phản hồi sơ bộ và cung cấp dẫn chứng. Giáo viên kiểm tra, điều chỉnh và duyệt. Mô hình này phù hợp với các bài toán đánh giá năng lực, nơi kết quả cần dựa trên dữ liệu nhưng vẫn cần đọc hiểu bối cảnh và đưa ra phán đoán sư phạm.

Thứ ba, hệ thống chuyển kết quả phân tích thành cơ hội luyện tập. Đây là điểm khác biệt quan trọng giữa một công cụ chấm bài và một hệ thống hỗ trợ học tập. Một kết quả chỉ có giá trị giáo dục khi giúp học sinh biết cách cải thiện. VietSpeak AI liên kết nhiệm vụ, rubric, phản hồi và lần thực hiện tiếp theo trong cùng một quy trình.

Các đóng góp này phù hợp với định hướng đánh giá vì sự tiến bộ của người học. Công nghệ hướng đến việc mở rộng số lần thực hành, cá nhân hoá phản hồi và giảm thời gian xử lí lặp lại. Giáo viên vẫn cần tổ chức các hoạt động nghe, nói và trao đổi trực tiếp trên lớp để học sinh chuyển những gì đã luyện qua video vào giao tiếp thực.

3.7. Điều kiện triển khai và hướng hoàn thiện

Việc triển khai VietSpeak AI cần bảo đảm ba nhóm điều kiện.

Nhóm thứ nhất là điều kiện kĩ thuật. Học sinh cần có thiết bị ghi hình, âm thanh đủ rõ và kết nối phù hợp để gửi bài. Nhà trường cần có quy trình xử lí khi tệp lỗi, âm thanh quá nhỏ hoặc video chưa đạt yêu cầu. Có thể tổ chức ghi bài tại lớp; với nhiệm vụ không cần quan sát hình ảnh, có thể cho nộp âm thanh nếu quy trình tiếp nhận hỗ trợ và dùng bộ tiêu chí tương ứng. Người lớn chỉ hỗ trợ thao tác. Tỉ lệ tự hoàn thành, lỗi thao tác và mức trợ giúp cần được đo riêng theo khối lớp.

Nhóm thứ hai là điều kiện chuyên môn. Tổ chuyên môn cần thống nhất cách giao nhiệm vụ, cách sử dụng rubric, cách đọc phản hồi và cách xử lí trường hợp hệ thống nhận diện chưa chính xác. Nhà trường có thể bắt đầu từ một khối lớp, cùng xem một số sản phẩm, thống nhất cách phản hồi rồi mở rộng dần.

Nhóm thứ ba là điều kiện quản trị dữ liệu. Nhà trường cần có sự đồng thuận của phụ huynh, phân quyền truy cập, quy định thời hạn lưu trữ và cơ chế bảo vệ dữ liệu. Hình ảnh, tiếng nói và sản phẩm của học sinh cần được sử dụng đúng mục đích giáo dục.

Các hạn chế chính gồm dữ liệu triển khai chưa được mô tả đủ để đánh giá tính đại diện, chưa xác nhận tập kiểm thử độc lập, chưa có chỉ số nhất quán giữa người chấm và phân tích sai số theo khối lớp hoặc giọng địa phương. Dữ liệu hiện tập trung vào sản phẩm nói; nghe hiểu, tương tác trực tiếp, mức tự thao tác và sự phụ thuộc vào phản hồi AI cần được kiểm chứng riêng. Nghiên cứu tiếp theo cần đo đầu–cuối với nhóm so sánh, dùng người chấm không biết nhóm và thời điểm đo, kết hợp nhiệm vụ giao tiếp trực tiếp.

4. Kết luận và kiến nghị

Nghiên cứu đã trình bày quá trình xây dựng và triển khai VietSpeak AI, hệ thống trí tuệ nhân tạo đa phương thức hỗ trợ rèn luyện và đánh giá nói và nghe ở tiểu học. Hệ thống được phát triển từ yêu cầu của chương trình, tổ chức thành 15 dạng nhiệm vụ và 45 bộ rubric theo khối lớp, đồng thời tích hợp 12 chủ đề văn hoá Huế.

VietSpeak AI phân tích ba nguồn dữ liệu gồm nội dung lời nói, âm thanh và hình ảnh. Kết quả phân tích tạo thành dự thảo phản hồi định hướng hành động. Giáo viên cần kiểm tra dẫn chứng, điều chỉnh và duyệt trước khi học sinh tiếp nhận để luyện lại. Mô hình này tạo ra sự phân công rõ ràng giữa hệ thống, học sinh và giáo viên, trong đó AI hỗ trợ các thao tác phân tích lặp lại còn giáo viên giữ vai trò phán đoán chuyên môn.

Số liệu triển khai ban đầu phản ánh phạm vi sử dụng và gợi ý khả năng hỗ trợ xử lí bài nộp. Tuy nhiên, các chỉ số vận hành còn cần kiểm chứng về cách tính, dữ liệu đối chiếu và thời gian kiểm duyệt; chúng chưa chứng minh độ chính xác, khả năng khái quát hoặc hiệu quả phát triển năng lực nói và nghe.

Đóng góp chính của VietSpeak AI nằm ở việc kết nối bốn thành phần trong một quy trình thống nhất: nhiệm vụ học tập, rubric theo chương trình, phân tích đa phương thức và phản hồi cho lượt luyện tiếp theo. Cách tiếp cận này đề xuất một hình thức tích hợp AI vào đánh giá thường xuyên; hiệu quả đối với cơ hội thực hành và khối lượng nhận xét cần tiếp tục được kiểm chứng.

Trong giai đoạn tiếp theo, VietSpeak AI cần ưu tiên nâng cao chất lượng phản hồi, đánh giá sâu các nhiệm vụ nghe hiểu và tương tác, theo dõi tiến bộ của người học qua nhiều lần thực hiện và kiểm tra độ ổn định của hệ thống với các nhóm học sinh có giọng địa phương, điều kiện thiết bị và mức độ thành thạo khác nhau.

TÀI LIỆU THAM KHẢO

Bộ Giáo dục và Đào tạo. (2018). *Chương trình giáo dục phổ thông môn Ngữ văn* (ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018).

Bộ Giáo dục và Đào tạo. (2020). *Thông tư số 27/2020/TT-BGDĐT ngày 04 tháng 9 năm 2020 ban hành Quy định đánh giá học sinh tiểu học.* https://vanban.chinhphu.vn/?pageid=27160&docid=201006

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading.* ASCD.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., & Picard, R. W. (2013). MACH: My automated conversation coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 697–706). Association for Computing Machinery. https://doi.org/10.1145/2493432.2493502

Huu, T. T., Pham, V. T., Nguyen, T. T. T., & Dao, T. L. (2023). Mispronunciation detection and diagnosis model for tonal language, applied to Vietnamese. In *Proceedings of Interspeech 2023* (pp. 1014–1018). https://doi.org/10.21437/Interspeech.2023-364

Mealings, K., Miles, K., & Buchholz, J. M. (2025). A methodological review of listening comprehension tests for primary school children. *International Journal of Listening, 39*(1), 17–50. https://doi.org/10.1080/10904018.2023.2229368

Nguyễn Thị Quế. (2024). Nghiên cứu đề xuất khung mô tả năng lực nói và nghe tiếng Việt của học sinh lớp 3. *Tạp chí Giáo dục, 24*(số đặc biệt 1), 24–28. https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/1549

Rudner, M., Lyberg-Åhlander, V., Brännström, J., Nirme, J., Pichora-Fuller, M. K., & Sahlén, B. (2018). Listening comprehension and listening effort in the primary school classroom. *Frontiers in Psychology, 9*, Article 1193. https://doi.org/10.3389/fpsyg.2018.01193

UNICEF. (2025). *Guidance on AI and children* (Version 3.0). https://www.unicef.org/innocenti/reports/policy-guidance-ai-children

Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication, 51*(10), 883–895. https://doi.org/10.1016/j.specom.2009.04.009